The Purification and N-terminal Amino Acid Sequence Analysis of the High Molecular Weight Gluten Polypeptides of Wheat

Peter R. SHEWRY, J. Michael FIELD, Audrey J. FAULKS, Saroj PARMAR, Benjamin J. MIFLIN

Biochemistry Department, Rothamsted Experimental Station, Harpenden, U.K.

Mary D. DIETLER, Ellen J-L. LEW, and Donald D. KASARDA

Food Proteins Research Unit, Western Regional Research Centre, Berkeley, U.S.A.

Correspondence to P.R. Shewry. Rothamsted Experimental Station, Harpenden, Herts. U.K.

SUMMARY

Nine high molecular weight gluten polypeptides have been purified from cultivars of bread wheat (Triticum aestivum). They have broadly similar amino acid compositions with 34-39 mol % glx, 12-17 mol % pro and 14-19 mol % gly. Some differences in the contents of other amino acids were apparent, notably cys varied from 0.4 to 1.5 mol %. On the basis of these compositions, the peptide maps with V8 protease and N-terminal amino acid sequences the polypeptides could be divided into four groups. Two of these contained polypeptides encoded by structural loci on chromosomes 1A and 1B respectively while the other two probably represented the products of two subfamilies of genes at a single locus on chromosome 1D. The N-terminal amino acid sequences were homologous, but differed in substitutions of single amino acids and in deletions of 3 and 7 amino acids. Two polypeptides purified from Triticum monococcum and Aegilops squarrosa had N-terminal sequences closely related or identical to those of the 1A polypeptides and one group of 1D polypeptides respectively. Two cys residues were present in the N-terminal sequences of all the subunits, but the distances between these were affected by the deletions and varied from 7 to 14 residues. The results are discussed in relation to the ability of the polypeptides to form elastic disulphide-linked aggregates, and the importance of these in the structure and functionality of gluten.

INTRODUCTION

Gluten is the viscoelastic proteinaceous mass which remains when dough made from wheat flour is washed to remove starch and soluble proteins. The proteins present in gluten are chiefly storage proteins derived from

endosperm protein bodies [1]. These have the common properties of being insoluble in aqueous solvents but soluble, although sometimes only after reduction of disulphide bonds, in mixtures of alcohols (e.g. ethanol, propan-1-ol or propan-2-ol) and water, and of being rich in glutamine and proline but poor in charged amino acids. They can thus be defined as prolamins. Classification of gluten proteins is difficult and controversial [see 2,3], but a simple division may be made between those which are present in aggregates (classically called glutenins) and those which are monomeric (gliadins).

The functional properties of wheat dough (i.e. the ability to be baked into leavened bread) are largely dependent on the gluten and in particular in its property of visco-elasticity. The exact physical-chemical properties of gluten and the contributions made to visco-elasticity by each of the individual components is not known. A widely held view is that the aggregated components are especially important [see 3,4], and in particular breadmaking quality has been related to the presence of high molecular weight glutenin subunits (HMW subunits). These are present in larger quantities in the grain of wheat than in those of other cereals (barley, rye) which make only a poor dough, if at all [3]

It has been proposed that the glutenin subunits are joined by disulphide bonds into long linear polymers with only a limited amount of branching and that during gluten formation these interact to form an elastic network [5]. Preliminary analyses of the HMW subunits indicate that they possess at least some of the necessary properties [6,7], notably they have a small number of cysteine residues which appear to be located close to one or both ends of the chains and a high content of glycine, the latter possibly conferring chain flexibility.

Wheat cultivars differ in their suitability for breadmaking (baking quality), and in some cases this is correlated with the absence or presence

of specific HMW subunits [8]. The genetics of wheat proteins is complicated by the fact that wheat is an allohexaploid comprised of three genomes (called A, B and D), each being seven pairs of chromosomes. The HMW subunits are coded for by loci (called Glu-1) on the long arms of chromosomes 1A, 1B and 1D [9]. Each locus appears to code for more than one polypeptide and, like those for other cereal storage proteins [see 3], is probably complex. A number of alleles at each locus are present in wheat cultivars and Payne et al. [10] have numbered the individual subunits separated by sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE) and shown that subunits 1 and 2* are coded for by chromosomes 1A, subunits 6-9 and 13-20 by chromosome 1B and subunits 2, 3, 5, 10 and 12 by chromosome 1D.

In this paper we report the purification and characterization of a number of HMW subunits from cultivars of bread wheat (Triticum aestivum) and from diploid species thought to be related to the ancestors of the A and D genomes. These results give further information on the genetic and biochemical relationships within this important group of proteins.

MATERIALS AND METHODS

Chemicals

Stock solutions of 8M urea were deionized on a column of Biorad AG501-X8D mixed bed ion-exchange resin. 4-Vinylpyridine (Sigma) was redistilled and stored under N_2 at -20°C. V8 protease (Staphylococcus aureus protease, V8) was from Miles Laboratories Ltd.

Protein Extraction

Seed of bread wheat (Triticum aestivum) cultivars Highbury, Maris Butler (spring) Copain, Brigand, Sentry and Cheyenne (winter) were obtained from commercial suppliers or from the experimental farm at Rothamsted. Seed of Triticum monococcum and Aegilops squarrosa were obtained from the Plant Breeding institute, Cambridge, and multiplied in pots at Rothamsted. The seeds were milled to pass a 0.7 mm sieve. The meal was first extracted by stirring for 2 h at 4°C with 5 ml/g of 70% (v/v) ethanol + 1% (v/v) 2-mercaptoethanol. The supernatant contained mainly gliadins and was discarded. A second fraction, enriched in HMW subunits, was then extracted by stirring for 18 h at 20°C with 5 ml/g of 50% (v/v) propan-1-ol + 2% (v/v) 2-mercaptoethanol + 1% (v/v) acetic acid. The supernatant was neutralized with NaOH and diluted with 2 volumes of 1.5M NaCl. After standing at 4°C for 18 h the precipitated prolamins were collected by centrifugation and lyophilized. They were then reduced and pyridylethylated [11]

Total prolamin fractions for electrophoresis were prepared from cultivars of T. aestivum by stirring 2 g milled grain for 1 h at 20°C with 20 ml of 50% (v/v) propan-1-ol + 2% (v/v) 2-mercaptoethanol + 1% (v/v) acetic acid. After centrifugation the supernatant was removed and the

extraction repeated with fresh solvent. The supernatants from the two extracts were bulked and the prolamins precipitated, reduced and pyridylethylated as described above. Total prolamins from single seeds of T. monococcum and Ae. squarrosa were prepared as described previously [12]

Protein Purification

1 g protein was dissolved in 50 ml of 0.01M acetic acid containing 6M urea, applied to a 90 x 4.4 cm column of Sephacryl S300 and eluted with the same solvent. The leading fractions containing the HMW subunits were bulked, dialyzed against distilled water and lyophilized. The bulked fractions from several columns were dissolved in 50 ml of 10 mM glycine/acetate buffer, pH 4.6, containing 3M urea and applied to a 30 x 2.2 cm column of CM cellulose equilibrated in the same buffer. The column was eluted with a 2:1 gradient of 0 - 0.25M NaCl in the same buffer (total volume 1200 ml). Fractions containing HMW subunits were again bulked, dialyzed and lyophilized. For the preparation of unalkylated proteins 2-mercaptoethanol was added to the buffers used to dissolve the proteins (1%, v/v) and elute the columns (0.1% v/v). Some HMW subunits were obtained in a pure state by this procedure. With others it was necessary to separate fractions from the ion exchange chromatography by preparative isoelectric focusing in a flat bed of Ultradex granulated gel as described by Field et al. [7]. The protein preparations from this procedure were chromatographed on Sephadex G75 in 0.1M acetic acid to remove large ampholyte molecules.

Protein Separation

Protein fractions were dissolved in 8M urea containing 1% SDS and separated by SDS-PAGE on 12.5% acrylamide gels at pH 8.9 as described previously [13]. For separation of peptides the concentration of acrylamide

was increased to 20%. The proteins were also dissolved in 6M urea containing aluminium lactate buffer (pH 3.2) and separated by PAGE in the same buffer in slab gels containing 6% acrylamide and 3M urea (lactate PAGE) [14]. Isoelectric focusing (IEF) (pH range pH 3.5-10) in slab gels containing 5% acrylamide was as described previously [13].

Amino Acid Analysis

1 mg protein was hydrolyzed for 21 h at 110°C under N_2 with 1 ml of 5.7N HCl containing 0.1% 2-mercaptoethanol. Amino acids were determined with an LKB 4400 analyser.

Determination of Amidation Level

The amidation level was determined by carbodiimide modification of the free carboxyl groups as described by Carraway and Koshland [15], but with the substitution of norleucine methyl ester for glycinamide [16].

Peptide Preparation

Preparations were incubated with cyanogen bromide (CNBr), trypsin, chymotrypsin and V8 protease as described previously [17], with incubation times of 24 h, 8 h, 2 h, and 2 h respectively.

N-terminal Sequencing

N-Terminal amino acid sequencing was carried out with a Beckman automatic sequencer, model 890B (updated including cold trap), and Beckman 1 M Quadrol program 112078. Phenylthiohydantoin (PTH) amino acids were determined by high performance liquid chromatography on a Waters C_{18} reverse-phase column and with Waters chromatography equipment (model 6000A solvent delivery system, model 720 system controller, model 730 data module, model 450 variable wavelength detector set at 269 nm) according to

the method of Bhown et al., [18]. The solvent system was: A, 0.04 M sodium acetate, pH 4.15; B, 100% methanol; a linear gradient from 20% B to 50% B was applied in 15 min, holding at 50% B for 10 min. PTH valine and PTH methionine were separated by thin layer chromatography [19].

RESULTS AND DISCUSSION

Purification and electrophoretic Analysis of HMW Polypeptides

A number of cultivars containing a range of HMW polypeptides were selected for study (Fig. 1). The cultivars together contained almost all the HMW subunits which are commonly found in commercially-grown varieties of wheat [10], including subunits 1, 5 and 10 which are reported to be associated with baking quality [8].

Prolamin fractions were separated by gel filtration on Sephacryl S300 followed by ion exchange chromatography on CM cellulose. Some fractions from the latter separation contained single subunits, others contained mixtures. The latter were separated again by preparative IEF in granulated gel. Eighteen fractions, containing a total of nine different subunits, were considered to be sufficiently pure for further characterization. SDS-PAGE analyses of these are shown in Fig. 2. This showed that all contained a single major subunit band, although small amounts of other subunits were present in some of the preparations. Some also contained trace amounts of components with very slow mobilities. These were not apparently present in the total prolamin fractions (Fig. 1 tracks a,e,j,o,r,t) and their identity is not known. Also shown in Fig. 2 (track z) is an HMW component from barley which is genetically and chemically homologous with the HMW subunits of wheat [7,19]. This component, called 'D' hordein, was purified from the mutant barley line Risø 1508 as described by Kreis et al., [20].

The yields from the purification procedure varied considerably; 200 g milled grain gave about 5 g of the enriched prolamin preparation, while yields of the purified subunits varied from less than 10 mg of subunits from IEF to up to 50 mg of components from CM cellulose.

All the subunits specified by the A and D genomes were successfully

purified, but only two B genome subunits (6 and 7) out of a total of 8. The D genome coded subunits fall into two distinct groups with different mobilities on SDS-PAGE. Payne et al. [10] showed that both groups were controlled by the same locus on chromosome 1D and called the slower subunits (2,3,5) the 1Dx group and the faster ones (10,12) the 1Dy group. All commonly grown varieties contain one subunit of each group. Similarly the subunits controlled by chromosome 1B can be divided into the slower 1Bx (6,7,17) and faster 1By (8,9,18) groups, with cultivars having either one subunit of each group or one 1Bx subunit only. The presence of three 1B subunits in the total prolamin fraction of Highbury (Fig. 1, track b) is due to heterogeneity in the cultivar. Two seed types are present, one with subunit 1Bx6 and the other with subunits 1Bx17 and 1By18. One of the two subunits controlled by chromosome 1A (1,2*) is present in some cultivars only. Subunit 1A2* has similar mobility on some SDS-PAGE systems to subunit 1D2. Molecular weights of 95-145,000 have been determined for HMW subunits by SDS-PAGE [21] (see also Fig. 1). Results from sedimentation equilibrium centrifugation indicate that these are over-estimates. Hamauzu et al. [22] reported a mol. wt. of 44,000 for a mixture of 'glutenin' polypeptides while we recently determined a mol. wt. of 69,600 for a pure preparation of subunit 2 [7].

A number of the purified subunit preparations were also analyzed by electrophoresis at pH 3.2 (lactate PAGE) (Fig. 3A) and analytical IEF (Fig. 3B). Lactate PAGE showed a predominant band for each subunit although some slower bands were also present. Where several preparations of the same subunit were separated they gave identical or closely similar patterns (see Fig. 3A tracks f,g,h and n,o,p). The three 1Dx subunits (2, 3 and 5) showed similar slow mobilities while the two 1Dy subunits (10 and 12) again had similar mobilities but were much faster (Fig. 3A tracks i-p). The 1Bx and 1A subunits were slightly faster than the 1Dx group. The

'D' hordein component showed similar mobility to the 1Dy components. The mobility at pH 3.2 is determined mainly by the content of basic amino acids, including the S-β-pyridylethylcysteine formed by the alkylation.

The presence of S-β-pyridylethylcysteine also affects the isoelectric points of the proteins and hence the patterns shown on IEF (Fig. 3B). This showed the presence of several major bands in all the preparations. A similar degree of polymorphism is observed in two-dimensional (IEF/SDS-PAGE) analyses of total prolamin fractions [3, 23,24] indicating that it is not generated by chemical modification of the subunits during purification. Clear differences in pI were observed between some of the subunits, the 1Bx, 1Dy and D hordein components all having higher pIs than the 1Dx subunits. Different preparations of the same subunit again gave similar or identical patterns (Fig. 3B, tracks l-n).

Characterization of the Purified HMW Subunits

Amino Acid Analysis

Amino acid compositions of most of the subunit preparations are presented in Table 1. There is generally good agreement between the results obtained for different preparations of the same subunit, the differences which are present probably resulting at least partly from variation in the purity of the preparations. Although all the preparations are rich in glx, pro and gly, some differences in the proportions of those amino acids are observed. Carbodiimide modification showed that 90% of the recovered glutamate and aspartate in subunit 2 were present as amides. This agrees with analyses of gliadins and hordeins, which have shown amidation levels of 90% or greater [25, 26]. There were also differences in the proportions of some other amino acids, notably aspartate, isoleucine, phenylalanine and arginine. Similar variation in the amino acid compositions of a number of pure and mixed HMW preparations

was observed by Khan and Bushuk [6]. There was also variation in the proportion of cysteine, notably about 1.2 to 1.5 mol % in the 1Dy subunits (10 and 12) compared to about 0.5 mol % in the 1Dx subunits (2,3 and 5). On the basis of molecular weights of 69,600 and 63,000 determined for subunits 1Dx2 and 1Dy12 respectively by sedimentation equilibrium ultracentrifugation [7 and authors unpublished results] and the amino acid compositions reported in Table 1, it can be calculated that these two subunits contain about 3-4 and 7 residues of cysteine per mol. Only traces of methionine (less than 0.5 mol %) were detected, but some losses of this amino acid during acid hydrolysis can be expected.

Peptide Mapping

Treatment of the purified subunit preparations with CNBr and analysis of the products by SDS-PAGE showed single major bands with mobilities only slightly faster than those of the uncleaved subunits (results not shown). This is consistent with the low contents of methionine determined by amino acid analysis (Table 1) and also indicates that the methionine residue(s) are not located in the central part of the polypeptide chain. Enzymic digestion with trypsin gave very limited proteolysis of the subunits, while digestion by chymotrypsin was rapid, with many of the fragments too small to be separated on polyacrylamide gels. When the subunits were digested with V8 protease they gave a range of peptides with molecular weights within the fractionation range of a 20% SDS-PAGE gel (Fig. 4). Where different preparations of the same subunit were digested they always gave identical or closely similar peptide maps (Fig. 4, tracks c,d and j,k,l). Also, the patterns observed were reproducible between separate digests of the same subunit preparation. The three 1Dx subunits (2, 3 and 5) gave peptide maps which were clearly related (Fig. 4, tracks c-f) and different from those of the 1Dy subunits 10 and 12 (Fig. 4, tracks g,h). The two 1Bx

N-terminal Amino Acid Sequences

N-terminal amino acid sequences were determined using a Beckman 890B automatic sequencer. A number of the subunit preparations were completely blocked to Edman degradation, although other preparations of the same subunits were not. The yields from these were variable and generally low, between 1 and 3 nmoles of the N-terminal glutamic acid residue being released from each mg of protein. If a mean molecular weight of 66,000 is assumed (from sedimentation equilibrium ultracentrifugation) this represents a yield of between 7 and 20%. The blocking and variable yields presumably resulted from cyclization of the N-terminal glutamic acid residues to pyrrolidone carboxylic acid during the low pH purification procedure. Because of this blocking it was not possible to determine N-terminal amino acid sequences of all the different purified HMW subunits.

The basic sequence was established by analyzing three preparations of subunit 1Dx2. Preparations from Copain and Highbury were analyzed for 35 cycles, the Copain preparation being run twice. A third preparation from Brigand was analyzed for 14 cycles and then aborted due to low yields. The sequence determined from these analyses is shown in Table 2. The other subunits were only analyzed once each, for a varying number of cycles due to differences in the amounts of the preparations available for analysis and the yields. Although we did not have sufficient protein to run all the analyses in duplicate, the sequences obtained were, with the exception of the 1Bx7 component, confirmed by being observed in two or more homologous subunits.

The sequences obtained for the 1A1, 1Bx7, 1Dx2, 1Dx5, 1Dy10 and 1Dy12 subunits are shown in Table 2. These are aligned to maximise homology,

resulting in gaps in these for the 1A1, 1Bx7, 1Dy10 and 1Dy12 components. These gaps are of three amino acids in the 1A and 1Dy subunits and seven in the 1Bx subunit, and presumably result from deletions in the genomic DNA encoding the proteins. The sequences observed for the 1Dx subunits (2 and 5) are, allowing for our inability to positively identify residues at some positions, identical for the first 35 residues. Similar sequences for the first 25 residues were observed for the two 1Dy subunits (10 and 12), except for a deletion of 3 amino acids, the substitution of arginine for glutamate at position 6 and the tentative identification of glutamate instead of lysine at position 23 of subunit 10 only. The three amino acid deletion in the 1A1 subunit of Copain was in a different position to that in the 1Dy subunits. Apart from this it differed from the 1Dx sequence only in the substitution of glycine for glutamate at position 6 and a mixture of glycine and proline for leucine at position 17. The only 1B subunit analyzed (7) also had glycine at position 6 and glutamic acid instead of glutamine at position 15. The seven amino acid deletion overlapped the two deletions present in the 1Dy and 1A subunits. Taken together there is a high degree of homology between the sequences of the various subunits. Only one position, 6, is very variable with the presence of an unchanged (glycine), acidic (glutamate) or basic (arginine) amino acid in different subunits. The presence of deletions between positions 15 and 24 may indicate that this region is less conserved.

Two cysteine residues were present in all the subunit preparations, although the distance between these was affected by the deletions and consequently varied from 14 residues in the 1Dx subunits to 11 residues in the 1A1 and 1Dy subunits and 7 residues in the 1Bx7 subunit.

Comparison of the sequences with the amino acid compositions of the whole subunits (Table 1) shows that the amino acids are not randomly distributed. The proportions of charged amino acids in the N-terminal

region are very high, with 3 or 4 basic amino acids and up to 8 glutamate residues in the first 25. Conversely glycine was only present at one position out of the first 25 (with the exception of the 1A1 subunit) although it accounted for 14-21 mol % of the total residues (Table 1). Similarly proline (12-21 mol %) was only detected at one position in one subunit and tyrosine (3.5 to 7.2 mol %) and threonine (about 3 mol %) were not detected at all. Analyses of other prolamin groups including hordein of barley [17, 27] and zein of maize [28, 29] also show that the amino acid compositions of the N-terminal and C-terminal regions are atypical compared to those of the whole proteins. Previous analysis of the 'D' hordein component of barley has shown that it is blocked to Edman degradation [7].

Purification and N-terminal Amino Acid Sequences of HMW Subunits from Triticum monococcum and Aegilops squarrosa

The progenitors of the A and D genomes of hexploid bread wheat are thought to be a Triticum species (possibly T. boeoticum or T. urartu) and Aegilops squarrosa (a related wild grass) respectively [30]. The mobilities on SDS-PAGE of the HMW subunits of T. urartu and T. boeoticum (usually one band) and Ae. squarrosa (two bands) are also similar to those of the A and D genome - coded subunits of bread wheat [31]. The origin of the B genome is not known but it has been suggested that it is polyphyletic [32]. We therefore purified small amounts of one HMW subunit each from T. monococcum (the domesticated form of T. boeoticum) and Ae. squarrosa. SDS-PAGE of these components showed single bands with mobilities similar to those of the 1A and 1Dx subunits of T. aestivum (Fig. 2). The Ae. squarrosa component also had an identical N-terminal amino acid sequence to the 1Dx subuits 2 and 5 (Table 2) while the N-terminal sequences of the 1A1 and T. monococcum subunits differed at only one position. This provides further evidence that Ae.

squarrosa and T. monococcum are related to the progenitors of the A and D genomes respectively.

GENERAL DISCUSSION

The results reported here indicate that the genes for the HMW subunits of the A, B and D genomes of bread wheat have evolved from a single ancestral gene. This divergence has involved both small deletions and substitutions in the region of the gene coding for the N-terminus of the protein. Presumably larger insertions or deletions must be present in other parts of the genes to account for the differences in molecular weights of the polypeptides. Divergence has also occurred within the genes present on chromosome 1D to give two subfamilies, coding for the 1Dx and 1Dy subunits. These groups of subunits have different N-terminal amino acid sequences, V8 protease peptide maps and amino acid compositions. Payne et al. [10] suggested that the 1Dx subunits, 1Bx subunits and 1A subunits form one homologous subfamily; similarly the 1Dy subunits, 1By subunits and 1Ay subunits (the latter found only in diploid species) form a second subfamily.

Because recombination between the 1Dx and 1Dy subunits has not been observed the results strongly indicate that the Glu-D1 locus is complex. The alternative hypothesis, that the differences are due to post-translational modification of a single polypeptide chain, is unlikely because of the presence of amino acid substitutions and deletions. Differences arising at the post-transcriptional level are also unlikely because, although the deletion of three amino acids could result from variation in the processing of mRNA, this is unlikely to account for the substitutions. The locus may also code for a number of polymorphic variants of both 1Dx and 1Dy subunits, which is indicated by the presence of isomers with different isoelectric points (Fig. 3B). In this respect

the complexity of the locus is comparable to that of the Hor 2 locus of barley which codes for the B hordein group of storage prolamins. This locus specifies a number of polypeptides varying in molecular weight and isoelectric point, and these can be classified into groups which appear to be encoded by two sub-families of mRNAs [33]. The present study does not allow us to draw any conclusions about the complexity of the Glu-A1 and Glu-B1 loci.

The presence of two cysteine residues close to the N-termini of all the subunits is compatible with the hypothesis that they are assembled into long head-to-tail polymers by disulphide bonds. This theory also requires the presence of at least one cysteine in the C-terminal region, and evidence for this has recently been obtained by sequencing cloned cDNA sequences which hybridize to mRNAs for HMW subunits [34]. It is not know whether the deletions in the N-terminal regions, which affect the distances between the cysteine residues, also affect the ability of these cysteines to become involved in inter- or intra-molecular disulphide bonds. Preliminary predictions of the secondary structures of these regions indicate that they are α-helical, which makes it unlikely that the two cysteine residues would form disulphide bonds with each other (Tatham, Shewry and Miflin, unpublished results). Further structural predictions must await more sequence information and physical chemical studies. In relation to baking quality Payne et al. [8] have shown that subunits 1Dx5 and 1Dy10 are associated with good quality and 1Dx2 and 1Dy12 with poor. Although no differences were observed between subunits 2 and 5 or 10 and 12 in N-terminal amino acid sequences, there was some variation in amino acid compositions and V8 protease peptide maps. The significance of these differences and their possible relationship to quality factors can only be assessed by further detailed studies.

ACKNOWLEDGEMENTS

The authors are grateful to S. Smith for amino acid analyses, Dr. C. N. Law (PBI, Cambridge) for supplying seed of T. monococcum and Ae. squarrosa and Dr. P. I. Payne (PBI) for valuable advice and discussions throughout the study. The work at RES was supported by European Economic Community grant No. 470, a grant from the Home Grown Cereals Authority, and NATO grant No. 0324/82.

FIGURE LEGENDS

Fig. 1 A. SDS-PAGE of reduced and pyridylethylated (PE) total prolamin fractions from T. aestivum, T. monococcum and Ae. squarrosa. a-f are from T. aestivum cultivars Cheyenne, Highbury, Copain, Sentry, Brigand and Butler respectively; g and h are from T. monococcum and Ae. squarrosa; i is molecular weight standards; 1, bovine serum albumin (66,000); 2, ovalbumin (45,000); 3, lactate dehydrogenase (36,200); 4, trypsinogen (24,000); 5, β lactoglobulin (18,400); 6, lysozyme (14,300).
B. Diagram to show chromosomal control of HMW subunits present in cultivars of T. aestivum. a-f as in part A.

Fig. 2 SDS-PAGE of purified HMW subunits and total prolamin preparations.
A. from T. aestivum and barley. a, total prolamin from cv. Brigand; b-d, subunits 1Dx2, 1Bx6 and 1Dy12 from cv. Brigand; e, total prolamins from cv. Copain; f-i subunits 1A1, 1Dx2, 1Bx7 and 1Dy12 from cv Copain; j, total prolamins from cv Sentry; k-n, subunits 1A1, 1Dx3, 1Bx7 and 1Dy12 from cv Sentry; o, total prolamins from cv Highbury; p,q, subunits 1Dx2 and 1Dy12 from cv Highbury; r, total prolamins from cv Butler; s, subunit 1Dx2 from cv Butler; t, total prolamins from cv Cheyenne; u-x, subunits 1Dx5, 1A2*, 1Bx7 and 1Dy10 from cv Cheyenne; y, total prolamins from barley mutant Risø 1508; z, D hordein from Risø 1508.
B. from T. monococcum and Ae. squarrosa. a and b, total prolamins and purified HMW subunit from T. monococcum; c and d, total prolamins and purified HMW subunit from Ae. squarrosa.
2 and 12 indicate the migration of subunits 1Dx2 and 1Dy12. These have been reported to have molecular weights by SDS-PAGE of 136,000 and 95,000 respectively [21].

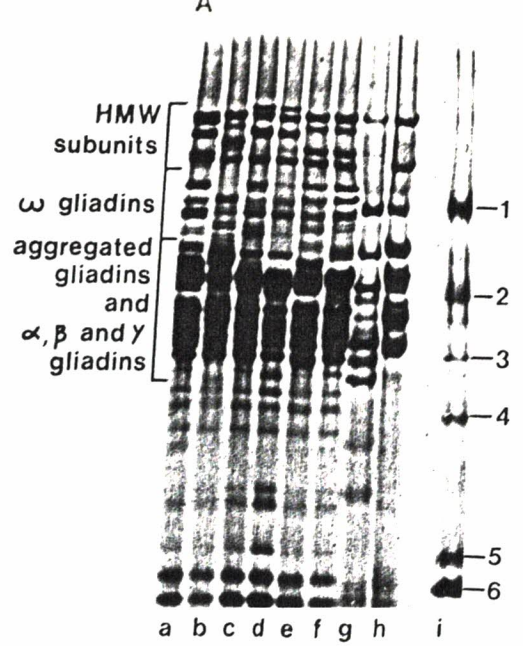
Fig. 3 A. lactate-PAGE of purified HMW subunits
a, HMW-enriched prolamin preparation from cv. Brigand; b, D hordein (Risø 1508); c, 1A1 (Copain); d, 1A2* (Cheyenne); e, 1Bx6 (Brigand); f-h, 1Bx7 (Cheyenne, Sentry and Copain respectively); i and j, 1Dx2 (Highbury and Copain respectively); k, 1Dx3 (Sentry); l, 1Dx5 (Cheyenne); m, 1Dy10 (Cheyenne); n-p, 1Dy12 (Copain, Brigand and Butler respectively).
B. IEF (pH range 3.5-10) of purified HMW subunits.
a, o total prolamins from cv Brigand; b, D hordein (Risø 1508); c, 1A1 (Copain); d, 1A2* (Cheyenne); e, 1Bx6 (Brigand); f, 1Bx7 (Sentry); g, 1Dx2 (Copain); H, 1Dx3 (Sentry); i, 1Dx5 (Cheyenne); j, 1Dy10 (Cheyenne); k, 1Dy12 (Brigand).
l-n are 1Dy12 from Copain, Sentry and Brigand respectively.
a-k were run on a separate gel to l-o.

Fig. 4 SDS-PAGE of HMW subunits after digestion with V8 protease
a, 1A1 (Sentry); b, 1A2* (Cheyenne); c and d, 1Dx2 (Butler and Copain); e, 1Dx3 (Sentry); f, 1Dx5 (Cheyenne); g, 1Dy 10 (Cheyenne); h, 1Dy12 (Brigand); i, 1Bx6 (Brigand); j-l, 1By7 (Cheyenne, Copain and Sentry). The markers to the left show the positions of the major bands in the protease preparation. These were not observed in the digests.
1-5 indicate the positions of molecular weight markers: 1, bovine serum albumin (66,000); 2 ovalbumin (45,000); 3 trypsinogen (24,000); 4, β-lactoglobulin (18,400); 5, lysozyme (14,300).

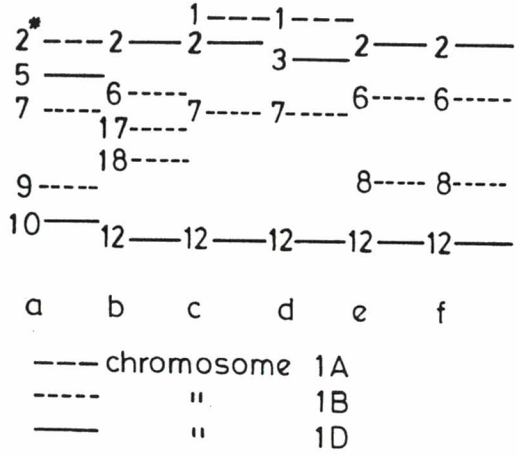
REFERENCES

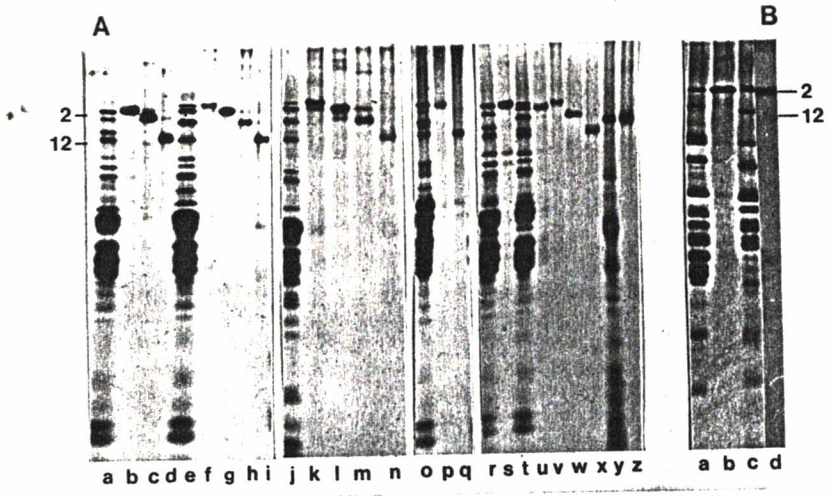
1. Field, J.M., Shewry, P.R., Burgess, S.R., Forde, J., Parmar, S., & Miflin, B.J. (1983) J. Cereal Sci. 1, 33-41.
2. Kasarda, D.D., Bernardin, J.E., & Nimmo, C.C. (1976) In Advances in Cereal Science and Technology (Pomeranz, Y. ed.), pp. 158-236. AACC. St. Paul, Minnesota, USA.
3. Miflin, B.J., Field, J.M., & Shewry, P.R. (1983) In Plant Proteins (Daussant, J., Mossé, J. & Vaughan, J. eds.), pp. 255-319. Academic Press, New York.
4. Wall, J.S. (1979) In Recent Advances in the Biochemistry of Cereals (Laidman, D.L. & Wyn Jones, R.G. eds.), pp. 275-312. Academic Press, New York.
5. Ewart, J.A.D. (1977) J. Sci. Food Agric. 28, 191-199.
6. Khan, K., & Bushuk, W. (1979) Cereal Chem. 56, 505-512.
7. Field, J.M., Shewry, P.R., Miflin, B.J., & March, J.F. (1982) Theor. Appl. Genet. 62, 329-336.
8. Payne, P.I., Corfield, K.G., Holt, L.M., & Blackman, J.A. (1981) J. Sci. Food Agric. 32, 51-60.
9. Payne, P.I., Holt, L.M., Worland, A.J., & Law, C.N. (1982) Theor. Appl. Genet. 63, 129-138.
10. Payne, P.I., Holt, L.M., & Law, C.N. (1981) Theor. Appl. Genet. 60, 229-236.
11. Friedman, M., Krull, L.H., Cavins, J.F. (1970) J. Biol. Chem. 245, 3868-3871.
12. Shewry, P.R., Faulks, A.J., Pratt, H.M., & Miflin, B.J. (1978) J. Sci. Food Agric. 29, 847-849.
13. Shewry, P.R., Ellis, J.R.S., Pratt, H.M., & Miflin, B.J. (1978) J. Sci. Food Agric. 29, 433-441.
14. Shewry, P.R., Autran, J.C., Nimmo, C.C., Lew, E.J-L., & Kasarda, D.D. (1980) Nature 286, 520-522.
15. Carraway, K.L. & Koshland, D.E. (1972) Meth. Enzym. 25B, 616-623.
16. Field, J.M., Rodger, G.W., Hunter, J.C., Serafini-Fracassini, A., & Spina, M. (1978) Archs. Biochem. Biophys. 11, 705-713.
17. Shewry, P.R., Lew, E.J-L., & Kasarda, D.D. (1981) Planta 153, 246-253.
18. Bhown, A.S., Mole, J.E., Weissenger, A., & Bennett, J.C. (1978). J. Chromatog 178, 532-535.
19. Shewry, P.R., Finch, R., Parmar, S., Franklin, J., & Miflin, B.J. (1983) Heredity 50, 179-189.
20. Kreis, M., Shewry, P.R., Forde, B.G., Rahman, S., Bahramian, M.B., & Miflin, B.J. (1983) Submitted to Theor. Appl. Genet.
21. Payne, P.I., Law, C.N., & Mudd, E.E. (1980) Theor. Appl. Genet. 58, 113-120.
22. Hamauzu, Z., Kamazuka, Y., Kamazawa, H., & Yonezawa, D. (1975) Agric. Biol. Chem. 39, 1527-1531.
23. Holt, L.M., Astin, R., & Payne, P.I. (1981) Theor. Appl. Genet. 60, 237-243.
24. Brown, J.W.S., Law, C.N., Worland, A.J., & Flavell, R.B. (1981) Theor. Appl. Genet. 59, 361-371.
25. Shewry, P.R., Field, J.M., Kirkman, M.A., Faulks, A.J., & Miflin, B. J. (1980) J. Exp. Bot. 31, 393-407.
26. Ewart, J.A.D. (1981) J. Sci. Food Agric. 32, 572-578.
27. Forde, B.G., Kreis, M., Bahramian, M.B., Matthews, J., Miflin, B.J., Thompson, R.D., Bartels, D., & Flavell, R.B. (1981) Nucl. Acids

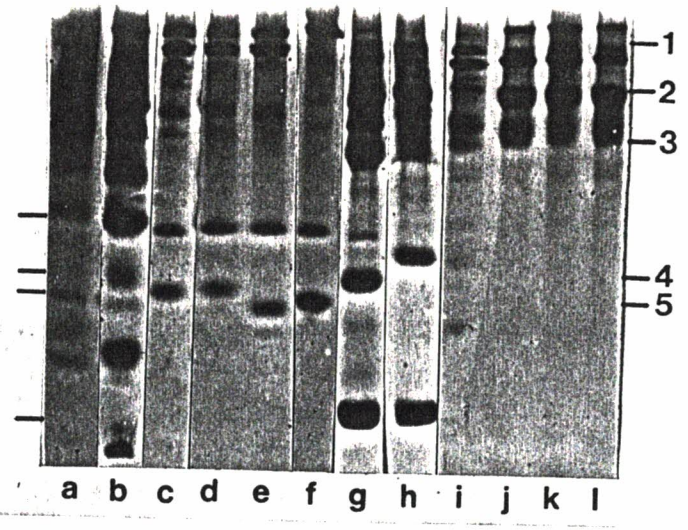
B.J. Manuscript [illegible]

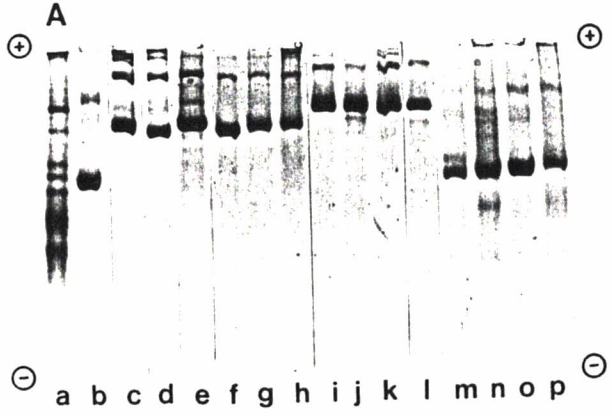


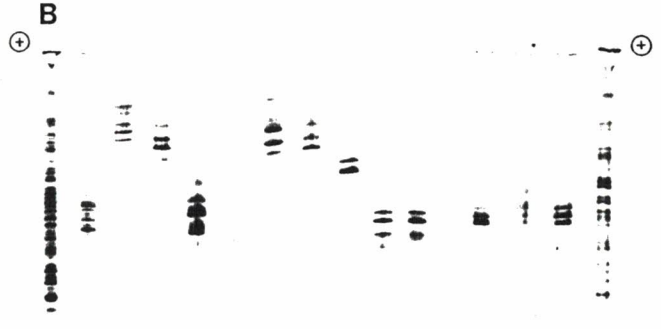
B




A
B
2
12
2
12
a b c d e f g h i j k l m n o p q r s t u v w x y z
a b c d


1
2
3
4
5
a b c d e f g h i j k l


A
a b c d e f g h i j k l m n o p


B


a b c d e f g h i j k l m n o

Table 1 *Amino acid compositions of the purified HMW subunits*

	1A 1 Copain	1A 1 Sentry	1A 2* Cheyenne	1Bx 6 Brigand	1Bx 7 Cheyenne	1Bx 7 Sentry	1Bx 7 Copain	1Dx 2 Copain	1Dx 2 Highbury	1Dx 2 Brigand	1Dx 3 Sentry	1Dx 5 Cheyenne	1Dy 10 Cheyenne	1Dy 12 Highbury	1Dy 12 Copain
Asx[a]	1.0	1.0	0.9	0.8	0.6	0.7	1.1	0.6	0.6	0.6	0.6	0.6	0.8	1.3	1.4
Thr	2.8	2.9	3.0	3.3	3.6	3.4	3.2	2.8	2.9	2.9	2.9	3.0	3.4	3.1	3.1
Ser	6.4	6.2	7.0	8.2	8.5	8.3	7.2	5.8	5.7	5.7	5.9	6.1	5.9	6.5	6.1
Glx[b]	36.7	37.4	39.2	36.3	36.8	37.7	35.7	37.7	35.6	35.6	37.4	38.6	33.7	35.4	34.4
Pro	16.8	16.2	13.8	13.1	12.6	13.0	15.4	16.3	15.5	15.6	15.5	13.5	14.0	14.4	14.9
Gly	15.6	13.7	16.0	18.0	17.8	17.2	18.6	18.8	20.4	20.8	19.0	17.9	17.7	14.2	13.7
Ala	2.3	2.6	2.4	3.1	3.2	3.2	3.1	2.8	3.1	2.9	2.9	3.2	3.2	3.5	3.6
Cys[c]	0.6	1.0	0.7	0.7	0.5	0.5	0.5	0.5	0.6	0.6	0.4	0.6	1.5	1.2	1.4
Val	2.2	2.5	1.7	1.5	1.5	1.6	1.6	1.7	1.5	1.6	1.7	1.8	2.8	2.6	3.3
Met	t	0.2	0.1	0.1	t	t	t	t	t	0.1	t	t	0.4	0.2	0.2
Ile	1.3	1.5	0.9	0.8	0.7	0.9	0.8	0.6	0.6	0.5	0.5	0.6	1.0	2.0	2.1
Leu	5.1	5.4	5.0	3.5	3.0	2.9	2.9	4.3	4.6	4.5	4.5	4.7	4.4	4.6	5.2
Tyr	4.7	4.7	5.3	6.4	7.2	6.6	5.8	5.5	5.7	5.8	5.8	6.0	5.0	4.3	3.5
Phe	1.0	1.3	0.6	0.7	0.5	0.7	0.6	0.5	0.5	0.5	0.3	0.4	0.8	1.9	1.8
His	0.6	0.7	0.6	0.5	0.5	0.6	0.6	0.4	0.5	0.5	0.5	0.5	1.9	1.7	1.9
Lys	0.6	0.7	0.7	0.8	0.7	0.6	0.8	0.7	1.1	0.9	0.7	0.9	1.3	1.1	1.2
Arg	2.1	2.1	2.2	2.2	2.3	2.1	2.1	1.2	1.0	0.9	1.2	1.5	2.2	2.0	2.2

[a] aspartate + asparagine

[b] glutamate + glutamine

[c] determined as S-β-pyridylethylcysteine

Results are the mean of duplicate hydrolyses and determinations.

Table 2 *N-terminal amino acid sequences of the purified HMW subunits*

The sequences are aligned to maximise homology, resulting in gaps in these for the 1A1, 1B7, 1D10, 1D12 and *T. monococcum* components. The boxes indicate regions of sequence identity.

Genome control	Subunit[a] No.	Species cultivar[b]	Residue Number[c] 1 5 10 15 20 25 30 35
D	2	(Highbury (Copain (Brigand	E G E A S E Q L Q C E R E L Q E L Q E R E L K A C Q Q V P D (E) Q L Z D
D	5	Cheyenne	E G E A S E Q L Q C E X[e] E L Q E L Q E X E L K A C Q (Q) V X D (E) (Q) L E D
D	10	Cheyenne	E G E A S R Q L Q C E R E L Q[d]E[d] X X L (E) A C
D	12	Copain	E G E A S R Q L Q C E R E L Q[d]E[d] X Q L (K) A C (Q) (Q) V
D		*Aegilops squarrosa*	E G E A S E Q L Q C E R E L Q E L Q E X E L K A C
A	1	Copain	E G E A S G Q L Q C E X[e] E L Q E G/P X L K A (C) Q Q V
A		*Triticum monococcum*	E G E A S G Q L Q C E R E L E E X X L K A C
B	7	Copain	E G E A S G Q L Q C E X[e] E L E A (C) Q Q V X D Q Q L

[a] nomenclature of Payne et al. [10]; [b] cultivars of *Triticum aestivum*; [c] notation follows standard single letter abbreviations: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; G, glycine; K, lysine; L, leucine; P, proline, Q, glutamine; R, arginine; S, serine; Z, glutamine or glutamate; X, unidentified. [d] these residues could correspond to positions 15 and 16 or positions 18 and 19. [e] these residues probably correspond to arginine, which is often difficult to identify positively.

Residues in parentheses are tentative identifications only