

Rothamsted Repository Download

A - Papers appearing in refereed journals

Forde, J., Malpica, J-M., Halford, N. G., Shewry, P. R., Anderson, O. D., Greene, F. C. and Miflin, B. J. 1985. The nucleotide sequence of a HMW glutenin subunit gene located on chromosome 1A of wheat (*Triticum aestivum* L.). *Nucleic Acids Research*. 13 (19), pp. 6817-7339.

The publisher's version can be accessed at:

- <https://dx.doi.org/10.1093/nar/13.19.6817>

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/8q129>.

© 11 October 1985, Oxford University Press (OUP).

The nucleotide sequence of a HMW glutenin subunit gene located on chromosome 1A of wheat (*Triticum aestivum* L.)

J. Forde^{1*}, J.-M. Malpica^{1,2}, N. G. Halford¹, P. R. Shewry¹, O. D. Anderson³, F. C. Greene³ and B. J. Mifflin¹

¹Rothamsted Experimental Station, Harpenden AL5 2JQ, UK, and ³USDA Western Regional Research Centre, 800 Buchanan Street, Albany CA 94710, USA

Received 23 August 1985; Accepted 12 September 1985

ABSTRACT

A cloned 8.2 kb *Eco*RI fragment has been isolated from a genomic library of DNA derived from *Triticum aestivum* L. cv. Cheyenne. This fragment contains sequences related to the high molecular weight (HMW) subunits of glutenin, proteins considered to be important in determining the elastic properties of gluten. The cloned HMW subunit gene appears to be derived from chromosome 1A. The nucleotide sequence of this gene has provided new information on the structure and evolution of the HMW subunits. However, hybrid-selection translation experiments suggest that this gene is silent.

INTRODUCTION

The high molecular weight subunits of glutenin (HMW subunits) are one of three sub-groups of prolamin storage proteins found in the seeds of barley, wheat and rye and are considered to be important in imparting elastomeric properties to gluten, the visco-elastic protein component of dough (see reviews 1-3). They are encoded by loci (Glu-1) on chromosome 1 of each of the three genomes (A, B, and D) of hexaploid bread wheat (*Triticum aestivum* L.) (3). The three homoeologous loci encode different numbers of HMW subunits in different wheat cultivars: in general Glu-A1 specifies either zero or one, Glu-B1 either one or two and Glu-D1 two major subunit proteins (4-6). The genes in loci expressing more than one subunit have been shown to be closely linked (4-7). Despite this observed variation in expression at the protein level, each genome appears to contain the same number of genes when analysed by hybridization of a HMW subunit cDNA sequence to restriction fragments of genomic DNA (8). It is not known why the Glu-A1 genes are so often silent, but a survey of 28 sets of homoeologous loci has shown that the proportion of silent genes in the A genome is almost double that in the B genome and three times that in the D genome (9).

The variation in the HMW subunit compositions of different cultivars is

associated with variation in bread-making quality (3) and there is considerable interest in determining how these proteins influence the properties of gluten. Partial sequences of the proteins have been derived from direct N-terminal amino-acid sequencing (10) and by deduction from cDNA sequences (8,11). Computer prediction of secondary structure based on these sequences, which is supported by circular dichroism spectroscopy, suggests that a considerable proportion of the molecule consists of β -turns (12,13). It has been suggested that the repetitive β -turns in the central domain form an elastic β -spiral, similar to that proposed for elastin (14), and that these elastic monomers are assembled into polymers by intermolecular disulphide bonds between the cysteine residues at the N- and C-termini (13). Comparison of the amino acid sequences of the non-repetitive portions of the HMW subunits with those of other cereal seed storage proteins, and with sequences of proteins present in dicotyledonous seeds, has provided information on the evolutionary relationships of seed proteins (15). However the conclusions that can be drawn, concerning both the evolution of the HMW subunits and their complete structure, are limited by the incompleteness of the sequence data: for example the full extent of the repeat region and the positions of all the cysteine residues are unknown. We have therefore isolated genomic clones related to the HMW subunits and we report here the entire nucleotide sequence of a gene from chromosome 1A. This particular gene was chosen because of the association of the chromosome 1A encoded subunit in Cheyenne (subunit 2*) with good bread-making quality (3), and also because some understanding of the lack of expression of genes located in the A genome might be obtained.

METHODS

Plant Material

Aneuploid seeds of the cultivar Cheyenne, monosomic for chromosomes 1A, 1B and 1D and nullisomic for chromosome 1A were kindly provided by Dr. Rosalind Morris, University of Nebraska. To check the composition of the HMW subunits in these lines, proteins were extracted from single seeds (16) and analysed by SDS-PAGE (17). The results were in agreement with the known chromosomal location of the genes for the individual HMW subunits of Cheyenne (18).

Screening of a Wheat Genomic Library

A genomic library of Triticum aestivum L. (cv. Cheyenne) was prepared as previously described (19). A portion (ca. 5×10^5 recombinant phage) of

the amplified library was screened (20) using the ^{32}P -labelled (21) cDNA insert from a HMW subunit clone pTaE-c256 (11). The probe was hybridized to plaque lifts on nitrocellulose filters for 24 h at 65°C in 5 x Denhardt's (22) solution, 5 x SSC, 0.1% SDS. (SSC is 0.15M NaCl, 0.15M trisodium citrate.) The filters were washed once in the same solution without the probe, three times with 2 x SSC, 0.1% SDS and once with 0.5 x SSC, 0.1% SDS before being air-dried and autoradiographed for 18 h. Of sixty hybridizing clones, six were plaque-purified and DNA was prepared by a plate-lysate method (23), digested with EcoRI and separated in 0.8% agarose. Phage DNA was co-electrophoresed with an EcoRI digest of Cheyenne genomic DNA prepared from either etiolated shoots (method of A. Viotti, personal communication) or endosperm tissue (24). A HindIII digest of phage lambda DNA was used to provide size markers.

Southern Blots and Hybridizations

After electrophoresis the DNA was treated with dilute HCl (25) and prepared for transfer to Pall Biodyne A according to the recommended procedures (Pall Filtration Process Ltd., Portsmouth, England). Pre-hybridization was performed at 65° C for 2 h in 5 ml of 5 x Denhardt's solution, 5 x SSPE, 0.2% SDS and 100 µg/ml denatured herring sperm DNA. (SSPE is 0.18 M NaCl, 0.01 M $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ pH 7.7, 1 mM EDTA.) Hybridization was for 48 h at 65° C in 5 ml of the same solution containing 0.5 µg of ^{32}P -labelled DNA. The filter was washed in 5 mM $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ pH 7, 1 mM EDTA, 0.2% SDS, firstly briefly, then 3 times for 30 min prior to autoradiography.

Nucleotide Sequence Analysis

The gene was sequenced using two approaches. Firstly, PstI restriction fragments and HindIII/HincII fragments were sub-cloned into M13mp18 or mp19 (26). Secondly, Bal31 deletions (27) were generated using CsCl purified plasmid DNA preparations (28) as described by Forde et al. (29).

The sequences of all the sub-clones were determined by the dideoxy-sequencing procedures of Sanger et al. (30). The sequences were assembled and analysed using the Staden programmes (31,32) operating on a VAX 11/750 computer.

Hybrid-selection and Translation of mRNA

20 µg of single-stranded recombinant M13 phage DNA were loaded onto nitrocellulose filters and hybridized, as described by Forde et al. (33), to 40 µg of wheat endosperm poly A⁺ RNA (prepared as in (34)). After

hybridization, the filters were washed at 60° C with 1 x SSC, 0.1% SDS (10 x 2 min washes), followed by 2 washes with 1 x SSC without SDS. The hybridization-selected RNAs were translated using the reticulocyte lysate (supplied by Amersham and used as instructed) with 5 μ Ci L-[4,5- 3 H]leucine (159 Ci/mmol) and 5 μ Ci L-[2,3,4,5- 3 H]proline (114 Ci/mmol) per 10 μ l incubation). The translation products were analysed by SDS-PAGE (17) and fluorography (35).

RESULTS

Chromosomal Location and Restriction Mapping of Cloned Genomic Fragments

A Southern blot of an *Eco*RI digest of wheat DNA (cv. Cheyenne) was

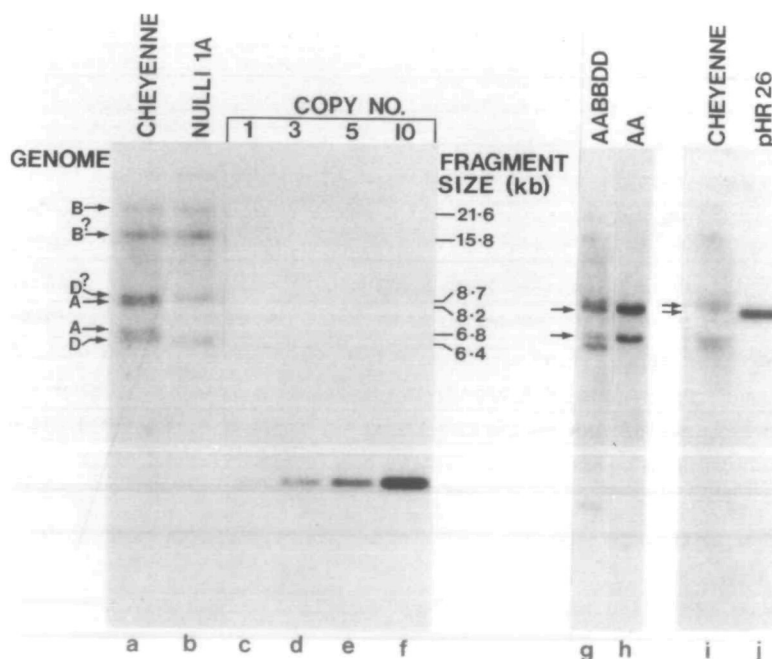


Fig. 1. Southern blot analyses of wheat genomic DNAs and a wheat genomic clone. The 32 P-labelled insert of the HMW subunit cDNA clone pTaE-c256 (11) was hybridized to the filters. The lanes contained *Eco*RI digested DNA from the following sources: lanes a and i, *T. aestivum* cv. Cheyenne, 20 μ g; lane b, an aneuploid line of Cheyenne nullisomic for chromosome 1A, 20 μ g; lanes c to f contained *Bam*HI-linearized pTaE-c256 representing 1, 3, 5 and 10 copies of the hybridizing sequence per haploid genome (20 μ g of hexaploid wheat DNA is equivalent (in moles) to 3.65×10^{-6} μ g of pTaE-c256). Lane g, *T. aestivum*, cv. Cheyenne, 15 μ g; lane h, *T. monococcum*, 5 μ g; lane j contained the *Eco*RI digested DNA from the plasmid pHR26 (see Fig. 2). The genomic origin assigned to the fragments is indicated.

hybridized to the labelled cDNA insert from a previously characterized HMW subunit clone, pTaE-c256 (11). Six fragments hybridized (Fig. 1, lane a) and these fragments were assigned to the different genomes on the basis of fragments present in nullisomic and monosomic lines of Cheyenne. Comparison of DNA from euploid Cheyenne with that from a line nullisomic for chromosome 1A shows that the 6.8 kb and 8.2 kb EcoRI fragments are absent from the latter (Fig. 1, compare lanes a and b). These fragments must therefore be derived from chromosome 1A. Fragments of similar size are present in the DNA from T. monococcum, which is a cultivated diploid containing the A-genome (Fig. 1, arrowed in lanes g and h). The genomic origin of the other EcoRI restriction fragments in Cheyenne has been tentatively assigned (Fig. 1, lane a) on the basis of fragments present in lines monosomic for chromosome 1B and 1D (results not shown). The comparative intensity of the hybridization to the reconstructions (Fig. 1, compare lane a with lanes c-f) suggests that there are 1-3 copies of HMW subunit gene per EcoRI fragment and thus a total of 2-6 copies per haploid genome.

A genomic library of DNA from the cultivar Cheyenne was screened for clones containing sequences related to HMW subunits. Three clones were identified, each containing an EcoRI fragment of 8.2 kb, and the insert from one of them was sub-cloned into pUC8 to give clone pHR26. The insert from pHR26 had a mobility similar to, but not identical with, the chromosome 1A derived 8.2 kb EcoRI fragment of Cheyenne (Fig. 1, arrowed bands in lanes i and j). This difference in mobility is probably due to the different amounts of DNA in the two lanes. Restriction analysis, Southern blotting and hybridization to pTaE-c256 suggested that the gene was entirely contained within a 3.5 kb SmaI/BamHI fragment (data not shown). This fragment was therefore sub-cloned into pUC8 to give pHSB26 (Fig. 2). Further detailed analysis was conducted on HindIII/HincII fragments sub-cloned into M13mp18 and mp19. The detailed restriction maps and the strategy used to sequence pHSB26 are shown in Fig. 2.

Nucleotide and Amino Acid Sequence

The complete nucleotide sequence and the deduced amino acid sequence of the coding portion of the gene are shown in Fig. 3. Analysis of the DNA sequence indicates that there are no intervening sequences present in the gene. An unusual feature of the coding sequence is a TAG termination codon starting at base 1832 (Fig. 3). The region containing this codon was found in three separate M13 sub-clones and was sequenced at least three times on each strand to exclude sequencing artefacts. We have presented the deduced

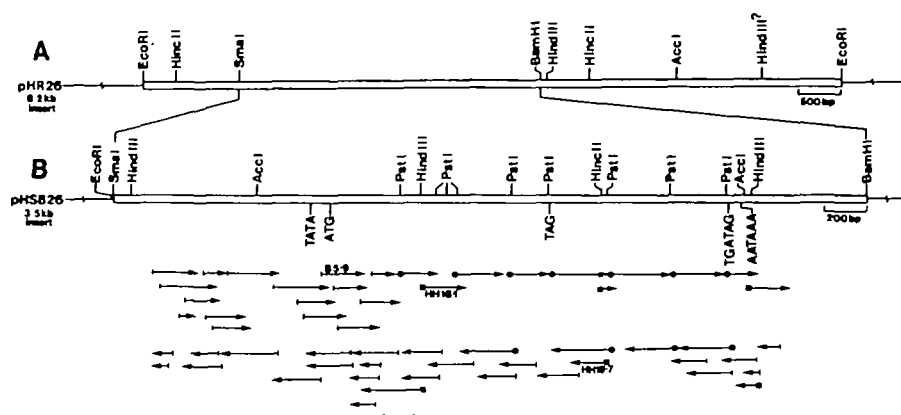


Fig. 2. Restriction map and sequencing strategy of a genomic clone related to the HMW subunits of glutenin.

(A) Experimentally determined restriction map of pHR26.

(B) Restriction map of a sub-clone of the 3.5 kb *SmaI/BamHI* fragment from pHR26. The *SmaI/BamHI* fragment was sub-cloned in pUC8 and designated pHSB26. Preliminary information on the position of the multiple *PstI* sites within this fragment was obtained by analysing *HindIII/HincII* fragments sub-cloned into M13mp18 and mp19 (26). The replicative forms were digested with *PstI* and the restriction fragments separated on polyacrylamide (6%) gels (63) prior to blotting (64) and hybridization to pTaE-c256. The regions of pHSB26 sequenced are indicated by arrows below the map. Each arrow represents a separate M13 sub-clone and is coded according to the method of generation: arrows with circles indicate *PstI* fragments; arrows with squares indicate *HindIII/HincII* fragments; arrows with vertical bars indicate Bal31 deletions made as follows. Clone pHSB26 was cut with *BamHI*, digested with Bal31, then with *EcoRI* and ligated to *BamHI/SalI* cut M13mp19. Deletions proceeding in the other direction were obtained using a *SmaI/PstI* sub-clone (ca. 1.5 kb) in pUC8. In this the plasmid was cut with *PstI*, digested with Bal31, cut with *HindIII* and ligated to a *HindIII/SmaI* cut vector M13mp18. Three sub-clones are labelled and mark those used for hybrid-selection of mRNA (see Fig. 4). Clone M13B5.9 extends from base 778 to base 1141 (see Fig. 3). Clones M13HH18.1 and M13HH19.7 represent a *HindIII/HincII* fragment in both orientations and span the sequence between bases 1242 and 2070.

The position of putative transcriptional control elements and of codons initiating and terminating translation of mRNA are indicated.

amino acid sequence beyond the TAG codon through to position 2620, which is then followed by adjacent TGA and TAG codons (Fig. 3). The predicted COOH-terminal amino acid sequence prior to these two stop codons (A S Q-COOH) agrees with the COOH-terminal sequence of a purified HMW subunit and that deduced from the nucleotide sequence of the cDNA clone, pTaE-c256 (11). The remainder of the deduced amino acid sequence between the internal TAG and

the TGA codon is very similar to that determined previously for pTaE-c256. In particular sequences related to the two consensus repeat motifs P G Q G Q Q and G Y Y P T S L Q Q (marked with broken and solid arrows respectively in Fig. 3 and encoded by nucleotides 1190 to 2494), were also observed in the cDNA clone (11). The sequence E G E A S starting at residue 22 is identical to that determined by direct sequencing of the NH₂-termini of several HMW subunits (10). This suggests that residues 1-21 form a signal peptide that is cleaved post-translationally, a result consistent with the synthesis of the HMW subunits on the rough endoplasmic reticulum (2,34). The predicted signal peptide has characteristics common to other signal sequences (36).

The nucleotide sequence shown in Fig. 3 extends for 814 bases 5' to the ATG codon and 295 bases 3' to the coding sequence. In the 5' flanking region the sequence TATAAAA is present between 85 and 91 bases upstream from the ATG codon; this agrees in position and sequence with the expectations for a 'TATA' box in plant and animal genes (37,38). Sequences similar to the 'CCAAT' box in animal genes (39) are also present; CCAT which begins at position 696, CAAAT at 620 and CCAAT at 582 (119-233 bases upstream from the ATG codon). The 5' flanking region also contains a series of three imperfect direct repeats of 20 nucleotides (Fig. 3, marked A) that have the consensus sequence GTGAGTCATAGCATACATAT. This sequence shows considerable homology to one half of the '-300 element' that is strongly conserved upstream from other prolamin genes (29). These repeats are flanked by a direct repeat of another sequence (marked B). Several inverted repeats are present and two are indicated in the figure. In the 3' flanking region, a consensus polyadenylation signal sequence (boxed in Fig. 3) is present 52 bases from the double stop codon and another sequence AATAAT is present a further 50 bases downstream, beginning at base 2728.

Hybrid-selection-translation

Normally the codon UAG functions as a stop codon. However recent work has shown that in at least one species of higher plant the two major cytoplasmic tRNAs^{TYR} can act as UAG suppressors (40). Thus, if the HMW subunit gene present in pHSB26 is expressed in vivo then the mRNA should direct the synthesis of a polypeptide of molecular weight 37,292 or, if the stop codon is read through, 65,229. Three sub-clones were used to hybrid-select mRNA from poly A⁺ RNA isolated from the membrane-bound polysomes of developing endosperms of the cultivars Cheyenne and Sentry. The released mRNAs were translated in vitro and the products separated by SDS-PAGE and

Downloaded from <https://academic.oup.com/nar/article-abstract/13/19/6817/2577579> by Periodicals Assistant - Library user on 20 May 2019

visualized by fluorography (Fig. 4). The HMW subunits present in the two cultivars have been numbered according to the classification of Payne et al. (5) and those encoded by the B and D genomes have been classified as x- and y-types (5,7; see Fig. 4). Where one chromosome 1A-encoded subunit occurs in cultivars, it is considered to be of the x-type (5). Clone M13HH18.1, which contains the repetitive portion of the gene, selected mRNAs directing the synthesis of all major HMW subunits (Fig. 4, lanes d and j); some low molecular weight polypeptides were also synthesized but there is no evidence of a major polypeptide of apparent M_r between 30,000 and 60,000. In contrast, clone M13B5.9, which encodes the non-repetitive N-terminal portion of the coding sequence and some of the 5' flanking sequence (Fig. 2), selected mRNA that specified one major polypeptide with a mobility equivalent to subunits 9 and 10 in Cheyenne and subunit 12 in Sentry (Fig. 4, lanes f and l). An accurate M_r for these subunits is not known: subunit 10 has been reported to be in the range of 84,000 to 118,000 by SDS-PAGE (41) and subunit 12 is estimated at 63,000 by sedimentation equilibrium ultracentrifugation (42). It is possible that factors which promote read-through of the amber codon are present in the reticulocyte lysate (40,43,44) and may also be present in the calf thymus tRNA preparation. However, it would be expected that a proportion of the polypeptides synthesized in vitro would terminate at the internal UAG. The apparent absence of a polypeptide of M_r 37,000 suggests either that the corresponding mRNA is absent or is not translated in vitro.

DISCUSSION

Gene Structure, Copy Number and Evolution of the HMW Subunits

The results indicate that the HMW subunits of glutenin are encoded by between 6-18 genes in the hexaploid bread wheat cultivar Cheyenne and suggest that these genes belong to a small multigene family as previously shown for the cultivar Chinese Spring (8). Whether there is significant variation in the copy number of genes for the HMW subunits between cultivars, as has been reported for the A-gliadin genes (45), is not known

Fig. 3. Nucleotide and derived protein sequence of clone pHSB26. Putative transcriptional control elements are boxed. Stop codons are marked by asterisks. In the 5' flanking regions arrows labelled A and B indicate two different direct repeats: the arrow labelled x marks an inverted repeat. Another sequence (arrow labelled y) is found inverted in the opposite 3' flanking region. In the coding region solid arrows mark a sequence based on the nonapeptide repeat G Y Y P T S L Q Q; broken arrows mark another sequence based on the hexapeptide repeat P G Q G Q Q.



Downloaded from <https://academic.oup.com/nar/article-abstract/13/1/9/6817237> by Periodicals Assistant - Library user on 20 May 2019

Downloaded from <https://academic.oup.com/nar/article-abstract/13/1/9/6817237> by Periodicals Assistant - Library user on 20 May 2019

fragment is absent in the cultivar Chinese Spring and replaced by a 16.5 kb fragment (8). Recent results have indicated that in this cultivar there has been an insertion of approximately 7 kb at a position equivalent to base 1718 in pHSB26 (see Fig. 3; N. P. Harberd, personal communication).

The apparent TAG stop codon at bases 1832-1834 (Fig. 3) occurs at the 3' of a 27 base repeat unit, in a position normally occupied by the glutamate codon CAG, and it is probably the result of a pyr-pyr substitution. In the absence of this substitution, the gene would have a 560 amino acid coding region devoid of introns. The nucleotide sequence of another HMW subunit gene (accompanying paper) also lacks introns, making the coding regions of this family of genes perhaps the longest so far reported for a higher eukaryote not to contain intervening sequences. Other cereal storage protein genes (38, 45) also lack introns, unlike most higher eukaryotic genes including those for legume seed storage proteins (46,47), and in contrast to Gilbert's suggestion (48) that genes with repeating structures are likely to contain introns.

The pattern of repeats shown in Fig. 3 suggests that the sequence organization of the HMW subunits has arisen during evolution by a series of intragenic duplications of two adjacent but unrelated ancestral units of 18 nucleotides and 27 nucleotides, in a manner similar to the model proposed for the evolution of the repetitive sequences of satellite DNA (49). However, it is not clear why the shorter repeat unit is so often tandemly repeated whereas the longer repeat unit is not. The origin of much of the sequence from amplification of two primordial sequences probably accounts for the strong bias in the codon usage for several of the amino acids. For example 85% of all proline codons are CCA. This bias is not evident in other plant or animal genes (see (50) for review).

It has been suggested that the prolamins, which are unique to the Gramineae, are of relatively recent evolutionary origin (51). The HMW subunit genes possess several characteristics that suggest they are the most recently evolved of the prolamin gene families. Firstly, their copy number is the lowest for the three groups of prolamins (8,24,52,53). Secondly, there appears to be less variation in the length of the restriction fragments related to the HMW prolamin genes than in those related to the S-rich or S-poor prolamins (8,53,54). Finally, the repeated sequences show a considerable degree of conservation - much greater than that seen in the S-rich prolamins (see (50) for review).

Though apparently of recent origin, new evidence (15) suggests that the

prolamins have evolved from another much older family of seed proteins. The proposed ancestral seed storage protein gene had three domains A, B and C which are still present in a super-family of salt-soluble seed proteins that includes wheat α -amylase inhibitor, barley trypsin inhibitor and the 2S storage proteins of certain dicotyledonous seeds (15,50). In the S-rich prolamins these three regions are fused with an N-terminal domain that consists largely of repeated sequences (50,55). In contrast, the repeats in the HMW subunits (which are different from those in the S-rich prolamins) are between the A and C regions. The B region, to date, had not been found. Analysis of the complete sequence in Fig. 3 by the DIAGON computer program using a graphic matrix (56), has indicated that the B region is present in the HMW subunit gene between residues 62 and 98 (50) (encoded by nucleotides 998 to 1108).

Is this A-genome gene expressed?

Termination codons which reduce the length of the open reading frame have been found in genes of other multigene families (57-60). Their presence does not mean the gene is not expressed; in the maize storage protein genes, nonsense mutations result in truncated zein polypeptides so that sequences hybridizing to members of the heavy chain class but encoding polypeptides of the light chain class are present (58). The results of the translation of hybrid-selected mRNAs (Fig. 4) suggest that the HMW subunit gene present in pHR26 does not produce a truncated polypeptide. The selection of a mRNA giving a single major polypeptide of equivalent mobility to subunit 12 (which in its mature form has an M_r of 63,000 by sedimentation equilibrium ultracentrifugation (42)) could be interpreted in terms of the synthesis of the 65,229 molecular weight polypeptide predicted by the longer coding sequence. However, in the cultivar Cheyenne, there is no evidence for an in vivo polypeptide of this mobility and encoded by chromosome 1A. Two reported analyses (18,61), and our own unpublished data using aneuploid lines of Cheyenne, fail to show any major chromosome 1A-encoded polypeptide of a mobility equivalent to that of subunits 9 and 10. It is possible that a minor band may be present but obscured by subunits 9 and 10 but this has not been shown. It would also be expected that if a transcript from this gene was selected by clone M1385.9 (Fig. 2), then the translation of a proportion of it would terminate at the internal UAG. Although the rabbit reticulocyte in vitro protein synthesis system can read through an internal UAG in TMV RNA, a major proportion of the polypeptides formed terminate at that codon (40,43). A more likely

explanation for our results is that clone M13B5.9 preferentially selects mRNA for subunit 10, which is encoded by a 10y allele (5,18): the nucleotide sequence of M13B5.9 is 92% homologous to the same region of a putative 10y gene from Chinese Spring (accompanying paper) and in particular a region of 120 nucleotides shows 95% homology. Furthermore, the amino acid composition predicted by pHS826 is much closer to that determined (10) for subunit 10 than that for 2*, particularly in the number of basic residues. This suggests that clone M13B5.9 may be useful as a differential probe to identify future genomic clones.

In conclusion, it is unlikely that the gene we have sequenced produces a functional mRNA and it is probably a silent gene from the A-genome encoding a y-type HMW subunit. The reasons for the apparent silence are not clear. It is possible that the gene is transcribed: Goeddel *et al.* (62) have described an interferon pseudogene, containing termination codons and a frameshift mutation, which is still transcribed. Fig. 3 shows that the 5' and 3' flanking regions contain the major known consensus sequences. However, only part of the conserved sequence of the putative regulatory '-300 element' (29) is present. The 5' flanking sequence of the 10 encoded HMW subunit gene from cv. Chinese Spring (accompanying paper) does not extend far enough to enable us to compare this region of the two genes. However, over their region of overlap, the two sequences are 84% homologous, and comparison using DIAGON (56) analysis (not shown) reveals no major insertions or deletions. Only one region of 50 bp in both sequences, extending from -415 to -376 relative to the ATG codons, is highly divergent.

Protein Structure

Irrespective of whether or not it is expressed, the coding sequence of pHS826 may be expected to be very similar to those of other, expressed HMW subunit genes. The sequence suggests that there are three domains within the HMW subunits: an N-terminal domains of about 100 amino acids, a repetitive domain of over 400 amino acids and a C-terminal domain of about 40 residues. A number of significant observations can be made. Firstly, the nonapeptide and hexapeptide repeats make up the majority of the molecule (75% of the final protein). Based on computer predictions of partial cDNA sequences (12,13) it is likely this repetitive portion is in the form of β -turns. Secondly, the sequence confirms the predicted presence of cysteine residues at the N- and C-termini and the absence of any in the centre of the molecule (13,42). The sequence predicts 6 cysteine residues, five within

the first 60 amino acids from the N-terminus and one located 13 residues from the C-terminus. Thirdly, as previously indicated by short amino- and carboxyl-terminal sequences (10,11), the charged residues are concentrated in these regions. Thus of the 62 charged residues, a third (mostly acidic) are in the N-terminal domain and a further 8 (mostly basic) are in the C-terminal region. Each of these three characteristics may be expected to be relevant to the behaviour of the HMW subunits in gluten and to support a recent hypothesis for their elastomeric properties (12,13). This is based upon the formation of glutenin polymers in which the individual HMW subunits (containing a high proportion of their backbone in the form of an elastic β -spiral) are linked head-to-tail by disulphide bonds between their N- and C-terminal cysteines.

ACKNOWLEDGEMENTS

The authors would like to thank Mrs. B. Buxton, Miss A. Heyworth and Mr. J. Franklin for their excellent technical assistance, and Dr. B. G. Forde for many helpful discussions. We thank Drs R. D. Thompson, D. Bartels and N. P. Harberd for communicating their results prior to publication. Finally we gratefully acknowledge the financial support provided by NATO Grant 324/82 (to P.R.S.), the AFRC New Initiatives Programme (to N.G.H.), and the Juan March Foundation (to J-M.M.).

²Present address: Dept. de Genetica, I.N.I.A. Madrid 35, Spain

*To whom correspondence should be addressed

REFERENCES

1. Wall, J.S (1979) in *Recent Advances in the Biochemistry of Cereals*, Laidman, D.L. and Wyn Jones, R.G. Eds. pp.275-311, Academic Press, London.
2. Mifflin, B.J., Field, J.M. and Shewry, P.R. (1983) in *Seed Proteins*, Daussant, J., Mosse, J. and Vaughan, J. Eds. pp.255-319, Academic Press, London.
3. Payne, P.I., Holt, L.M., Jackson, E.A. and Law, C.N. (1984) *Phil. Trans. R. Soc. Lond. B* 304, 359-371.
4. Lawrence, G.J. and Shepherd, K.W. (1980) *Aust. J. Biol. Sci.* 33, 221-233.
5. Payne, P.I., Holt, L.M. and Law, C.N. (1981) *Theor. Appl. Genet.* 60, 229-236.
6. Galili, G. and Feldman, M. (1983) *Theor. Appl. Genet.* 66, 77-86.
7. Payne, P.I., Holt, L.M., Thompson, R.D., Bartels, D., Harberd, N.P., Harris, P.A. and Law, C.N. (1983) *Proc. Sixth Int. Wheat Genetics Symp.* pp.827-834, Kyoto.
8. Thompson, R.D., Bartels, D., Harberd, N.P. and Flavell, R.B. (1983) *Theor. Appl. Genet.* 67, 87-96.
9. Garcia-Olmedo, F., Carbonero, P., Aragoncillo, A. and Salcedo, G. (1978) *Experientia* 34, 332-333.

- 10 Shewry, P.R., Field, J.M., Faulks, A.J., Parmar, S., Mifflin, B.J., Dietler, M.D., Lew, E.J.-L. and Kasarda, D.D. (1984) *Biochim. Biophys. Acta* 788, 23-34.
- 11 Forde, J., Forde, B.G., Fry, R.P., Kreis, M., Shewry, P.R. and Mifflin, B.J. (1983) *FEBS Letts.* 162, 360-366.
- 12 Tatham, A.S., Shewry, P.R. and Mifflin, B.J. (1984) *FEBS Letts.* 177, 205-208.
- 13 Tatham, A.S., Mifflin, B.J. and Shewry, P.R. (1985) *Cereal Chem.* In press.
- 14 Venkatachalam, C.M. and Urry, D.W. (1981) *Macromol.* 14, 1225-1229.
- 15 Kreis, M., Forde, B.G., Rahman, S., Mifflin, B.J. and Shewry, P.R. (1985) *J. Mol. Biol.* 183, 499-502.
- 16 Payne, P.I., Corfield, K.G. and Blackman, J.A. (1979) *Theor. Appl. Genet.* 55, 153-159.
- 17 Laemmli, U.K. (1970) *Nature* 227, 681-685.
- 18 Payne, P.I., Law, C.N. and Mudd, E.E. (1980) *Theor. Appl. Genet.* 58, 113-120.
- 19 Anderson, O.D., Litts, J.C. and Green, F.C. (1985) In preparation.
- 20 Benton, W.D. and Davis, R.W. (1977) *Science* 196, 180-182.
- 21 Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
- 22 Denhardt, D.T. (1966) *Biochem. Biophys. Res. Commun.* 23, 641-646.
- 23 Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Cold Spring Harbour Laboratory Manual*.
- 24 Kreis, M., Shewry, P.R., Forde, B.G., Rahman, S. and Mifflin, B.J. (1983) *Cell* 34, 161-167.
- 25 Wahl, G.M., Stern, M. and Stark, G.R. (1979) *Proc. Natl. Acad. Sci. USA* 76, 3683-3687.
- 26 Norrander, J., Kempe, T. and Messing, J. (1983) *Gene* 26, 101-106.
- 27 Guo, L.-H., Yang, R.C.A. and Wu, R. (1983) *Nucl. Acids Res.* 11, 5521-5540.
- 28 Clewell, D.B. (1972) *J. Bacteriol.* 110, 667-676.
- 29 Forde, B.G., Heyworth, A., Pywell, J. and Kreis, M. (1985) Manuscript submitted.
- 30 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
- 31 Staden, R. (1982) *Nucl. Acids Res.* 10, 4731-4751.
- 32 Staden, R. (1984) *Nucl. Acids Res.* 12, 505-519.
- 33 Forde, B.G., Kreis, M., Bahramian, M.B., Matthews, J.A., Mifflin, B.J., Thompson, R.D., Bartels, D. and Flavell, R.B. (1981) *Nucl. Acids Res.* 9, 6689-6707.
- 34 Forde, J. and Mifflin, B.J. (1983) *Planta* 157, 567-576.
- 35 Chamberlain, J.P. (1979) *Anal. Biochem.* 98, 132-135.
- 36 Watson, M.E.E. (1984) *Nucl. Acids Res.* 12, 5145-5164.
- 37 McKnight, S.L. and Kingsbury, R. (1982) *Science* 217, 316-324.
- 38 Messing, J., Geraghty, D., Heidecker, G., Hu, N.-T., Kridl, J. and Rubenstein, I. (1983) In *Genetic Engineering of Plants: an agricultural perspective*, Kosuge, T., Meredith, C.P. and Hollaender, A. Eds., pp. 211-227, Plenum Press, New York.
- 39 Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C., Proudfoot, N.J. (1980) *Cell* 21, 653-668.
- 40 Beier, H., Barciszewska, M., Krupp, G., Mitnacht, R. and Gross, H.J. (1984) *EMBO J.* 3, 351-356.
- 41 Bunce, N.A.C., White, R.P. and Shewry, P.R. (1985) *J. Cereal Sci.* 3, 131-142.

- 42 Field, J.M., Shewry, P.R., Mifflin, B.J. and March, J.F. (1982) *Theor. Appl. Genet.* 62, 329-336.
- 43 Pelham, H.R.B. (1978) *Nature* 272, 469-471.
- 44 Beier, H., Barciszewska, M. and Sickinger, H.-D. (1984) *EMBO J.* 3, 1091-1096.
- 45 Anderson, O.D., Litts, J.C., Gautier, M.-F. and Greene, F.C. (1984) *Nucl. Acids Res.* 12, 8129-8144.
- 46 Slightom, J.L., Sun, S.M. and Hall, T.C. (1983) *Proc. Natl. Acad. Sci. USA* 80, 1897-1901.
- 47 Schuler, M.A., Schmitt, E.S. and Beachy, R.N. (1982) *Nucl. Acids Res.* 10, 8225-8245.
- 48 Gilbert, W. (1985) *Science* 228, 823-824.
- 49 Flavell, R.B. (1980) *Ann. Rev. Plant Physiol.* 31, 569-596.
- 50 Kreis, M., Shewry, P.R., Forde, B.G., Forde, J. and Mifflin, B.J. (1985) in *Oxford Surveys of Plant Molecular and Cell Biology*, Mifflin, B.J. Ed., Vol. 2, pp.253-317, Oxford University Press, Oxford.
- 51 Kasarda, D.D., Autran, J.C., Lew, E.J.-L., Mimmo, C.C. and Shewry, P.R. (1983) *Biochim. Biophys. Acta* 747, 138-150.
- 52 Harberd, N.P., Bartels, D. and Thompson, R.D. (1985) *Mol. Gen. Genet.* In press.
- 53 Shewry, P.R., Bunce, N.A.C., Kreis, M. and Forde, B.G. (1985) *Biochem. Genet.* 23, 391-404.
- 54 Bunce, N.A.C., Forde, B.G., Kreis, M. and Shewry, P.R. (1985) *Seed Sci. Technol.* In press.
- 55 Forde, B.G., Kreis, M., Williamson, M.S., Fry, R.P., Pywell, J., Shewry, P.R., Bunce, N. and Mifflin, B.J. (1985) *EMBO J.* 4, 9-15.
- 56 Staden, R. (1982) *Nucl. Acids Res.* 10, 2951-2961.
- 57 Proudfoot, N.J., Gil, A. and Maniatis, T. (1982) *Cell* 31, 553-563.
- 58 Spena, A., Viotti, A. and Pirrotta, V. (1982) *EMBO J.* 1, 1589-1594.
- 59 Kridl, J.C., Vieira, J., Rubenstein, I. and Messing, J. (1984) *Gene* 28, 113-118.
- 60 Kastern, W. (1985) *Gene* 34, 227-233.
- 61 Galili, G. and Feldman, M. (1983) *Theor. Appl. Genet.* 66, 77-86.
- 62 Goeddel, D.V., Leung, D.W., Dull, T.J., Gross, M., Lawn, R.M., McCandliss, R., Seeburg, P.H., Ullrich, A., Yelverton, E. and Gray, P.W. (1981) *Nature* 290, 20-26.
- 63 Peacock, A.C. and Dingman, C.W. (1967) *Biochem.* 6, 1818-1827.
- 64 Smith, G.E. and Summers, M.D. (1980) *Anal. Biochem.* 109, 123-129.