Volume 12 Number 21 1984 Nucleic Acids Research

Nucleic acid sequence and chromosome assignment of a wheat storage protein gene

Olin D.Anderson, James C.Litts, Marie-Francoise Gautier[1] and Frank C.Greene

Food Proteins Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Western Regional Research Center, Albany, CA 94710, USA

Received 30 May 1984; Revised and Accepted 28 September 1984

ABSTRACT

A cloned gliadin gene was isolated from a wheat genomic library, and 2.4 kb of its primary sequence determined. The gene, α-1Y, was found by Southern analysis to be located on chromosome 6A, and its derived amino acid sequence identifies it as a member of the A-gliadin subgroup of α-gliadins located on the short arm of that chromosome. α-1Y is apparently functional, and contains consensus TATA and CAAT boxes, and poly-adenylation signals. This gliadin gene has no introns, and its noncoding flanking regions contain several short repeats and inverted sequences. The gene is contained in a 6.2 kb EcoRI genomic fragment whose apparent copy number varies in different wheat cultivars.

INTRODUCTION

The protein nutritional quality and unique rheological properties (dough-forming abilities) of wheat flour are determined largely by its principal storage proteins, the gliadins and glutenins (1). The gliadins are monomeric proteins of 30,000 - 78,000 molecular weight, and are characterized by low electrostatic charge density, poor solubility in dilute salt solutions, and good solubility in alcohol:water mixtures. They comprise a multigene family which has evolved by gene duplication and divergence from ancestral genes (1). They have been historically assigned to α, β, γ, and ω classes based on electrophoretic mobility, and more than 40 gliadin components can be detected by two-dimensional gel electrophoresis (2,3). The complete α-gliadin class, and most components of the β class are coded at loci located on the short arms of group 6 chromosomes of wheat; the complete ω class and most components of the γ class are coded at loci on the short arms of group 1 chromosomes (3,4). There is close linkage among the genes at each locus, and in intervarietal crosses, they are inherited largely as nonrecombinant groups (3,4,5). The analysis of subfamily structure at each gliadin locus is incomplete because insufficient sequence and hybridization information is available, but at least one such

grouping has been recognized on the basis of the specific aggregation properties of its gene products, which are termed A-gliadins (1). This subfamily is coded at the 6A locus, has α-mobility, and based on two-dimensional gel electrophoresis, contains at least 7 members (6).

Gliadin genes are expressed in the seed endosperm, under developmental control, probably at the level of transcription (7). A full complement of gene products is detectable at 6-9 days after fertilization, suggesting that the genes are coordinately activated (7,8. Greene unpublished). Gliadin biosynthesis occurs in association with membranes (9), directed by long-lived mRNAs (7), and the presence of an N-terminal leader sequence has been confirmed (10,11). Sequence analysis of gliadin proteins and cloned gliadin cDNAs have yielded information on the coding regions of some members of this gene family (11,12), but no genomic sequences have been reported.

The evidence for close genetic linkage and coordinated expression of gliadin genes is consistent with a physical clustering in the wheat genome, and with the presence of similar control sequences in the genes. In order to investigate these facets of gene control further, we are pursuing a study of the gliadin loci in the wheat genome. The present report describes the isolation and structural analysis of a cloned gliadin gene coded at the 6A locus.

MATERIALS AND METHODS

Materials

Restriction enzymes were from Bethesda Research Labs, New England Biolabs and P-L Biochemicals. T4 ligase, DNA polymerase I, X-Gal, Protease K, and acrylamide were from Bethesda Research Labs. Nitrocellulose was from Schleicher & Schuell. Sequencing reaction mixtures, and DNA polymerase I Klenow fragment were from Bethesda Research Labs, and P-L Biochemicals. Hybridization primers and probe primers were from P-L Biochemicals. The X-ray film used was XAR-5 from Kodak. ^{32}P-dATP, dCTP, TTP and dGTP (>400 Ci/mmol), ^{35}S-dATP (>1000 Ci/mmol), and Gene-Screen Plus hybridization membrane were from New England Nuclear. Low-melting agarose was from FMC. Zeta-Probe membrane was from Bio-Rad.

Isolation of gliadin genomic clones

Gliadin genomic sequences were isolated from a wheat (Triticum aestivum, cultivar Yamhill) library (13) constructed in Charon 32 (14) using DH1 (15) as host. Similar clones have been isolated from a cultivar Cheyenne library constructed by us (unpublished) in the vector Sep6-Lac5 (E. Meyerowitz, unpublished). Screening of gliadin clones was according to the methods of

Benton and Davis (16). The probes for all library screenings were restriction fragments of the gliadin cDNA clone pTO-A10 (11). Plasmid subcloning of lambda inserts was accomplished by ligating an EcoRI digest of cloned DNA with EcoRI restricted RVIΔ7 DNA (described in (17)), or plasmid pUC8 (23). The Yamhill clone Yam-2 yielded the subclone pYAZ-28 (in RVIIΔ7), and the Cheyenne clone Chey-5 was the source of DNA for the subclone pChey-56 (in pUC8).

Analysis of gliadin clone YAM-2

M13 phage subcloning was performed by ligating fragments of the 6.2 kb (Figs 2 and 4) insert digested with four-base recognition restriction enzymes (Alu I, Hae III, Rsa I), with Sma I restricted vectors, or Sau 3A digested insert and Bam HI restricted vectors mp8-11 (18) to yield four sets of subclones. Coding region subclones were identified using a pTO-A10 probe. In some cases a sequenced clone was used to make a hybridization probe to isolate an overlapping sequence from a different subset of clones.

Sequencing reactions were by the dideoxy-procedure of Sanger et al. (18). Hybridization probe was prepared as described by Hu and Messing (19). The conditions for both reactions were those suggested by P-L Biochemicals and Bethesda Research Labs.

Blots were performed as described by Southern (20), using nitrocellulose or nylon membranes under the following conditions: blots containing genomic and clone DNA were prehybridized for 48 hours and 16 hours, respectively, at 68°C in 1 M NaCl, 50 mM Tris 7.5, 5 mM EDTA, 200 μg/ml sheared denatured salmon sperm DNA. Labelled gliadin DNA was added to fresh prehybridization buffer, and the blots incubated at 68°C for 24 hours (clone DNA) or 72 hours (genomic DNA). Blots were then washed once each at 68°C with 5 mM EDTA, 0.1% SDS plus the following: 2x SSC, 0.5x SSC, 0.1x SSC. Nick-translation of DNA fragments was according to Rigby et al. (21).

Stabilization energies of potential secondary structures were estimated according to the rules given by Tinoco et al. (22).

DNA Isolation

Single-stranded M13 DNA was prepared according to Messing and Vieira (23). M13 double-stranded DNA and plasmid DNA was isolated by the alkaline-SDS method of Birnboim and Doly (24) as described by Maniatis et al., (25). When necessary, the supercoiled DNA was further purified using CsCl equilibrium gradients or hydroxylapaptite.

Wheat nuclei were prepared by modifications of the procedure of Luthe and Quatrano (26) using ethidium bromide as suggested by Kislev and Rubinstein (27). DNA was isolated from nuclei by the Proteinase-K method

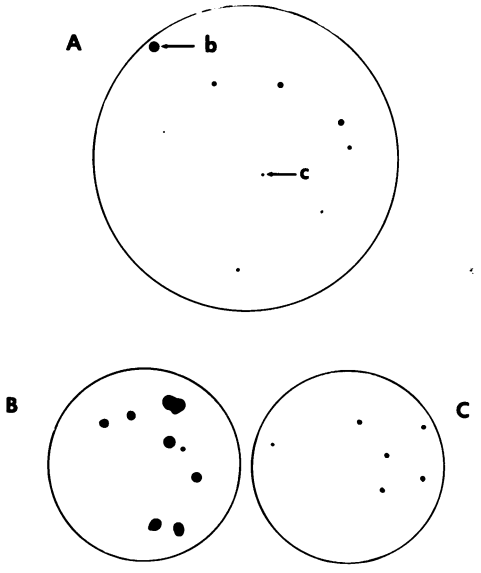

Figure 1. Screening of a wheat genomic library. The wheat lambda library was screened by the method of Benton and Davis (16). After the plaques were transferred to nitrocellulose filters, they were probed as described in Materials. Filters are to scale, and the plaques in B and C are the same size. A) 10,000 pfu of the total library on a 150 cm plate. Most filters contained only 1-3 detectable signals. B) 20 pfu of the 3rd plating of signal b on an 88 cm plate. C) 24 pfu of the 3rd plating of signal c on an 88 cm plate.

of Blin and Stafford (28). The isolation is described in more detail in Litts et al. (in preparation).

Specific DNA fragments were isolated from low-melting agarose as described by Weislander (29). DNA ligations, Cell transformations, lambda growth and lambda DNA isolation were all performed by the procedures described in Maniatis et al. (25).

RESULTS

In our initial screen, approximately 600,000 plaques from a wheat library (cultivar Yamhill) were probed with the labelled gliadin cDNA clone pTO-A10 (11). Figure 1A is an autoradiogram from a plate showing several positive clones displaying varying signal intensity. From 120 such plaques, 20 were carried through two additional cycles of purification to isolate single clones (Figure 1 B & C). These further cycles established that the different signal intensities were not due to plaque size, but likely

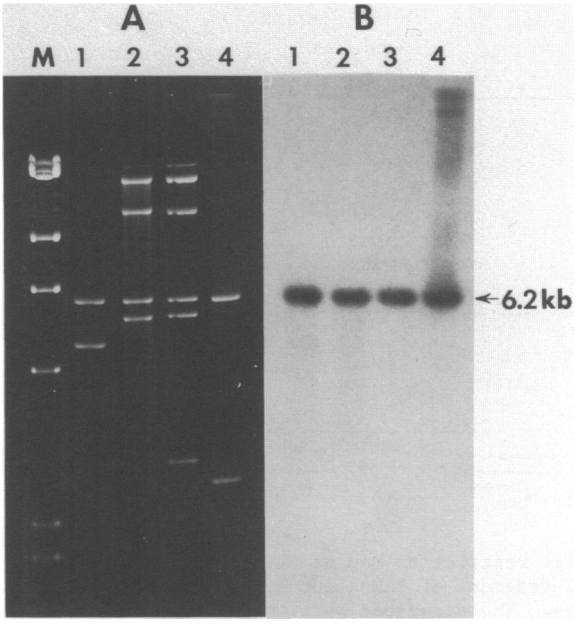

Figure 2. EcoRI restriction enzyme analysis of gliadin clones. DNA electrophoresed on a 1% agarose gel, stained with ethidium bromide (A) and probed with the [^{32}P]-labelled gliadin cDNA pTO-A10 (B). M: Hind III digest of lambda DNA. Lane 1, plasmid pYAZ-28 (6.2 kb fragment of YAM-2 subcloned into the plasmid RVIIΔ7); Lane 2, lambda YAM-2; Lane 3, lambda Chey-5; Lane 4, plasmid pCUR-56 (6.2 kb fragment of Chey-5 subcloned into plasmid pUCB).

due to different degrees of homology of each clone with the cDNA gliadin probe. This result would be expected since the gliadins are a multigene family of evolutionarily related, but distinct members (1,12).

From 12 genomic clones giving strong signals to the gliadin cDNA probe, one of the strongest, YAM-2, was chosen for further analysis. When YAM-2 DNA was isolated and subjected to EcoRl restriction, the wheat insert yielded fragments of 5.5 and 6.2 kb (Figure 2A, lane 2), clearly separated from the lambda arms of approximately 11 kb and 19 kb. Only the 6.2 kb fragment hybridized with the gliadin cDNA probe. This fragment was subcloned into plasmid RVIIΔ7 for further analysis (lane 1). A clone (Chey-5) isolated from the Cheyenne library is shown in lanes 3 & 4 for comparison.

A partial restriction map of the gliadin related 6.2 kb insert from YAM-2 is shown in Figure 3. The gliadin related sequence is approximately centered within the EcoRI fragment, between two Nco I sites 1 kb apart. The 6.2 kb

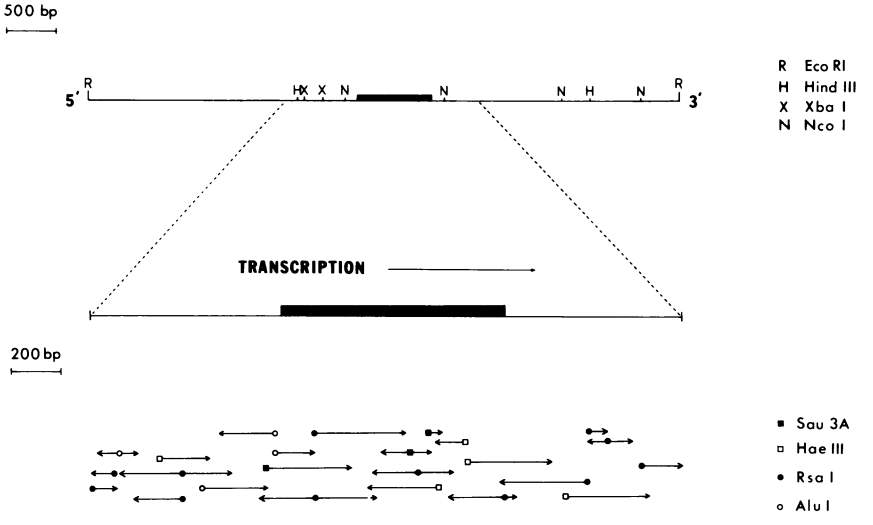

Figure 3. Partial restriction map of the YAM-2 6.2 kb insert. The map of the 6.2 kb EcoRI fragment of YAM-2 into pYAZ-28 was determined for the indicated enzymes. The sequenced portion is shown expanded. Arrows below the map indicate the specific sequences determined with M13 subclones.

insert was restricted with four-base recognition restriction enzymes and the resultant fragments subcloned into M13 as described in Materials and the central portion of the 6.2 kb fragment sequenced as shown in Figure 3. Overlapping clones were assembled into the final sequence given in Figure 4.

Position +1 of the coding sequence of the YAM-2 gliadin was assigned because it is the only potential initiator codon for an open reading frame. There is no indication of introns interrupting the coding region. The nucleotide sequence codes for a protein of 286 amino acids and a molecular weight of 32,912. The amino acid sequence derived from this sequence is 97% homologous to the A-gliadin protein amino acid sequence determined by Kasarda et al. (11). This confirms its identity as a member of the A-gliadin subfamily of gliadin storage protein genes. In addition, the YAM-2 sequence shares 93% and 96% nucleotide homology with the gliadin cDNA clones pGliA-42 and pTO-A10, respectively (11). A characteristic of these gliadins is the presence of two polyglutamine regions, of which the 5' polyglutamine region seems more conserved in length than the 3' one. In YAM-2, for example, the 5' polyglutamine has the sequence $(CAG)_9$-$(CAA)_9$, compared to $(CAG)_6$-$(CAA)_{12}$ in the gliadin cDNA clone pTO-A10. The nonrandom distribution of CAG and CAA may indicate mechanisms controlling codon usage within the region.

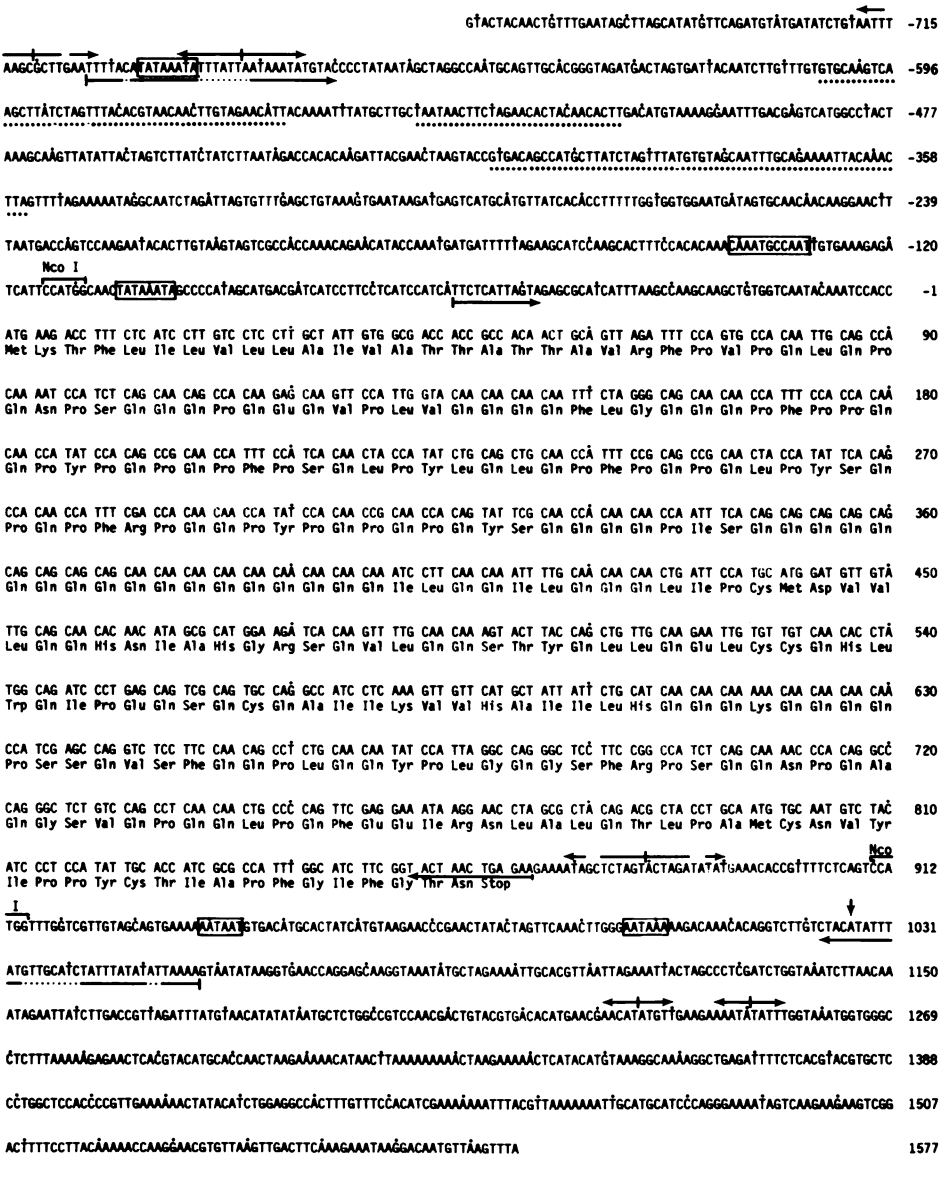

Figure 4. Sequence of a 2346 base region of the YAM-2 6.2 kb insert. The primary sequence is shown along with the translation of the open reading frame. Putative control elements are boxed. The dotted lines indicate 80% homologous repeated segments of the 5' flanking region. Overlined sequenced are direct inversions. Underlined sequences in the noncoding regions indicate inversions of sequences found in the opposite flanking region. An arrow points to the polyadenylation site of cDNA clones (11).

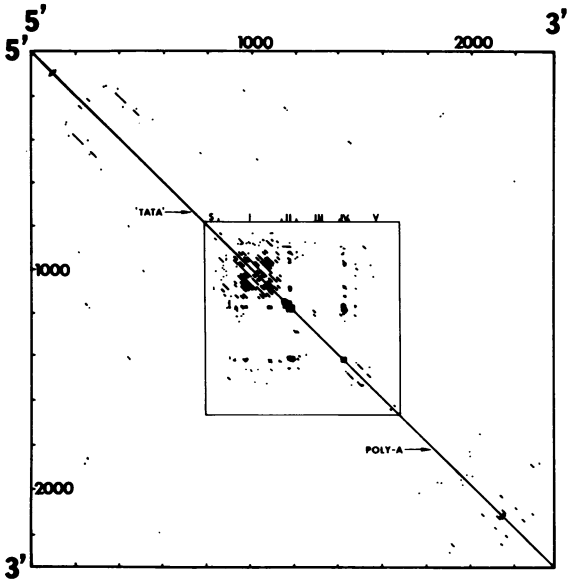

Figure 5. Homology matrix of the 2346 base sequence with itself. A homology matrix was plotted of the entire sequenced portion of YAM-2. A homology criterion of 14 bases out of 20 was used. The coding region is boxed and the domains of the A-gliadin protein (11) are labelled. The presumptive 'TATA' and polyadenylation sites are indicated.

The 3' noncoding region of most messengers contains a putatitive polyadenylatin signal related to AAUAAA (30). Two such sites seem to be common in plant genes reported thus far (31, 32). These sequences have been shown to be necessary for proper polyadenylatin of mRNAs (33). The 3' noncoding region of the YAM-2 sequence is 98% homologous to the 3' ends of two gliadin cDNAs reported by Kasarda et al. (11). The 3' region of all 3 sequences contain 2 potential polyadenylation signals, centered, in α-1Y at +941 (AAATAAT) and +998 (AATAAA).

The 5' end of the coding region of α-1Y is established by the potential nonsence codon at -70 followed in frame at +1 by the only start codon (ATG) allowing correct reading of the following gliadin sequence. Nearby, in the 5' upstream sequence, are several sequences related to presumptive control elements discussed by Breathnach and Chambon (34). A 'TATA' sequence of TATAAAT, matching the consensus sequence of TATA$\frac{A}{T}A\frac{A}{T}$ is found at position -104 (Figure 4 and Figure 7A). Thirty-seven bases further upstream from the 'TATA' at -141 is the sequence CAAATGCCAAT which contains two potential 'CAAT' like elements.

In order to examine the internal sequence homologies within YAM-2, the sequenced portion was analyzed via homology matrices. Homologies from 60-90% and windows from 20 to 50 bases were used, with the 70% homology at a 20 base window shown in Figure 5 showing the main, consistant features of the analysis. Within the coding sequence, the five domains of the A-gliadin primary sequence (11) and a signal sequence are discernible with the following characteristics: S. The leader sequence coding for a 20 amino acid signal peptide with little external homology. 1. A 300 base region with extensive internal and little external homology. 2. The first polyglutamine region. 3. A 200 base fragment with limited internal homologies. 4. The second poly-glutamine region. 5. A 200 bases 3'-terminal sequence with some internal homology in its 5' portion.

The matrix also points to several short homologies in the flanking regions. The 5' noncoding sequence contains a 300 base region from about -600 to -310 with several internal homologies, the highest of which is a 56 base repeat of 80% homology starting at -589 and -395. A third sequence, of 28 bases and starting at -539, shares as much as 82% homology with the first two sequences. In addition, there are two sequences of 31 bases sharing 77% homology (at -511 and -321). The 3' flanking region contained no significant external homologies, and short (10-15 bases) internal homologies mainly in its more distal sequence from the coding region.

Three potentially significant inverted repeat structures occur in the noncoding sequences (see Figure 4). The first is centered at -690. It is comprised of two contiguous nine base direct inversions, one with eight of nine bases matching, and the other of nine perfect matches. This region also has the potential to form a cruciform-like structure with a 3' sequence at +1023 as shown in Figure 7C. A second interesting sequence occurs at +1230. Here, within 26 bases, are two 5-base direct inversions, a ten base perfect repeat of the sequence following the coding terminator codon, and a nine base inversion of the sequence beginning at -745. Finally, at +852 is a 13 base sequence, including the termination codon, which is an imperfect inversion of the sequence at -59, which includes one of the two potential mRNA start sites. This pair is of interest because it involves two important locations, the beginning of transcription and the termination of translation. What functional significance this relationship has is at present unknown, but we note that several other plant genes have short inversions involving the termination codon and a sequence between the TATA box and the initiator codon (31,32,35,36) although the

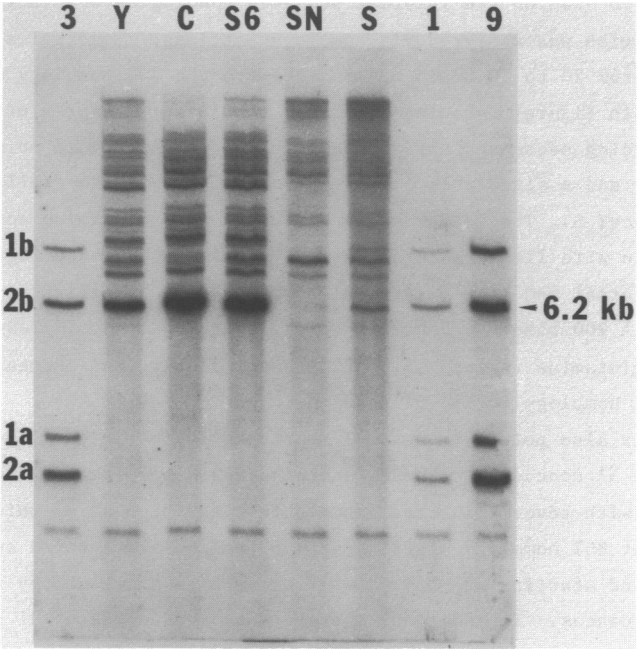

Figure 6. Southern analysis of wheat genomic DNA with a gliadin probe. Fifteen micrograms of total nuclear DNA of the indicated wheat cultivars was digested with EcoRI and electrophoresed on a 0.7% agarose gel. The gel was blotted and probed, as described, with the ^{32}P labelled 1.1 kb Nco I fragment of YAM-2 containing the entire coding region of the gene. Y, Yamhill, C, Cheyenne; S6, Chinese Spring with a Cheyenne 6A substitution; SN, Chinese Spring nulli-6A-tetra-6B; S, Chinese Spring. Control bands; 2a, 3.0 Hind III fragment of YAM-2 containing the entire coding region of the gliadin gene α-1Y; 2b, 6.2 kb EcoRI fragment of YAM-2; 1a and 1b, derived from the clone YAM-1.

degree of homology is not always as great as with the present sequence.

Southern blot analysis was employed in determinations of copy numbers and chromosomal locations of the gliadin genes. Blots of total Yamhill and Cheyenne DNA probed with the Nco I - Nco I coding region fragment of α-1Y revealed a series of hybridizing bands from an intense 6.2 kb band to fainter, higher molecular weight bands of up to 20 kb (Figure 6). Similar blot patterns were obtained using probes derived entirely from within the coding region of gliadin cDNA clone pTO-A10 (data not shown), indicating that the pattern represents gliadin-related gene fragments. In addition, these patterns have been consistently observed under digestion conditions in which both time and enzyme/DNA ratios were varied, indicating that they represent limit digestions. Yamhill (Y), Cheyenne (C) and Chinese Spring (S) all contain

A) TATA Box Region

```
Consensus
sequence (34)   GNGTATA A A NGNNG (9-17 bases) PNNNPAPPPP
                      T T

Gliadin         AACTATAAATAGCCC  (10 bases)  CGATCATCCT
sequence          -104
```

B) CAAT Box Region

Gene	Sequence	Reference
Gliadin	CAAAT-G-CCAAT	
Zein	CAAAAT---CGAAAAT	(32)
Zein	CAAAAT---CCAAAAT	(53)
Zein	CAAAAT---CCAAAAAAAAT	(41)
Zein	CAAAAT---CCAAAAAAAT	(51)
Leghemoglobin	CAT-A-CCAT	(54)
Phaseolin	CAAATCT-CCAT	(31)
Wheat Histone	CACT---CCAT	(36)
Soybean Lectin	GATAAA---CAACT	(38)
Soybean Actin	CAGGTT---CAAAACT	(35)

C)

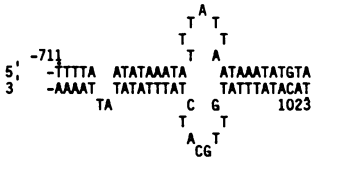

Figure 7. Specific sequences within the 2346 base fragment. The putative 'TATA' of the gliadin gene alpha-1Y is compared to the consensus sequence of Breathnach and Chambon (34). N; purine or pyrimidine: P; pyrimidine. B) The 'CAAT' sequence at -131 of gliadin gene α-1Y is compared to similar reported sequences in other plant genes. C) The secondary structure that could potentially form between the sequences at -711 and +1022.

the 6.2 kb band, though with different intensities, suggesting different copy numbers of 6.2 kb gliadin sequences. Copy number estimates were based on reconstructions using standard bands 2a & 2b derived from genomic clone YAM-2 mixed with a blank hybridizing background of sea urchin DNA. YAM-2 has been shown to be a member of the 6.2 kb gliadin gene group. Partial sequencing (unpublished) has established that YAM-1, which contains a gliadin gene within a 7.8 kb EcoRI fragment, belongs to a subfamily of gliadins closely related to, but distinct from the subfamily of YAM-2. The copy number estimates for the 6.2 kb band indicate 1-3 copies in Chinese Spring, 15-20 copies in Cheyenne, and an intermediate number in Yamhill. These estimates assume close homology with the gene α-1Y (isolated as clone YAM-2) whose coding region was used to probe the blot. EcoRI fragments of lower homology (such as YAM-1) would yield a lower apparent copy number. This appears to be the case with sequence at 6.2 kb in Chinese Spring nulli-6A-tetra-6B whose intensity is equivalent to about 0.1 copy per genome. When the

Cheyenne 6A chromosome is substituted into Chinese Spring (37) (Figure 3, lane S6), a series of gliadin hybridizing bands similar to those of Cheyenne, including the intense 6.2 kb band, appear over the Chinese Spring background. Based on these results, the 6.2 kb band, plus most of the higher molecular weight EcoRI gliadin fragments, are assigned to the 6A chromosome.

DISCUSSION

As part of our study of the developmental expression of the wheat endosperm genes, we are characterizing the structure and distribution of members of the gliadin multigene family. In the present report, we have described the isolation and determination of the primary structure of an A-gliadin gene, a member of the α-gliadin group. This gene is found within a 6.2 kb EcoRI fragment isolated from a wheat genomic library constructed from DNA of cultivar Yamhill (13). We designate this gene 'α-1Y' (first A-gliadin characterized from the α gliadin group in cultivar Yamhill). As far as can be determined from sequence data, α-1Y is a functional gliadin gene. A translation of the sequence yields a continuous read from initiator to terminator with no nonsense or premature termination codons. In addition, the flanking sequences possess consensus control sequences associated with functional genes (discussed below).

The Southern analysis shown in Figure 6 associates an intensely hybridizing 6.2 kb EcoRI band with those cultivars containing the 6A chromosome known to encode for the A-gliadin group of proteins. The restriction map and hybridization data for this clone indicates that the coding sequence of this gene is at least 2 kb distant from adjacent A-gliadin genes on one flank and 8 kb on the other.

The coding sequence of YAM-2 is similar, but not identical to the previously sequenced cDNA clones pGliA-42 and pTO-A10 (11). These three genes are similar enough to be placed in a related sub-family of the α-gliadins different from other gliadins such as the sequence reported for a cDNA clone by Bartels and Thompson (12). The α gliadins and the Bartels and Thompson clone do, however, share 70% homology in the 3' noncoding sequence between the stop codon and the polyadenylation site (Figure 8) consistent with their presumed divergence from a common ancestral gene. This region of the YAM-2 sequence has the potential for forming three stabilized secondary setm-loop structures, two of which start near the 3' border of putative polyadenylation/cleavage signals. The γ-gliadin sequence has the potential for forming two stabilized secondary stem-loop structures, 3' to polyadenylation signals in positins equivalent to those

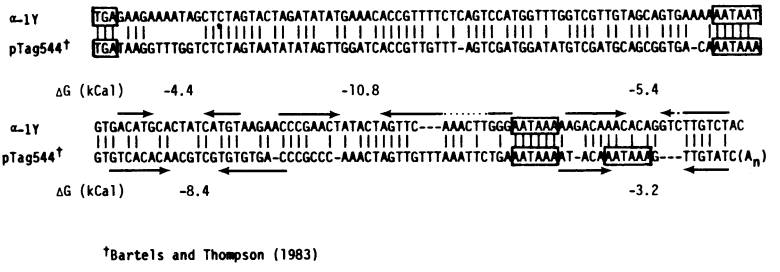

Figure 8. Homology between α-1Y and a γ-gliadin. A comparison of the 3' noncoding sequences of the γ-gliadin clone pTAG544 (12) and the A-gliadin genomic clone α-1Y. Several gaps in each sequence were introduced to allow maximal alignment. The sequences involved in potential secondary structures are indicated with arrows. Energies of formation for each potential structure are shown.

in YAM-2. The positions of these putative structures are apparently conserved, though the sequences themselves have diverged. The second putative polyadenylation signal of pTAG544 forms a portion of its distal secondary structure. Vodkin et al. (38) have observed a potential stem-loop structure near the putative polyadenylatin/cleavage signal of a soybean lectin gene, and Schuler et al. (39) have reported that regions of potential secondary structure occur on 5' and 3' sides of the putative polyadenylation/ cleavage site in soybean 7S storage protein genes. The significance of such alignments is not clear, but may be relevant to the suggestion by Montell et al. (40) that transcript cleavage signals may be complex ones, involving specific RNA secondary or tertiary structure in addition to the AAUAAA sequence.

The major mRNAs found in 20-25 day developing wheat seeds code for 32,000-34,000 dalton polypeptides. Okita and Greene (10) report isolating a 14 S RNA fraction which encodes for 2 size classes of gliadin precursors of approximately 34,000 and 36,000 daltons (the in vitro translation products include a 2000 dalton leader sequence). α-1Y codes for a protein precursor of 32,912 daltons, placing it as a member of the smaller class, The cDNA clone pGlia42 codes for a protein precursor of 36,500 daltons and may be a member of the larger class.

One significant structural characteristic of this gliadin gene is its lack of introns. In this respect it is similar to the zein genes (32,41) and the soybean lectin gene (38) and different from maize alcohol dehydrogenase genes (42) maize and soybean actin genes (43) and the legume storage proteins phaseolin (31) glycinin (44) and conglycinin (39) all of which contain introns. Although the CAG/G(T,G) consensus sequence characteristic of splice

junctions (34) is present in three positions in the coding sequence of this A-gliadin gene; 670-674, 681-685, 721-725 (Fig. 4), no actual introns are indicated. The gliadin genes are considered to have resulted, in part, from duplications of shorter ancestral sequences (6,11,12), but the present information suggests either that the evolution of A-gliadin genes did not involve intron/exon structures, or that such structures were eliminated during the evolution (see 45,46,47).

The α-1Y sequence contains all of the recognized consensus control regions. The sequence CAAATGCCAAT located 37 bases upstream from the 'TATA' box is particularly interesting in that such 'CAAT-CAAT' structures are present in several, but not all, plant genes thus far reported (Figure 7B). Further sequences from a wider variety of species and genes are needed to establish the distribution and variability of this region. The precise functional role of specific portions of this region is yet to be determined, particularly in light of conflicting reports as to its functional importance in in vitro mutated genes (48,49,50).

Langridge and Feix (51) have reported two promoter regions in a zein gene, yielding transcripts of two lengths. An examination of the α-1Y sequence shows a 'TATA'-rich region at approximately -730 bp, in addition to the TATA at -104. Interestingly, the more distal 'TATA' region includes a direct 9 base inverted repeat, similar to the zein reported by Langridge and Feix, but lacking the internal loop of the zein. In addition, Langridge and Feix found a 15 base direct repeat. α-1Y, instead, contains a larger region with several 70-80% homologies which may be the remnants of ancient duplications.

This distal 'TATA' is part of the first region of inversions mentioned earlier (centered at -690). In addition, the sequence from the 3' part of this inversion begins at +1023, the polyadenylation site in 2 cDNA clones (11). The potential significance of these sequences must await further analysis of the gliadin multigene family and transcriptional studies to delineate those sequences necessary for gene activity.

We have also isolated gliadin clones from a wheat genomic library (cultivar Cheyenne) constructed (Anderson et al., in preparation) in vector lambda Sep6-Lac5 (E. Meyerowitz; unpublished, see (25)). The wheat cultivars Yamhill and Cheyenne show similar patterns in the A-gliadin regions in 2-D PAGE of seed proteins (unpublished). One of the Cheyenne clones, Chey-5, is similar to YAM-2 and is also shown in Figure 2. Copy number estimates made from Southern blots of wheat cultivars Yamhill and the related Cheyenne (Figure 3, and (52)) indicate that there are 1-20 copies of 6.2 kb EcoRI

gliadin fragments in the wheat A genome, with the exact number dependent on each specific wheat cultivar. The results of the Southern analysis shown on Figure 6 indicates potential changes in the 6.2 kb gliadin sequences, either by expansion or diminution of the total number of members of this group. Further study will delineate if this group represents a duplication and possible divergence within a contiguous locus on the 6A chromosome. Support for a group of similar, but distinct genes comes from a further restriction analysis (unpublished) of Chey-5 and YAM-2 which indicates similar, but not identical restriction patterns for the two clones. These questions of duplications and divergence will be resolved as we expand our study into the rest of the A-gliadin gene sub-family.

ACKNOWLEDGMENTS

We wish to thank M. Murray and J. Slightom (Agrigenetics) for sending us the wheat genomic library. The Chinese Spring nullisomic-6A-tetra-6B was generously provided by E. R. Sears (37).

[1]Permanent address: Laboratoire de Technologie des Cereales, Institut National de la Recherche Agronomique, 9 Place Viala, 34060 Montpellier, France

REFERENCES

1. Kasarda, D. D., Bernardin, J. E., and Nimmo, C. C. (1976) Wheat Proteins, In Advances in Cereal Science and Technology, Vol. 1, Y. Pomeranz, ed., pp. 158-236, American Association of Cereal Chemists, St. Paul, Minnesota.
2. Wrigley, C. W. and Shepard, K. W. (1973) Ann. N.Y. Acad. Sci. 209, 154-162.
3. Mecham, D. K., Kasarda, D. D., and Qualset, C. O. (1978) Biochem. Genet. 16, 831-853.
4. Payne, P. I., Jackson, E. A., Holt, L. M., and Law, C. N. (1984) Theor. Appl. Genet. 67, 235-243.
5. Sozinov, A. A., and Poperelya, F. A. (1980) Ann. Technol. Agric. 29, 229-245.
6. Kasarda, D. D. (1980) Ann. Technol. Agric. 29, 151-173.
7. Greene, F. C. (1983) Plant Physiol. 71, 40-46.
8. Mecham, D. K., Fullington, J. G., and Greene, F. C. (1981) J. Sci. Fd. Agric. 32, 773-780.
9. Greene, F. C. (1981) Plant Physiol. 68, 778-783.
10. Okita, T. W., and Greene, F. C. (1982) Plant Physiol. 69, 834-839.
11. Kasarda, D. D., Okita, T. W., Bernardin, J. E., Baecker, P. A., Nimmo, C. C., Lew, E. J.-L., Dietler, M. D., and Greene, F. C. (1984) Proc. Natl. Acad. Sci. U.S.A. in press.
12. Bartels, D., and Thompson, R. D. (1983) Nucl. Acids Res. 11, 2961-2977.
13. Murray, M. G., Kennard, W. C., Drong, R. F., and Slightom, J. L. (1984) Gene (in press).
14. Loenen, W. A., and Blattner, F. (1983) Gene 26: 171-179.
15. Hanahan, D. (1983) J. Mol. Biol. 166: 557-580.
16. Benton, W. D., and Davis, R. W. (1977) Science 196: 180-182.
17. Lynn, D. A., Angerer, L. M., Bruskin, A. M., Klein, W. H., and Angereer, R. C. (1983) Proc. Natl. Acad. Sci. USA 80: 2656-2660.
18. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74: 5463-5467.

19. Hu, N.-T., and J. Messing, J. (1982) Gene 17: 271-278.
20. Southern, E. (1975) J. Mol. Biol. 98: 503-517.
21. Rigby, P. W. J., Dreckmann, M., Rhodes, C., and Berg, P. (1977) J. Mol. Biol. 113: 237-251.
22. Tinoco Jr., I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. E., Crothers, D. M., and Gralla, J. (1973) Nature New Biology 246: 40-41.
23. Messing, J., and Vieira, J. (1982) Gene 19: 269-276.
24. Birnboim, H. C., and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
25. Maniatis, T., Fritsch, E. F., J. Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Press, NY
26. Luthe, D. S., and Quatrano, R. (1980) Plant Physiol. 65, 305-308.
27. Kislev, N., and Rubenstein, I. (1980) Plant Physiol. 66, 1140-1143.
28. Blin, N., and Statford, D. W. (1976) Nucleic Acids Res. 3: 2303-2308.
29. Weislander, L. (1979) Anal. Biochem. 98: 305-309.
30. Proudfoot, N. J., and Brownlee, G. G. (1976) Nature 263, 211-214.
31. Slightom, J. L., Sun, S. M., and Hall, T. C. (1983) Proc. Natl. Acad. Sci. USA 80, 1897-1901.
32. Pederson, K., Devereux, J., Wilson, D. R., Sheldon, E., and Larkins, B. A. (1982) Cell 29, 1015-1026.
33. Fitzgerald, M., and Shenk, T. (1981) Cell 24, 251-260.
34. Breathnach, R., and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
35. Shah, D. M., Hightower, R. C., and Meagher, R. B. (1982) Proc. Natl. Acad. Sci. USA 79, 1022-1026.
36. Tabata, T., Sasaki, K., and Iwabuchi, M. (1983) Nucleic Acids Res. 11, 5865-5875.
37. Sears, E. R. (1954) The Aneuploids of Common Wheat. Res. Bull. (Missouri Agric. Expt. Sta.) Vol. 572.
38. Vodkin, L. O., Rhodes, P. R., and Goldberg, R. B. (1983) Cell 34, 1023-1031.
39. Schuler, M. A., Schmitt, E. S., and Beachy, R. N. (1982) Nucleic Acids Res. 10, 8225-8245.
40. Montell, C., Fischer, E. F., Caruthers, M. H., and Beck, A. J. (1983) Nature 305, 600-605.
41. Spena, A., Viotti, A., and Pirrotta, V. (1982) EMBO Journal 1, 1589-1594.
42. Peacock, W. J., Dennis, E. S., Ellis, J., Finnegan, E. J., Gerlach, W. L., Llewellyn, D., and Sachs, M. M. (1984) J. Cellular Biochem. Abstracts 8B, 15.
43. Shah, D. M., Hightower, R. C., and Meagher, R. B. (1983) J. Mol. Appl. Genet. 2, 111-126.
44. Fischer, R. L., and Goldberg, R. B. (1982) Cell 29, 651-660.
45. Zakut, R., Shani, M., Givol, D., Nueman, S., Yaffe, D., and Nudel, U. (1982) Nature 298, 857-859.
46. Go, M. (1983) Proc. Natl. Acad. Sci. USA 80, 1964-1968.
47. Eiferman, F. A., Yound, P. R., Scott, R. W., and Tilghman, S. M. (1981) Nature 294, 713-718.
48. McKnight, S. L., and Kingsbury, R. (1982) Science 217, 316-324.
49. Grosveld, G. C., deBoer, E., Shewmaker, C. K., and Flavell, R. A. (1982) Nature 295, 120-126.
50. Dierks, P., van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J., and Weissmann, C. (1983) Cell 32, 695-706.
51. Langridge, P., and Feix, G. (1983) Cell 34, 1015-1022.
52. Litts, J. C., Anderson, O. D., Okita, T. W., and Greene, F. C. (1983) Plant Physiol. 72, (Suppl.), 2.
53. Spena, A., Viotti, A., and Pirrotta, V. (1983) J. Mol. Biol. 169, 799-811.
54. Brisson, N., and Verma, D. P. S. (1982) Proc. Natl. Acad. Sci. USA 79, 4055-4059.