Theor Appl Genet (1997) 95:59–65

O. D. Anderson · F. C. Greene

The α-gliadin gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes

Received: 1 October 1996 / Accepted: 20 December 1996

Abstract The derived amino-acid sequences of all reported α-gliadin clones are compared and analyzed, and the patterns of sequence change within the α-gliadin family are examined. The most variable sequences are two polyglutamine domains. These two domains are characteristic features of the α-gliadin storage proteins and account for most of the variation in protein size of this otherwise highly conserved protein family. In addition, their encoding DNA sequences form microsatellites. Single-base substitutions in the α-gliadin genes show a preponderance of transitions, including the C to T substitution which contributes to the generation of stop codons, and consequently to the observation that approximately 50% of the α-gliadin genes are pseudogenes. In one unusual gene, a microsatellite has expanded to 321 bp as compared to the normal 36–72 bp, and may result from similar mechanisms that produce polyglutamine-associated genetic diseases in humans. A comparison of the 27 reported sequences show several α-gliadin gene subfamilies, at least some of which are genome specific.

Key words α-Gliadin · Microsatellites · Sequence variation · Pseudogenes · Polyglutamine

Introduction

The α-gliadins are monomeric prolamines (cereal seed-storage proteins high in proline and glutamine). They, and the closely related γ-gliadins, are the most abundant wheat seed proteins, and when it is considered that wheat is relatively high in protein content and is the first or second (year-to-year) highest contributor to the human diet, this makes the α-gliadins among the most consumed proteins by humans. An unfortunate aspect of this human consumption is that the α-gliadins seem to be the major initiators of coeliac disease, an often severe dietary syndrome that effects as many as 1 person in 300 (Shewry et al. 1992).

The number of α-gliadin proteins synthesized is highly variable, although there has been uncertainty in estimating the number of proteins and genes. Lafiandra et al. (1984) have resolved at least 16 major α-gliadin spots by 2-D PAGE of protein extracts from cv Cheyenne seed. This number is considerably less than the estimated 150 genes (Anderson et al. 1997). Among the possible explanations for this discrepancy are that many of the family members are pseudogenes and/or that single protein bands/spots could originate from multiple genes. An examination of RFLP patterns and the sequences of flanking DNA indicate that the α-gliadin gene family is composed of subfamilies of closely related genes (Anderson et al. 1997). At least one subfamily consists of pseudogenes which contain one or more stop codons and which have lost most of the 3′ untranslated sequence, including polyadenylation signals and sites (Anderson 1991).

The α-gliadin protein primary structure is diagramed in Fig. 1. A 20 amino-acid-residue signal peptide is cleaved post-translationally, leaving a mature protein of approximately 250 amino-acid residues. The N-terminal repetitive region is composed of imperfect repeats of 7–14 amino-acid residues, followed by a polyglutamine domain, a unique region, a second polyglutamine domain, and finally a C-terminal unique sequence. The exact higher-order structure of these polypeptides is not known, but the repetitive region may form an extended structure in contrast to the more compact, disulfide bond-stabilized remainder of the

Communicated by G. E. Hart

O. D. Anderson (✉) · F. C. Greene[1]
USDA, ARS, Western Regional Research Center,
800 Buchanan Street, Albany, CA 94710, USA

Present address:
[1] USDA, ARS, North Atlantic Area, 600 East Mermaid Lane, Philadelphia, PA 19118, USA

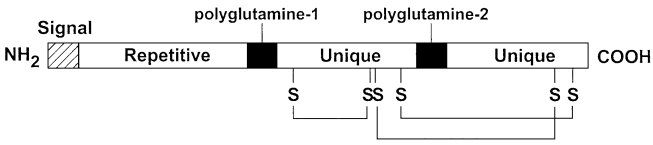


Fig. 1 α-Gliadin protein structure. The general structure of an α-gliadin protein is diagramed. The signal peptide is indicated by the *striped box*. *Filled boxes* are the two polyglutamine domains encoded by microsatellites. The *open boxes* are the repetitive and unique domains. *S* = positions of cysteine residues. Intramolecular disulfide bonds are drawn as determined by Müller and Wieser (1995)

protein (Cole et al. 1981). The six conserved cysteine residues form three intramolecular disulfide bonds (Müller and Wieser 1995).

The present report uses the availability of the extensive α-gliadin sequence information to examine the DNA-encoded protein sequences of this family. Results are presented of analyses of sequence conservation and variability, the structure of a repetitive domain, the structure of microsatellites encoding polyglutamine domains, and the high percentage of pseudogenes within this gene family.

Materials and methods

All currently reported α-gliadin clones are described and referenced in Anderson et al. (1997). Amino-acid and DNA sequences were aligned using the Clustal Analysis option of the Megalign module of Lasergene software (DNAstar, Inc.) and displayed as linear sequence alignments and phylogenetic trees. This algorithm uses an examination of all pairs of sequences to cluster the sequences into groups. The microcomputer implementation is derived from Higgins and Sharp (1989). Further alignment adjustments for sequence display were performed manually. Patterns of base substitutions were determined by comparing all clones at each sequence position. Substitutions were scored if 1 sequence out of 27 had a different base than the others or if 2–4 related sequences had a different base. All sequences were as originally reported except for the sequence of clone A26 (Okita et al. 1985) from which a 20-bp duplication was edited. This segment, as published, generates a frameshift and a stop codon by an exact duplication of an adjoining sequence and adds a sequence not present in any other α-gliadin. Three of seven cDNA sequences of the original report contain similar aberrations unique to this source. We believe the source cDNA library contained a high percent of cloning artifacts that could not be recognized at that time.

Results and discussion

α-Gliadin amino-acid sequence variations

All known derived α-gliadin protein sequences are aligned in Fig. 2. The polyglutamine domains were not included since they are hypervariable and are discussed below in detail. The main region of variability of the amino-acid alignment in Fig. 2 is centered within the repetitive domain and can be reduced to two main

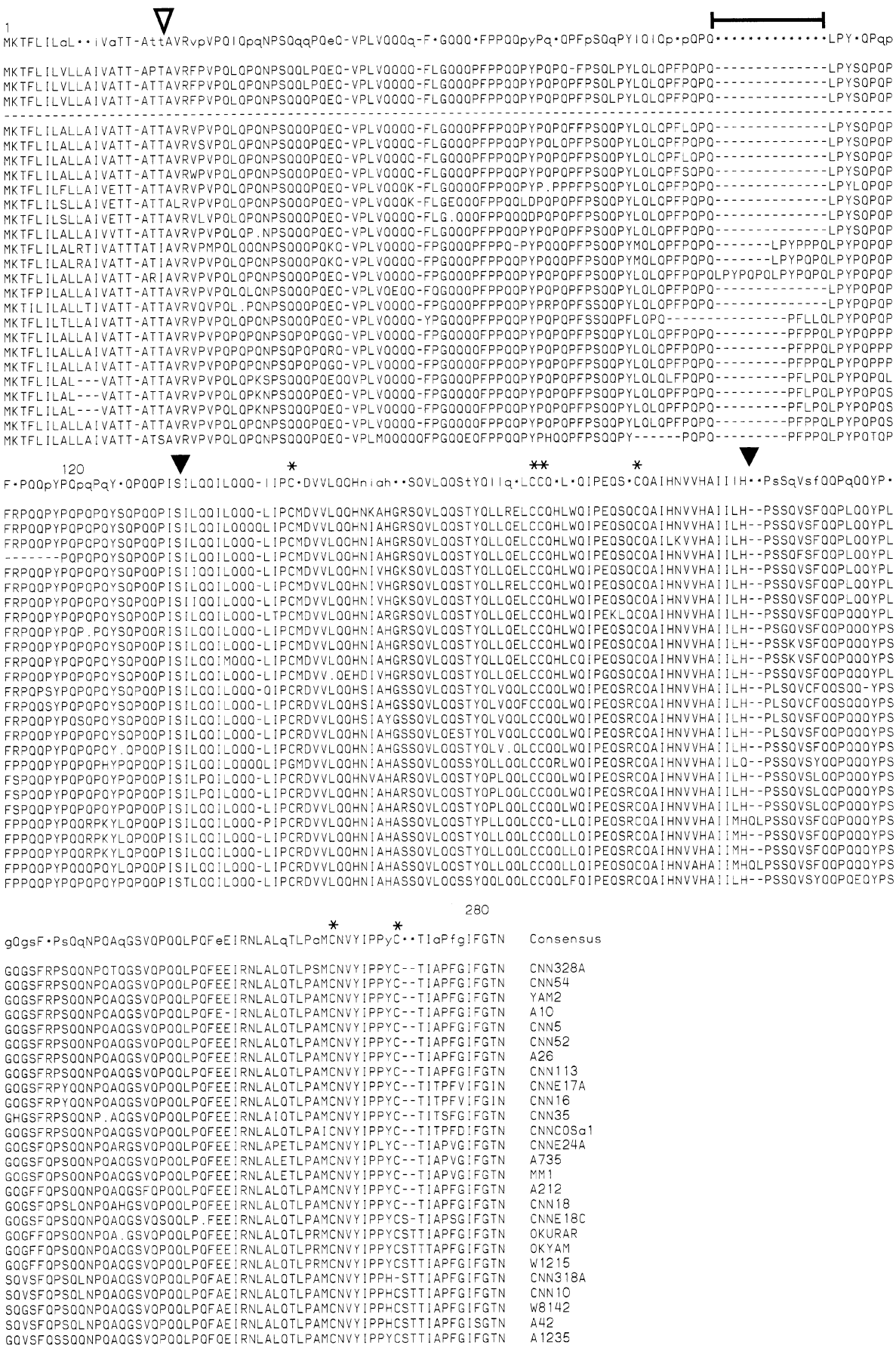
Table 1 Percent of DNA base changes within the α-gliadin genes for all possible base combinations

From \ To	A	T	C	G
A	–	4.5	4.5	9.1
T	6.3	–	10.2	3.4
C	9.1	24.4	–	6.3
G	13.1	6.8	2.3	–

features: one set of α-gliadins contains an extra repeat composed of $PF^{P}_{L}PQ$ (CNNE18C to A1235) and three clones show one (CNNE24A and A735) or two (MM1) duplications of the sequence LPYPQP. On the basis of the protein sequence, the internal composition of the repeats has been suggested to be composed of two repeats: PQPQPFP and PQQPY (Shewry and Tatham 1990). Although the exact determination of what constitutes a repeat unit is partially subjective, we believe that the DNA sequences suggest a single repeat motif based on the codon series CCA $T^{T}_{A}T$ CC^{A}_{G} CAR (see Fig. 3 for one example), where CAR represents a 3–6 glutamine codon-rich region: mainly CAA plus CAG, with the remainder as CC^{A}_{G} proline codons or codons one-base-change removed from those four codons. This α-gliadin consensus repeat is similar to, but distinct from, those of the wheat γ-gliadin and low-molecular-weight glutenin proteins (Anderson, unpublished). Presumably the patterns of the repeats have diverged subsequent to the separation of the gliadin gene families, similar to the manner in which specific DNA sequences diverge after gene duplication.

The remaining primary structure of the α-gliadins is relatively conserved. In the consensus sequence of Fig. 2, 70% of the residue positions are identical in at least 26 of the 27 sequences. An additional 21% of the residue positions are identical in 22–25 sequences, and 9% of the residue positions are conserved in 21 or fewer sequences. Of these last two groups, position variation is limited to two possible residues, except for two positions where three amino-acid residues are possible. All amino-acid differences can be attributed to single DNA base changes, or to sequence changes involving complete codons, such that frameshifts are not introduced even in the pseudogene members of the α-gliadin gene family.

Fig. 2 Amino-acid sequence alignment of α-gliadin clones. Sequences were aligned after deleting the polyglutamine regions (*filled arrowheads*). Additional alignment adjustments were performed manually. The *open arrowhead* marks the junction of the signal peptide. *Asterisks* mark the cysteine residues, and the *bar* shows the variable region of the repetitive domain. A consensus amino-acid sequence is given above the alignment. *Uppercase letters* indicate consensus positions where 26–27 of the 27 sequences have an identical residue in that position. *Lowercase* indicates 22–25 identical out of 27, and *dots* indicate that 21, or fewer, of the sequences have the same residue in that position. Stop codons are indicated as *periods*


1
120
280
Consensus
CNN328A
CNN54
YAM2
A10
CNN5
CNN52
A26
CNN113
CNNE17A
CNN16
CNN35
CNNCOSa1
CNNE24A
A735
MM1
A212
CNN18
CNNE18C
OKURAR
OKYAM
W1215
CNN318A
CNN10
W8142
A42
A1235

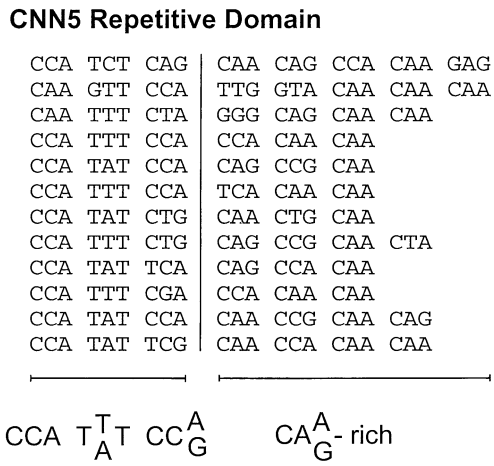


Fig. 3 Repetitive domain motif structure. The DNA sequence of the CNN5 repetitive domain is arranged by codons and suggested repeats are arrayed vertically. A consensus structure is given below. The *vertical line* separates the conserved first three codons of each repeat motif from the variable-length glutamine-rich part of the repeat

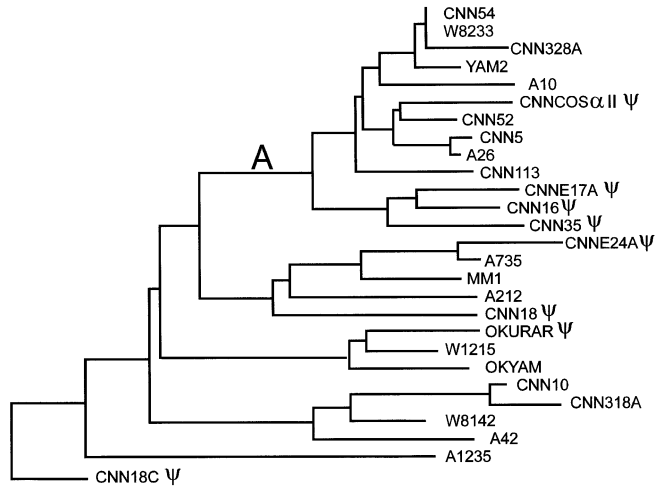


Fig. 4 Phylogenetic tree of α-gliadin DNA coding sequences. Relatedness among α-gliadin sequences was drawn using the Clustal Analysis module of Lasergene software (DNAstar, Inc.). The sequences compared were from the start codon to the stop codon, but minus the two microsatellites

Most α-gliadins contain six conserved cysteine residues that form intramolecular disulfide bonds. Figure 2 shows five examples of gliadins with odd numbers of cysteine residues. Clones CNNE24A and A735 have an additional cysteine created by a serine-to-cysteine residue change at position 210, and clone CNN35 has a tryptophan-to-cysteine substitution at position 181. Clone CNNE18C loses a cysteine through a cysteine-to-glycine change at position 146, and clone CNN318A is missing the final cysteine (position 268). This has implications for covalent participation of some gliadins in the protein matrix of doughs since Kasarda et al. (1987) theorize that gliadins with odd numbers of cysteines become available to join the disulfide-crosslinked gluten matrix and function as polymer chain terminators.

Several of the sequences in Fig. 2 are so similar that it may not be possible to distinguish unique proteins by commonly used physical/chemical procedures; e.g., CNN54 and YAM2 differ by a single leucine/proline, CNN10 differs from W8242 in an additional LQ sequence plus single V → G and L → Q substitutions, and OKURAR differs from W1215 mainly by a single Q → R substitution. Thus, it is likely that proteins encoded by such genes are observed in the same band/spot in PAGE analyses.

Phylogenetic tree of α-gliadin clones

The relatedness of α-gliadin genes was assessed in the phylogenetic tree shown in Fig. 4 and generally agrees with the branch associations and the protein sequence alignments in Fig. 2. The only exception is clone CNNE18C which was distal from the other DNA sequences in Fig. 4, but is placed between clones CNN18 and PA212 in a protein-sequence-tree construction (data not shown, but similar to Fig. 2). The cause of this displacement is a higher percentage of synonymous single base changes for CNNE18C than for other clones, but the reason for this different pattern of single base changes is unknown.

The branch labelled A in Fig. 4 is assigned to chromosome 6A based on the restriction fragments of all the genomic clones in the branch (Anderson et al. 1997). However, CNN18 and OKURAR are also from the A genome but are located on different branches from the other A-genome clones. Clone CNN10 has a chromosome-6B preliminary assignment and CNE24A a preliminary assignment to chromosome 6D. Further work is needed to confirm these last two assignments and to determine if other branches coincide with chromosome assignments.

Patterns of base change

The large number of α-gliadin DNA sequences now available allows an analysis of the pattern of single base changes within a single plant gene family. A DNA alignment of the 27 α-gliadin clones was used to determine the occurrence of all possible base changes and the results are shown in Table 1. The C → T transition was the most common, occurring in 24.4% of the total changes. Next most common were the other three possible transitions (T → C, A → G, and G → A) and the various transversions. The majority of DNA changes are base substitutions, with either purine-to-purine and pyrimidine-to-pyrimidine transitions, or purine-pyrimidine transversions. A random distribution of substitutions would result in 33.3% transitions.

Table 2 Polyglutamine-encoding microsatellites within the α-gliadin gene DNA sequence

Clone[a]	Microsatellite-1[b]	Microsatellite-2[b]
CNN328A	GGGGGGGAAAAAAAAAAAA	AAA aaa AAAAA
CNN54	GGGGGGGGGAAAAAAAAA	AAA aaa AAAA
YAM2	GGGGGGGGGAAAAAAAAA	AAA aaa AAAA
A10	GGGGGGAAAAAAAAA gaa AA	AAAA aaa AAAAA
CNN5	GGGGGGGAAAAAAAAAAAA	AAA aaa AAAA
CNN52	GGGGGGAAAAAAAAAAAAAA	AAA aaa AAAA
W8233	GGGGGGAAAAAAAAAAAAAA	AAA aaa AAAA
A26	GGGGGGGAAAAAAAAAAAA	AAA aaa AAAA
CNN113	GGGGGGAAAAAAAAAAA gaa AA	AAAA aaa AAA
CNNE17A	GAA taa AAAAAAAA	AA aaa AAAAA
CNN16	GAA taa AAAAAAAAA	AAA aaa AA gaa AAA
CNN35	GAA taa AAAAAAAAAA	AAA aaa AA cac A
CNNCOSαII	GGGGGGGAAAAAAAAA	AA taa aaa AAAA
CNNE24A	GGGGGAAAAAAAAAAG	AAAAAAAAAAAAAAAAAAAA
A735	GGGGGGGGAAAAAAG	AAAAAAAAAAAAAA
MM1	GGGGGAAAAAAA aaa AAAAAAAG	AGAAAAAAAAAA
A212	GGGGGGGAAAAAAAA	AACCCAAAAAAAAAA
CNNE18C	GAA gcg AA gca AAAAAA taa AAAA aaa AAAAA	AAAAAAAAAAAAAAAAAAAA
OKURAR	GAA gca AAAAAAAAAAGAAAAAAA	AAGGAAA
OKYAM	GAA gca AAAAAAAAAAGAAAAAAA	AAAAAAA
W1215	GAA gca AAAAAAAAAAGAAAAAAA	AAG cga AAA
CNN318A	GAAA gca AAAAAAAAAAAAAA	AA gaa AAAG ttg AAAAGG ctg AAAAAAAAAAAAAA
CNN10	GAA gca AAAAAAAAAAAAAAAA	AA gaa AAAG ttg AAAAAGA ctg AAAAAAAAAAAA
W8142	GAA gca AAAAAAAAAAAAAAAA	AA gaa AAAG ttg AAAAAGA ctg AAAAAAAAAAAA
A42	GAA gca AAAAAAAAAAAAAA	AAAAAAA gaa A aaa AG ttg AAAAAAAGA ctg AAAAAAAAAA
A1235	GAA gca AAAAAAAA	AAAAAA
CNN18[c]	G ctg GGAAAAA tag AAAGAA aaa GGAAAGGGAAAAAGAA cac GAAG cac AGG tag AAAAA	
	GGGGAAAAAAAAGGGGGGGGGGGGGAAAAAAAAAAAAAAAAGGGGGGGGGAAAAAA taa AAA	

[a] Clone references and GENBANK accession numbers are given in Anderson et al. (1996)
[b] G = CAG, A = CAA (both glutamine codons). C = CAC codon (histidine codon). Lower-case letters are the three bases of all other codons; e.g., AAA aaa AAAA represents three CAA glutamine codons, a single lysine codon, and four more CAA glutamine codons
[c] The first polyglutamine region of CNN18 contains 116 codons (left) and the second contains five codons (right)

However, it is known that transitions occur approximately 59% of the time, with the $C \cdot G \rightarrow T \cdot A$ base-pair substitution the most common, at approximately 39% of total substitutions (Gojobori and Grauer 1982; Li et al. 1982). The $C \rightarrow T$ transition has been theorized to predominate because of the ability of 5-methyl-cytidine to be incorrectly replicated as a thymidine (Gojobori and Grauer 1982). Methylation at the 5-position of cytidine is the most common modified DNA base, and is particularly important in plants where as much as 20% of the total residues can be methylated. Table 1 shows that transitions occur in 57% of substitutions, in agreement with numbers from animal systems. There is also a 26% imbalance toward $G \cdot C \rightarrow A \cdot T$ shifts which would imply a tendency of the DNA sequences to become more AT-rich if there were no counterbalancing mechanism. The $C \rightarrow T$ transition is particularly important to pseudogene generation, as discussed in the next section.

Pseudogenes

A number of cereal pseudogenes have been reported, including pseudogenes for two wheat high-molecular-weight glutenins (Forde et al. 1985; Harberd et al. 1987), and a γ-gliadin (Rafalski 1986). Heidecker and Messing (1986) estimated that perhaps half of the zein genomic fragments are pseudogenes since there are twice as many copies in DNA RFLP patterns as there are spots on 2-D protein gels. Many zein pseudogene sequences have been reported: one zein subfamily of 15–20 members contains only 3–4 active genes (Liu and Rubenstein 1992). Eight of the 20 known α-gliadin genomic sequences are evidently pseudogenes. Two additional Cheyenne α-gliadin genomic clones were only partially sequenced but contained one or more stop codons within the polyglutamine domains (Anderson, unpublished).

Modiano et al. (1981) noted that codon usage is not random in the human globins. Those codons which could mutate to a nonsense codon with a single base change are used relatively infrequently. This option is limited for the prolamines since both glutamine codons become stop codons if a $C \rightarrow T$ transition occurs. Heidecker and Messing (1986) note that zeins include about 32% codons that can become stops with a single base change, mainly due to the high percent of glutamine codons (CAA and CAG). They calculate that about

6.8% of all zein single base changes will result in the generation of a stop codon. The situation is similar for the α-gliadins; i.e. 43% of CNN5's codons are potential stops. If the results of Table 1 are used to estimate stop codon generation frequency it can be calculated that 6.9% of single base changes within the coding sequence of CNN5 will generate a premature stop. If CNN18 were an active gene, with 47% glutamine, 55.3% of the residues could become stops with a single base change. The relatively rapid changes in gene family members and composition, via mechanisms such as unequal crossing-over and gene conversion, and the selection pressure due to a functional role of α-gliadin proteins, may prevent the entire gene family from inactivating.

The high percentage of pseudogenes in the α-gliadin gene family and the conservation of amino-acid sequences (see above) seem sufficient explanations for the apparent discrepancy between protein and gene estimates of α-gliadin family size. Lafiandra et al. (1984) used 2-D protein electrophoresis to detect 16 major group-6 chromosome-encoded spots for Cheyenne and 17 for Chinese Spring. Additional spots were too faint to be assigned but may originate from genes expressed at lower levels. Based on the coding-sequence data reported in this paper, it is likely that at least some of the spots include proteins encoded by multiple genes. There are at least several closely related sub-families, some of whose members are so similar at the amino-acid level that physical/chemical separation methods would be unlikely to resolve them. Such similarity in α-gliadin gene family members may be due to the relatively rapid rate of change of this gene family, presumably by combinations of duplications and deletions of individual genes and blocks of genes (D'Ovidio et al. 1991). It is also possible that relatively recent duplication events can result in multiple, distinct, genes which code for identical proteins.

Microsatellite structure and variation

A major characteristic of all α-gliadin proteins is the presence of two polyglutamine domains encoded by microsatellite-like sequences (Fig. 1). The two codons for glutamine, CAA and CAG, are not randomly distributed in the α-gliadin polyglutamine domains, but tend to occur in homomeric runs of single codons (Table 2). Occurrences of non-glutamine codons can be accounted for mainly by single base changes in glutamine codons (CAA to TAA, CAA to GAA, etc.) except for the codon GCA (alanine) which occurs in nine microsatellite-1 sequences.

Microsatellites are known to be hypervariable, and these regions are the most variable among the α-gliadin genes. For example, the clones CNN16 and CNN35 have only two sites of sequence-length difference in over 3500 base pairs (Anderson 1991), and both occur by codon number variation in the two microsatellite

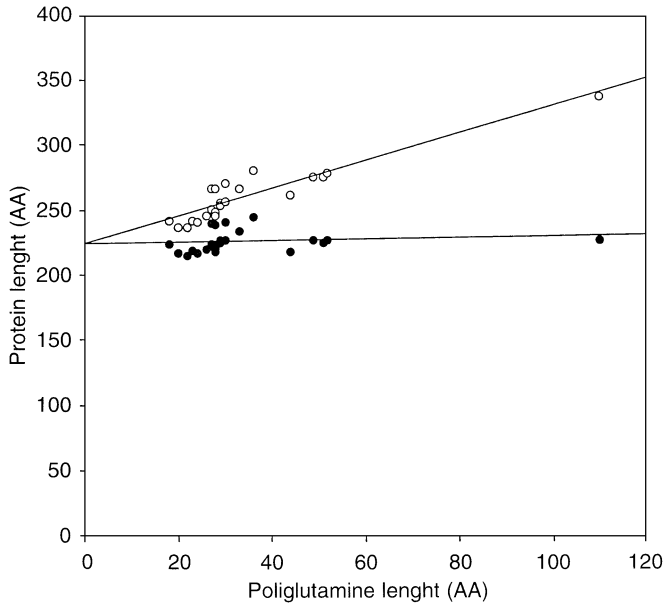


Fig. 5 α-Gliadin protein size variation is mainly due to different microsatellite length variation. The total amino-acid-residue length of the two polyglutamine regions for each clone is plotted against the total number of residues in each coding region and the total amino-acid residues minus the polyglutamine domains. *Open circles* represent the complete coding sequences, and *closed circles* represent the coding sequences minus the polyglutamine domains

domains. Moreover, the microsatellite variation accounts for most of the difference in protein size among α-gliadins (Fig. 5).

An exception to the above pattern of microsatellite structure was found in clone CNN18. Microsatellite-1 DNA is 107 codons in length, five or more times the size of that in other genes, while microsatellite-2 DNA is five codons long, the smallest of known genes. Especially prominant is the non-random arrangement of glutamine codons, with a preference for runs of the single codon CAA interspersed with CAG runs. Seven of the 107 codons in CNN18 microsatellite-1 are not glutamine codons, but six of them could be derived from a glutamine codon with a single base change.

Although CNN18 is a pseudogene, with five stop codons, active α-gliadins with similarly expanded microsatellite regions may provide an explanation for the observations of Kasarda et al. (1987) and Harberd et al. (1985) who reported α-gliadins of 40–50 kDa.

Among the factors influencing the stability of simple repeats are repeat length and homogeneity of the sequence (Wells 1996). Simple-sequence DNA often undergoes slippage-mispairing during DNA replication (reviewed in Albertini et al. 1982; Moore 1983; Tautz et al. 1986). Such repeats have also been identified as hotspots of recombination (Wahls et al. 1990). If this recombination involves unequal crossing-over, expansions and contractions of the original sequence

can occur. Simple repeats may play a role in the homogenization of repetitive DNA arrays by mediating near-equal cross overs or gene conversions. These mechanisms seem responsible for the tendency for changes to spread to adjacent repeats of a basic sequence, a process that has been referred to as preferential homogenization (Lassner and Dvorak 1986), and has been proposed to result from constrained sister-chromatid exchange (Jeffreys et al. 1985).

Homogenization in the α-gliadin microsatellites would have the, perhaps fortuitous, effect of helping to suppress stop codons, by either eliminating or multiplying stops. In the latter case, the gene was already defective and thus additional stops would have no functional effect.

An understanding of the mechanisms of α-gliadin microsatellite variation extends beyond interest in the evolution of this gene family. Coding-sequence microsatellite variation is associated with many hereditary diseases, particularly expansions of the glutamine codon CAG in several human neurodegenerative disorders (reviewed by Jennings 1995). Homopolymeric glutamine stretches are also found in many transcription factors (Gerber et al. 1994). In these cases, normal alleles contain 30 or fewer glutamine codons, while larger polyglutamines are associated with disease states. In two such examples, the increase in polyglutamine length leads to increased binding of a transcription factor to an enzyme involved in energy production, resulting in cell death or impairment (Burke et al. 1996). The α-gliadins will form fibrillar aggregates under specific conditions (Kasarda 1980) but the role of the polyglutamine domains in the in vivo self-association of the α-gliadins is not understood.

References

Albertini AM, Hofer M, Calos MP, Miller JH (1982) On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. Cell 29 : 319–328

Anderson OD (1991) Characterization of members of a pseudogene subfamily of the wheat α-gliadin storage protein genes. Plant Mol Biol 16 : 335–337

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

Burke JR, Enghild JJ, Martin ME, Jou Y-S, Myers RM, Roses AD, Vancel JM, Strittmatter WJ (1996) Huntingtin and DRPLA proteins selectively interact with the enzyme GAPDH. Nature Medicine 2 : 347–350

Cole EW, Fullington JG, Kasarda DD (1981) Grain protein variability among species of *Triticum* and *Aegilops*: quantitative SDS-PAGE studies. Theor Appl Genet 60 : 17–30

D'Ovidio R, Lafiandra D, Tanzarella OA, Anderson OD, Greene FC (1991) Molecular characterization of bread wheats lacking the entire cluster of chromosome 6A-controlled gliadin components. J Cereal Sci 14 : 125–129

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

Gerber H-P, Seipel K, Georgiev O, Höfferer M, Hug M, Rusconi S, Schaffner W (1994) Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science 263 : 808–811

Gojobori T, Li W-H, Graur D (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol 18 : 360–369

Harberd NP, Bartels D, Thompson RD (1985) Analysis of the gliadin multigene loci in bread wheat using nullisomic-tetrasomic lines. Mol Gen Genet 198 : 234–242

Harberd NP, Flavell RB, Thompson RD (1987) Identification of a transposon-like insertion in a *Glu-1* allele of wheat. Mol Gen Genet 209 : 326–332

Heidecker G, Messing J (1986) Structural analysis of plant genes. Annu Rev Plant Physiol 37 : 439–466

Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS 5 : 151–153

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. Nature 314 : 67–73

Jennings C (1995) How trinucleotide repeats may function. Nature 378 : 127

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

Kasarda DD, Adalsteins AE, Laird NF (1987) γ-Gliadins with α-type structure coded on chromosome 6B of the wheat (*Triticum aestivum* L.) cultivar 'Chinese Spring.' Proc 3rd Int Workshop Gluten Proteins, pp 20–29

Lafiandra D, Kasarda DD, Morris R (1984) Chromosomal assignment of genes coding for the wheat gliadin protein components of the cultivars 'Cheyenne' and 'Chinese Spring' by two-dimensional (two-pH) electrophoresis. Theor Appl Genetics 68 : 531–539

Lassner M, Dvorak J (1986) Preferential homogenization between adjacent and alternate subrepeats in wheat rDNA. Nucleic Acids Res 14 : 5499–5512

Liu C-N, Rubenstein I (1992) Molecular characterization of two types of 22 kilodalton α-zein genes in a gene cluster in maize. Mol Gen Genet 234 : 244–253

Li W-H, Wu C-I, Luo C-C (1982) Non-randomness of point mutations as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J Mol Evol 21 : 58–71

Modiano G, Battistuzzi G, Motulsky AG (1981) Non-random patterns of codon usage and nucleotide substitutions in human α- and β-globin genes: an evolutionary strategy reducing the rate of mutations with drastic effects. Proc Natl Acad Sci USA 78 : 1110–1114

Moore G (1983) Slipped-mispairing and the evolution of introns. Trends Biochem Sci 41 : 411–414

Müller S, Wieser H (1995) The location of disulphide bonds in α-type gliadins. J Cereal Sci 22 : 21–27

Okita TW, Cheesbrough V, Reeves CD (1985) Evolution and heterogeneity of the α-/β-type and γ-type gliadin DNA sequences. J Biol Chem 260 : 8203–8213

Rafalski JA (1986) Structure of wheat gamma-gliadin genes. Gene 43 : 221–229

Shewry PR, Tatham AS (1990) The prolamin storage proteins of cereal seeds: structure and evolution. Biochem J 267 : 1–12

Shewry PR, Tatham AS, Kasarda DD (1992) Cereal proteins and coeliac disease. In: Marsh MN (ed) Coeliac disease. Blackwell Scientific Publications, London, pp 305–348

Tautz D, Trick M, Dover GA (1986) Cryptic simplicity in DNA is a major source of genetic variation. Nature 322 : 652–656

Wahls WP, Wallace LJ, Moore PD (1990) Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells. Cell 60 : 95–103

Wells RD (1996) Molecular basis of genetic instability of triplet repeats. J Biol Chem 271 : 2875–2878