Biochimica et Biophysica Acta, 747 (1983) 138–150
Elsevier

BBA 31687

N-TERMINAL AMINO ACID SEQUENCES OF ω-GLIADINS AND ω-SECALINS

IMPLICATIONS FOR THE EVOLUTION OF PROLAMIN GENES

DONALD D. KASARDA, JEAN-CLAUDE AUTRAN *, ELLEN J.-L. LEW, CHARLES C. NIMMO and PETER R. SHEWRY **

Food Proteins Research Unit, Western Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, Berkeley, CA 94710 (U.S.A.)

(Received October 21st, 1982)
(Revised manuscript received June 13th, 1983)

Key words: Gliadin; Amino acid sequence; Protein evolution; Prolamin; (Wheat, Rye)

Two new types of N-terminal amino acid sequence have been found for ω-gliadins from tetraploid and hexaploid wheats. A third type of sequence, which we have recently reported for an ω-gliadin from diploid wheat and for the 'C' hordeins of barley, has now been found for a group of components from rye (ω-secalins) and for ω-gliadins of tetraploid and hexaploid wheats. The sequences provide evidence that all these ω-type prolamins are related and derived in part from gene duplications, followed by divergence of the duplicated genes by point mutations and by insertion and deletion of portions of the gene DNA. Five-residue amino acid sequences that include mainly glutamine and proline occur a number of times in the N-terminal sequences of ω-type prolamins and may indicate that the genes for these prolamins have been formed through repeated duplication of short DNA sequences in which the codons for glutamine and proline predominated. The prolamins of wheat and related species seem to be unique to the grass family of flowering plants, which appeared late in the evolutionary process. The genes coding for these proteins may be the most recently evolved family of genes.

Introduction

Cereal grain storage proteins are usually complex mixtures of components. The components of the mixture from any one species are often similar to one another in composition, structure and properties, but the mixtures from different species, for example, wheat and maize, may differ considerably. This is probably a consequence of the degree of genetic closeness of the species being compared [1,2]. The similar components within a given mixture have probably resulted from gene duplications followed by divergence of the duplicated genes to produce distinguishable (by electrophoretic or chromatographic methods) protein components [3–5]. Post-translational modifications may play a role in generating prolamin polymorphism [6], but there is little evidence that they make a major contribution.

Amino acid sequencing has great potential for analysis of the evolutionary relationships among storage protein components and the genes coding for them, but this technique is only beginning to be applied in the study of these proteins. So far, no cereal grain storage protein has been sequenced completely, even though the complete sequence of

* Permanent address: Laboratoire de Technologie des Cereales, Institute National de la Recherche Agronomique, 9, Place Viala, 34060 Montpellier, France.
** Permanent address: Rothamsted Experimental Station, Harpenden, Herts., AL5 2JQ, U.K.
Abbreviation: PTH, phenylthiohydantion.

more than a thousand proteins has been determined [7]. This delay has largely resulted from difficulties encountered in purifying sufficient amounts of single components from complex mixtures of highly similar components that are poorly soluble in many commonly used aqueous buffer systems. The problem is circumvented to a considerable degree by recent developments in nucleic acid sequencing of cloned DNA, which has resulted in complete sequences of zein proteins [8,9], but actual amino acid sequences remain important for the interpretation of the DNA-based sequences.

Despite this lag in amino acid sequencing, initial studies carried out within the last few years have provided important new information about the relationship of components within a species and some indications of relationships between species. The 40 or more components of the wheat gliadins [10,11] have been shown to fall mainly into two groups, the α-type and the γ-type, on the basis of their N-terminal sequences [3,4], and these same two groups have been shown to make up most of the equivalent proteins of other species of *Triticum* and *Aegilops* [4]. As a consequence of the essential homology of many of the prolamin components in a mixture from a single species, considerable information can be obtained about the genetic relationships among species from N-terminal sequencing of unfractionated prolamin mixtures prepared from them [4,12–15]. Sequencing of purified protein components has contributed a better understanding of the relationships, including variations, that occur within the mixture [3,5,16–19].

In this paper, we report the purification, amino acid compositions, molecular weights by SDS-polyacrylamide gel electrophoresis, and N-terminal amino acid sequences of ω-gliadins from diploid, tetraploid and hexaploid wheats, and ω-type secalins (ω-secalins) from rye. Two new types of N-terminal sequence have been found, and a third type that had previously been reported for an ω-gliadin from diploid wheat [19] and for the 'C' hordeins of barley [5,13–15] has been shown to be homologous with that of ω-secalins. This latter sequence has also been found in tetraploid and hexaploid wheats. We discuss the implications of these sequences for the evolution of the prolamins of barley, rye and wheat, and speculate on the possibility that short repeating sequences observed in some of the sequences indicate the mechanism by which a gene (or genes) ancestral to those coding for the prolamins of these species arose

Materials and Methods

Plant materials. Grain of diploid wheat (*Triticum monococcum*, accession 5317), which was from the same lot used previously [19], and of *Aegilops squarrosa* was ground in a Wiley mill for extraction of gliadin proteins. Grain of the *T. aestivum* cultivars 'Chinese Spring' and 'Justin', of the *T. durum* cultivar 'Mindum', and of the *Secale cereale* cultivar 'Frontier' was ground in a Brabender Quadrumat, Jr., mill; only the endosperm fractions were retained for protein extraction. Semolina of the *T. durum* cultivar 'Edmore' was obtained and used directly for protein extractions without further grinding. Seeds of the nullisomic-1D tetrasomic-1A line of 'Chinese Spring' [20] were ground in a mortar and pestle for protein extraction.

Prolamin preparations. Several similar methods were used in the preparation of prolamin mixtures from the various cultivars. Flour or ground grain was extracted with 60% (v/v) ethanol/water or 50% (v/v) propan-1-ol/water solutions at ratios of extractant to solid of 5 : 1 or 10 : 1. Extractions were at room temperature (approx. 20°C). In some cases, two extractions were made and the supernatants were combined after centrifugation. Prolamins were precipitated from the supernatants by the addition of 0.25 M sodium chloride. The mixture was allowed to stand overnight at 4°C and the precipitated protein was collected by centrifugation at about $25\,000 \times g$. The protein was then dissolved in a small volume of 0.01 M acetic acid or in 8 M urea, dialyzed thoroughly against distilled water at 4°C, and then freeze-dried.

Column chromatography. The ω-gliadin components were separated initially by ion-exchange chromatography on Whatman CM-52 according to the procedure of Booth and Ewart [21] with some minor modifications. The column was equilibrated with a buffer that was 5 mM sodium acetate/1 M dimethylformamide (DMF), pH 3.5 (in some experiments, pH 3.7). Gliadin proteins were dis-

solved in the column buffer, clarified by centrifugation, and applied to the column. Proteins were eluted with a gradient of 5 to 100 mM sodium acetate in the column buffer. Fractions were examined by lactate-polyacrylamide gel electrophoresis (see Electrophoresis) and those containing the desired components were dialyzed against distilled water and freeze-dried.

A different CM-52 ion-exchange procedure was used for the initial purification of rye prolamins and for further purification of some partially purified ω-gliadin fractions. In this procedure, the buffer was 10 mM glycine-acetate/4 M urea, pH 4.6. The protein was eluted with a concave quadratic gradient of 0 to 100 mM sodium chloride in the column buffer [5].

The gliadin preparation from *Ae. squarrosa* was fractionated only by gel filtration chromatography on Sephadex G-100 (Pharmacia) in 0.001 M hydrochloric acid. This was because only a few grams of seed were available and as a consequence very small amounts of protein were obtained.

Some protein components were obtained sufficiently pure from ion-exchange chromatography, but others required further purification by gel filtration chromatography on columns of Sephacryl S-300 (Pharmacia) in 3 M urea adjusted to pH 4.0 with acetic acid. These latter fractions were desalted by gel filtration chromatography on columns of Sephadex G-15 (Pharmacia) in 0.1 M acetic acid, and then freeze-dried.

Electrophoresis. Fractions were analysed by polyacrylamide gel electrophoresis in aluminium lactate buffer, pH 3.2. In this system, migration is determined mainly by the net positive charge on the protein molecules. The procedures were those of Kasarda et al. [22] and Mecham et al. [11]. Components were numbered in order of increasing mobility. Fractions were also separated on the basis of their apparent molecular weights by polyacrylamide gel electrophoresis in buffers containing sodium dodecyl sulfate (SDS-polyacrylamide gel electrophoresis) as described by Shewry et al. [5].

Amino acid analysis. Samples (usually about 1 mg) were hydrolyzed under nitrogen for 21 h at 110°C with 2 ml of 6 N hydrochloric acid (sometimes with 0.1% mercaptoethanol included in the acid). Amino acids were determined with a Durrum D-500 amino acid analyzer. Tryptophan was not determined. Glutamine and asparagine were present in the hydrolysate as glutamic and aspartic acids, but 94–98% of these acids are amidated in the native proteins [21,23,24].

N-Terminal amino acid sequencing. Sequencing was carried out with a Beckman Instruments model 890b (updated) automatic amino acid sequencer and Beckman DMAA program 111374 ('Justin' components ω-1 and ω-5) or Beckman 0.1 M Quadrol programs 011576 or 121178 (all other components). Phenylthiohydantoin (PTH) derivatives of the amino acids resulting from each cycle of the Edman degradation [25] were determined by gas chromatography [26] supplemented with thin-layer chromatography (TLC) on polyamide sheets [27] and silica gel plates [28] for 'Justin' components ω-1 and ω-5, or by high-performance liquid chromatography (HPLC) according to the method of Bhown et al. [29] for all other fractions except the two-component mixture from *Ae. squarrosa*, in which case PTH-amino acids were only analyzed by gas chromatography. PTH-valine and PTH-methionine were not resolved by the HPLC method, and so were distinguished by the method of Jepson and Sjoquist [28].

C-terminal amino acid sequencing. Carboxypeptidase Y (Pierce Chemical Co.) was used for C-terminal amino acid sequencing as described by Schmitt [30] and by Shewry et al. [5].

Results

Fractions: description and electrophoretic analysis

The fractions purified for this study are described in Table I, which lists the species and cultivar, the chromosomal location of the genes coding for the component or fraction (when known), and a code number for each preparation that corresponds to the numbers used in Fig. 3 for the sequences. The 'C' hordein component (code 1) and the ω-1 gliadin components (codes 4, 5) from *T. monococcum*, which were sequenced simultaneously as a mixture, have been described earlier [19]; they are included for comparison with the fractions prepared in the course of the present study. We have designated our components 1, 2, 3 and so on in relation to their distance from the origin in the respective seed electrophoretic pat-

TABLE I

DESCRIPTION OF PREPARATIONS CORRESPONDING TO AMINO ACID SEQUENCES OF FIG. 3

n.d., not determined.

Code	Species	Genome composition	Cultivar	Component	M_r by SDS-PAGE ($\times 10^{-3}$)	Chromosomal control
1.	*Hordeum vulgare*	HH	Julia	C-1	57	barley-5 [53]
2.	*Secale cereale*	RR	Frontier	ω-1	48, 51	rye-1 [54]
3.	*Secale cereale*	RR	Frontier	ω-total	48, 51, 53	rye-1 [54]
4.	*Triticum monococcum*	AA	U.M.	ω-1	44	1A [20]
5.	*Triticum monococcum*	AA	U.M.	ω-1	44	1A
6.	*Triticum monococcum*	AA	U.M.	ω-2	63	1A
7.	*Triticum monococcum*	AA	U.M.	ω-3	63	1A
8.	*Triticum durum*	AABB	Edmore	ω-2	52	not known
9.	*Triticum durum*	AABB	Mindum	ω-2	n.d.	not known
10.	*Triticum aestivum*	AABBDD	Chinese Spring	ω-2	59	1D [31]
11.	*Triticum aestivum*	AABBDD	Justin	ω-2	59	1D? [11]
12.	*Triticum durum*	AABB	Edmore	ω-1	51	not known
13.	*Triticum durum*	AABB	Mindum	ω-1	n.d.	not known
14.	*Triticum aestivum*	AABBDD	Chinese Spring	ω-1	57.5	1D [31]
15.	*Triticum aestivum*	AABBDD	Justin	ω-1	57.5	1D? [11]
16.	*Triticum aestivum*	AABBDD	Justin	ω-5	74	1B? [11]
17.	*Aegilops squarrosa*	DD	—	ω-1, 2	n.d.	1D

tern, but the same designation does not indicate the same mobility when different species or accessions are compared.

The lactate-polyacrylamide gel patterns of most of our preparations are shown in Fig. 1 with the exception of the ω-2 component from *T. monococcum* (code 6) and the ω-1 and ω-2 components (codes 9, 13) from the durum (tetraploid) wheat 'Mindum'. Small amounts of these proteins were obtained in purified form and none was available when the gel intended for Fig. 1 was being run. These preparations had been analyzed by lactate-polyacrylamide gel electrophoresis in the course of their purification. The position of the ω-2 from *T. monococcum* is marked with an arrow in the pattern of the gliadin mixture from this accession in Fig. 1.

All the preparations described in Table I had low electrophoretic mobilities corresponding to ω-gliadins. Most of the preparations consisted of a single major component when analyzed by lactate-polyacrylamide gel electrophoresis (Fig. 1). Exceptions were the ω-secalin mixture (code 3), which was intended to represent the total mixture of ω-secalins found in rye, and the preparation from *Ae. squarrosa* (code 17), which consisted of two major components; the ω-2 gliadin from 'Chinese Spring' had a minor component corresponding to about 10% of the major component. The preparations from 'Mindum' (codes 9, 13), which were not included in Fig. 1, also contained mainly single components, one equivalent in mobility to the slowest band of the gliadin mixture from this cultivar and the other slightly faster in mobility.

The lactate-polyacrylamide gel pattern of total gliadins extracted from seeds of the nullisomic-1D tetrasomic-1A aneuploid line of 'Chinese Spring' [20] is shown in Fig. 1, where the pattern of the aneuploid is compared with that of normal 'Chinese Spring'. The bands corresponding to the ω-1 and ω-2 gliadins are missing from the aneuploid, which is in accordance with the assignment of the structural genes for these proteins to chromosome 1D of 'Chinese Spring' [31].

SDS-polyacrylamide gel electrophoresis patterns of most of our preparations are shown in Fig. 2 and the apparent M_r values determined from the analyses are given in Table I. The 'Mindum' fractions and the preparation from *Ae. squarrosa* were not analyzed by SDS-polyacrylamide gel electrophoresis. As in lactate-polyacrylamide gel electrophoresis, each preparation

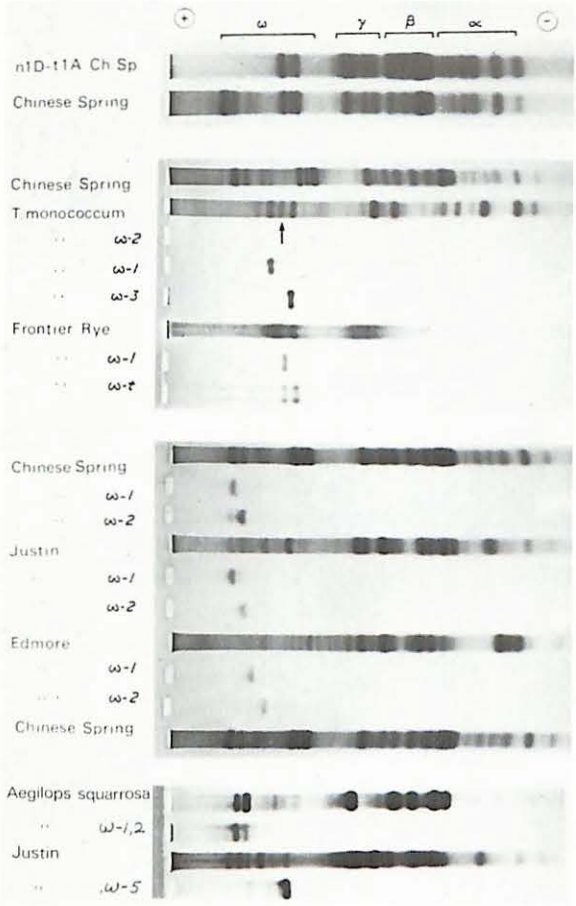

Fig. 1. Lactate-polyacrylamide gel electrophoresis (pH 3.2) of purified ω-gliadin components and fractions and of total gliadin fractions from *T. monococcum*, *Ae. squarrosa*, rye (*S. cereale*) cultivar Frontier, bread wheat (*T. aestivum*) cultivars Chinese Spring and Justin and durum wheat (*T. durum*) cultivar Edmore. The position of the ω-2 component from *T. monococcum* is indicated by an arrow. The pattern of the total gliadin fraction from the nullisomic 1D-tetrasomic 1A aneuploid line of Chinese Spring is also shown. The α-, β-, γ- and ω-gliadin regions correspond to the usual assignments of electrophoretic mobilities for gliadin patterns of common wheats. The nomenclature used for the purified components corresponds to that used in Table I.

consisted mainly of a single component with the exceptions of the ω-1 secalin, which gave two components, the ω-secalin mixture (ω-t), which gave three components, and the ω-5 component of 'Justin', which showed minor amounts of high-molecular-weight contaminants.

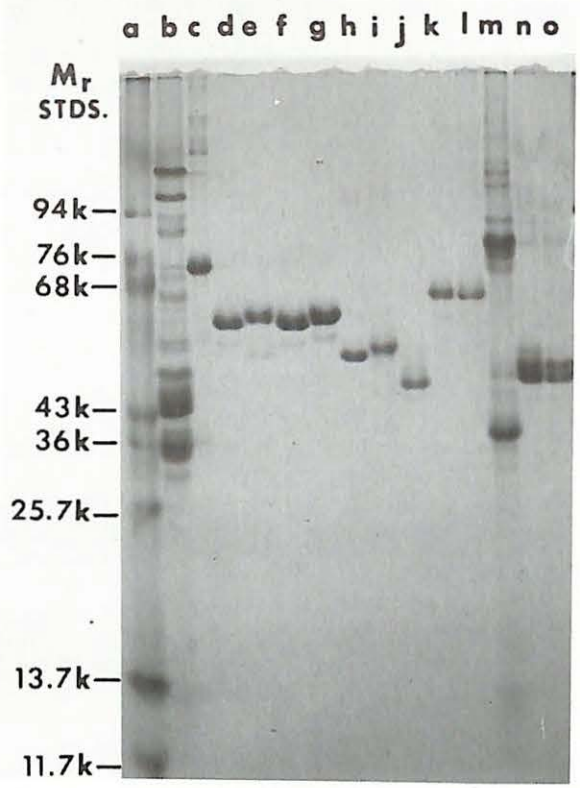

Fig. 2. SDS-polyacrylamide gel electrophoresis (pH 8.9) of purified ω-gliadin components and fractions. Molecular weights of standards are given in units of 10^3 (abbreviated as k). a, molecular weight standards (phosphorylase *b*, 94000; transferrin, 76000; bovine serum albumin, 68000; ovalbumin, 43000; lactate dehydrogenase, 36000; chymotrypsinogen A, 25000; ribonuclease, 13700; cytochrome *c*, 11700; b, total gliadin from Justin; c, ω-5 gliadin from Justin; d, ω-1 gliadin from Justin; e, ω-2 gliadin from Justin; f, ω-1 gliadin from Chinese Spring; g, ω-2 gliadin from Chinese Spring; h, ω-1 gliadin from Edmore; i, ω-2 gliadin from Edmore; j, ω-1 gliadin from *T. monococcum*; k, ω-2 gliadin from *T. monococcum*; l, ω-3 gliadin from *T. monococcum*; m, total secalins from Frontier; n, total ω-secalins from Frontier; o, ω-1 secalin from Frontier.

Amino acid compositions

The amino acid compositions of some of the preparations are given in Table II. Previously published results [19] for a 'C' hordein and the ω-1 gliadin from *T. monococcum* are included in the table for comparison. All analyses were carried out in duplicate and standard deviations for the two analyses are given in parentheses.

The composition data indicated an overall similarity of the components — with the possible

TABLE II

AMINO ACID COMPOSITIONS OF WHEAT ω-GLIADINS AND HOMOLOGOUS FRACTIONS FROM BARLEY ('C' HORDEIN) AND RYE (ω-SECALIN)

Values given in mol percent. Standard deviations of duplicate measurements given in parentheses following each value. Tryptophan not determined. The numbering code is from Table I. Data for Julia hordein and *T. monococcum* gliadin are from Ref. 19.

Amino acids	1 'C'-1 Julia hordein	2 ω-1 Frontier secalin	4,5 ω-1 *T. mono.* gliadin	12 ω-1 Edmore gliadin	14 ω-1 C.S. gliadin	10 ω-2 C.S. gliadin	15 ω-1 Justin gliadin	16 ω-5 Justin gliadin
Asp [a]	0.83	0.51 (0.06)	1.18	0.33 (0.08)	0.5 (0.5)	0.77 (0.07)	0.19 (0.07)	0.6 (0.06)
Thr	0.99	1.0 (0.2)	1.08	1.18 (0.05)	1.7 (0.3)	1.67 (0.03)	1.68 (0.03)	0.75 (0.02)
Ser	2.61	4.11 (0.04)	4.89	6.3 (0.2)	5.7 (0.2)	5.9 (0.2)	5.61 (0.04)	3.67 (0.09)
Glu [a]	41.1	42.9 (0.5)	45.2	43.8 (0.2)	43.1 (0.4)	42.8 (0.03)	43.70 (0.05)	53.4 (0.4)
Pro	31.9	30.6 (0.2)	26.2	27 (2)	29.4 (0.5)	29.11 (0.04)	30.0 (0.4)	20.1 (0.5)
Gly	0.41	1.4 (0.2)	0.89	1.36 (0.07)	1.10 (0.05)	1.09 (0.00)	1.0 (0.1)	1.07 (0.07)
Ala	0.69	0.59 (0.08)	1.57	0.53 (0.02)	0.5 (0.2)	0.61 (0.04)	0.35 (0.02)	0.63 (0.08)
Val	1.11	0.97 (0.06)	0.80	0.72 (0.02)	0.42 (0.06)	0.40 (0.04)	0.35 (0.09)	0.4(0.1)
Met	0.22	0.0	0.13	0.0	0.0	0.0	0.0	0.0
Cys	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0.
Ile	3.02	2.43 (0.02)	2.22	2.55 (0.06)	1.65 (0.08)	1.58 (0.05)	1.6 (0.1)	3.7 (0.1)
Leu	4.31	3.92 (0.04)	4.70	4.9 (0.1)	4.0 (0.1)	4.11 (0.01)	3.9 (0.1)	3.5 (0.1)
Tyr	2.29	1.37 (0.04)	1.17	1.51 (0.08)	1.48 (0.01)	1.47 (0.00)	1.53 (0.08)	0.75 .01)
Phe	9.03	7.4 (0.1)	8.09	8.3 (0.3)	9.0 (0.1)	8.88 (0.01)	8.94 (0.04)	8.7 (0.4)
His	0.60	0.5 (0.1)	0.56	0.93 (0.04)	0.70 (0.06)	0.8 (0.1)	0.55 (0.04)	1.26 (0.04)
Lys	0.0	0.30 (0.0)	0.13	0.33 (0.02)	0.27 (0.01)	0.27 (0.01)	0.32 (0.00)	0.48 (0.01)
Arg	0.84	1.79 (0.06)	1.04	0.39 (0.04)	0.4 (0.2)	0.61 (0.03)	0.32 (0.00)	0.93 (0.06)

[a] Includes amide form.

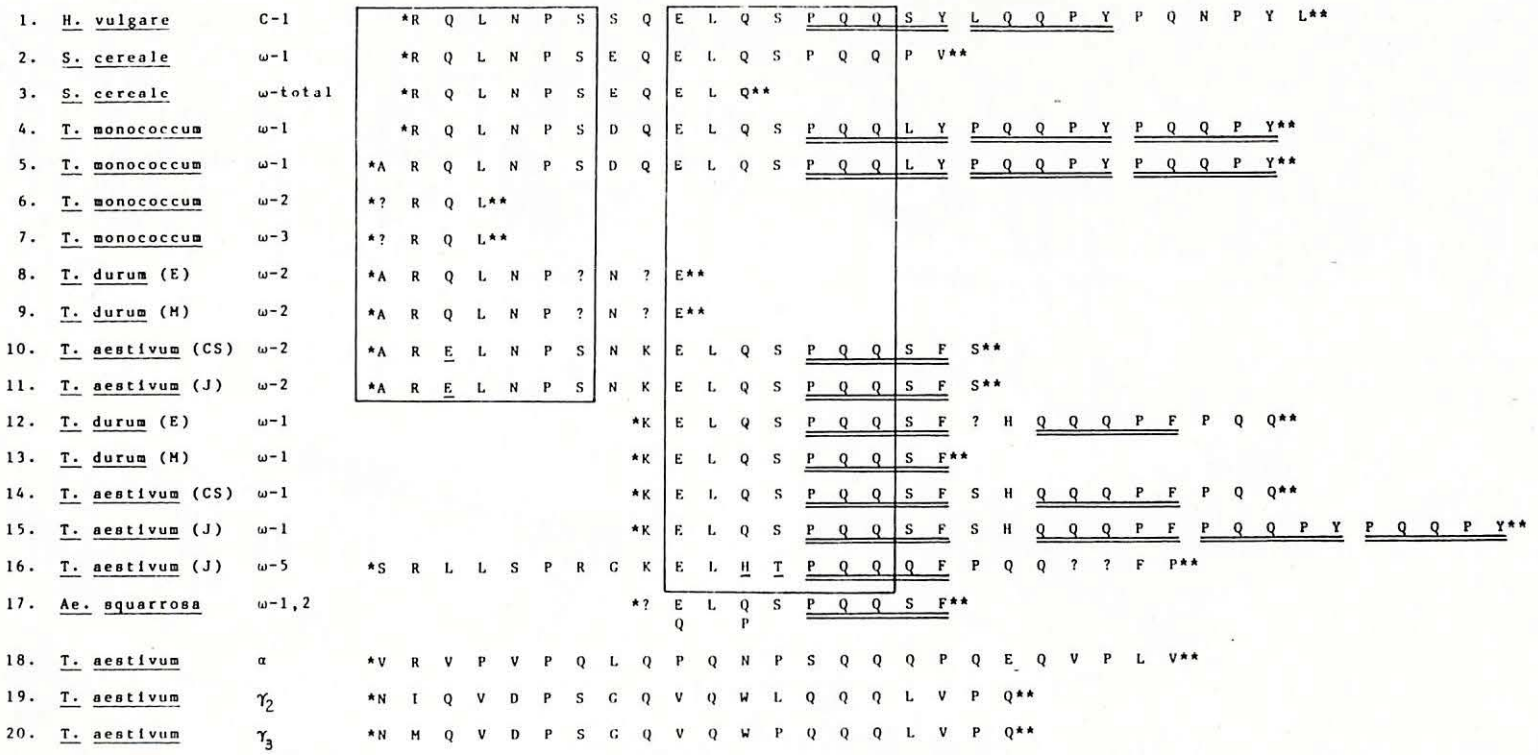

Fig. 3. N-terminal amino acid sequences of purified ω-gliadin components and fractions. Code numbers refer to code of Table I. Sequences for α-gliadins [3,16] and γ-gliadins (M.D. Dietler and D.D. Kasarda, unpublished results) are included for purposes of comparison. Possible repeating sequences are underlined twice. Likely regions of homology are enclosed in boxes; deviating residues likely to have resulted from single base changes in the codon of the principal amino acid are underlined. * Indicates first residue of N-terminal sequence. ** Indicates last residue identified in repeated cycles of Edman degradation. Single-letter code for amino acids: A, alanine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; Y, tyrosine; V, valine, W, tryptophan.

exception of the ω-5 component from 'Justin', which had about 10% more glutamine (plus glutamic acid) and about 10% less proline than the other components. This is in accord with the results of Charbonnier [23], who found that the faster-moving ω-gliadins from the cultivar 'Cappelle', which had similar mobilities in lactate electrophoresis to our ω-5 component, also had similar compositions to our component. The slower-moving gliadin components studied by Charbonnier [23] and the C secalins of Charbonnier et al. [32] had compositions similar to our other ω-gliadin and ω-secalin components.

All these ω-type prolamins had high contents of glutamine (42–53 mol percent), proline (20–31 mole percent) and phenylalanine (7–9 mol percent). Together, these three amino acids represent more than 75 mol percent of all the amino acids. This is rather more than in α, β- and γ-gliadins [33] and in 'B' hordeins [34], where the combined amounts of these three amino acids are closer to 60 mol percent. The ω-type prolamins are also notable for having no cysteine or cystine and no or little methionine [5,21,23].

N-terminal amino acid sequences

The sequences we obtained for our preparations are given in Fig. 3, where they are compared with previously published sequences [19] for a 'C' hordein component (code 1) and the ω-1 gliadin from *T. monococcum* (codes 4, 5). The sequences of the ω-1 (code 15) and ω-5 (code 16) components of 'Justin' and the ω-1 component (code 12) of 'Edmore' were based on duplicate analyses, whereas other sequences in this study represent single analyses because of lack of purified proteins (an elaborate purification procedure frequently yields 1 mg or less of purified ω-gliadin from 100 g of flour — when successful at all — and we analyzed as much as 10 mg of purified component in many single analyses). In most cases, we sequenced equivalent proteins from different cultivars (for example. ω-2 of 'Justin' and ω-2 of 'Chinese Spring') and when we did, identical sequences were always obtained. We considered this sufficient confirmation of the results. The sequences shown in Fig. 3 were aligned to improve homologies.

Initial yields were in the range 60–80% for all analyses except for the 'Mindum' ω-2 gliadin (code 9) and the *T. monococcum* ω-2 and ω-3 gliadins (codes 6, 7), which were 30, 10 and 5% respectively. Yield analysis was based on initial sample weight and an estimated M_r for the proteins. The analyses of the three samples that gave low yields were all carried out on submilligram amounts of protein. Because it is difficult to weigh such samples accurately and contamination of them by lint, salts, moisture and other materials can be significant, and because we did not attempt to deal with these problems in a systematic way, the yields may not be as low as they appear. We cannot rule out the possibility, however, that some of the protein in these samples was blocked at the N-terminus so that the sequence we obtained was not representative of the protein evident in the gel pattern, but of a contaminating protein. Even so, the presence of the observed sequence is of importance in itself.

Also, with the *T. monococcum* samples (ω-2 and ω-3), extremely low levels of PTH-amino acids were analyzed (less than 1 nmol), which increased the possibility of error. In these two analyses, several amino acids were noted in the first cycle. Although alanine or arginine would have been expected, these did not seem to be present, but rather threonine, tyrosine and proline were apparently present. Because artifactual peaks have occasionally been noted in the first cycle analyses, we did not feel that these apparent identifications were necessarily accurate. Cycles 2, 3 and 4 yielded single amino acids, arginine, glutamine and leucine, respectively, and we were reasonably confident of those identifications. Clearly, further work needs to be done to clarify the situation with regard to the types of ω-gliadins found in *T. monococcum*.

The analysis of the preparation from *Ae. squarrosa* (code 17), which was separated into two components by lactate-polyacrylamide gel electrophoresis (Fig. 1), gave several strong peaks in the first cycle. PTH-amino acids were determined only by gas chromatography for this one analysis. Because this method does not determine PTH-lysine and PTH-arginine well (both of which are possible N-terminal amino acids for the mixture) and because we suspected that one or more of the peaks in the first cycle were artifacts, we made no identification for the first cycle. In cycles 2 and 4, only two major peaks were evident in the chromato-

grams and further sequencing indicated only one sequence, as can be seen in Fig. 3.

C-terminal sequencing

An attempt was made to determine the C-terminal amino acid sequences of four of our gliadin components (codes, 10, 12, 14, 15) by digestion with carboxypeptidase Y. No amino acids were released from these components even though an analysis of a 'C' hordein component carried out at the same time gave a normal digestion in accordance with previous results for 'C' hordeins [5]. Digestion of the ω-secalin component from rye (code 2) yielded valine, but no other amino acids. These proteins seem to be resistant to digestion with carboxypeptidase Y, but this needs further investigation. It is of interest that valine is also the C-terminal amino acid of 'C' hordein [5].

Discussion

Types of N-terminal sequence

We have found three new types of N-terminal amino acid sequence for ω-gliadins and have found that two previously described N-terminal sequences, characteristic of the 'C' hordeins of barley [5,13–15,19] and of an ω-gliadin from *T. monococcum* [19], are also present in ω-secalins from rye and in some ω-gliadins of tetraploid and hexaploid wheats. These five types of N-terminal sequence for ω-gliadins, 'C' hordeins and ω-secalins are apparently variants of one main type, as we shall discuss below. These results extend the number of types of gliadin sequences to three, the α-type and γ-type [3,4] and the ω-type. We shall designate the variants of the ω-type sequences the KEL-variant, the RQL-variant, the ARQ-variant, the ARE-variant and the SRL-variant on the basis of the single-letter symbols for the first three amino acids of their sequences (see Fig. 3 for an explanation of the single-letter code). On this basis, the α-type sequence might be designated VRV-type and the γ-type sequences would have NIQ-variants or NMQ-variants.

The KEL-variant

This sequence was found for the slowest-moving gliadin component (lactate-polyacrylamide gel electrophoresis) of each of two varieties of *T. aestivum* and each of two varieties of *T. durum*. It is probably also characteristic of an ω-gliadin from *Ae. squarrosa,* although we failed to identify an N-terminal lysine residue for reasons discussed in the Results section. The presence of the KEL-variant sequence in *Ae. squarrosa*, considered the donor of the D genome to *T. aestivum* [35], is in accord with the assignment of the genes coding for the corresponding protein in 'Chinese Spring' to chromosome 1D in this cultivar [31]. Because *T. durum* is a tetraploid wheat of the AABB type and has no D genome, proteins with the KEL-variant sequence in *T. durum* must be coded by genes located on chromosomes of the A or B genomes. This might reflect introgression [36,37] of some genetic material from *Ae. squarrosa* into the A or B genomes of *T. durum*, which has been considered a possibility [38,39]. The KEL-variant sequence was apparently absent from *T. monococcum; T, monococcum* is closely related to the donor of the A genome to *T. durum* and *T. aestivum* [35].

Another possibility, however, is that the KEL-variant sequence arose in conjunction with the divergence of *Triticum* and *Aegilops* so that this type of sequence might be found in most (possibly all) species of *Aegilops*, but not in species of *Triticum*. This hypothesis would locate the genes coding for KEL-variant proteins in the B and D genomes (probably derived mainly from species of *Aegilops*) of polyploid wheats. Further experimental work is needed to test these speculations.

The SRL-variant

This sequence was present in an ω-gliadin from 'Justin' that gave a relatively intense, fast-moving band in the ω-region of the lactate-polyacrylamide gel electrophoresis patterns. The more intense staining relative to the other ω-gliadins of 'Justin' may reflect a relatively high concentration of basic groups (lysine, histidine and arginine) that bind dye molecules rather than a greater amount of this protein. This higher concentration of basic groups (about twice that of the ω-1 gliadin from 'Justin') is reflected in the amino acid composition (Table II) and in the presence of one lysine, one histidine, and two arginine residues among the first 12 residues of the N-terminal sequence (Fig. 3).

The SRL-variant ω-gliadin of 'Justin' is probably coded by a gene located on chromosome 1B by

analogy with a band of identical mobility in the electrophoretic pattern of the cultivar 'Cheyenne' [11]. It is possible that this sequence is specific for the B genome, but this possibility needs further investigation.

The RQL-, ARQ-, and ARE-variants

The essential homology of these three variants seems fairly certain when the sequences are compared (Fig. 3). One of the variants of this type (RQL) has been found in barley and *T. monococcum* [19], and we now report finding this variant in rye. We have found the other two variants in tetraploid wheats (ARQ) and hexaploid wheats (ARE). Because barley and then rye are thought to have diverged from some common line before it gave rise to *Triticum* and *Aegilops* [40,41] and because the RQL-type is characteristic of the 'C' hordeins of barley and the ω-secalins of rye, it may be the ancestral type.

Homologies among components

Two regions of probable sequence homology among most of the components examined so far are enclosed in boxes in Fig. 3. These regions consist of six or seven residues in each case. The first box (codes 1–11) represents a homology that is probably characteristic of all proteins with RQL-, ARQ- or ARE-variations of the ω-type N-terminal sequence. The second box may represent a homology shared by all ω-type prolamins, which differ from other prolamins in wheat, rye and barley by having greater proportions of glutamine and proline in their compositions and by having no cystine or cysteine [19,21,23]. We failed to extend all of our sequences far enough to include this region, but it was evident in 11 of the 17 different preparations sequenced and probably present in one other component. This latter component (the single representative of the SRL-variant) differed at two of the seven residues proposed to be homologous, but both of the differences correspond to single base changes in the codons of the homologous sequence.

The ARE-variant type genes may have given rise to KEL-variant genes by deletion of part of the DNA from the 5′ end of the gene so that transcription was initiated at a site 24 bases farther into the gene. The situation is complicated by the probability that these proteins are synthesized as larger precursors [43,44], and modified post-translationally by cleavage of a signal peptide from the N-terminal end of the polypeptide chain. The difference between ARE-variant proteins and KEL-variant proteins might be related to a change in the site of post-translational cleavage of a signal peptide rather than a change in the site at which transcription is initiated, but these possibilities cannot be evaluated at present.

It seems likely on the basis of the sequences presented here that all ω-type prolamines have arisen from a common ancestral gene. Homologous genes of different species and duplicated genes within a species have diverged from one another not only through point mutations, but also by insertions and deletions of portions of the gene DNA. The differences in M_r that we found for our components (Table I), which ranged in apparent M_r from $44 \cdot 10^3$–$74 \cdot 10^3$, could have resulted from insertions and deletions of portions of the gene DNA — perhaps through unequal crossing-over, as has been suggested by Lawrence and Shepherd [45] in regard to genes coding for glutenin proteins, or through corrections of mismatched DNA that do not involve crossing-over. Insertion and deletion may be an especially important mechanism for the development of diversity among genes coding for gliadins and related storage proteins; such a mechanism could explain why there are so many components with apparently different M_r values [39], but with similar N-terminal and C-terminal amino acid sequences [4,5].

Variable regions

In contrast to the regions of apparently homologous sequence, there are positions (posssibly regions) where many of the components differ in sequence. For example, following the alignment we have chosen for the sequences of Fig. 3, the amino acid corresponding to residue 7 in the sequence of the 'C-1' hordein is serine, whereas it is glutamic acid in the components from rye, aspartic acid in the components from *T. monococcum*, and asparagine in the components from *T. durum* and *T. aestivum*. Other positions and regions with lesser variation may be noted in Fig. 3. The most likely

explanation of these variable positions is that they are less functionally important than other positions so that the organism has tolerated point mutations at the corresponding codons in the gene more readily. This implies that there are functional constraints on the evolution of genes for gliadins [46] and related prolamins even though the relationship between structure and function may be less crucial for these proteins than for many other proteins, such as enzymes, as has been pointed out by Booth and Ewart [21].

Repeating sequences

The five-residue amino acid sequence -proline-glutamine-glutamine-proline-tyrosine- (P-Q-Q-P-Y) is repeated in the ω-1 sequence from *T. monococcum* (codes 4, 5; Fig. 3) and in the ω-1 sequence from 'Justin' (code 15). In addition, other similar five-residue sequences with differences that correspond mostly to single base changes in the codons are present in many of the sequences shown in Fig. 3 (-Q-Q-Q-P-F-, -P-Q-Q-Q-F-, -P-Q-Q-S-F-, -P-Q-Q-L-Y-, -P-Q-Q-S-Y and -L-Q-Q-P-Y-); these sequences are underlined in Fig. 3. Although no exact five-residue homologies to these sequences are present in the N-terminal sequences of α- and γ-type gliadins (Fig. 3), some similar sequences, such as S-Q-Q-Q-P- in α-type gliadins and P-Q-Q-Q-L in γ_2-gliadin, may be noted.

Repeated duplications of short DNA sequences that included the codons corresponding to the amino acids of these five-member seqences (predominantly glutamine and proline) within the context of an existing gene, or followed by introduction of transcription controls, may have given rise to an ancestral gene for the ω-type prolamins discussed here and for other high-glutamine, high-proline type storage proteins [46]. It is also possible that such a process occurred a number of times with an independent occurrence giving rise to zein genes, for example, which also seem to exhibit repeated sequences [8,9]. This is a plausible mechanism for the concentration of similar codons into a single gene, as is necessary to code for proteins containing up to 75 mol percent glutamine and proline in the case of ω-gliadins. A similar process has probably been involved in the formation of some other maize proteins [47].

We note that amplifications of DNA sequences along with their translocation or deletion from the genome have been frequent events in the evolution of higher plants [48]. Although these processes have been studied mainly in connection with non-coding DNA [48], similar mechanisms may have contributed to the formation of genes for many proteins, including collagen [49] and the serum albumin family [50]. Although the genes for collagen arose more than 500 million years ago [49], the genes coding for gliadins and similar prolamins may be very young genes, having developed recently in the course of evolution. This suggestion is based on the occurrence of such high-glutamine, high-proline prolamins only in the seeds (caryopses) of grasses combined with indications that the grass family (Gramineae) of the flowering plants (Angiosperms) is of recent origin. Fossil evidence of grass evolution is sparse, but indicates that grasses had become established by 25–40 million years ago [51]. Angiosperms appeared about 100 million years ago [52]. If genes for gliadins and similar prolamins originated in connection with the development of grasses, the time of their origin should be limited approximately by these dates. Even allowing for a preceding period of development twice as long, the time of origin would be within about 200 million years — still recent compared to the origin of genes such as those for collagen.

Further sequencing will be necessary to determine the extent to which repeated sequences occur in ω-type prolamins. We were unable to extend our N-terminal sequences beyond about 28 residues at best; we think that this results from limitations inherent in the nature of these proteins, because we can sequence other proteins, sperm whale myoglobin, for example, at least twice as far. The absence of extensive repetition of a simple peptide sequence in these proteins would not indicate the incorrectness of our hypothesis that gliadin genes evolved through the duplication of simple sequences. This process probably occurred in a stepwise manner with larger, possibly diverged, sequences being duplicated in later steps, and eventually involved duplication of complete genes [46].

Acknowledgements

We thank W.M. Fitch for helpful discussion, E.R. Sears for the aneuploids of 'Chinese Spring', B.L. Jones for the sample of *T. monococcum*, V. Youngs for the sample of 'Edmore', and A. Noma for amino acid analyses. P.R.S. and J.-C.A. were visiting scientists in receipt of NATO Postdoctoral Research Fellowships. This work was supported in part by USDA Competitive Research Grant No. 59-2068-0-1-526-0. Reference to a company or product name is for purposes of information only and does not imply approval or recommendation by the U.S. Department of Agriculture to the exclusion of others that may also be suitable.

References

1 Kasarda, D.D. (1981) in Food, Nutrition, and Evolution: Food as an Environmental Factor in the Genesis of Human Variability (Walcher, D. and Kretchmer, N., eds.), pp. 201–216, Masson Publishing, New York
2 Bietz, J.A. (1982) Biochem. Genet. 20, 1039–1053
3 Bietz, J.A., Heubner, F.R., Sanderson, J.E. and Wall, J.S. (1977) Cereal Chem. 54, 1070–1083
4 Autran, J.-C., Lew, E.J.-L., Nimmo, C.C. and Kasarda, D.D. (1979) Nature 282, 527–529
5 Shewry, P.R., Lew, E.J.-L. and Kasarda, D.D. (1981) Planta 153, 246–253
6 Burr, F.A. (1979) in Seed Protein Improvement in Cereals and Grain Legumes, Vol. I, pp. 159–164, IAEA, Vienna.
7 Walsh, K.A., Ericsson, L.H., Parmelee, D.C. and Titani, K. (1981) Annu. Rev. Biochem. 50, 261–284
8 Geraghty, D., Pfeiffer, M.A., Rubenstein, I. and Messing, J. (1981) Nucleic Acids Res. 9, 5163–5174
9 Pederson, K., Devereux, J., Wilson, D.R., Sheldon, E. and Larkins, B.A. (1982) Cell 29, 1015–1026
10 Wrigley, C.W. and Shepherd, K.W. (1973) Ann. N.Y. Acad. Sci. 209, 154–162
11 Mecham, D.K., Kasarda, D.D. and Qualset, C.O. (1978) Biochem. Genet. 16, 831–853
12 Bietz, J.A., Paulis, J.W. and Wall, J.S. (1979) Cereal Chem. 56, 327–332
13 Bietz, J.A. (1981) Cereal Chem. 58, 83–85
14 Schmitt, J.M. and Svendsen, I. (1980) Carlsberg Res. Commun. 45, 143–148
15 Shewry, P.R., March, J.F. and Miflin, B.J. (1980) Phytochemistry 19, 2113–2115
16 Kasarda, D.D., Da Roza, D.A., and Ohms, J.I. (1974) Biochim. Biophys. Acta 351, 290–294
17 Kasarda, D.D., Nimmo, C.C. and Bernardin, J.E. (1974) in Coeliac Disease: Proc. 2nd Int. Symp. (Hekkens, W.T.J.M. and Pena, S., eds.), pp. 25–36, Stenfert Kroese, Leiden
18 Patey, A.L., Evans, D.J., Tiplady, R., Byfield, P.G.H. and Matthews, E.W. (1975) Lancet ii, 718
19 Shewry, P.R., Autran, J.-C., Nimmo, C.C., Lew, E.J.-L. and Kasarda, D.D. (1980) Nature 286, 520–522
20 Sears, E.R. (1954) Res. Bull. 572, Missouri Agricultural Experiment Station, Columbia, MO
21 Booth, M.R. and Ewart, J.A.D. (1969) Biochim. Biophys. Acta 181, 226–233
22 Kasarda, D.D., Bernardin, J.E. and Qualset, C.O. (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 3646–3650
23 Charbonnier, L. (1974) Biochim. Biophys. Acta 359, 142–151
24 Ewart, J.A.D. (1981) J. Sci. Food Agric. 32, 572–578
25 Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91
26 Pisano, J.J., Bronzert, T.J. and Brewer, H.B., Jr. (1972) Anal. Biochem. 45, 43–59
27 Kulbe, K.D. (1974) Anal. Biochem. 59, 564–573
28 Jeppson, J.-O. and Sjoquist, J. (1967) Anal. Biochem. 18, 264–269
29 Bhown, A.S., Mole, J.E., Weissinger, A. and Bennett, J.C. (1978) J. Chromatogr. 148, 532–535
30 Schmitt, J.M. (1979) Carlsberg Res. Commun. 44, 431–438
31 Shepherd, K.W. (1968) in Proc. 3rd Int. Wheat Genet. Symp. (Finley, K.W. and Shepherd, K.W., eds.), pp. 86–96, Australian Acad. Sci., Canberra
32 Charbonnier, L., Terce-Laforgue, T. and Mossé, J. (1981) J. Agric. Food Chem. 29, 968–973
33 Kasarda, D.D., Bernardin, J.E. and Nimmo, C.C. (1976) in Advances in Cereal Science and Technology (Pomeranz, Y., ed.), pp. 158–236, American Association of Cereal Chemists, St. Paul, MN
34 Shewry, P.R., Field, J.M., Kirkman, M.A., Faulks, A.J. and Miflin, B.J. (1980) J. Exp. Bot. 31, 393–407
35 Morris, R. and Sears, E.R. (1967) in Wheat and Wheat Improvement (Quisenberry, K.S. and Reitz, L.P., eds.), 19–87, American Society of Agronomy, Madison, WI
36 Zohary, D. and Feldman, M. (1962) Evolution 16, 44–61
37 Vardi, A. (1973) in Proc. 4th Int. Wheat Genet. Symp. (Sears, E.R. and Sears, L.M.S., ed.), pp. 131–141, Agricultural Experiment Station, University of Missouri, Columbia, MO
38 Mann, S.S. (1975) in Prairie: A Multiple View (Wali, M.K., ed.), pp. 255–281, University of North Dakota Press, Grand Forks, ND
39 Cole, E.W., Fullington, J.G. and Kasarda, D.D. (1981) Theor. Appl. Genet. 60, 17–30
40 Vedel, F., Lebacq, P. and Quetier, F. (1980) Theor. Appl. Genet. 58, 219–224
41 Smith, D.B. and Flavell, R.B. (1974) Biochim. Genet. 12, 243–256.
42 Shewry, P.R., Field, J.M., Lew, E.J.-L. and Kasarda, D.D. (1982) J. Exp. Bot. 33, 261–268
43 Greene, F.C. (1981) Plant Physiol. 68, 778–783
44 Matthews, J.A. and Miflin, B.J. (1980) Planta 149, 262–268
45 Lawrence, G.J. and Shepherd, K.W. (1980) Aust. J. Biol. Sci. 33, 221–233
46 Kasarda, D.D. (1980) Ann. Technol. Agric. (France) 29, 151–173
47 Esen, A., Bietz, J.A., Paulis, J.W. and Wall, J.S. (1982) Nature 296, 678–679

48 Flavell, R. (1980) Annu. Rev. Plant Physiol. 31, 569–596
49 Bornstein, P. and Sage, H. (1980) Annu. Rev. Biochem. 49, 957–1003
50 Ohno, S. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 7657–7661
51 Thomasson, J.R. (1980) Iowa State J. Res. 54, 301–317
52 Cronquist, A. (1968) The Evolution of Flowering Plants, Houghton Mifflin, Boston, MA
53 Solari, R.M. and Favret, E.A. (1971) in Barley Genetics II (Nilan, R., ed.), pp. 23–31, Washington State University Press, Pullman, WA
54 Shepherd, K.W. (1973) in Proc. 4th Int. Wheat Genet. Symp. (Sears, E.R. and Sears, L.M.S., eds.), pp. 745–760, Agricultural Experiment Station, University of Missouri, Columbia, MO.