Biochem. J. (1990) 267, 1–12 (Printed in Great Britain)

REVIEW ARTICLE

The prolamin storage proteins of cereal seeds: structure and evolution

Peter R. SHEWRY* and Arthur S. TATHAM†

*Department of Agricultural Science, University of Bristol, AFRC Institute of Arable Crops Research, Long Ashton Research Station, Bristol BS18 9AF, and †AFRC Institute of Arable Crops Research, Rothamsted Experimental Station, Harpenden AL5 2JQ, U.K.

INTRODUCTION

Prolamins are the major storage proteins in most cereal seeds, and as such are an important source of dietary protein for man and his livestock. In addition the prolamins of wheat are the major components of gluten, the properties of which determine the quality of wheat flour for various technological processes including breadmaking [1].

Gluten was one of the first proteins to be studied, by Beccari in 1745 [2]. Subsequent work showed that it could be divided into two fractions that were soluble or insoluble in aq. alcohol [3,4], and that similar alcohol-soluble proteins were present in rye and barley [5,6]. This laid the basis for the systematic studies of T. B. Osborne, carried out around the turn of the century and described in [7].

Osborne recognized that the alcohol-soluble proteins of cereal seeds constituted a distinct group, and coined the name prolamin to reflect their high contents of proline and amide nitrogen (now known to be derived from glutamine). In wheat the prolamins are called gliadins (based on the name originally proposed in 1819 [4]), while in other cereals they are given trivial names based on their Latin generic names: zein of maize (*Zea mays*), hordein of barley (*Hordeum vulgare*) and secalin of rye (*Secale cereale*). The alcohol-insoluble fraction of wheat gluten is called glutenin and the corresponding fractions from other species glutelins.

Although prolamins and glutelins were initially thought to be distinct groups of proteins, we now know that many glutelin proteins are closely related structurally to prolamins, but are not soluble in aq. alcohols because they form high-M_r polymers stabilized by inter-chain disulphide bonds. Because the individual reduced subunits of the polymers are alcohol-soluble (they are in fact often extracted in aq. alcohols in the presence of a reducing agent) and rich in proline and glutamine they are now also considered to be prolamins (see [1]).

Prolamins are deposited in discrete protein bodies in the developing starchy endosperm, and have no known function apart from storage. They generally account for about half of the protein present in the mature grain, exceptions being oats and rice in which smaller amounts of prolamins (about 10% of the total protein) are accompanied by other storage proteins (see [8]). Prolamin fractions also consist of many individual proteins, which can usually be classified into a small number of groups or families. Although all are soluble (as individual subunits) in aq. alcohols, they vary considerably in their other properties, notably M_r and pI values and amino acid compositions. For example, the combined proportions of the two most characteristic amino acids, glutamine and proline, range from about 30 to 70 mol%.

With the exception of rice, the major cereals fall into two groups. The temperate cereals comprise barley, wheat, rye and oats, and the tropical cereals maize, sorghum and millets. Only the prolamins of wheat, barley and maize have been studied in detail at the molecular and physicochemical levels, and we will therefore focus on these. We will, however, also draw comparisons with the prolamins of other species where appropriate.

THE PROLAMINS OF MAIZE AND RELATED SPECIES

Although only the zeins of maize have been studied in detail, available evidence indicates that sorghum [9,10], pearl millet [10] and Job's tears (*Coix lachryma-jobi*) [11] contain essentially similar groups of prolamins. SDS/PAGE of a total zein preparation extracted in the presence of a reducing agent shows several groups of bands with M_r values of about 10000, 14000, 16000, 19000, 22000 and 28000 (Fig. 1*a*). Although the classification and nomenclature of zeins have been a contentious issue, the recent availability of complete amino acid sequences enables us to classify them into four major groups, which differ in their amino acid compositions (Table 1). Whereas all groups are fairly rich in glutamine and proline and devoid of lysine and tryptophan, they differ in their contents of other amino acids such as methionine (high in β- and γ-zeins), cysteine and histidine (both high in γ-zeins). These groups will now be discussed in turn, using the nomenclature of Esen [12].

α-Zeins

The α-zeins comprise the major bands of M_r 19000 and 22000, which can each be separated into a number of components by isoelectric focusing (Fig. 1*b*) or electrophoresis at low pH. They together account for about 75–85% of the total zein fraction [12], and consist of monomers and oligomers [13]. Although most of the oligomers are readily soluble in aq. alcohols, some are only extracted as subunits after reduction of disulphide bonds.

A number of complete amino acid sequences derived from cloned cDNAs and genes [14–16] show that the M_r-19000 and M_r-22000 subclasses consist of about 210–220 and 240–245 residues respectively, with true molecular masses of about 23000–24000 and 26500–27000. They also show that the members of these subclasses have a similar structure, which is summarized in Fig. 1(*c*). Unique domains of about 36–37 and 10 residues are present at the *N*- and *C*-termini respectively, separated by a repetitive domain consisting of blocks of about 20 residues. These repeats are degenerate, and it is not possible to recognize an unambiguous consensus motif (see Fig. 1*c*). They are rich in leucine and alanine, accounting for the unusually high proportions of these residues in the whole zeins (Table 1). The complete amino acid sequences also show the presence of only one (the M_r-22000 zeins) or two (the M_r-19000 zeins) cysteine residues, which are always in the *N*-terminal domain.

In Fig. 1(*c*) the size difference between the M_r-19000 and M_r-

Abbreviations used: i.e.f., isoelectric focusing, STM, scanning tunnelling microscopy; TNS, 2-*p*-toluidinylnaphthalene-6-sulphonate.

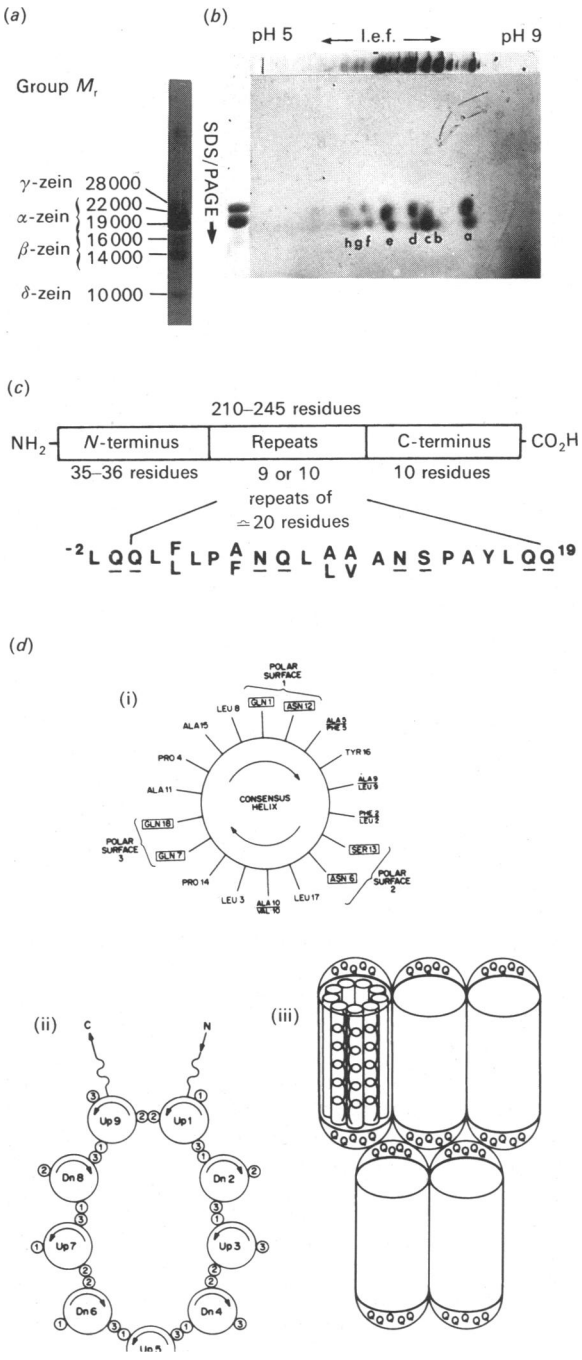


Fig. 1. Structures of zeins of maize

(*a*) SDS/PAGE of a reduced and pyridylethylated total zein fraction, showing the α-, β-, γ- and δ-zeins. (*b*) Two-dimensional analysis (i.e.f.–SDS/PAGE) of reduced and pyridylethylated α-zein, showing polymorphism, a–h indicate components that give clearly resolved bands on one-dimensional i.e.f. (taken from [23]. (*c*) Summary of the amino acid sequences of the M_r-19000 and M_r-22000 α-zeins, based on Argos *et al.* [17], except that the M_r-20000 group is shown as having an additional repeat unit, rather than an insertion in the *C*-terminal region. Polar residues in the consensus repeat sequence are underlined. (*d*) The structural model for α-zeins proposed by Argos *et al.* [17]. (i) Helical wheel representation of the consensus repeat motif. (ii) A possible nine-helix protein with the axes orthogonal to the plane of the Figure. The hydrogen-bonding polar residue segments (shown as small circles) are numbered as in (i); 'up' and 'down' indicate directions of the anti-parallel helices. (iii) A model for the arrangement of zein proteins within and between planes, with hydrogen bonding between Gln residues in the turn regions between helices. Taken with permission from Argos *et al.* [17].

22000 α-zeins is shown to result from the insertion of an addition repeat unit at the *C*-terminal end of the repetitive domain of the latter group, making a total of ten repeat units compared to nine in the M_r-19000 group. This interpretation is not accepted by Argos and co-workers [17], who consider the insertion in the M_r-22000 group to form a *C*-terminal turn region.

The secondary structure contents of purified α-zeins have not been determined, but analyses by optical rotary dispersion and circular dichroism of fractions enriched in α-zeins have shown high contents of α-helix (≃ 45–60%) and low contents of β-sheet when dissolved in aq. alcohols [17,18].

Despite the availability of extensive protein sequences, there has only been one attempt to use this information to develop a structural model for the α-zeins. This model of Argos *et al.* [17] is based on the presence of nine homologous repeat units in the M_r-19000 and M_r-22000 zeins (see above). They proposed that each repeat unit forms an α-helix (Fig. 1*d*, part i), the nine helices being arranged in an anti-parallel ring stabilized by intramolecular hydrogen bonding between polar residues on adjacent helices (Fig. 1*d*, part ii). Further polar residues on the helices form intermolecular hydrogen bonds with other zein molecules in the same plane, while the glutamine residues at the ends of the repeat units form turn regions that could form hydrogen bonds with zeins in adjacent planes (Fig. 1*d*, part iii). This is an attractive model as it allows regular packing of the molecules within the protein bodies, but is supported by little experimental evidence. The amount of α-helix is consistent with the higher values determined by c.d. spectroscopy, but most of the putative α-helices contain one or two residues of proline, which is considered to be a helix breaker [19]. In addition, the protein would be only slightly asymmetric, which is not consistent with early physical studies of the conformations of zein fractions and subfractions. These all indicated a more extended rod-shaped conformation, with precise axial ratios varying from about 7:1 to 25:1 depending on the zein fraction and on the method used [20–22].

β-Zein

This consists of the M_r-14000 and M_r-16000 bands, which can be separated into several individual polypeptides by i.e.f. [23]. It accounts for about 10–15% of the total zein fraction [12], but is only extractable in the presence of a reducing agent.

Two cDNAs and one gene encoding β-zeins of M_r about 15000 have been characterized [24,25]. The encoded proteins consist of 160 residues with true M_r values of about 17500, and are rich in methionine (18 residues) and cysteine (7 residues). At least some of the latter are presumably involved in interchain disulphide bonds, resulting in the insolubility of the native protein in aq. alcohols. They show no sequence similarity to the α-zeins and, unusually for prolamins, do not contain repeated sequences. It is not possible to recognize a distinct domain structure on the basis of the sequences, but the methionine residues are clustered, most notably between positions 121 and 132 (six methionines out of twelve residues).

C.d. studies of a purified β-zein fraction showed little α-helix, with most of the sequence being β-sheet (calculated as 33% and 55% using two different deconvolution procedures) and aperiodic structure (β-turn and random coil) [24].

γ-Zeins

γ-Zeins consist of the M_r-28000 band, which has also been called glutelin-2 [26] or reduced soluble protein [23]. The latter name refers to the fact that it is only extracted in the presence of a reducing agent, but is readily soluble in water when reduced. It accounts for about 5–10% of total zein [12], and i.e.f. shows a number of components with basic pI values [23].

Table 1. Amino acid compositions (residues/mol) of total zein and individual zein proteins: *a*-zein (M_r 19000, M_r 22000), *β*-zein (M_r 15000), *γ*-zein (M_r 28000) and *δ*-zein (M_r 10000)

Based on [24, 28, 31, 115, 116, 117].

	Total zein	α-Zein		β-Zein	γ-Zein	δ-Zein
		M_r 19000 (A30)	M_r 22000 (pZ22.1)	(M_r 15000)	(M_r 28000)	(M_r 10000)
Asp }		0	0	1	0	1
Asn }	3.9	10	12	3	0	3
Thr	3.2	5	8	4	9	5
Ser	5.9	15	16	8	8	8
Glu }		1	1	3	2	0
Gln }	20.7	41	50	26	30	15
Pro	13.0	23	22	14	51	20
Gly	3.8	5	4	14	13	4
Ala	11.3	29	34	22	10	7
Cys	1.9	2	1	7	15	5
Val	4.5	5	15	4	15	5
Met	1.9	0	5	18	1	29
Ile	2.9	9	8	1	4	3
Leu	15.3	43	44	16	19	15
Tyr	3.9	8	8	14	4	1
Phe	3.8	13	9	0	2	5
His	2.1	2	3	0	16	3
Lys	0.2	0	0	0	0	0
Arg	1.7	2	2	5	5	0
Trp	nd.	0	0	0	0	0
Amino acid residues		213	242	160	204	129
Absolute M_r		23329	26532	17279	21824	14431

Several cloned DNAs for γ-zeins have been characterized [26–29]. Most encode proteins of 204 residues with true M_r values of about 21800, and with a clear domain structure. Unique *N*-terminal and *C*-terminal regions of 11 and 156 residues respectively flank a repetitive domain consisting of eight conserved hexapeptides with the sequence Pro-Pro-Pro-Val-His-Leu. The hydrophilic nature of this domain may be responsible for the solubility of the reduced protein in water. Prat *et al.* [26] have recognized further regions within the *C*-terminal domain, including a Pro-Xaa region (with proline at every second position) between residues 70 and 91, and a cysteine-rich region between residues 92 and 148. Prat *et al.* [29] also isolated a cDNA encoding a γ-type zein consisting of only 164 residues with a true M_r of about 17800. In contrast to the other γ-zeins the repetitive domain consisted of only two complete hexapeptides and two truncated hexapeptides (of five and three residues).

Wu *et al.* [18] determined the secondary structure content of a water-soluble zein fraction consisting predominantly of γ-zein by using o.r.d. and c.d. spectroscopy. They calculated the content of α-helix as between 19 and 32% and of β-sheet as between 11 and 34%, depending on the procedure used. More detailed studies of the repetitive domain were reported by Pons *et al.* [30], using ^1H, ^{13}C and combined $^{14}C/^1H$ chemical shift correlation two-dimensional n.m.r. spectroscopy of a synthetic peptide (Boc Val-His-Leu-Pro-Pro-Pro-OH). This was in a random coil all-*trans* configuration when dissolved in ^2H_2O and dimethyl sulphoxide, but adopted a poly(L-proline) II-like conformation in non-polar solvents.

δ-Zeins

The recently reported sequence of a gene encoding an M_r 10000 zein [31] shows that it is not closely related to any other prolamin. It is therefore logical to call this component γ-zein. It accounts for only a small proportion of the total zein fraction and, like the β- and γ-zeins, is only extracted under reducing conditions [12]. The protein consists of 129 residues, with a true M_r of about 14400. It does not contain any repeated sequences or have a clear-cut domain structure, but it is possible to recognize a central region that contains 17 of the 29 methionine residues, mostly as Met-Met doublets separated by two or three other residues. The conformation of δ-zein has not been studied.

Structural relationships of zeins

Comparison of the amino acid sequences of more than ten α-zeins of the M_r-19000 and -22000 size classes shows between 60 and 97% homology [12]. There is not, however, any homology with the other zein groups, or with prolamins from other species. Similarly the M_r-10000 δ-zeins appear to form a discrete group, although their high content of methionine and solubility properties would suggest a relationship to the β-zeins. The β-zeins are, in fact, related to the γ-zeins, both belonging to the prolamin superfamily defined by Kreis and coworkers [32,33]. These relationships will be discussed in more detail after reviewing the structures of the prolamins of temperate cereals.

THE PROLAMINS OF WHEAT, BARLEY AND RYE

The prolamin fractions of these three closely related species consist of complex mixtures of components with M_r values by SDS/PAGE ranging from about 30000 to 100000 (Fig. 2). Despite this high level of polymorphism, the individual components can be classified on the basis of their amino acid compositions (Table 2) and sequences into only three groups [34]: the sulphur-rich (S-rich), sulphur-poor (S-poor) and high-molecular-weight (HMW) prolamins (see Fig. 2).

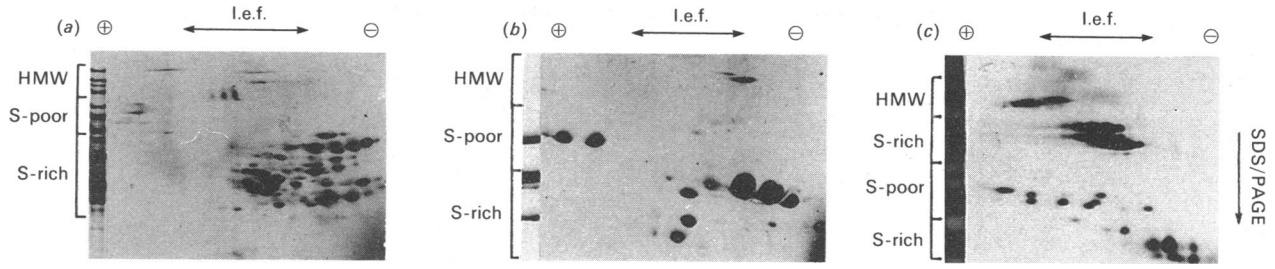


Fig. 2. Two-dimensional (i.e.f.–SDS/PAGE) analyses of reduced and pyridylethylated total prolamin fractions from (*a*) wheat (cv. Chinese Spring), (*b*) Barley (line P 12/1) and (*c*) rye (inbred line MPI 109)

One-dimensional SDS/PAGE separations are shown to the left of the two-dimensional separations. In wheat the HMW, S-poor and S-rich groups correspond to HMW subunits of glutenin, ω-gliadins and a mixture of α-gliadins, γ-gliadins and aggregated S-rich prolamins respectively. In barley the HMW, S-poor and S-rich groups correspond to D hordein, C hordein and a mixture of B hordein and γ-hordein respectively. In rye the HMW, S-poor and S-rich groups correspond to HMW secalins, ω-gliadins and γ-secalins (M_r-75000 and M_r-40000 groups) respectively (taken from [80], [108] and [109]).

Table 2. Amino acid compositions (residues/mol) of individual HMW, S-poor and S-rich prolamins

Based on sequences in [43, 88, 93, 111, 112]. The data for C hordein are calculated on the basis of the M_r (52570) determined by sedimention equilibrium ultracentrifugation, as described in [60]. Absolute M_r values are given for the other proteins. Abbreviation: n.d., not determined.

	HMW subunits		S-poor	S-rich			
	1Dx5	1Dy10	C hordein	γ-Type gliadin	α-Type gliadin	Aggregated gluten protein	B1 hordein
Asp	4	4		2	1	0	0
Asn	0	0	4.4	4	7	2	2
Thr	24	24	4.4	7	4	8	6
Ser	47	42	20.2	17	14	28	13
Glu	15	17		3	5	6	5
Gln	299	206	180.9	93	92	97	83
Pro	109	69	134.3	45	38	33	53
Gly	166	113	1.3	8	7	11	8
Ala	25	23	3.1	7	9	9	7
Cys	5	7	0	8	6	8	8
Val	14	16	4.4	15	13	17	17
Met	2	3	0.9	6	2	5	3
Ile	4	4	11.4	14	13	12	12
Leu	36	24	15.8	17	22	22	22
Tyr	46	34	10.0	3	9	4	7
Phe	2	2	38.6	14	11	11	13
His	4	13	4.8	5	6	3	4
Lys	6	7	0.9	3	1	1	2
Arg	10	13	3.5	2	5	6	7
Trp	9	6	n.d.	3	1	2	2
Amino acid residues	827	627	439	276	266	285	274
M_r	88128	67476	52570	31629	30820	32268	31444

THE HMW PROLAMINS

HMW subunits of wheat glutenin

The HMW subunits account for about 10% of gluten and consist, in European bread wheats, of three, four or five individual proteins [35]. Allelic variation in the number and properties of these subunits (Fig. 3*a*) is associated with variation in breadmaking quality [36,37], and this has provided a stimulus for their study. They are classified into two types on the basis of their M_r values and sequences (see below): x-types which migrate more slowly on SDS/PAGE and have true M_r values of about 83000–88000, and y-types which migrate faster and have true M_r values of about 67000–74000 [38,40]. Their amino acid compositions (Table 2) [39] are characterized by high contents of glycine (14–19 mol%), glutamine (37–39 mol%) and proline (12–14 mol%). Genes for seven different HMW subunits have been isolated and the amino acid sequences of the encoded proteins deduced [40–47]. The proteins vary in length from 627 to 827 residues, with true M_r values of 67500 to 88128.

All of the HMW subunits have a clear domain structure, with non-repetitive domains at the *N*-terminus (81–104 residues) and *C*-terminus (42 residues) flanking a central repetitive domain (Fig. 3*b*). The non-repetitive domains contain most or all of the cysteine residues, with three (x-type) or five (y-type) in the *N*-

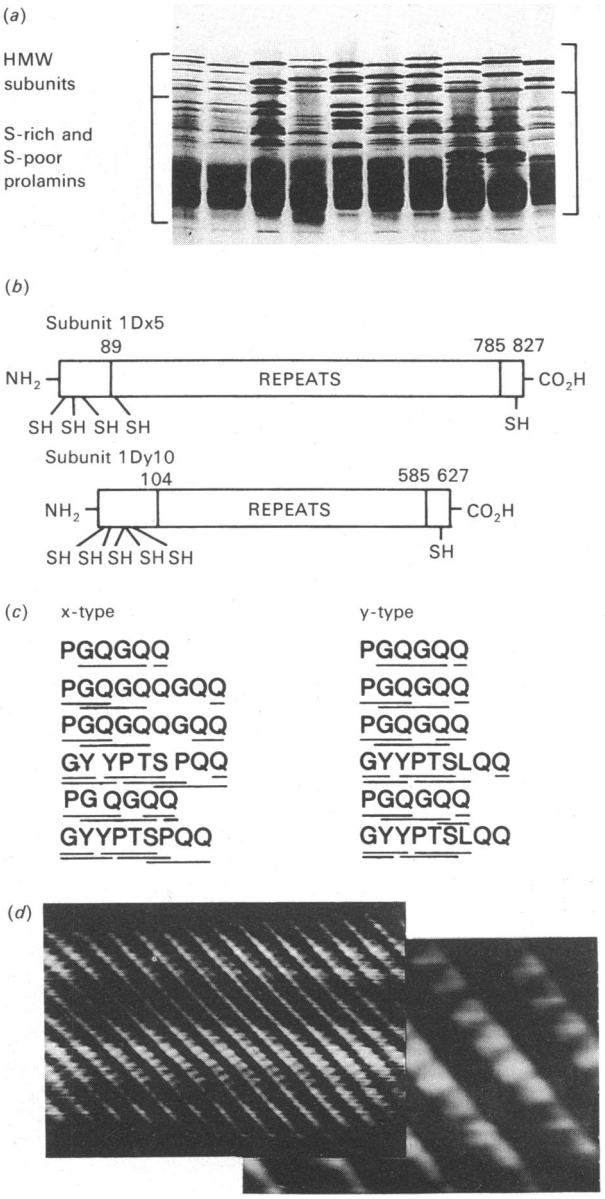


Fig. 3. HMW subunits of wheat glutenin

(*a*) Polymorphism in the SDS/PAGE patterns of HMW subunits in total protein extracts of 10 cultivars of bread wheat (taken from [40]). (*b*) Schematic structures of typical x-type and y-type HMW subunits. The positions of cysteine residues are indicated by SH (based on sequences reported in [43]). (*c*) Consensus repeat motifs of x-type and y-type HMW subunits, showing the positions of predicted β-turns (taken from [40]). (*d*) STM of HMW subunit 20 from *T. durum* (pasta wheat) cv. Bidi 17 (taken from [53]). Left, Fourier-filtered STM image (54.5 nm × 37.5 nm); right, higher magnification (13.7 nm × 9.4 nm).

terminal domain and one in the *C*-terminal domain. Differences in the *N*-terminal domain are due to substitutions and insertions/deletions, the x-type subunits have an 18-residue deletion which results in the loss of two cysteine residues. The *C*-terminal domains differ only in substitutions.

The repetitive domains are based on three different motifs. Hexapeptides (consensus Pro-Gly-Gln-Gly-Gln-Gln) and nonapeptides (consensus Gly-Tyr-Tyr-Pro-Thr-Ser-Pro/Leu-Gln-Gln) are present in both x- and y-type subunits and tripeptides (consensus Gly-Gln-Gln) in x-type subunits only. The hexapeptides often occur in tandem arrays, whereas the tripeptides and nonapeptides always occur in an interspersed fashion. Some subunits also have single cysteine residues towards the *C*-terminal or *N*-terminal end of the repetitive domain (Fig. 3*b*). It has been suggested that the repetitive domains may have evolved by a series of amplification and/or deletion events involving single and multiple blocks of residues (see [40]), and insertions and/or deletions of single residues or parts of blocks are rare.

Structure prediction indicates that the *N*- and *C*-terminal domains are predominantly α-helical, while the central repetitive domains form regularly repeated β-turns [48,49]. Turns are predicted to form both within repeat motifs and spanning the junctions between them (Fig. 3*c*). Some of the predicted turns are overlapping, but it is not known whether all form or only those with the highest probabilities.

C.d. spectroscopy of a group of HMW prolamins [48] from bread wheat and a single subunit from durum wheat [50] dissolved in trifluoroethanol showed α-helical like spectra. Deconvolution of the spectrum of the purified subunit gave about 30% α-helix, which was consistent with the *N*- and *C*-terminal domains being α-helically structured [50]. Little information on the conformation of the repetitive domain was obtained by c.d. spectroscopy of the whole protein, but the presence of β-turns was confirmed by analysis of synthetic peptides. C.d. and i.r. spectroscopy of a peptide corresponding to the hexapeptide repeat motif (Gly-Gln-Gln-Pro-Gly-Gln-Gly) indicated a type II β-turn [51]. Similar studies of two peptides corresponding to the consensus sequence and a major variant sequence at the junctions between hexapeptide and nonapeptide motifs (Gly-Gln-Pro/Gln-Gly-Tyr-Tyr-Pro-Thr-Ser-Pro) also indicated the presence of β-turns, but the type(s) could not be determined from the c.d. spectra due to absorption resulting from interaction of adjacent tyrosine residues [51].

Hydrodynamic studies of the purified pasta wheat subunit indicated that the molecule was rod-shaped [50], the dimensions varying from about 49 nm × 1.8 nm in 50% aq. (v/v) propan-1-ol ($\pm$0.01 M glycine as a counterion) to about 62 nm × 1.5 nm in trifluoroethanol (both at 30 °C).

A structural model was proposed for the HMW subunits, based on the spectroscopic and hydrodynamic studies [50]. This suggested that the regularly repeated β-turns in the central domain formed a loose spiral, which gave a rod-shaped structure to the whole protein. The only experimentally determined structure of this type is the β-spiral formed by a synthetic polypeptide based on a pentapeptide repeat motif of elastin [52]. This consists of regularly organized β-turns and has a diameter of some 1.7–1.8 nm with about 13.5 residues per turn of the helix. Assuming that it forms a spiral with similar dimensions to the β-spiral, the repetitive domain of the HMW subunit purified from pasta wheat would be some 44–45 nm × 1.8 nm. This agrees well with the dimensions calculated from the intrinsic viscosity of the whole protein [50]. Direct evidence for the formation of such a spiral is provided by recent scanning tunnelling microscopy (STM) studies of the pasta wheat subunit. STM of the protein deposited from trifluoroethanol, and imaged without coating, showed parallel rod-like structures with diagonal striations [53]. Fourier transformation gave a diffraction pattern which suggested a high degree of order, and filtering some of the non-periodic noise revealed a spiral structure with a diameter of about 1.95 nm and a pitch of about 1.45 nm (Fig. 3*d*). The lengths of the rods could not be determined as the *N*- and *C*-terminal domains did not image clearly. It appears, therefore, that the repetitive domains of the HMW subunits form a novel supersecondary structure based on repeated β-turns. Whether this structure is involved in the elastic mechanism of wheat gluten is unknown.

Wheat gluten has an unusual combination of two physical properties: elasticity which is associated with the polymeric glutenins, and viscous flow which is associated with the monomeric gliadins. The poor breadmaking performance of many European wheats results from low elasticity, and this is correlated with allelic variation in the HMW subunits which are present and in the amounts of high-M_r polymers that are formed.

The HMW subunits are present in the seed in polymers stabilized by interchain disulphide bonds. These polymers may also contain S-rich prolamin subunits (see below), and have M_r values ranging between about 1×10^6 to 1×10^7. In the HMW subunits the cross-linking sites (cysteine residues) are predominantly in the *N*- and *C*-terminal domains, which would allow the formation of head-to-tail polymers with some branching and cross-linking. Some of the cysteine residues may also form intrachain disulphide bonds, at least in the y-type subunits (N. Bulleid & R. B. Freedman, personal communication). The precise number and distribution of the cross-links would be expected to influence the elastic modulus of gluten, and differences of this type could be responsible for allelic variation in breadmaking quality. However, it has also been suggested that the spiral structure formed by the repetitive domains is intrinsically elastic [49], and that differences in the regularity of this structure are present in HMW subunit proteins associated with good and poor breadmaking quality [47].

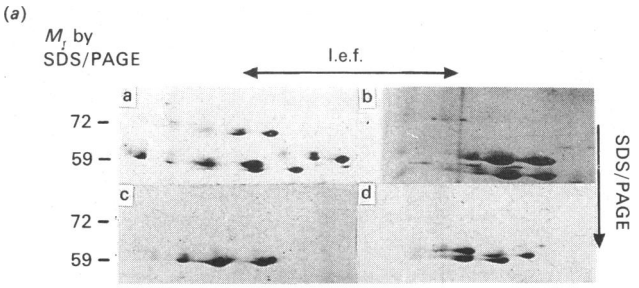
HMW prolamins of barley and rye

The HMW prolamins of barley (D hordein) and rye (HMW secalins) have not been characterized in detail, but available evidence indicates that they are closely related to the HMW subunits of wheat. Their amino acid compositions are rich in glycine, glutamine and proline, [54,55], while their *N*-terminal amino acid sequences show clear similarity to those of the HMW subunits of wheat glutenin [56]. A close relationship is also indicated by the cross-hybridization behaviour of cDNAs and genes for HMW prolamins from the three species ([57]; P. R. Shewry & A. S. Tatham, unpublished results). Preliminary c.d. spectroscopy of a HMW secalin fraction gave spectra which were almost identical to those of the HMW subunits of glutenin, indicating a similar conformation [58].

S-POOR PROLAMINS

C hordein of barley

C hordein consists of between four and 15 individual components, depending on the genotype (Fig. 4*a*) [59]. Their M_r values by SDS/PAGE vary from about 55000 to 70000, but the fractions which have been studied in detail (Fig. 4*a*, part a) have major components with M_r values of about 52000–54000 (as determined by sedimentation equilibrium ultracentrifugation [60] and intrinsic viscosity measurements in 5.9 M-guanidinium chloride [61]). Partial amino acid sequences determined directly [62–64] and deduced from cloned cDNAs [65,66] show unique *N*-terminal and *C*-terminal domains of 12 and 6 residues respectively, flanking a repetitive domain calculated (from the M_r values) as consisting of over 400 residues (Fig. 4*b*). This repetitive domain consists mainly of octapeptides (consensus Pro-Gln-Gln-Pro-Phe-Pro-Gln-Gln) with at least three related pentapeptides (consensus Pro-Gln-Gln-Pro-Tyr) close to the *N*-terminus. The consensus composition of the octapeptide motif (4 Gln, 3 Pro, 1 Phe) is similar to the proportions of these three residues in the whole protein (about 40 mol% Gln, 30 mol% Pro and 8–9 mol% Phe). Individual C hordein proteins contain no cysteine, 1 residue of methionine and 0–2 residues of lysine per mol [63].



(c)

NH_2–RQLNPSSQELQSPQQSY XFXQQ
LQQPY PQQPFPLQ
PQNPY PQQPFPQQ
L PQQI

(d)

	[η] (ml/g)	Axial ratio (a/b)	Molecular dimensions (nm) Length	Diameter
0.1 M-CH_3 COOH	16.8	12.1	28.2	191
70% (v/v) aq. ethanol/0.01 M-NaCl	26.8	16.2	34.6	174
Trifluoroethanol	44.2	23.4	42.9	150

Fig. 4. C hordein of barley

(*a*) Two-dimensional electrophoresis (i.e.f–SDS/PAGE) of C hordein fractions from four cultivars of barley: a, cv. Sundance; b, cv. Athos; c, cv. Keg; d, cv. Jupiter (taken from [59]). (*b*) Schematic structure of a typical C hordein, based on the molecular mass and partial amino acid sequences (see [111]). (*c*) The amino acid sequences of the *N*-terminus and adjacent peptapeptide repeats (left) and of a random chymotryptic peptide (right), showing the positions of predicted β-turns (taken from [64]). (*d*) The intrinsic viscosity determined for C hordein dissolved in a range of solvents and the calculated axial ratios and molecular dimensions (taken from [61]).

Secondary structure predictions indicate β-turns to be present in the pentapeptide and octapeptide repeat regions (Fig. 4*c*) [64]. Two overlapping turns are predicted in the octapeptide region, within the repeat motifs and spanning the junctions between adjacent motifis. C.d. spectra of C hordein dissolved in 70% (v/v) aq. ethanol at temperatures between 6 and 86 °C were consistent with a structure rich in β-turns [64], while solution and solid-state n.m.r. showed a regular structure with the majority of the resonances assignable to the major residues present in the repeat motifs [67]. N.m.r. spectroscopy also showed that the proline residues were in the *trans* conformation, which is consistent with their presence in β-turns [67].

More detailed information on the structure of the repetitive domain has been given by analysis of a synthetic peptide corresponding to the repeat motif (Gly-Gln-Pro-Gln-Gln-Pro-Phe-Pro-Gln-Gly) [68]. C.d. spectroscopy indicated type I/III β-turns in trifluoroethanol at +22 °C or in water at +70 °C, but

showed a random-coil like structure in water at +22 °C. Cryogenic c.d. studies in ethanediol/water (2:1, v/v) or methanol/glycerol (9:1, v/v) indicated that the β-turn structure at higher temperatures was in equilibrium with a poly(L-proline) II conformation at low temperatures. The position of the equilibrium was also influenced by the solvent, the βI/III turn structure being favoured in solvents of low dielectric constant. On the basis of these observations it was suggested that the random coil-like spectrum observed in water at room temperature resulted from a mixture of the two conformational states rather than true random coil.

Hydrodynamic measurements in 0.1 M-acetic acid and 70% (v/v) aq. ethanol (±0.01 M-NaCl) at temperatures between 15 and 50 °C showed a rod-like structure, with dimensions varying from 36 nm × 1.7 nm to 26.5 nm × 2.0 nm [61] (Fig. 4*d*). The molecule was more extended in trifluoroethanol, about 42.9 nm × 1.5 nm at 30 °C. It was suggested the repetitive domain of C hordein formed a loose spiral structure, similar to that discussed above for the HMW subunits of glutenin. Assuming that C hordein consists of about 440 residues (calculated from the M_r), a β-spiral similar to that formed by the synthetic polypentapeptide based on elastin [52] would result in a molecule with dimensions of about 29.4 nm × 1.7 nm, which is consistent with those determined experimentally.

S-poor prolamins of wheat and rye

A number of S-poor prolamins of wheat (ω-gliadins) and rye (ω-secalins) have been characterized [69]. The ω-gliadins had M_r values by SDS/PAGE of 44000–78000, and the ω-secalins M_r values of about 48000–53000. In all cases glutamate + glutamine, proline and phenylalanine together accounted for about 80 mol%, but the proportions of these residues varied. Whereas the ω-secalins and the ω-gliadins encoded by the A and D genomes of wheat had proportions of glutamate + glutamine (≃ 40 mol%) and proline (≃ 30 mol%) similar to those of C hordein, the ω-gliadins encoded by chromosome 1B of wheat contained more glutamate + glutamine (50 mol%) and less proline (20 mol%). This must indicate a different repeat structure. Comparison of the *N*-terminal amino acid sequences also showed some variation (especially in the chromosome 1B-encoded ω-gliadins), but all were clearly homologous [69].

The c.d. spectra of ω-gliadins and ω-secalins in 70% (v/v) ethanol are similar to those of C hordein, indicating that they have similar β-turn-rich structures [58,70]. The Fourier transform i.r. spectra of ω-gliadins dissolved in ^2H_2O at pH 3 provide further evidence for the presence of β-turns [71], with a strong maximum at 1658 cm^{-1} attributed to β-turns and a weaker maximum at 1623 cm^{-1} to linear β-structure. In addition, Popineau & Pineau [72] have studied the gel filtration behaviour and TNS binding of ω-gliadins under a range of conditions of concentration, pH and salt, and concluded that the hydrodynamic volume is higher than would be expected for a compact globular molecule.

S-RICH PROLAMINS

The S-rich prolamins are the quantitatively major group in barley, wheat and rye, accounting for about 80% of the total fraction. Most have M_r values between about 36000 and 44000 by SDS/PAGE, and consist of 250–300 residues. However, one group of secalins of rye have higher M_r values, about 75000 by SDS/PAGE [73] (see Fig. 2*c*). They all appear to have a similar basic structure, with an *N*-terminal domain of proline-rich repeats and a *C*-terminal domain containing most, or all, of the cysteine residues. In some cases a short unique *N*-terminal sequence also precedes the repetitive domain. There is considerable diversity in their detailed structures, and at least three distinct families can be recognized (see [32] for details).

The γ-type S-rich prolamins

These may correspond most closely to the ancestral type of S-rich prolamin, as they are present in temperate meadow grasses [74] as well as in wheat [75], barley [76,77] and rye [73]. Most are monomers with intrachain disulphide bonds, but polymeric types may also occur in all three species ([78–80]; P. R. Shewry & A. S. Tatham, unpublished results for barley).

The amino acid sequence of a typical γ-type gliadin is summarized in Fig. 5. A unique *N*-terminal domain of 12 residues precedes a series of repeats based on the consensus motif Pro-Gln-Gln-Pro-Phe-Pro-Gln. The proline-poor non-repetitive domain contains all of the cysteine and most of the charged residues.

C.d. spectroscopy of a γ-gliadin fraction dissolved in 70% (v/v) aq. ethanol showed a far-u.v. spectrum typical of a protein rich in α-helix [70], and this was confirmed by similar analyses of a purified single component [81]. Deconvolution of the c.d. spectra indicated an α-helical content of some 30–35% [81,82], and structure prediction from available amino acid sequences indicated that the α-helix was mainly present in the *C*-terminal domain, while the *N*-terminal repetitive domain contained regularly repeated β-turns [58]. These predictions have been confirmed by c.d. studies of purified peptides corresponding broadly to the two domains [81]. The peptide corresponding to the proline-rich repetitive domain behaved like the S-poor prolamin, C hordein (see above). It showed a βI/III turn-rich structure at high temperatures, which was in equilibrium with a poly(L-proline) II structure at lower temperature. These two conformations were probably present in equilibrium at room temperature, with the spectrum tending towards the βI/III turn type. In contrast the peptide corresponding to the non-repetitive domain was predominantly α-helical, and conformationally stable to heating and cooling [81]. These results indicate that the two domains adopt different conformations. Whereas the *N*-terminal domain may adopt a rod-like conformation similar to C hordein, the *C*-terminal domain is probably compact and globular. This is consistent with the results of Popineau & Pineau [72], who showed that the whole γ-gliadin was more compact than the S-poor ω-gliadin, but less compact than the S-rich α-type gliadins (see below).

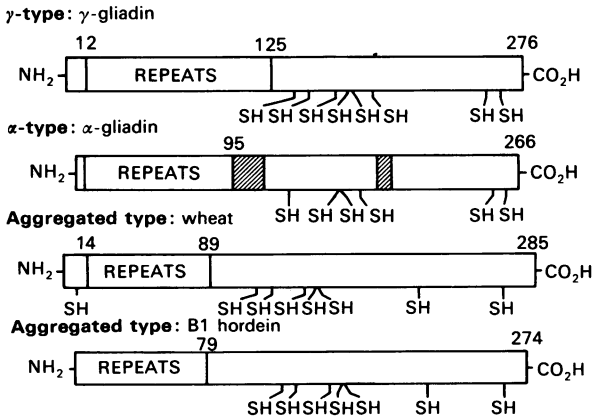


Fig. 5. Schematic comparison of the structures of typical S-rich prolamins

Short unique *N*-terminal sequences are present in all except B1 hordein. The positions of cysteine residues are shown by SH. The poly-Gln regions in the α-type gliadin are shaded (based on sequences in [88], [93], [111] and [112]).

Preliminary c.d. studies of γ-secalins and γ-type hordeins of barley indicate that they have similar conformations to the γ-type gliadins [58].

The α-type S-rich prolamins

These are monomeric, and are present only in wheat and in closely related species of *Triticum* and *Aegilops* [83], not in rye or barley. Although they have a similar structural organization to the γ-gliadins (Fig. 5), the repetitive domain is less conserved, consisting of interspersed repeats based on two consensus motifs: Pro-Gln-Pro-Gln-Pro-Phe-Pro, and Pro-Gln-Gln-Pro-Tyr. In addition, the *C*-terminal domain contains two regions which are rich in glutamine, which are called poly-Gln regions (see Fig. 5).

One subgroup of α-gliadins, called A gliadin, has been studied in detail by Kasarda and his co-workers in Albany, California. These have the unusual property of aggregating under certain conditions of pH (≃ 5) and ionic strength (5 mM-KCl) to form fibrils of diameter up to 8 nm and 300–400 nm in length [84,85]. These fibrils dissociate to individual globular protein subunits when the pH and/or ionic strength are reduced. C.d. and o.r.d. studies of the dissociated subunits at pH 5 indicated 33–34% α-helical structure, and this remained constant when 5 mM salt was added to promote fibril formation [86]. Hydrodynamic studies of the protein (1 mM-acetic acid/10 mM-NaCl, pH 4) indicated a compact globular molecule of axial ratio 3.4:1 [87].

I.r. and c.d. studies of peptides prepared from A-gliadin indicated that the *N*-terminal domain contained predominantly β-turns with lower contents of β-structure, α-helix and unordered structure [71]. Peptides corresponding to the *C*-terminal domain showed major absorbances associated with α-helix, with lower contents of other structural types [71]. Structure prediction confirms that the *N*-terminal domains of α-gliadins are rich in β-turns, but shows that their arrangement is less regular than in the repetitive domains of the γ-type gliadins and in the S-poor C hordein [49,58]. It therefore appears unlikely that a regular helical structure would form, which is consistent with their more compact structure [72].

The aggregated S-rich prolamins

These occur in barley (B hordeins) and wheat ['low molecular weight' (LMW) subunits of glutenin], but not apparently in rye. As their name implies, they are present only in polymers stabilized by interchain disulphide bonds. Although their unique *C*-terminal domains are closely related, their repetitive domains are more variable (Fig. 5).

The aggregated S-rich prolamins of wheat have a short non-repetitive *N*-terminal sequence, which may contain a single cysteine residue, followed by a proline-rich repetitive domain of interspersed repeats based on two consensus heptapeptide motifs: Pro-Gln-Gln-Pro-Pro-Pro-Phe-Ser and Gln-Gln-Gln-Gln-Pro-Val-Leu [78,79,88,89] (Fig. 5). Their c.d. spectra are similar to those of α-type and γ-type gliadins, showing α-helical contents of about 35% [82,90]. Structure prediction indicates that the α-helix is mainly located in the *C*-terminal domain, while the *N*-terminal domain contains β-turns, although those are less abundant than in the corresponding domains of other S-rich prolamins [58].

The B hordeins have been studied in detail at the molecular and genetic levels, and provide an excellent illustration of an aspect of the S-rich prolamins which we have not yet considered: their high degree of polymorphism (Fig. 6). Two-dimensional analyses (Fig. 6*a*) of B hordein fractions from a range of genotypes showed variation in the numbers, M_r values and pI values of B hordein components, and CNBr peptide mapping of the major components eluted from the gels showed three types of pattern [91]. Subsequent studies showed that these corresponded

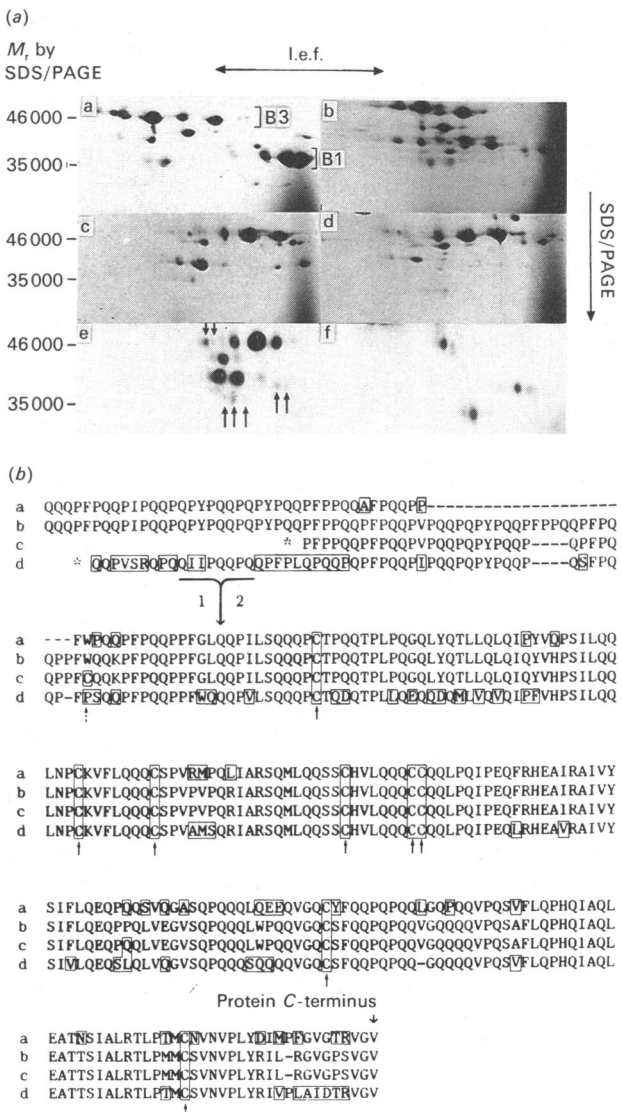


Fig. 6. Polymorphism of B hordeins of barley

(*a*) Two-dimensional analyses of B hordein fractions from six cultivars and lines of barley: a, cv. Sundance; b, cv. Athos; c, cv. Keg; d, cv. Jupiter; e, cv. Carlsberg II; f, Risø mutant 56. The brackets in a indicate the B1 and B3 bands discussed in the text. Although the major proteins present in a–f are B hordeins, some of the minor components are γ-type hordeins. Risø 56 (f) is a γ-ray-induced mutant line derived from Carlsberg II (e). A chromosome mutation has resulted in the loss of all the genes that encode B hordeins, and all the proteins present in the gel separation are γ-type hordeins (see [76] and [113]). The corresponding proteins in the wild-type parent line Carlsberg II are indicated by arrows in e. (Taken from [114].) (*b*) Amino acid sequences of individual B hordeins, showing the molecular basis for the polymorphism shown in (*a*). a and b are complete amino acid sequences of two hordeins of the B1 family, deduced from the nucleotide sequences of cloned genes from cultivars Carlsberg II and Sundance respectively. c and d are partial amino acid sequences of proteins of the B1 and B3 subfamilies respectively, deduced from cDNAs from cv. Sundance. The asterisks indicate the *N*-terminal residues of the incomplete sequences. The junction between the *N*-terminal (1) and *C*-terminal (2) domains is shown by the large arrow. The individual proteins differ in amino acid substitutions and in insertion/deletions varying in size from 1 to 23 residues. In addition the B3 type protein (d) differs from the B1 types (a–c) in the organization of the *N*-terminal region. Cysteine residues are conserved (indicated by small solid arrows) with the exception of an additional cysteine in the *N*-terminal domain of sequence c (small broken arrow). (Based on [65], [93], [94] and [114].)

to the products of two subfamilies of cDNAs (and genes), with one of the subfamilies giving two types of peptide map in different genotypes [92]. These correspond to the B1 and B3 hordein bands present in SDS/PAGE separations of hordein fractions from the cultivar Sundance (Fig. 6*a*, part a), and are usually referred to as B1 and B3 types.

Comparison of the amino acid sequences of individual B hordeins [65,93,94] shows the molecular basis for this polymorphism (Fig. 6*b*). The individual B1-type hordeins differ in amino acid substitutions and small insertions/deletions, while the B3 hordein differs additionally in the organization of the repetitive domain. It is also possible to recognize some regions that are conserved, particularly around the cysteine residues. B1 hordein does not have a unique *N*-terminal sequence, and the repetitive domain consists of a mixture of penta-, hexa-, hepta- and octapeptides, all based on the consensus pentapeptide motif Pro-Gln-Gln-Pro-Xaa [93] (Figs. 5 and 6*b*). Preliminary c.d. studies and structural predictions indicate that the repetitive and non-repetitive domains have similar conformations to those in the aggregated S-rich prolamins of wheat, but with a higher content of β-turns due to the presence of more regularly repeated sequences [58,82].

S-Rich prolamins: a summary

The different types of S-rich prolamins have a number of characteristics in common. All have a clear domain structure, and appear to be compactly folded molecules. In addition they are unusually stable to thermal denaturation, which is probably due in part to intrachain disulphide bonds [70,81,90]. The proline-rich repetitive domains are rich in β-turns, their regularity dependent on the degree of conservation of the repeat motifs. In the γ-gliadin the repeat motifs are highly conserved and the turns may be sufficiently regular to form a loose spiral structure similar to that described for the S-poor prolamins. γ-Gliadins have indeed been reported to have a higher hydrodynamic volume than other gliadins [72] which supports the hypothesis of an extended structure for the domain. In contrast, the repetitive domains of the α-type gliadins and aggregated S-rich prolamins of wheat are based on interspersed repeats of two poorly conserved motifs and the predicted β-turns are distributed irregularly, and interspersed with other structures. Further studies are required to understand the detailed structures of this complex group of prolamins.

EVOLUTIONARY RELATIONSHIPS OF THE S-RICH, S-POOR AND HMW PROLAMINS

Comparison of the amino acid sequences of the prolamins of the Triticeae shows that all are structurally related [32,33] (Figs. 7*a* and 8).

The proline-rich repetitive domains of the S-rich prolamins all contain motifs based on the Pro-Gln-Gln-Xaa tetrapeptide: Pro-Gln-Gln-Pro-Phe-Pro-Gln (γ-gliadins and γ-secalins), Pro-Gln-Gln-Pro-Phe-Pro-Gln-Gln (γ-hordeins), Pro-Gln-Gln-Pro-Tyr (α-gliadins), Pro-Gln-Gln-Gln-Pro-Pro-Phe-Ser (aggregated S-rich prolamins of wheat) and Pro-Gln-Gln-Pro-Xaa(Xaa)(Xaa) (B1 hordein). These are all related, to varying extents, to the Pro-Gln-Gln-Pro-Tyr and Pro-Gln-Gln-Pro-Phe-Pro-Gln-Gln repeat motifs present in the S-poor C hordein. The logical conclusion is that these domains have a common evolutionary origin. In contrast, the three repeat motifs present in the HMW prolamins of wheat (Gly-Tyr-Tyr-Pro-Thr-Ser-Pro/Leu-Gln-Gln, Pro-Gly-Gln-Gly-Gln-Gln and Gly-Gly-Gln) are not related to those present in the S-rich and S-poor groups.

The non-repetitive *C*-terminal domains of the S-rich prolamins contain three conserved regions [32,33]. These vary in length from 20 to 35 residues, and are called A, B and C in Fig. 7(*a*). They also contain most of the cysteine residues. They are flanked and separated by more-variable regions, which are homologous within but not always between the different families of S-rich prolamins. Regions A, B and C are also related to each other, indicating that they may have arisen from an ancient triplication of a single domain consisting of about 30 residues.

Sequences related to regions A, B and C are also present in the HMW prolamins, but in this case regions A and B are in the non-repetitive *N*-terminal domain and region C in the *C*-terminal domain (Figs. 7*a* and 8). In addition, the short unique sequence of six residues at the *C*-terminus of C hordein shows weak homology with region C (Fig. 7*a*).

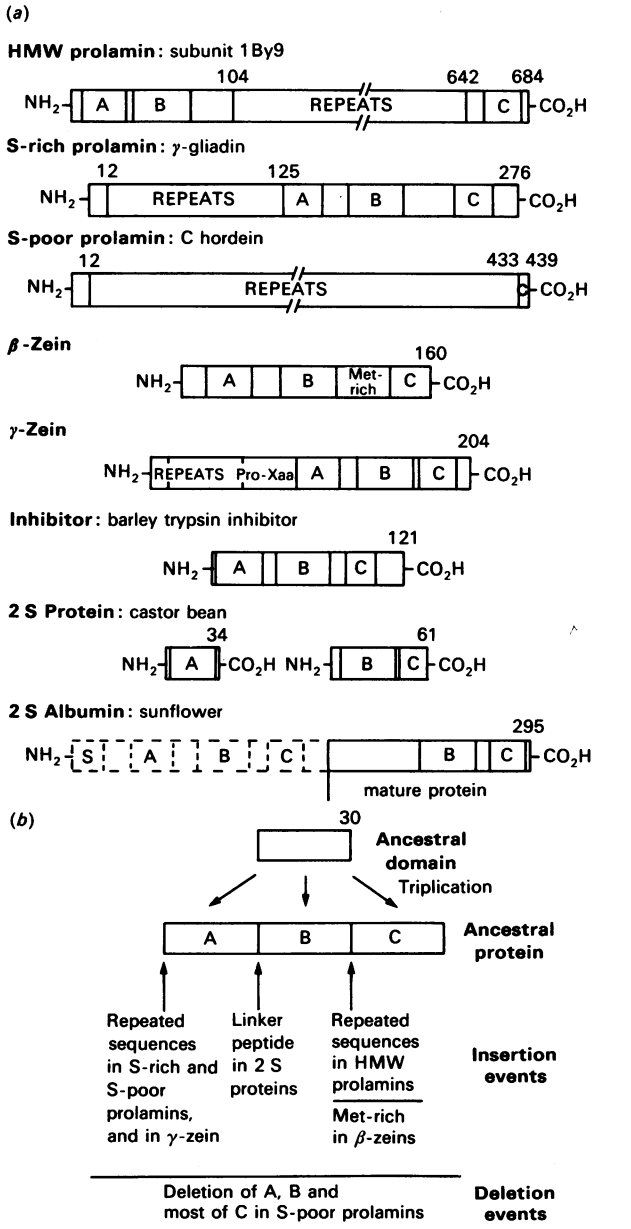


Fig. 7. (*a*) Homologous regions in prolamins and other seed proteins, and (*b*) hypothetical events in the evolution of prolamins and related proteins from a common ancestral protein

In (*a*) the putative signal sequence of the sunflower albumin (labelled S) is shown because the processing pattern of this protein is still uncertain. (Based on amino acid sequences in [25], [27], [44], [95], [101], [105], [110] and [111]).

It was proposed on the basis of these comparisons that all three groups of prolamins evolved from a short ancestral protein of about 90 residues and consisting of regions A, B and C. The major events in the HMW and S-rich prolamins have been the insertion of sequences between and flanking these regions, notably extensive repeated sequences between regions B and C in the HMW prolamins, and of unrelated repeated sequences at the *N*-terminal end of region A in the S-rich group. Finally, the S-poor prolamins such as C hordein are proposed to have evolved from the S-rich by further amplification of the repeated sequences followed by deletion of all of the non-repetitive domain except a vestige of region C.

RELATIONSHIPS OF THE PROLAMINS OF THE TRITICEAE TO OTHER SEED PROTEINS

Regions related to A, B and C are also present in three other groups of seed proteins. The positions of these regions within the proteins are shown in Fig. 7(*a*), and their sequences compared in detail in Fig. 8.

β- and *γ*-zeins of maize

The M_r-15000 *β*-zeins do not contain repeated sequences but the methionine-rich region discussed above is located between regions B and C (see Fig. 7*a*). In the M_r-28000 *γ*-zein the repetitive domain is at the *N*-terminal end of region A, but is separated from it by a further sequence of about 40 residues which includes the Pro-Xaa region (Fig. 7*a*). The cysteine-rich region recognized by Prat *et al.* [26] includes parts of regions A and B, and the sequence between them.

Inhibitors of hydrolytic enzymes

These are a group of inhibitors of serine proteases and *α*-amylase, at least some of which are bifunctional. They are present in the mature seeds of several species of cereals, including barley [95], wheat [96], rye [97], maize [98] and finger millet [99]. They consist of about 120 residues, and correspond essentially to regions A, B, and C with only limited regions of inserted sequences.

2 S storage proteins

The seeds of many plants contain storage albumins or globulins with $s_{20,w}$ values of about 2 S. At least some of these are related structurally to the prolamins. In oilseed rape [100], castor bean [101], *Arabidopsis* [102], brazil nut [103] and lupin [104] these have similar structures, the mature proteins consisting of two subunits associated by interchain disulphide bond(s). The small subunits at about 25–40 residues contain region A, and the large

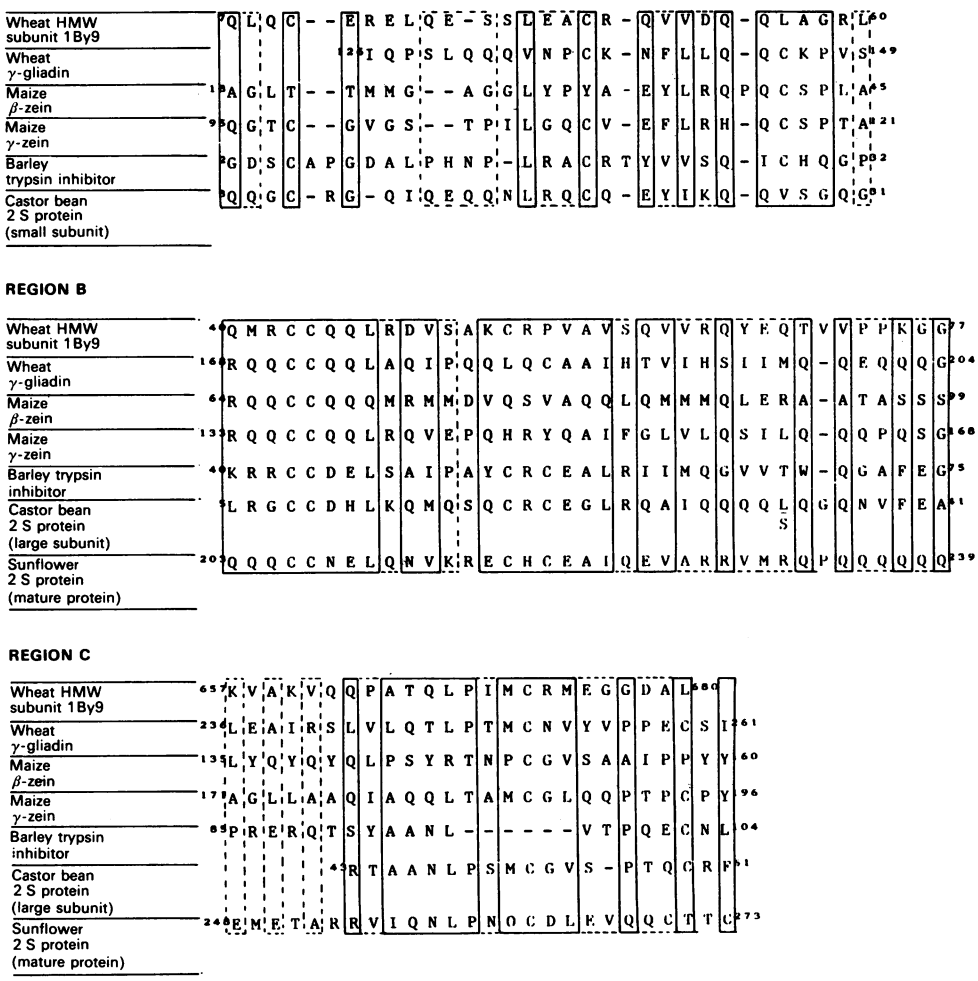


Fig. 8. Comparison of the amino acid sequences of regions A, B and C present in the protein types shown in Fig. 7(*a*)

Only the mature sunflower 2 S protein is shown, and C hordein is omitted. The sequences are aligned manually to maximize the similarity, resulting in some gaps. Well-conserved regions (identical residues in three or more proteins) are boxed with solid lines, less well-conserved regions (identical residues in two proteins) with broken lines. (Sequences are taken from [25], [27], [44], [95], [101], [105] and [111].)

subunits of about 70–80 residues regions B and C (Fig. 6*a*). Isolation of genomic and cDNA clones shows that the two subunits are synthesized as a single precursor, followed by proteolytic cleavage in a linker region between regions A and B.

Regions corresponding to A, B and C are also present in a different type of storage albumin from sunflower [105]. Although cloned cDNAs encode a protein containing two regions related to B and C and one region related to A, the mature protein appears to correspond to the last 130 residues, and therefore only contains one copy each of regions B and C. The reason for this unusual processing pattern and the fate of the *N*-terminal part of the protein are not known.

On the basis of these comparisons Kreis and co-workers [32,33,92,106] proposed that the prolamins of the Triticeae and the three groups of proteins discussed above formed a protein superfamily with limited sequence homology, as defined by Dayhoff [107]. The major events in the evolution and divergence of these groups are summarized in Fig. 6(*b*). Aspects of prolamin evolution are also discussed in more detail in a recent review article [106].

CONCLUSIONS: REDEFINING PROLAMINS

We now know that the cereal prolamins are not all related to each other, but that most are related to a range of other seed proteins. Nevertheless, it is still possible to define them as a group on the basis of one property: their *insolubility* in the native state in water or dilute salt solutions. With one exception, all of the reduced subunits of polymeric prolamins are also insoluble in these solvents but, like the native monomeric subunits, are soluble in alcohol/water mixtures. These properties may result from the presence of repeated sequences (in α-zeins and the prolamins of the Triticeae) or methionine-rich regions (in β- and γ-zeins). The exception is the δ-zeins which are water-soluble as reduced subunits but, like the β-zeins, γ-zeins and many S-rich and HMW prolamins, form alcohol-insoluble and water-insoluble polymers stabilized by interchain disulphide bonds. Insolubility under physiological conditions is probably essential for effective packaging of the protein into protein bodies in the developing cereal endosperm, and the associated solubility in alcohol/water mixtures may be a fortuitous consequence of the mechanism adopted by the plant. Their characteristic amino acid compositions, including the high contents of proline, glutamine and, in certain types, methionine, aromatic amino acids and glycine, are also consequences of the structural mechanisms that ensure the correct solubility properties. We can, therefore, redefine prolamins on the basis of currently available evidence as 'storage proteins that are deposited in developing endosperms of cereals, are rich in glutamine and proline, and are insoluble in the native state in water or dilute solutions of salts'.

We are indebted to our colleagues and co-workers who have contributed to the work discussed here, especially to Dr. Martin Kreis who carried out much of the work on the evolutionary relationships of cereal seed proteins, and to Dr. J. M. Field [ICI Seeds (UK) Ltd.] who supplied Fig. 3(*a*). We are also grateful to Mrs Linda Castle for preparing the figures.

REFERENCES

1. Miflin, B. J., Field, J. M. & Shewry, P. R. (1983) in Seed Proteins (Daussant, J., Mossé, J. & Vaughan, J., eds.), pp. 255–319, Academic Press, London
2. Beccari, J. (1745) De Bononiensi Scientarium et Artium, Instituto Atque Academia Commentarii 2 (Part 1), 122
3. Parmentier, A. A. (1773) Examin Chimique des Pommes de Terre, dans Lequel On Trait des Parties Constituents du Blé, Didot Dejeaune, Paris
4. Taddei, G. (1819) Brugnatelli (2) **2**, 360–361
5. Einhof, H. (1805) Neues Allgen. J. Chem. **5**, 131–153
6. Einhof, H. (1806) Neues Allgen. J. Chem. **6**, 62–98
7. Osborne, T. B. (1924) The Vegetable Proteins, 2nd edn., Longmans, Green and Co., London
8. Shewry, P. R. & Miflin, B. J. (1985) Adv. Cereal Sci. Technol. **5**, 1–83
9. Beitz, J. A. (1983) Biochem. Genet. **20**, 1039–1053
10. Taylor, J. R. N., von Benecke, R. & Carlsson, F. H. H. (1989) J. Cereal Sci. **9**, 169–177
11. Ottoboni, L. M. M., Leite, A., Targon, M. L. N., Crozier, A. & Arruda, P. (1989), in the press
12. Esen, A. A. (1987) J. Cereal Sci. **5**, 117–128
13. Paulis, J. W. (1981) Cereal Chem. **58**, 542–546
14. Messing, J. (1987) in Genetic Engineering (Rigby, P. W. J., ed.) vol. 6, pp. 1–46, Academic Press, London
15. Rubenstein, I. & Geraghty, D. E. (1986) Adv. Cereal Sci. Technol. **8**, 297–315
16. Elliston, K. O., Imran, S. & Messing, J. (1988) Plant Mol. Biol. Reporter **6**, 22–26
17. Argos, P., Pedersen, K., Marks, M. D. & Larkins, B. A. (1982) J. Biol. Chem. **257**, 9984–9990
18. Wu, Y. V., Paulis, J. W., Sexson, K. R. & Wall, J. S. (1983) Cereal Chem. **60**, 342–344
19. Chou, P. Y. & Fasman, G. D. (1978) Annu. Rev. Biochem. **47**, 251–276
20. Elliot, M. a. & Williams, J. W. (1939) J. Am. Chem. Soc. **61**, 718–725
21. Foster, J. F. & Edsall, J. T. (1945) J. Am. Chem. Soc. **67**, 617–625
22. Williams, J. W. & Watson, C. C. (1938) Cold Spring Harbour Symp. Quant. Biol. **26**, 208–214
23. Wilson, C. M., Shewry, P. R. & Miflin, B. J. (1981) Cereal Chem. **58**, 275–281
24. Pedersen, K., Argos, P., Naravana, S. V. L. & Larkins, B. A. (1986) J. Biol. Chem. **261**, 6279–6284
25. Marks, M. D., Lindell, J. S. & Larkins, B. A. (1985) J. Biol. Chem. **260**, 16451–16459
26. Prat, S., Cortadas, J., Puigdoménech, P. & Palau, J. (1985) Nucleic Acids. Res. **5**, 1493–1504
27. Boronat, A., Martinez, M. C., Reina, M., Puigdoménech, P & Palau, J. (1986) Plant Sci. **47**, 95–102
28. Wang, S.-Z. & Esen, A. (1986) Plant Physiol. **81**, 70–74
29. Prat, S., Pérez-Grau, L. & Puigdoménech, P. (1987) Gene **52**, 41–49
30. Pons, M., Feliz, M., Celma, C. & Giralt, E. (1987) Magn. Reson. Chem. **25**, 402–406
31. Kirihara, J. A., Petri, J. B. & Messing, J. (1988) Gene **71**, 359–370
32. Kreis, M., Shewry, P. R., Forde, B. G., Forde, J. & Miflin, B. J. (1985) in Oxford Surveys of Plant Molecular Biology (Miflin, B. J., ed.), vol. 2, pp. 253–317, Oxford University Press, Oxford
33. Kreis, M., Forde, B. G., Rahman, S., Miflin, B. J. & Shewry, P. R. (1985) J. Mol. Biol. **183**, 499–502
34. Shewry, P. R., Miflin, B. J. & Kasarda, D. D. (1984) Philos. Trans. R. Soc. London Ser. B **304**, 297–308
35. Lawrence, G. J. & Shepherd, K. W. (1980) Aust. J. Biol. **33**, 221–233
36. Payne, P. I., Corfield, K. G. & Blackman, J. A. (1979) Theor. Appl. Genet. **55**, 153–159
37. Payne, P. I., Corfield, K. G., Holt, L. M. & Blackman, J. A. (1981) J. Sci. Food Agric. **32**, 51–60
38. Payne, P. I., Holt, L. M. & Law, C. N. (1981) Theor. Appl. Genet. **60**, 229–236
39. Shewry, P. R., Field, J. M., Faulks, A. J., Parmar, S., Miflin, B. J., Dietler, M. D., Lew, E. J.-L. & Kasarda, D. D. (1984) Biochim. Biophys. Acta **223**, 315–322
40. Shewry, P. R., Halford, N. G. & Tatham, A. S. (1989) in Oxford Surveys of Plant Molecular and Cell Biology (Miflin, B. J., ed)., vol. 6., Oxford University Press, Oxford, in the press
41. Anderson, O. D. & Greene, F. C. (1989) Theor. Appl. Genet. **77**, 689–700
42. Forde, J., Malpica, J. M., Halford, N. G., Shewry, P. R., Anderson, O. D., Greene, F. C. & Miflin, B. J. (1985) Nucleic Acids. Res. **13**, 6817–6832
43. Anderson, O. D., Greene, F. C., Yip, R. E., Halford, N. G., Shewry, P. R. & Malpica-Romero, J. M. (1989) Nucleic Acids Res. **17**, 461–462
44. Halford, N. G., Forde, J., Anderson, O. D., Greene, F. C. & Shewry, P. R. (1987) Theor. Appl. Genet. **75**, 117–126
45. Thompson, R. D., Bartels, D. & Harberd, N. P. (1985) Nucleic Acids. Res. **13**, 6833–6846

46. Sugiyama, T., Rafalski, A., Peterson, D. & Soll, D. (1985) Nucleic Acids Res. **13**, 8729–8737
47. Flavell, R. B., Goldsborough, A., Robert, L., Schnick, D. & Thompson, R. D. (1989) Bio/Technology, in the press
48. Tatham, A. S., Shewry, P. R. & Miflin, B. J. (1984) FEBS Lett. **177**, 205–208
49. Tatham, A. S., Miflin, B. J. & Shewry, P. R. (1985) Cereal Chem. **62**, 405–411
50. Field, J. M., Tatham, A. S. & Shewry, P. R. (1987) Biochem. J. **247**, 215–221
51. Tatham, A. S., Drake, A. F. & Shewry, P. R. (1990), in the press
52. Venkatachalam, C. M. & Urry, D. W. (1981) Macromolecules **141**, 1225–1232
53. Miles, M. J., Carr, H. J., McMaster, T., Belton, P. S., Morris, V. J., Field, J. M., Shewry, P. R. & Tatham, A. S. (1990), in the press
54. Field, J. M., Shewry, P. R., Miflin, B. J. & March, J. (1982) Theor. Appl. Genet. **62**, 329–336
55. Kreis, M., Shewry, P. R., Forde, B. G., Rahman, S., Bahramian, M.B. & Miflin, B. J. (1984) Biochem. Genet. **22**, 231–255
56. Shewry, P. R., Tatham, A. S., Pappin, D. J. C. & Keen, J. (1988) Cereal Chem. **65**, 510–511
57. Forde, J., Forde, B. G., Fry, R., Kreis, M., Shewry, P. R. & Miflin, B. J. (1983) FEBS Lett. **162**, 360–366
58. Tatham, A. S., Belton, P. S. & Shewry, P. R. (1990) Adv. Cereal Sci. Technol. **10**, in the press
59. Shewry, P. R., Bunce, N. A. C., Kreis, M. & Forde, B. G. (1985) Biochem. Genet. **23**, 391–404
60. Shewry, P. R., Field, J. M., Kirkman, M. A., Faulks, A. J. & Miflin, B. J. (1980) J. Exp. Bot. **31**, 397–401
61. Field, J. M., Tatham, A. S., Baker, A. M. & Shewry, P. R. (1986) FEBS Lett. **200**, 76–80
62. Shewry, P. R., Autran, J.-C., Nimmo, C. C., Lew, E. J.-L. & Kasarda, D. D. (1980) Nature (London) **286**, 520–522
63. Shewry, P. R., Lew, E. J.-L. & Kasarda, D. D. (1981) Planta **153**, 246–253
64. Tatham, A. S., Drake, A. F. & Shewry, P. R. (1985) Biochem. J. **226**, 557–562
65. Forde, B. G., Kreis, M., Williamson, M., Fry, R., Pywell, J., Shewry, P. R., Bunce, N. A. C. & Miflin, B. J. (1985) EMBO J. **4**, 9–15
66. Rasmussen, S. K. & Brandt, A. (1986) Carlsberg Res. Commun. **51**, 371–379
67. Tatham, A. S., Shewry, P. R. & Belton, P. S. (1985) Biochem. J. **236**, 617–620
68. Tatham, A. S., Drake, A. F. & Shewry, P. R. (1989) Biochem. J. **259**, 471–476
69. Kasarda, D. D., Autran, J.-C., Lew, E. J.-L., Nimmo, C. C. & Shewry, P. R. (1983) Biochim. Biophys. Acta **747**, 138–150
70. Tatham, A. S. & Shewry, P. R. (1985) J. Cereal Sci. **3**, 103–113
71. Purcell, J. M., Kasarda, D. D. & Wu, C.-S. C. (1988) J. Cereal Sci. **7**, 21–32
72. Popineau, Y. & Pineau, F. (1988) Lebens. Wiss. Technol. **21**, 113–117
73. Shewry, P. R., Field, J. M., Lew, E. J.-L. & Kasarda, D. D. (1982) J. Exp. Bot. **33**, 261–268
74. Shewry, P. R., Smith, S. J., Lew, E. J.-L. & Kasarda, D. D. (1986) J. Exp. Bot. **37**, 633–639
75. Scheets, K. & Hedgcoth, C. (1988) Plant Sci. **57**, 141–150
76. Shewry, P. R., Kreis, M., Parmar, S., Lew, E. J.-L. & Kasarda, D. D. (1985) FEBS Lett. **190**, 61–64
77. Cameron-Mills, V. & Brandt, A. (1988) Plant Mol. Biol. **11**, 449–461
78. Shewry, P. R., Miflin, B. J., Lew, E. J.-L. & Kasarda, D. D. (1983) J. Exp. Bot. **34**, 1403–1410
79. Tao, H. P. & Kasarda, D. D. (1989) J. Exp. Bot., in the press
80. Shewry, P. R., Parmar, S. & Miflin, B. J. (1983) Cereal Chem. **60**, 1–6
81. Tatham, A. S., Masson, P. & Popineau, Y. (1989) J. Cereal Sci. **11**, 1–13
82. Shewry, P. R., Field, J. M. & Tatham, A. S. (1987) in Cereals in a European Context (Morton, I. D., ed.), pp. 421–437, Ellis Horwood, Chichester
83. Autran, J.-C., Lew, E. J.-L., Nimmo, C. C. & Kasarda, D. D. (1979) Nature (London) **282**, 527–529
84. Bernardin, J. E., Kasarda, D. D. & Macham, D. K. (1967) J. Biol. Chem. **242**, 445–450
85. Kasarda, D. D., Bernardin, J. E. & Thomas, R. S. (1967) Science **155**, 203–205
86. Kasarda, D. D., Bernardin, J. E. & Gaffield, W. (1968) Biochemistry **7**, 3950–3957
87. Cole, E. W., Kasarda, D. D. & Lafiandra, D. (1984) Biochim. Biophys. Acta **787**, 244–251
88. Okita, T. W., Cheesbrough, V. & Reeves, C. D. (1985) J. Biol. Chem. **260**, 8203–8213
89. Colot, V., Bartels, D., Thompson, R. & Flavell, R. (1989) Mol. Gen. Genet. **216**, 81–90
90. Tatham, A. S., Field, J. M., Smith, S. J. & Shewry, P. R. (1987) J. Cereal Sci. **51**, 203–214
91. Faulks, A. J., Shewry, P. R. & Miflin, B. J. (1980) Biochem. Genet. **19**, 841–858
92. Kreis, M., Rahman, S., Forde, B. G., Pywell, J., Shewry, P. R. & Miflin, B. J. (1983) Mol. Gen. Genet. **191**, 194–200
93. Forde, B. G., Heyworth, A., Pywell, J. & Kreis, M. (1985) Nucleic Acids Res. **131**, 7327–7339
94. Brandt, A., Montembault, A., Cameron-Mills, V. & Rasmussen, S. K. (1985) Carlsberg Res. Commun. **50**, 333–345
95. Odani, S., Koide, T. & Ono, T. (1983) J. Biol. Chem. **259**, 7998–8003
96. Kashlan, N. & Richardson, M. (1981) Phytochemistry **20**, 1781–1784
97. Lyons, A., Richardson, M., Tatham, A. S. & Shewry, P. R. (1987) Biochim. Biophys. Acta **915**, 305–313
98. Mahoney, W. C., Hermodson, M. A., Jones, B., Powers, D. D., Corfman, R. S. & Reeck, G. R. (1981) J. Biol. Chem. **259**, 8412–8416
99. Campos, F. D. A. & Richardson, M. (1983) FEBS Lett. **152**, 300–304
100. Crouch, M. L., Tembarge, K. M., Simon, A. E. & Ferl, R. (1983) J. Mol. Appl. Genet. **2**, 273–283
101. Sharief, F. S. & Li, S. S.-L. (1982) J. Biol. Chem. **257**, 14753–14759
102. Krebbers, E., Herdies, L., de Clercq., A., Seurinck, J., Leemans, J., van Damme, J., Segura, M. Gheysen, G., van Montague, M. & Vandekerchkove, J. (1988) Plant Physiol. **89**, 859–866
103. Ampe, C., Van Damme J., Castro, L. A. B., Sampaio, M. J. A. M. & Van Montague, M. (1986) Eur. J. Biochem. **159**, 597–604
104. Lilley, G. G. & Inglis, A. S. (1986) FEBS Lett. **195**, 231–241
105. Allen, R. D., Cohen, E. A., Vonder Haar, R. A., Adams, C. A., Ma, D. P., Nessler, C. L. & Thomas, T. L. (1987) Mol. Gen. Genet. **210**, 211–218
106. Kreis, M. & Shewry, P. R. (1989) BioEssays **10**, 201–207
107. Dayhoff, M. O. (1978) Atlas of Protein Sequence and Structure, vol. 5, suppl. 3.
108. Shewry, P. R., Parmar, S., Franklin, J. & White, R. (1988) Genet. Res. **51**, 5–12
109. Shewry, P. R., Parmar, S. & Pappin, D. J. C. (1987) Biochem. Genet. **215**, 309–325
110. Shewry, P. R. & Tatham, A. S. (1988) Comments Agric. Food Chem. **1**, 71–93
111. Bartels, D., Altosaar, I., Harberd, N. P., Barker, R. F. & Thompson, R. D. (1986) Theor. Appl. Genet. **72**, 845–853
112. Kasarda, D. D., Okita, T. W., Bernardin, J. E., Baecker, P. A., Nimmo, C. C., Lew, E. J.-L., Dietler, M. D. & Greene, F. C. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 4712–4715
113. Kreis, M., Shewry, P. R., Forde, B. G., Rahman, S. & Miflin, B. J. (1983) Cell **34**, 161–167
114. Kreis, M., Williamson, M. S., Forde, J., Clarke, J., Buxton, B., Pywell, J., Marris, C., Gallois, P. & Shewry, P. R. (1987) Plant Physiol. Biochem. **25**, 291–302
115. Bright, S. W. J. & Shewry, P. R. (1983) CRC Crit. Rev. Plant Sci. **1**, 49–93
116. Geraghty, D., Peifer, M. A., Rubenstein, I. & Messing, J. (1981) Nucleic Acids Res. **9**, 5163–5174
117. Marks, M. D. & Larkins, B. A. (1982) J. Biol. Chem. **257**, 9976–9983