PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Structural and Evolutionary Relationships of the Prolamin Storage Proteins of Barley, Rye and Wheat

P. R. Shewry, B. J. Miflin and D. D. Kasarda

Phil. Trans. R. Soc. Lond. B 1984 **304**, 297-308
doi: 10.1098/rstb.1984.0025

Phil. Trans. R. Soc. Lond. B **304**, 297–308 (1984)
Printed in Great Britain

The structural and evolutionary relationships of the prolamin storage proteins of barley, rye and wheat

By P. R. Shewry[1], B. J. Miflin[1] and D. D. Kasarda[2]

[1]*Biochemistry Department, Rothamsted Experimental Station, Harpenden, Hertfordshire, U.K.*
[2]*U.S.D.A. Western Regional Research Centre, Albany, California* 94710, *U.S.A.*

[Plate 1]

The major endosperm storage proteins of barley, wheat and rye are soluble in aqueous alcohols, either native or after the reduction of disulphide bonds, and can be defined as prolamins. They can be divided into three groups on the basis of their chemical characteristics, notably their molecular mass and amino acid composition, and the chromosomal location of their structural genes. Two of the groups, the high molecular mass prolamins and the sulphur-poor (ω-gliadin-type) prolamins, show clear homology between the three species. The remaining prolamins are characterized by a high content of cysteine. In wheat this is a complex mixture of at least three groups of components that vary in their aggregation properties and N-terminal amino acid sequences. The precise chemical and genetic relationships of those components to each other and to the more clearly defined groups of sulphur-rich prolamins of rye and barley are still not completely understood.

Introduction

Wheat and barley were probably first domesticated in the near East and spread to Europe, forming the most important crops in these areas even before the start of historical records. Barley (*Hordeum vulgare*) was the primary cereal in early times, and archaeological remains show that it was harvested for food, if not cultivated, in the Egyptian Late Paleolithic, some 17–18000 years ago (Wendorf *et al.* 1979). The earliest cultivated wheats were emmer (tetraploid *Triticum dicoccum*) and einkorn (diploid *T. monococcum*), although these were superceded by hexaploid bread wheat (*T. aestivum*) by classical times (Harlan 1978). Rye (*Secale cereale*) was domesticated at a later date, possibly originating as a weed in wheat (Harlan 1978).

Currently wheat is the most widely grown cereal in the world, with an annual yield in excess of 400×10^6 t, and also the greatest single source of protein in the human diet. Barley and rye are currently the fourth and eighth most important cereals in the world respectively, although they are the primary crops in some countries in northern and north-western Europe. The most important use of wheat is baking bread, cakes and biscuits, but considerable quantities (35–40% in the U.K.) are used to feed livestock. Rye and barley are also important for animal feed, although significant amounts of these are used for baking and for malting and distilling respectively.

The grain proteins are important in determining the suitability of the grain for all these purposes. The nutritional quality of the proteins is determined by their amino acid compositions (see Bright & Shewry 1983), while the gluten proteins of wheat are also important in determining the baking characteristics (see Wall 1979; Miflin *et al.* 1983). The amounts, and

possibly also the characteristics, of proteins present in barley grain also affect the malting quality (see Palmer 1980; Miflin *et al.* 1983). Although these important nutritional and functional aspects of the grain proteins were not recognized when the storage proteins of wheat, barley and rye were first studied in the early 19th century (Einhof 1805, 1806; Taddei 1819) they have provided considerable stimulus to the modern study. Also, the recent use of cereal storage proteins as model systems for studying the organization and expression of plant genes has provided further important information on their chemistry and genetics.

In this paper we present a brief account of the major storage proteins of wheat, barley and rye. This information is then discussed in relation to the origin and evolution of the proteins. The review will be restricted to information derived from the approaches of protein chemistry and classical genetics. The molecular biology and genetics of these proteins form the subject of a separate article (Miflin *et al.*) in this symposium.

The classification of plant proteins

The modern study of seed proteins is founded on the work of T. B. Osborne that is summarized in his book *The vegetable proteins* (Osborne 1924). He classified proteins into groups (often called Osborne groups or fractions) on the basis of their extraction and solubility in series of solvents. The first two groups are proteins soluble in water (albumins) and dilute saline (globulins). The third group, which contains the major endosperm storage proteins of wheat, barley and rye, are defined by their solubility in alcohol–water mixtures and are called prolamins. Although early workers used 60–70% (by volume) ethanol, many now use propan-1-ol (50% by volume) or propan-2-ol (55% by volume) as these give more complete extraction. The prolamin fractions are given the trivial names gliadin (wheat and sometimes rye), secalin (rye) and hordein (barley).

Gluten is the proteinaceous mass that remains when wheat dough is washed to remove starch, albumins, and globulins. Conventionally gluten is considered to consist of two major protein fractions, gliadin and glutenin. Glutenins are not readily soluble in 60–70% (by volume) ethanol; however some components are soluble in aqueous alcohols, especially propan-1-ol or propan-2-ol mixtures and at elevated temperatures (Byers *et al.* 1983), and all the major polypeptides behave as prolamins after reduction of disulphide bonds. Similarly some prolamin-like polypeptides can be extracted from rye and barley by aqueous alcohols only in the presence of a reducing agent (Shewry *et al.* 1978*b*, 1983*e*). We regard these proteins as prolamins in which extensive disulphide bonding has resulted in limited solubility or total insolubility in aqueous alcohols. In wheat and rye, as in barley (Wilson *et al.* 1981), the true glutelins are probably a mixture of structural and metabolic proteins.

All prolamins, including the glutenins of wheat, are deposited in protein bodies in the developing endosperms (Miflin *et al.* 1981, 1983; Shewry *et al.* 1983*c*) and are not synthesized in any other tissues. They also have characteristic amino acid compositions being rich in *prol*ine and glut*amine* (hence the name prolamin) and low in charged amino acids, notably lysine which is consequently the nutritionally limiting amino acid in the whole grain. Electrophoretic analyses show that prolamin fractions are complex mixtures of polypeptides, with over 50 components present in wheat. There is also extensive variation in the numbers and electrophoretic properties of the components between cultivars (Shewry *et al.* 1978*a*, 1983*e*; Wrigley & Shepherd, 1973; Mecham *et al.* 1978). This polymorphism is illustrated in figure 1, plate 1, that shows two dimensional analyses of prolamin fractions from barley, wheat and rye.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

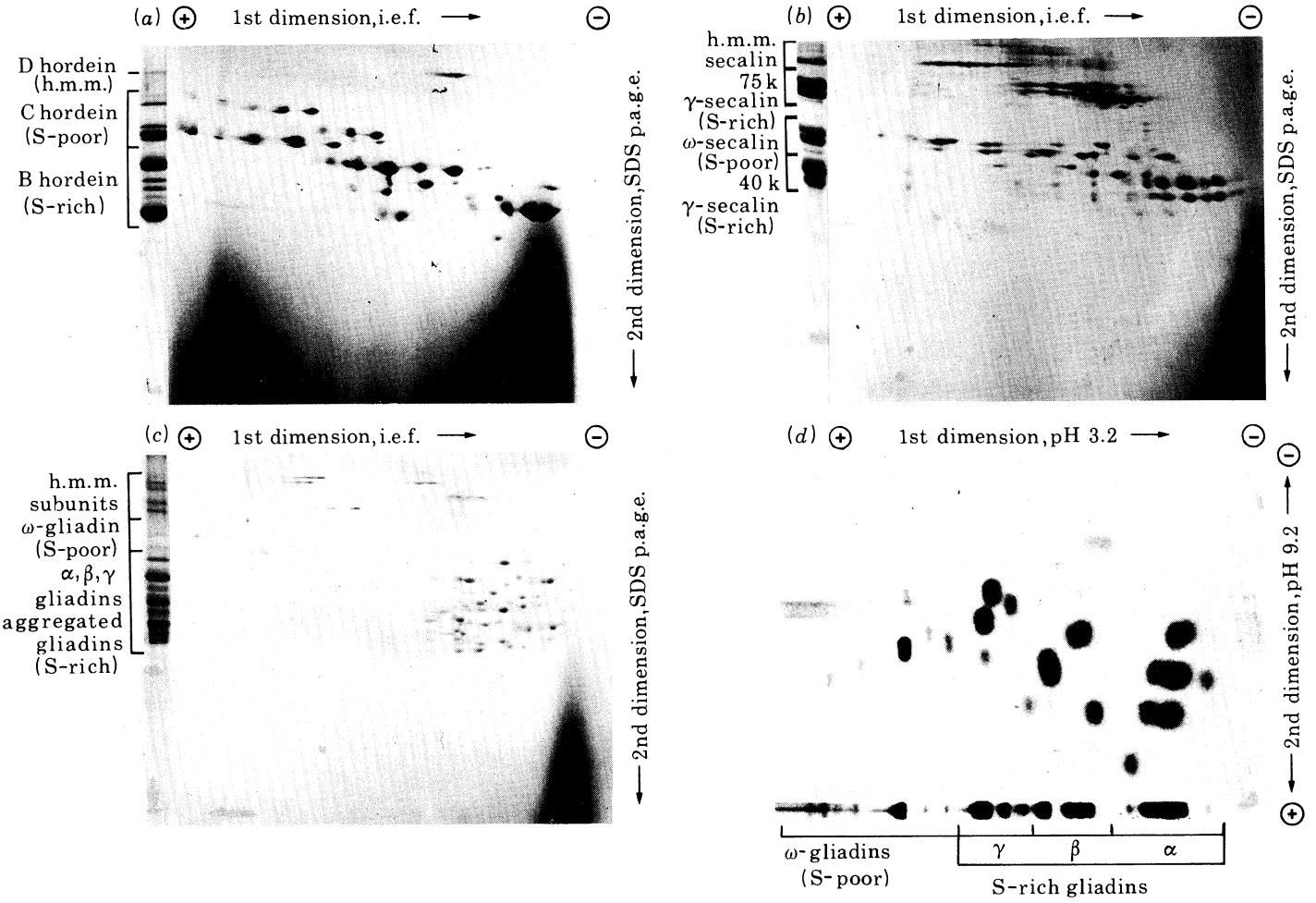

FIGURE 1. Two dimensional analyses of prolamin fractions. (*a*), (*b*) and (*c*), Reduced and pyridylethylated total prolamin fractions from barley, (c.v. Sundance), rye (homozygous line MPI 209) and wheat (c.v. Cheyenne) respectively, separated by the i.e.f. (pH range 5–9) and SDS p.a.g.e. system of Rahman *et al.* (1982). (*d*), Unreduced monomeric gliadins of wheat (c.v. Cheyenne) separated by the two pH systems of Mecham *et al.* (1978). Reproduced with permission from Kasarda (1980).

(*Facing p.* 298)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

We have recently proposed a classification of the prolamins of barley, rye and wheat (Miflin *et al.* 1983), based on their chemical characteristics and genetic control. These groups, called the high molecular mass (h.m.m.)† prolamins, S-poor prolamins and S-rich prolamins, are indicated in figure 1 and will now be discussed individually.

High molecular mass prolamins

These appear to be present entirely in aggregates stabilized by disulphide bonds, and in wheat are often called h.m.m. subunits of glutenin. Cultivars of *T. aestivum* have either one or no subunits controlled by the A genome, one or two subunits controlled by the B genome and two subunits controlled by the D genome (Payne *et al.* 1981). These subunits are controlled by loci on the long arms of the homoeologous (homologous but nonpairing) chromosomes 1A, 1B and 1D (Payne *et al.* 1982*b*) (figure 2). These loci have been designated *Glu-A1*, *Glu-B1* and *Glu-D1*. A single h.m.m. component, called D hordein, is present in European cultivars

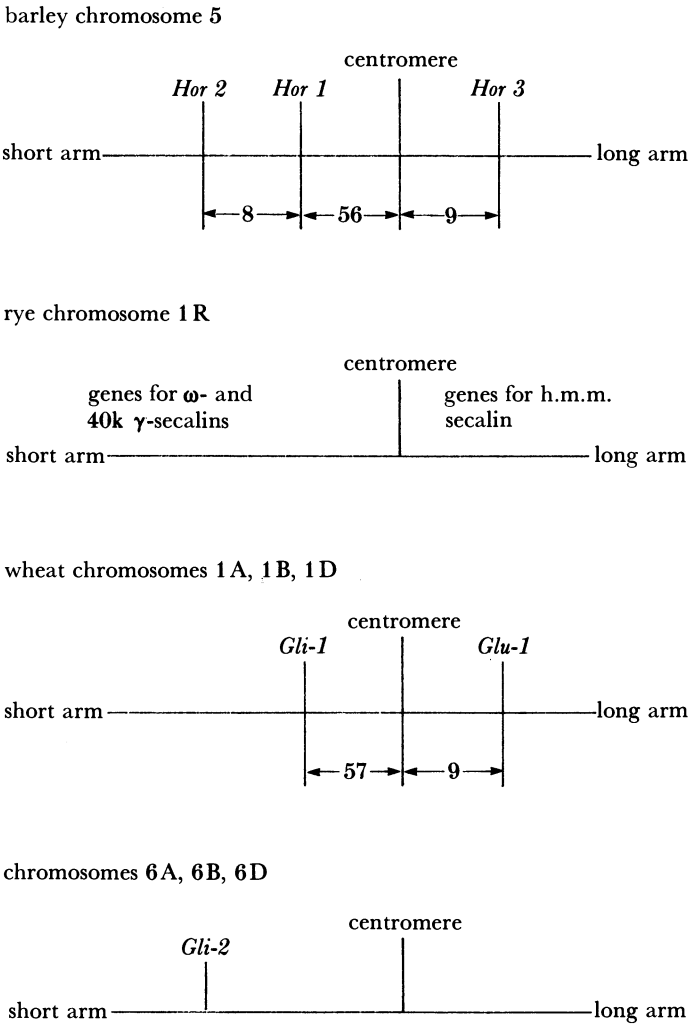

FIGURE 2. The chromosomal locations of structural genes for prolamins in barley, rye and wheat. The position of the centromere on barley chromosome 5 is provisional. Map distances are in centimorgans (cM). Data from Shewry *et al.* (1983*b*), Jensen (1983), and Payne *et al.* (1982*b*). A more detailed account of the chromosomal locations of wheat storage protein genes is given by Payne *et al.* (this symposium).

† These groups are the same as the HMW groups referred to in the author's previous papers.

of barley, although two components may be present in other lines (Shewry & Miflin 1982). It is controlled by the *Hor 3* locus, which is located on the long arm of chromosome 5 in a similar position to the *Glu-1* loci on the group 1 chromosomes of wheat (Shewry *et al.* 1983*b*) (figure 2). Prolamin fractions from inbred (homozygous) lines of rye usually give one or two h.m.m. bands on sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS p.a.g.e.) (Shewry *et al.* 1983*e* and unpublished results), although more bands may be present in fractions from heterozygous commercial grain. These are controlled by genes on the long arm of chromosome 1R (Lawrence & Shepherd 1981).

The h.m.m. prolamins have the highest subunit molecular masses (M_rs) of all the prolamins. These have been determined by SDS p.a.g.e. as 95000–145000 for wheat (Payne *et al.* 1980), 105000 for barley (Shewry & Miflin 1983) and above 100000 for rye (Shewry *et al.* 1983*e*). The M_rs determined by sedimentation equilibrium ultracentrifugation are much lower; 69600, 67600 and 54700 for purified components from wheat, rye and barley respectively (Field *et al.* 1982). This overestimation of M_rs by SDS p.a.g.e. has been reported for some other groups of prolamins of barley, rye and wheat and may be related to their high content of proline (see Field *et al.* 1982 for a discussion).

In all cases the apparently single bands separated by SDS p.a.g.e. can be separated into isomers by isoelectric focusing (see figure 1 and Shewry *et al.* 1983*a*). The amino acid compositions of the h.m.m. prolamins are characterized by a high content of glycine (14–20 mol %), which usually exceeds the content of proline (13–17 mol %) (table 1). Their content of cysteine (0.4–1.5 mol %) corresponds to several residues per mol, enabling them to form disulphide-stabilized aggregates.

Only limited amino acid sequence data is available for this group. D hordein is blocked to N-terminal Edman degradation, while the rye components have not been purified in sufficient amounts for sequence determination (Field *et al.* 1982). However, we have recently reported the N-terminal amino acid sequences of a number of h.m.m. subunits of wheat (table 2). Four sequence types were present that had a high degree of homology, but differed in amino acid substitutions and deletions of three or seven amino acids. Two of the types appeared to be characteristic of subunits coded by the A and B genomes respectively. The second two types were present in 1D-coded subunits, one being characteristic of the high molecular mass 1Dx subunits and the second of the low molecular mass 1Dy subunits as defined by Payne *et al.* (1981). All the sequences had two cysteine residues, the distance between these being affected by the deletions and consequently varying from seven to 14 residues.

S-poor prolamins

These form a clearly defined group of homologous proteins in the three species. The controlling genes for the ω-gliadins of wheat (designated *Gli-A1*, *Gli-B1* and *Gli-D1*) and C hordein of barley (designated *Hor 1*) are located in equivalent positions on the short arms of chromosomes 1A, 1B and 1D of wheat and 5 of barley (Payne *et al.* 1982*a*, *b*; Shewry *et al.* 1980*b*; Jensen *et al.* 1980) (figure 2). The precise locations of the genes that control ω-secalins are not known, although they have been reported to be present on the short arm of chromosome 1R (see Lawrence & Shepherd 1981).

The M_rs determined by SDS p.a.g.e. vary between about 44000 and 74000, with the major components usually around 50000–60000 (Shewry *et al.* 1981, 1983*e*; Kasarda *et al.* 1983). Sedimentation equilibrium ultracentrifugation of a total C hordein fraction gave a value of

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 1. Partial amino acid compositions (expressed as mol %) of prolamin components

	h.m.m.			S-poor			S-rich							
prolamin	barley	rye	wheat	barley	rye	wheat	barley	rye		wheat				
group species	D	h.m.m.	h.m.m.	C	ω-	ω-	B	40k	75k	α(A)	β_5	γ_1	γ_2	aggregated
component	hordeins	subunits	subunits	hordein	secalin	gliadin	hordein	γ-secalin	γ-secalin	gliadin	gliadin	gliadin	gliadin	gliadin
reference	1	2	3	4	5	5	4	6	6	7	8	8	8	9
glutamate + glutamine	29.6	34.0	34–39	41.2	42.9	43–53	35.4	34.8	40.3	38.1	39.6	41.7	39.1	38.4
proline	11.6	13.7	13–17	30.6	30.6	20–30	20.6	18.4	23.5	14.8	15.8	15.1	18.7	15.0
glycine	15.7	16.5	14–20	0.3	1.4	0.9–1.4	1.5	2.4	1.7	2.6	1.9	2.4	2.7	3.3
phenyl-alanine	1.4	0.9	0.3–1.9	8.8	7.4	8.1–8.9	4.8	5.3	5.4	3.8	3.7	3.7	5.2	4.7
cysteine	1.5	1.4	0.4–1.5	0	0	0	2.5	2.5	1.8	2.1	1.8	1.8	1.9	2.7
methionine	0.2	0.3	0–0.4	0.2	0.1	0–0.1	0.6	1.0	0.6	0.8	1.1	0.9	1.7	0.6
lysine	1.1	0.2	0.6–1.3	0.2	0.3	0.1–0.3	0.5	0.7	0.4	0.5	0.2	0.1	0.7	0.5

1, Kreis *et al.* 1983; 2, Field *et al.* 1982; 3, Shewry *et al.* 1983*a*; 4, Shewry *et al.* 1980*c*; 5, Kasarda *et al.* 1983; 6, Shewry *et al.* 1982; 7, Platt & Kasarda 1971; 8, Bietz *et al.* 1977; 9, Shewry *et al.* 1983*d*.

between 90 and 100% of the recovered glutamate residues are present in hordein and gliadin as glutamine (Charbonnier 1974; Shewry *et al.* 1980*b*; Ewart 1981).

Table 2. N-terminal amino acid sequences of prolamin components

h.m.m. prolamins	positions																									
	−1	1				5					10					15					20					25	reference
wheat (*T. aestivum*)																										
1A subunit 1		E	G	E	A	S	G	Q	L	Q	C	E	X	E	L	Q	E	G/P	X				L	K	A	C	1
1B subunit 7		E	G	E	A	S	G	Q	L	Q	C	E	X	E	L	E									A	C	1
1Dx subunits 2/5		E	G	E	A	S	E	Q	L	Q	C	E	X	E	L	Q	E	L	Q	E	X	E	L	K	A	C	1
1Dy subunits 10/12		E	G	E	A	S	R	Q	L	Q	C	E	R	E	L	Q					X	Q	L	E/K	A	C	1
S-poor prolamins																										
barley C hordein		R	Q	L	N	P	S	S	Q	E	L	Q	S	P	Q	Q	S	Y	L	Q	Q	P	Y				2, 3
rye ω-secalin		R	Q	L	N	P	S	E	Q	E	L	Q	S	P	Q	Q	P	V									4
wheat ω-gliadin																										
ancestral type†	A	R	Q	L	N	P	S	D	Q	E	L	Q	S	P	Q	Q	L	Y	P	Q	Q	P	Y	P	Q	Q	2
ancestral type‡	A	R	E	L	N	P	S	N	K	E	L	Q	S	P	Q	Q	S	F	S								4
variant type‡	S	R	L	L	S	P	R	G	K	E	L	H	T	P	Q	Q	Q	F	P	Q	Q	X	X	F	P		4
variant type‡									K	E	L	Q	S	P	Q	Q	S	F	S	H	Q	Q	Q	P	F	P	4
S-rich prolamins																										
Rye																										
40k γ-secalin		N	M	Q	V	G	P	S	G	Q	V	E	W	P	Q	Q	Q	P	L	P	Q						5
75k γ-secalin		N	M	Q	V	N	P	S	G	Q	V	Q	W	P	Q	Q	Q	P	F	P	Q						5
wheat (*T. aestivum*)																										
γ-type sequence¶		N	I/M	G	V	D	P	W	G	Q	V	Q	W	L/V	P	Q	Q	Q/L	V/Q	P	Q	L/Q	Z/Q				6
α-type sequence§		V	R	V	P	V	P	Q	L	Q	P	Q	N	P	S	Q	Q	Q	P	Q	E	Q	V	P	L	V	6, 7
aggregated gliadins		S	H	I	P	G	L	E	R	P	S	Q	Q	Q	Q	L											8

The sequences are aligned for maximum homology.

references: 1, Shewry *et al.* 1983*a*; 2, Shewry *et al.* 1980*a*; 3, Shewry *et al.* 1981; 4, Kasarda *et al.* 1983; 5, Shewry *et al.* 1982; 6, Bietz *et al.* 1977; 7, Kasarda *et al.* 1974; 8, Shewry *et al.* 1983*d*.

† Preparation from *T. monococcum*. This was a mixture of at least two components which differed in the absence or presence of the additional N-terminal alanine. ‡ Components from *T. aestivum*. ¶ Present in γ_2 and γ_3 gliadins. At variant positions the left-hand residue is present in γ_2-gliadin and the right in γ_3-gliadin. § Present in α_2, α_{8-12}, β_5 and γ_1-gliadins.

Standard single letter notations for amino acids are used: A, alanine; C, cysteine; 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; V, valine; W, tryptophan; Y, tyrosine; Z, glutamate or glutamine; X, undetermined.

52000 (Shewry *et al.* 1980*c*), but widely differing values of 27000 (Hamauzu *et al.* 1974) and 75000–79000 (Booth & Ewart 1969) have been reported for preparations of ω-gliadins.

Apart from the complete absence of cysteine and virtual absence of methionine the S-poor prolamins are characterized by high contents of glutamine (over 40 mol %), proline (over 30 mol % in most components) and phenylalanine (7 to 9 mol %). These three residues together account for about 80% of the total, and to a large extent determine the properties of the proteins.

N-terminal amino acid sequences are known for components from all three species (table 2). This confirms the chemical homology, but a number of differences are present. The sequences determined for several C hordein components and for ω-secalins are identical for the first 15 residues, apart from one amino acid substitution. A closely related sequence was reported for an ω-gliadin preparation from *T. monococcum*, although this was clearly a mixture of at least two components differing only in the absence or presence of an additional N-terminal alanine residue. This sequence type, with the additional alanine residue, is also present in some ω-gliadins from *T. aestivum* and *T. durum*; Kasarda *et al.* (1983) have suggested that it may be the ancestral type. Two variant types of N-terminal amino acid sequences have been reported for other ω-gliadins. Both differ from the ancestral type in the first eight residues (positions −1 to 7 in table 2) that are deleted in one variant and substituted extensively in the other. Some positions appear to vary more than others, for example five different amino acids have been found at position 6, four at position 16 and three at positions 17 and 18.

S-rich prolamins

Wheat contains three groups of S-rich prolamins that can be called α-type gliadins, γ-type gliadins (Autran *et al.* 1979) and aggregated gliadins (Shewry *et al.* 1983*d*). The precise chemical and genetic relationships of these to each other, and to the more clearly defined groups of S-rich prolamins of barley and rye, are not known. However, the similarity in the amino acid compositions of all these groups (table 1), and the fact that most have M_rs by SDS p.a.g.e. of between 30000 and 45000, suggests that they may be homologous.

Only one group of S-rich prolamins, called B hordein, is present in barley. This is encoded by a single structural locus (*Hor 2*) located about 7–8 cM distally to *Hor 1* on the short arm of chromosome 5 (Shewry *et al.* 1980*b*; Jensen *et al.* 1980; Jensen 1983) (figure 2). B hordeins have M_rs between 35000 and 46000 by SDS p.a.g.e. (Faulks *et al.* 1981) and 32000–35000 by sedimentation equilibrium ultracentrifugation (Miflin *et al.* 1983). Their amino acid compositions are characterized by the presence of about 35 mol % glutamate and glutamine, 20 mol % proline and 2.5 mol % cysteine (table 1). They are blocked to Edman degradation and N-terminal amino acid sequences have not been determined. Although some B hordein polypeptides are present in disulphide stabilized aggregates, others probably only have intramolecular disulphide bonds (Field *et al.* 1983).

Two groups of S-rich prolamins are present in rye. These have M_rs around 40000 and 75000 by SDS p.a.g.e. (33000 and 54000 by sedimentation equilibrium ultracentrifugation) and have been called 40k and 75k γ-secalins (Shrewry *et al.* 1982). The 40k group are controlled by genes on the short arm of chromosome 1R (figure 2) (see Lawrence & Shepherd 1981). The 40k group have similar amino acid compositions and M_rs to B hordein (table 1) and these are presumed to be homologous. The 75k γ-secalins are the only S-rich prolamins with M_rs greatly above 45000. Although they contain more glutamate and glutamine (40 compared to 35 mol %) and

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

proline (23.5 compared to 18 mol %) than the 40k group (table 1), their homology is clearly established by the presence of only two different residues out of the first 20 at the N-terminus (table 2). Whereas the 75k γ-secalins appear to be present only in aggregates stabilized by disulphide bonds, the 40k group probably only have intramolecular disulphide bonds (Shewry *et al.* 1983*e*; Field *et al.* 1983).

Several groups of S-rich prolamins of wheat can be recognized on the basis of their genetic control, aggregation properties and N-terminal amino acid sequences. The α-, β- and γ-gliadins are groups defined by their electrophoretic mobility at low pH (figure 1). They are present only as monomers with intramolecular disulphide bonds. They have M_rs of 30000–44000 by SDS p.a.g.e., 30000–37000 by sedimentation equilibrium ultracentrifugation (see Miflin *et al.* 1983 for references). A further group of S-rich prolamins are present only in aggregates stabilized by disulphide bonds. These have been given a number of names, but we prefer the term aggregated gliadins (Shewry *et al.* 1983*d*). They have slightly higher M_rs (around 40000–50000) than the α-, β- and γ-gliadins. All these groups have amino acid compositions that are generally similar to those of the S-rich prolamins of barley and rye (table 1), although there is generally less proline (around 15 mol %).

Genetic studies show that the S-rich prolamins of wheat are controlled by loci on the short arms of the group 1 and group 6 chromosomes. The *Gli-1* loci on the group 1 chromosomes appear to control most γ-gliadins in addition to ω-gliadins. Other γ-gliadins, α-gliadins and β-gliadins are controlled by additional loci (provisionally designated *Gli-2*) on the group 6 chromosomes (Wrigley & Shepherd 1973; Kasarda *et al.* 1976; Mecham *et al.* 1978; Sozinov & Poperelya 1982; Payne *et al.* 1982*a*, *b*). More recent studies by Payne and associates have shown that the aggregated gliadins are also encoded by genes on the short arms of the group 1 chromosomes, and that these genes are located close to or at the *Gli-1* loci (Jackson *et al.* 1983; Payne *et al.*, this symposium).

The monomeric gliadins encoded by the *Gli-1* and *Gli-2* loci also have distinct N-terminal amino acid sequences. The γ_2- and γ_3-gliadins encoded by *Gli-1* have a γ-type sequence that is clearly related to those of the γ-secalins of rye, while the α-gliadins, β_5-gliadin and γ_1-gliadin encoded by *Gli-2* have a distinctly different α-type sequence (table 2). A third sequence type is present in the aggregated gliadin subunits (table 2).

The origin and relationships of the groups of S-rich prolamins are not completely understood. The γ-type gliadins of wheat are clearly related genetically and chemically to the 40k γ-secalins of rye. Also, M. D. Dietler and D. D. Kasarda (unpublished) have purified a peptide from γ_3-gliadin and shown that its amino acid sequence is related to a sequence previously reported for the carboxy terminal end of B hordein (Schmitt & Svendsen 1980; Forde *et al.* 1981). This establishes that B hordein, γ_3-gliadin and the 40k γ-secalins are homologous and these three groups can probably be regarded as ancestral forms. The 40k γ-secalins of rye gave rise, it is presumed, to the 75k group by the insertion of one or more sequences. This must have affected the ability of the cysteine residues to form inter- or intramolecular disulphide bonds, with the result that the 75k γ-secalins are completely aggregated. In wheat, divergence has given rise to gliadins that have the ability to form disulphide stabilized aggregates, but are still encoded by genes close to those for γ-type gliadins in the *Gli-1* locus. The origin of the α-type gliadins is still uncertain. In barley the divergence into separate groups of aggregated and non-aggregated S-rich prolamins has not occurred, both types of polypeptide being present in B hordein.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The established and proposed relationships of the prolamin groups are summarized in figure 3.

Origin and evolution of prolamin genes

Prolamins have only been reported in the seed of the Gramineae, one of the most advanced families of flowering plants. They can therefore be regarded as a late evolutionary development.

It is probable that the genes encoding them originated by the duplication of short DNA

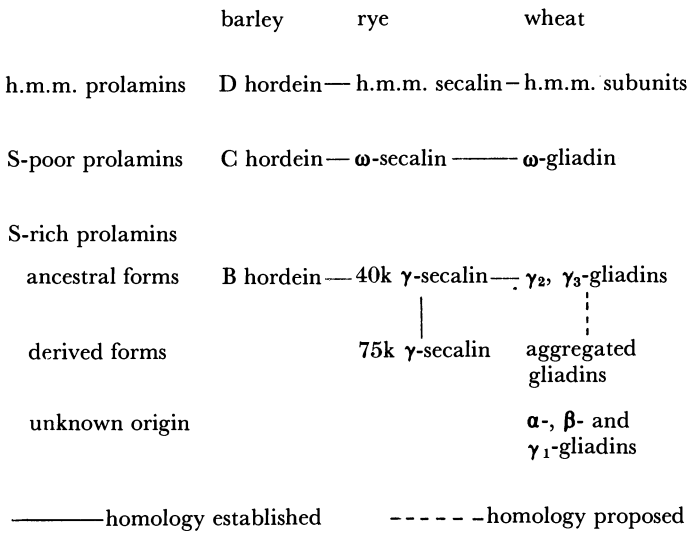

FIGURE 3. The proposed relationships of the prolamins of barley, rye and wheat.

sequences encoding peptides rich in glutamine and proline. Shewry *et al.* (1980*a*) reported the presence of a repeated pentapeptide (Pro-Gln-Gln-Pro-Tyr) at positions 18–23 and 24–28 of an ω-gliadin from *T. monococcum*. Related pentapeptides were subsequently found in the same region of other S-poor prolamins including C hordein (Kasarda *et al.* 1983). This led Kasarda (1980) to propose that pentapeptides were the basic building blocks of all groups of gliadins. More recently D. D. Kasarda (unpublished) has found evidence that the A gliadin group of α-gliadins contain some repeated blocks of about 12 residues.

We have since prepared two chymotryptic peptides from C hordein and determined their N-terminal amino acid sequences by automated Edman degradation. Both contained repeats of an octapeptide with a consensus sequence of Pro-Gln-Gln-Pro-Phe-Pro-Gln-Gln. The ratio of amino acids present in this peptide (4 Gln:3 Pro:1 Phe) is close to the molar ratio of these amino acids in C hordein (table 1), which is consistent with the origin of the main part of the polypeptide from the repeated octapeptide. The first five residues of the octapeptide are similar to the pentapeptides that are characteristic of the N-terminal regions of the S-poor prolamins. This suggests that one of the peptides was the ancestral form, with the second one derived from it by the addition or deletion of three residues.

The origin of a single polypeptide from two or more different ancestral peptides might provide a basis for the presence of distinct structural domains which is a characteristic of prolamins. Thus the N-terminal regions of S-rich, S-poor and h.m.m. prolamins all show distinct differences in amino acid composition compared to the whole proteins (table 3), with

more charged amino acids usually being present. The localization of charged amino acids in the N-terminal and C-terminal regions of S-poor and h.m.m. prolamins is also indicated by peptide mapping (Shewry *et al.* 1981, 1983*a*). An alternative explanation for the presence of structural domains is that these have some functional significance.

The ancestral prolamin genes generated by these tandem duplications have themselves duplicated and diverged to give complex families of genes that code for structurally related groups of polypeptides. The results discussed here show that this divergence has involved the

TABLE 3. COMPARISON OF AMINO ACID COMPOSITIONS OF WHOLE PROLAMINS AND N-TERMINAL REGIONS (25 AMINO ACIDS)

	mol %					
	Gln	Pro	Gly	Phe	Arg + His + Lys	Glu + Asp
h.m.m. wheat subunit 2						
total	33	15	20	0.5	2.6	3
N-terminus	20	0	4	0	12	32
S-poor barley C hordein						
total	37	32	0.4	9	1.4	4
N-terminus	32	16	0	0	4	4
S-rich wheat α-gliadin						
total	34	15	2.6	3.8	4.4	4
N-terminus	32	24	0	0	4	4

insertion and deletion of sequences as well as amino acid substitutions. The insertions and deletions may have arisen partly by unequal crossing over, which would be facilitated by the high frequency of some codons (for example, those for Gln and Pro) and the presence of repeated sequences. These could also result in mispairing of bases, leading to errors in replication of the DNA. These changes have resulted in complex mixtures of homologous proteins that vary widely in their molecular masses, isoelectric points and charge at low pH (see figure 1). The acceptability of such major changes in structure indicates that these proteins are not subject to strong selection pressures. This is in contrast to globulin storage proteins, in which N-terminal sequence homology has been retained between the 11 *S* globulin of soybean (a dicotyledonous plant) and the 12 globulin of oats (Walberg & Larkins 1983). This may be because globulins, unlike prolamins, are assembled into regular holoproteins and transported into protein bodies (see Boulter 1982). These requirements must place considerable constraints on the divergence of the polypeptides. In contrast the most important constraint on the divergence of prolamins is probably to maintain a low solubility in aqueous solutions that will ensure that the newly synthesized polypeptides aggregate into deposits that eventually become protein bodies. Hence they have few charged amino acids resulting in a low net charge at any pH, a high proportion of glutamine residues probably involved in hydrogen bonding and fairly high proportions of hydrophobic amino acids.

It is difficult to compare the rates of divergence of prolamins with those of other proteins because of the absence of any fossil record to enable the times of divergence of different species to be calculated. However, the major changes that have occurred between the prolamin

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

components of wheat, rye and barley, which probably diverged fairly recently, indicate that they are one of the most rapidly evolving groups of proteins yet reported.

We are grateful to our colleagues at Rothamsted and at Albany who have collaborated in the work and contributed valuable advice and discussions. These include, Dr J. M. Field, Ms S. Parmar, Ms A. Faulks, Dr S. Rahman, Ms J. Forde, Dr M. Kreis, Dr B. Forde (Rothamsted) and Dr J-C. Autran, Dr C. C. Nimmo, Ms M. Dietler and Ms E. Lew (Albany). In addition we wish to acknowledge the invaluable advice and assistance of the following from other institutes: Dr P. I. Payne, Dr R. A. Finch (P.B.I., Cambridge), Dr H. Doll, Dr J. Jensen (Risø) and Mr R. Pickering (W.P.B.S., Aberystwyth). This work has been supported in part by grants from the E.E.C. (number 470 under the plant protein programme) and N.A.T.O. (number 0324/82).

References

Autran, J-C., Nimmo, C. C., Lew, E. J-L. & Kasarda, D. D. 1979 N-terminal amino acid sequencing of prolamins from wheat and related species. *Nature, Lond.* **285**, 527–529.

Bietz, J. A., Huebner, F. R., Sandersen, J. E. & Wall, J. S. 1977 Wheat gliadin homology revealed through N-terminal amino acid sequence analysis. *Cereal Chem.* **54**, 1070–1083.

Booth, M. R. & Ewart, J. A. D. 1969 Studies on four components of wheat gliadins. *Biochim. biophys. Acta* **181**, 226–233.

Boulter, D. 1982 Biochemistry of storage protein synthesis and deposition in the developing legume seed. In *Advances in botanical research* vol. 9 (ed. H. W. Woolhouse), pp. 1–32. London: Academic Press.

Bright, S. W. J. & Shewry, P. R. S. 1983 Improvement of protein quality in cereals. *CRC Crit. Rev. Pl. Sci.* **1**, 49–93. the press.)

Byers, M., Smith, S. J. & Miflin, B. J. 1983 A quantitative comparison of the extraction of protein fractions from wheat grain by different solvents and of the polypeptide and amino acid composition of the alcohol-soluble proteins. *J. Sci. Fd Agric.* **34**, 447–462.

Charbonnier, L. 1974 Isolation and characterization of ω-gliadin fractions. *Biochim. biophys. Acta* **359**, 142–151.

Einhof, H. 1805 *Neus allgem. J. d. Chem.* **5**, 131.

Einhof, H. 1806 *Neus allgem. J. d. Chem.* **6**, 62.

Ewart, J. A. D. 1981 Determination of carboxyl groups in wheat and barley flour proteins. *J. Sci. Fd Agric.* **32**, 572–578.

Faulks, A. J., Shewry, P. R. & Miflin, B. J. 1981 The polymorphism and structural homology of storage polypeptides (hordein) coded by the *Hor-2* locus in barley (*Hordeum vulgare* L.). *Biochem. Genet.* **19**, 841–858.

Field, J. M., Shewry, P. R. & Miflin, B. J. 1983 Aggregation states of alcohol-soluble storage proteins of barley, rye, wheat and maize. *J. Sci. Fd Agric.* **34**, 362–369.

Field, J. M., Shewry, P. R., Miflin, B. J. & March, J. F. 1982 The purification and characterization of homologous high molecular weight storage proteins from grain of wheat, rye and barley. *Theor. appl. Genet.* **62**, 329–336.

Forde, B. G., Kreis, M., Bahramian, M. B., Matthews, J. A., Miflin, B. J., Thompson, R. D., Bartels, D. and Flavell, R. B. 1981 Molecular cloning and analysis of cDNA sequences derived from poly A^+ RNA from barley endosperm: identification of B hordein related clones. *Nucl. Acids Res.* **9**, 6689–6707.

Hamauzu, Z., Toyomasu, T. & Yonezawa, D. 1974 Molecular weight determination of wheat gliadin fractions in gel filtration by SDS-polyacrylamide gel electrophoresis and sedimentation equilibrium. *Agr. Biol. Chem.* **38**, 2445–2450.

Harlan, J. R. 1978 The origins of cereal agriculture in the old world. In *Origins of Agriculture* (ed. C. A. Reed), pp. 357–383. The Hague: Mouton Publishers.

Jackson, E. A., Holt, L. M. & Payne, P. I. 1983 Characterization of high-molecular weight gliadin and low-molecular weight glutenin subunits of wheat endosperm by two-dimensional electrophoresis and the chromosomal location of their controlling genes. *Theor. appl. Genet.* **66**, 29–37.

Jensen, J. 1983 Coordinators report: chromosome 5. *Barley Genet. Newsl.* **13**. (In the press.)

Jensen, J., Jorgensen, J. H., Jensen, H. P., Giese, H. & Doll, H. 1980 Linkage of the hordein loci *Hor 1* and *Hor 2* with the powdery mildew resistance loci *Mlk* and *Mla* on barley chromosome 5. *Theor. appl. Genet.* **58**, 27–31.

Kasarda, D. D. 1980 Structure and properties of α-gliadins. *Ann. Technol. Agric.* **29**, 151–173.

Kasarda, D. D., Autran, J. C., Lew, E. J-L., Nimmo, C. C. & Shewry, P. R. 1983 N-terminal amino acid sequences of ω-gliadins and ω-secalins: implications for the evolution of prolamin genes. *Biochim. biophys. Acta.* (In the press.)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Kasarda, D. D., Bernardin, J. E. & Qualset, C. O. 1976 Relationship of gliadin protein components to chromosomes in hexaploid wheats (*Triticum aestivum* L). *Proc. natn. Acad. Sci. U.S.A.* **73**, 3646–3650.

Kasarda, D. D., Da Roza, D. A. & Ohms, J. I. 1974 N-terminal sequence of α-gliadin *Biochim. biophys. Acta* **351**, 290–294.

Kreis, M., Shewry, P. R., Forde, B. G., Rahman, S., Bahramian, M. B. & Miflin, B. J. 1983 Molecular analysis of the effects of the mutant *lys 3a* gene on the expression of *Hor* loci in developing endosperms of barley (*Hordeum vulgare* L.). *Biochem. Genet.* (In the press.)

Lawrence, G. J. & Shepherd, K. W. 1981 Chromosomal location of genes controlling seed proteins in species related to wheat. *Theor. Appl. Genet.* **59**, 25–31.

Mecham, D. K., Kasarda, D. D. & Qualset, C. O. 1978 Genetic aspects of wheat gliadin proteins. *Biochem. Genet.* **16**, 831–853.

Miflin, B. J., Burgess, S. R., & Shewry, P. R. 1981 The development of protein bodies in the storage tissues of seeds. *J. exp. Bot.* **32**, 119–219.

Miflin, B. J., Field, J. M., & Shewry, P. R. 1983 Cereal storage proteins and their effects on technological properties. In *Seed proteins* (eds. J. Daussant, J. Mosse & J. Vaughan), pp. 255–319. London: Academic Press.

Osborne, T. B. 1924 *The vegetable proteins*, p. 154. London: Longmans Green & Co.

Palmer, G. H. 1980 The morphology and physiology of malting barleys. In *Cereals for food and beverages* (eds. G. E. Inglett & L. Munck), pp. 301–338. London: Academic Press.

Payne, P. I., Law, C. N. & Mudd, E. E. 1980 Control by homoeologous group 1 chromosomes of the high molecular weight subunits of glutenin, a major protein of wheat endosperm. *Theor. appl. Genet.* **58**, 113–120.

Payne, P. I., Holt, L. M., & Law, C. N. 1981 Structural and genetical studies on the high molecular weight subunits of wheat glutenin I. Allelic variation in subunits amongst varieties of wheat (*Triticum aestivum*). *Theor. appl. Genet.* **60**, 229–236.

Payne, P. I., Holt, L. M., Lawrence, G. J. & Law, C. N. 1982*a* The genetics of gliadins and glutenin, the major storage proteins of the wheat endosperm. *Qual. Pl. Fds Hum. Nutr.* **31**, 229–241.

Payne, P. I., Holt, L. M., Worland, A. J. & Law, C. N. 1982*b* Structural and genetical studies on the high molecular weight subunits of wheat glutenin III. Telocentric mapping of the subunit genes on the long arms of the homoeologous group 1 chromosomes. *Theor. appl. Genet.* **63**, 129–138.

Platt, S. G. & Kasarda, D. D. 1971 Separation and characterization of α-gliadin fractions. *Biochim. biophys. Acta.* **243**, 407–415.

Rahman, S., Shewry, P. R. & Miflin, B. J. 1982 Differential protein accumulation during barley grain development. *J. exp. Bot.* **33**, 717–728.

Schmitt, J. M. & Svendsen, I. 1980 Partial amino acid sequence from hordein polypeptide B1. *Carlsberg Res. Commun.* **45**, 549–555.

Shewry, P. R. & Miflin, B. J. 1982 Genes for the storage proteins of barley. *Qual. Pl. Fds Hum. Nutr.* **31**, 251–267.

Shewry, P. R. & Miflin, B. J. 1983 Characterization of and synthesis of barley seed proteins. In *Seed proteins. Biochemistry, genetics and nutritive value* (ed. W. Gottschalk & H. Muller), pp. 143–205. The Hague: Martinus Nijhoff.

Shewry, P. R., Ellis, J. R. S., Pratt, H. M. & Miflin, B. J. 1978*a* A comparison of methods for the extraction and separation of hordein fractions from 29 barley varieties. *J. Sci. Fd Agric.* **29**, 433–441.

Shewry, P. R., Hill, J. M., Pratt, H. M., Leggatt, M. M. & Miflin, B. J. 1978*b* An evaluation of techniques for the extraction and separation of hordein and glutelin from barley seed and a comparison of the protein composition of Bomi and Risø 1508. *J. exp. Bot.* **29**, 677–692.

Shewry, P. R., Autran, J-C., Nimmo, C. C., Lew, E. J-L. & Kasarda, D. D. 1980*a* N-terminal amino acid sequence homology of storage protein components from barley and a diploid wheat. *Nature, Lond.* **286**, 520–522.

Shewry, P. R., Faulks, A. J., Pickering, R. A., Jones, I. T., Finch, R. A. & Miflin, B. J. 1980*b* The genetic analysis of barley storage proteins. *Heredity* **44**, 383–389.

Shewry, P. R., Field, J. M., Kirkman, M. A., Faulks, A. J. & Miflin, B. J. 1980*c* The extraction, solubility and characterization of two groups of barley storage polypeptides. *J. exp. Bot.* **31**, 393–407.

Shewry, P. R., Lew, E. J-L. & Kasarda, D. D. 1981 Structural homology of storage proteins coded by the *Hor 1* locus of barley (*Hordeum vulgare* L.). *Planta* **153**, 246–253.

Shewry, P. R., Field, J. M., Lew, E. J-L. & Kasarda, D. D. 1982 The purification and characterization of two groups of storage proteins (secalins) from rye (*Secale cereale* L.). *J. exp. Bot.* **133**, 261–268.

Shewry, P. R., Field, J. M., Faulks, A. J., Parmar, S., Miflin, B. J., Dietler, M. D., Lew, E. J-L. & Kasarda, D. D. 1983*a* Purification and N-terminal amino acid sequence analysis of high molecular weight (HMW) gluten polypeptides of wheat. (Submitted.)

Shewry, P. R., Finch, R., Parmar, S., Franklin, J. & Miflin, B. J. 1983*b* Chromosomal location of *Hor 3*, a new locus governing storage proteins in barley. *Heredity* **50**, 179–189.

Shewry, P. R., Kreis, M., Burgess, S. R., Parmar, S. & Miflin, B. J. 1983*c* The synthesis and deposition of the storage prolamins (secalins) of rye. *Planta.* (In the press.)

Shewry, P. R., Miflin, B. J., Lew, E. J-L. & Kasarda, D. D. 1983*d* The purification and characterization of an aggregated gliadin fraction from wheat. *J. exp. Bot.* (In the press.)

Shewry, P. R., Parmar, S. & Miflin, B. J. 1983*e* The extraction, separation and polymorphism of the prolamin storage proteins (secalins) of rye (*Secale cereale* L.). *Cereal Chem.* **60**, 1–6.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Sozinov, A. A. & Poperelya, F. O. 1982 Polymorphism of prolamins and variability of grain quality. *Qual. Pl. Fds Hum. Nutr.* **31**, 243–249.

Taddei, G. 1819 *G. Fis. Chem. Stor. nat., Brugnatelli* **2**, 360.

Walberg, G. & Larkins, B. 1983 Structure and synthesis of 11*S* globulin from oat seeds. *Pl Physiol.* **72**, 161–165.

Wall, J. S. 1979 The role of wheat proteins in determining baking quality. In *Recent advances in the biochemistry of cereals* (ed. D. L. Laidman & R. G. Wyn Jones), pp. 275–312. London: Academic Press.

Wendorf, F., Schild, R., El Hadidi, N., Close, A. E., Kobusiewicz, M., Wieckoska, J., Issawi, B. & Haas, H. 1979 Use of barley in the Egyptian Late Palaeolithic. *Science, N.Y.* **205**, 1341–1347.

Wilson, C. M., Shewry, P. R., Faulks, A. J. & Miflin, B. J. 1981 The extraction and separation of barley glutelins and their relationship to other endosperm proteins. *J. exp. Bot.* **32**, 1287–1293.

Wrigley, C. W. & Shepherd, K. W. 1973 Electrofocusing of grain proteins from wheat cultivars. *Ann. N.Y. Acad. Sci.* **209**, 154–162.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

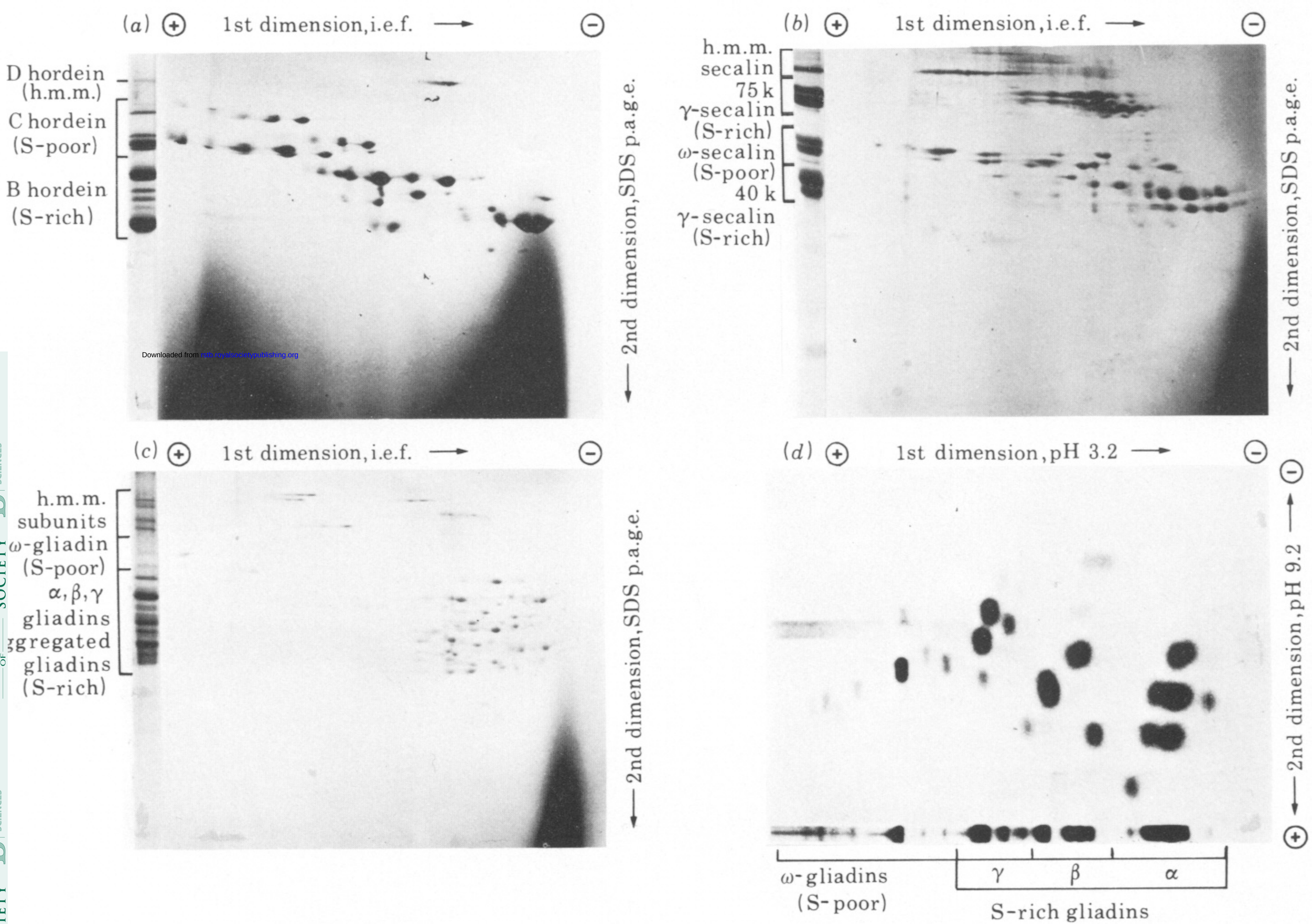

FIGURE 1. Two dimensional analyses of prolamin fractions. (*a*), (*b*) and (*c*), Reduced and pyridylethylated total prolamin fractions from barley, (c.v. Sundance), rye (homozygous line MPI 209) and wheat (c.v. Cheyenne) respectively, separated by the i.e.f. (pH range 5–9) and SDS p.a.g.e. system of Rahman *et al.* (1982). (*d*), Unreduced monomeric gliadins of wheat (c.v. Cheyenne) separated by the two pH systems of Mecham *et al.* (1978). Reproduced with permission from Kasarda (1980).