PETER J. LEA
JEAN-F. MOROT-GAUDRY (EDS.)

Plant Nitrogen

INRA EDITIONS

Springer

The Biochemistry and Molecular Biology of Seed Storage Proteins

Jean-Claude AUTRAN[1], Nigel G. HALFORD[2] and Peter R. SHEWRY[2]

Introduction

Economic importance of seed storage proteins

Most plants synthesise proteins in their organs of reproduction and propagation, such as seeds of gymnosperms and angiosperms. Storage proteins are usually located in two tissues. In dicotyledonous plants they may be located in the diploid cotyledons (exalbuminous), in the triploid endosperm (albuminous) or, occasionally, in both tissues. In monocotyledonous cereals they are primarily located in the triploid endosperm tissue. They are deposited in high amounts in the seed, in discrete deposits (protein bodies) and survive desiccation for long periods of time. In most cases, storage proteins lack any other biological activity and simply provide a source of nitrogen, sulphur and carbon skeletons for the developing seedling (Shotwell and Larkins 1989; Shewry 1995).

From the human point of view, seeds represent the most important plant tissue that is harvested and consumed, consequently, the economic importance of seed proteins is considerable. Seed storage proteins form the most important source of dietary proteins for humans as about 70% of the total intake comes directly from this source. In addition, seed proteins provide the major component of the diet of non-ruminant farm animals. Although proteins make up a relatively small proportion of the cereal grain (usually 7-15%, compared with up to 40% in legumes), cereals are the dominant world crops in terms of both dry matter production and protein pro-

1. INRA, Unité de Technologie des Céréales et des Agropolymères,
2, Place Viala
34060 Montpelier Cedex 01
France
E-mail: autran@ensam.inra.fr
2. IACR-Long Ashton Research Station
Bristol
BS41 9AF
UK
E-mail: peter.shewry@bbsrc.ac.uk

duction. For instance, the annual yields of the eight most important species (wheat, maize, rice, barley, sorghum, oats, millets and rye in order of decreasing importance) exceeded 1700 million metric tons in the 1995-1997 period, which corresponds to about 200 million tons of proteins, *i.e.* in theory a sufficient amount to meet the requirements of mankind. There has therefore been a considerable economic stimulus to the study of cereal proteins and, in particular, the storage proteins that account for 50% of the total (Shewry et al. 1994b). Of the remaining plant proteins (about 70 million tons), nearly all come from dicotyledonous seeds, especially the legumes (soybean, pea, peanut, bean, faba bean, lentil, chickpea, lupin etc.) and oilseeds (cottonseed, sunflower, oilseed rape etc.).

Because of their abundance, the storage proteins are largely responsible for the nutritional quality and technological properties of the seed. These aspects have, therefore, been the subject of considerable research since 1745, when Beccari is credited with having isolated gluten from wheat.

Nutritional Quality

Cereals and legumes are not only the major crops used to provide energy in food and feed, but they also supply most of the proteins consumed by humans and used for animal production. Cereals have some advantages in containing very few types of antinutritional factors, but their storage proteins have low nutritional value, as they are limiting in lysine besides having extremely high levels of non-essential amino acids (*e.g.* glutamine and proline). If the nutritionally inferior storage proteins of the two most important crops in the world, wheat and maize, could be converted into proteins with better nutritional value, it would certainly have a great impact on human nutrition in many areas as well as on animal production (Doll 1984). In contrast, the storage proteins of legume seeds have a much better balance of essential amino acids, although still limiting in methionine and cysteine. They often, however, also contain various types of antinutritional factors (*e.g.* trypsin inhibitors, phytohaemagglutin, α-galactosides, glucosinolates, alkaloids) which may be only partially removed by processing or plant breeding.

Much research has been devoted to increasing the amounts of seed proteins, and their contents of essential amino acids, to improve the nutritional quality of seeds. At first, this seemed possible because seed storage proteins can undergo major changes, as indicated by their high level of biochemical heterogeneity and genetic polymorphism. They may vary between wide limits (as the result of glycosylation, posttranslational processing, gene mutations and environmental effects), with such changes in protein composition being tolerated by the developing seed. However, it must be noted that seed storage proteins also possess certain essential properties that enable them to fulfil their physiological role (Spencer 1984). There may, therefore, be less flexibility than expected to tailor storage protein composition with a view to improve end-use quality. For instance, storage proteins are sequestered in protein bodies where they are not exposed to the proteinases responsible for the breakdown of metabolic proteins, and their structure is likely to contain information that determines their selective transport from the site of synthesis to the site of accumulation. Such constraints on seed storage proteins are also common to all secretory proteins made on the ER and processed in the endomembrane system, and the variation observed is restricted to very specific regions. Although we know the role of the typical leader sequence that directs the transport of the nascent polypeptide through the

membrane of the endoplasmic reticulum and into the lumen, we understand very little of the sequence requirements that specify the subsequent steps in the transport of storage proteins to, and their deposition in, the protein bodies (see below; Spencer 1984).

Use in the Food Industry

Setting aside nutritional considerations, proteins are used as food ingredients for their functional properties, *i.e.* to provide a certain specific function in the product. The proteins of most concern in the food industry are the storage proteins, although this is not to deny that some enzymes may also be important (Miflin et al. 1983). Most functional properties influence the sensory characteristics of food, especially texture, but they can also affect the physical behaviour of foods or food ingredients during processing (*e.g.* mixing, extrusion, fermentation, heating, drying, cooking) and storage. These properties are discussed in more detail in a later section.

Cereal seeds provide the raw material for two of mankind's oldest technologies: the baking of bread and the fermentation of alcoholic beverages. One of the questions that has challenged cereal chemists for a long time is why wheat protein is unique among cereals and other plant proteins in forming a dough with viscoelastic properties ideally suited to making leavened bread (Bushuk and MacRitchie 1989). Today, the basis of wheat "protein quality" remains poorly understood in detail, although most scientists agree on the fact that the ability to form a viscoelastic gluten depends on the capacity of wheat storage proteins to interact and to form polydisperse aggregates in an appropriate balance. In contrast, the storage proteins of barley tend to have, when in excess, a negative influence on endosperm disaggregation during the malting process and on brewing properties.

The functional properties of legume proteins relate mainly to their ability to stabilize emulsions or foams, and to impart textural attributes (Wright and Bumstead 1984). Proteins of legume seeds are often refined using dry (air-classification) or wet (alkaline extraction followed by acid precipitation) methods with selective removal or destruction of undesirable components. Concentrates or isolates are then processed to make meat substitutes or functional agents for the food industry using texturisation.

Although recent research has increased our knowledge of the components of storage proteins so that we are now in a better position to relate protein composition to functionality for specific end uses, the basic mechanisms that determine the functional properties are seldom clearly understood because of the complexity of the various food systems and of the processes by which the raw materials are transformed into end-products (Cheftel et al. 1985).

Interest in Seed Storage Proteins

Seed storage proteins, and especially wheat storage proteins, have been the subject of considerable research during several decades with the use of increasingly sophisticated analytical methods, leading to detailed knowledge of their structures and properties and to impacts on quality improvement through breeding, varietal identification and better control of technological processes. However, studies of cereal proteins have tended to lag behind studies of other more fashionable proteins such

as enzymes. Interestingly, there has been a great resurgence of interest in seed storage proteins over the past two decades. This is partly because the unusual features of storage proteins, such as their synthesis in large amounts in specific tissues at precise stages of development, have made them attractive for studies of cDNA cloning and gene regulation. In addition, the availability of complete amino acid sequences of many plant storage proteins and the recent development of transient expression and transformation systems have stimulated renewed interest in their biophysics and cell biology (Shewry 1995).

This Chapter Reflects All of These Interests

It aims to review our current knowledge on seed storage proteins, focusing on their biochemistry and molecular biology, including classification, structures, evolution, synthesis and deposition, biophysical properties, genetic manipulation and their impact on technological utilization. It is necessary to be selective in order to keep the chapter down to a reasonable size, but the aim is to give a both broad and up-to-date account of seed storage proteins.

Classification of Seed Proteins

Classification is an artificial process reflecting the purpose of the classifier (Boulter and Derbyshire 1978). Seed proteins may be therefore classified in a variety of ways (*e.g.* chemical structure, mechanisms of action, biological function, location, genetic relationships and the separation procedures employed in purification). The ideal classification may be based on the mechanism of protein action but we do not yet know enough about the detailed three-dimensional structures of plant proteins. Other systems have therefore been used, based mainly on separation procedures, biological function (storage, metabolic or structural proteins), physicochemical properties (electrophoretic mobility, contents of sulphur-containing amino acids) or genetics (gene location, duplication and divergence).

In fact, the classification of plant proteins was historically based on the work carried out around the turn of the century by Osborne (1907), who classified plant proteins into groups (called Osborne groups) on the basis of their extraction in a series of solvents: water (albumins), dilute salt solution (globulins), aqueous ethanol (prolamins) and dilute alkali or acid (glutelins). The two former groups essentially included metabolic (*e.g.* enzymatic) and storage proteins, whereas the two latter were mainly storage proteins. Prolamins are usually given trivial names based on the Latin generic name of the cereal, for example secalins in rye (*Secale cereale*), hordeins in barley (*Hordeum vulgare*) and zeins in maize (*Zea mays*). In wheat, the prolamin storage proteins are usually classified into two groups, gliadins and glutenins, which together form gluten. Whereas gliadins are monomeric proteins, glutenins are polymers consisting of disulphide-bonded polypeptides, so-called subunits.

Despite the paucity of knowledge of protein structure in Osborne's time, his classification has proved to be remarkably durable and still provides a framework for modern studies of cereal proteins (Shewry 1995). Although this classification is simple in concept, it has led to a considerable amount of confusion and dispute, as discussed by Shewry and Miflin (1985). The basis of these problems is that the extractability of proteins is affected by many factors including the physiological state

of the tissue. Coupled with this is the fact that many modifications to Osborne's original extraction procedures have been introduced by different workers, without always monitoring the fractions for purity and cross-contamination using electrophoresis and amino acid analysis (Shewry and Miflin 1985). Consequently, in the 1980s, following the proposal of Field et al. (1983), a majority of cereal protein chemists agreed to take into account physicochemical, molecular and genetic properties to redefine prolamins to include cereal storage proteins previously defined as both prolamins and glutelins (*i.e.* gliadins and glutenins of wheat). These proteins are discussed in detail below. In contrast, the glutelins were redefined to include only non-storage proteins (mainly enzymic and structural) and will not be discussed further.

The prolamins are unusual in being restricted to only one family of plants, the grasses, which include the cultivated cereals. This contrasts with the globulin and albumin storage proteins which have wider distributions (Table 1).

Table 1. Major groups of seed storage proteins and their distributions. (Shewry 1995)

Type	2S Albumins	Prolamins	7S Globulins	11S Globulins
Solubility	Water	Aq. alcohols (± reducing agents)	Dilute saline	Dilute saline
Major components	Brassicas Sunflower Castor bean Brazil nut	Cereal endosperms (wheat, barley, rye, maize)	Legumes Cottonseed	Most legumes Cucurbits Brassicas Endosperms of oats and rice Castor bean
Minor components		Oats and rice endosperm	Cereal embryos and aleurones	Wheat endosperm

The globulin storage proteins of seeds were historically separated from pea, soybean and faba bean by cryoprecipitation, differential salt solubility and heat coagulation. Two broad types, called legumins and vicilins, were recognized by Osborne (1924). An important technical advance was the introduction of ultracentrifugation (Danielsson 1949), which allowed characterisation of the main storage proteins, legumins and vicilins, with sedimentation coefficients (S_{20w}) of about 11S-12S and 7S-8S, respectively. These 7S and 11S storage globulins have similar characteristics, but vary widely in their relative amounts depending on the species. The 11S globulins are the most widely distributed (Table 1) whereas the 7S globulins are more restricted in distribution, being present in some legumes, cottonseed and in cereal embryos and aleurones. Finally, the 2S albumins represent a fourth major group of storage proteins, occurring in a range of dicotyledonous species, including oilseeds such as rapeseed, sunflower and castor bean (Youle and Huang 1981).

Structures and Characteristics of Storage Proteins

Storage proteins are probably ubiquitous in seeds. In the vast majority of cases they have no known function apart from providing nutrition (carbon, nitrogen and sulphur) to the developing seedling. It is probable, therefore, that the structures of

storage proteins are not as highly constrained as those of many other proteins, such as enzymes, although there is clearly a requirement that the protein should be efficiently synthesised, packaged, stored and mobilized during germination. Consequently, it is not surprising that storage proteins exhibit great variation in their structures and properties. Nevertheless, almost all of the storage proteins present in the seeds of major crops fall into four groups which derive from only two gene superfamilies.

Globulin Storage Proteins

Globulins are the most widespread type of storage protein and may well prove to be present in all angiosperm seeds, although varying in amount, properties and tissue distribution. Two types are recognised, with sedimentation coefficients of 7-8S and 11-12S. Both have been studied in most detail from legume seeds and are often called legumins (11S) and vicilins (7S), based on the taxonomy of the species from which they were first derived (family Leguminosae, tribe Viciae). These names are currently used for fractions from *Vicia faba* (field bean) and *Pisum sativum* (pea) but different names are used for *Phaseolus* (7S phaseolin) and soybean (11S glycinin, 7S β-conglycinin) globulins. Similarly, specific trivial names are often used for fractions from other plant groups, notably for 11S globulins, which are more widely distributed than 7S. These include cruciferin and helianthinin for 11S globulins of crucifers and sunflower, respectively (Casey 1999).

A typical 11S globulin has an M_r of about 300 000-400 000 and is a hexamer of six subunits (M_r about 60,000) associated by non-covalent forces. Each subunit is posttranslationally "nicked" by a specific proteinase to give acidic and basic polypeptides (M_r about 40 000 and 20 000, respectively), which are linked by a single disulphide bond. Thus, native legumins are broken down into six acidic and six basic polypeptides when treated with reducing agent under denaturing conditions (Casey 1999; Casey and Domoney 1999).

The 7S globulins differ from the 11S in being trimeric, with M_r typically about 150 000-190 000. The subunit M_r is typically about 50 000, but proteolytic processing can lead in some species to the generation of smaller polypeptides while further polymorphism can result from glycosylation. Thus, in pea, proteolysis of M_r 50 000 precursors at one or two sites gives rise to polypeptides ranging in M_r from about 12 000-33 000, some of which may be glycosylated, in addition to uncleaved precursor (Casey 1999; Casey and Domoney 1999). Unlike 11S globulins, the 7S globulins contain no disulphide bonds.

Although the 7S and 11S globulin subunits have little or no amino acid sequence identity, sophisticated alignments and structural predictions indicate that they are indeed related and this is supported by analyses of their three-dimensional structures. Thus, the acidic and basic polypeptides of the 11S globulin subunits appear to correspond to the *N*- and *C*-terminal regions, respectively, of the 7S globulin subunits (Argos et al. 1985; Lawrence et al. 1994). High-resolution 3-D structures have been determined for two 7S globulins, phaseolin (Lawrence *et al.* 1990, 1994) and canavalin from jack bean (Ko et al. 1993). Each subunit comprises two structurally similar units, each consisting of a β-barrel of antiparallel β-strands followed by α-helices which form loops (Fig. 1). The three subunits form trimers of dimensions 90 x 90 x 35 Å (phaseolin) and 80 x 80 x 40 Å (canavalin). Although the structures of 11S globulins have not been determined in such detail, preliminary studies of the trimeric proglycinin expressed in *E. coli* show a similar structure to that of 7S globulins (Utsumi et al. 1993; 1996), with the backbones of the two protein types being readily superimposed. The trimeric proglycinin also has similar dimensions (93 x 93 x 36 Å) to 7S globulins with two trimers being assembled within the vacuole to give a hexamer of about 110 x 110 x 80 Å (Badley et al. 1975).

The similar structures of the 7S and 11S globulins will presumably facilitate their regular packing within protein bodies, the two protein groups occurring together in many species. It also implies that the two groups have evolved from a common ancestor. The presence of two structurally similar units within each 7S/11S globulin subunit indicates that a short ancestral domain may have initially been duplicated to give two domains corresponding to the 7S *N*-terminus/11S acid chain and 7S *C*-terminus/11S basic chain and that this ancestral protein was then duplicated to give the ancestral 7S and 11S globulin subunits, as shown in Fig. 2 (Lawrence 1999).

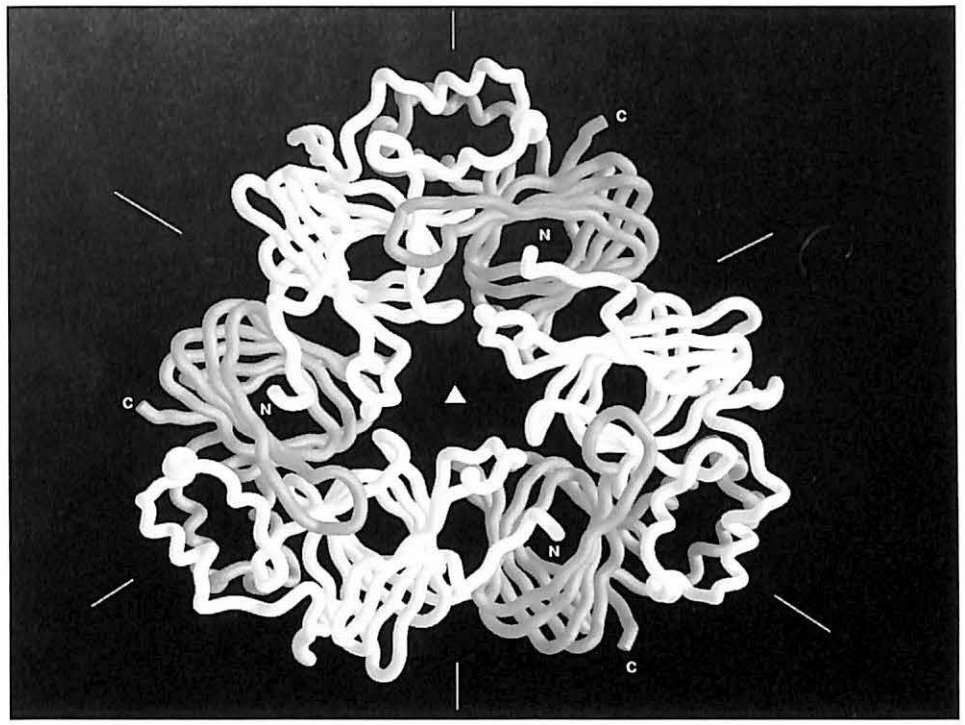

Fig. 1 : Schematic ("backbone-worm") representation of the phaseolin trimer, based on the X-ray structure of Lawrence et al. (1994). The *N* and *C* termini of each polypeptide are labelled. The location of the threefold axis perpendicular to the plane of the figure is indicated by the *white triangle*, whilst the locations of the pseudo-twofold axes are indicated by *white lines*. The latter axes lie in the plane of the paper and occur both as intrasubunit axes (relating the *N*- and *C*-terminal modules of the same polypeptide) and as inter-subunit axes (relating *N*- and *C*-terminal modules of neighbouring polypeptides). The polypeptide link between helix 4 and *C*-terminal strand A is absent in the structure. (Lawrence 1999)

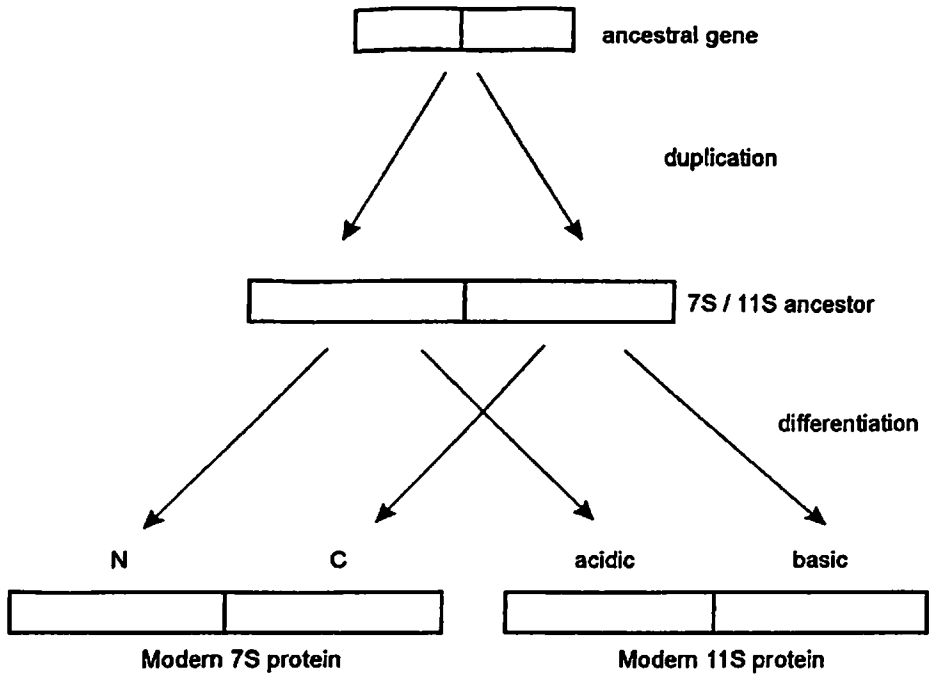

Fig. 2 : Possible evolutionary pathway for 7S and 11S globulins, based on an ancestral gene duplication (Gibbs et al. 1989; Lawrence et al. 1994; Shutov et al., 1995; Lawrence 1999).

In fact, more recent studies have shown that 7S and 11S globulins belong to an even more extensive superfamily of functionally diverse proteins found in plants and microbes. They include "germin" (oxalate oxidase) of germinating wheat embryos, spherulation-specific spherulins of myxomycetes (slime moulds), plant auxin binding proteins and various enzymes. Dunwell (1998) has coined the name "cupins" (latin for small barrel) for this superfamily to reflect the presence of a core β-barrel structure.

2S Albumins

Albumin storage proteins were initially defined as a group in 1981 when Youle and Huang (1981) isolated and characterized 2S albumin fractions from seeds of 12 botanically diverse species including two legumes but only one monocot (*Yucca*, Liliaceae). Detailed characterization has since been reported for 2S albumins from a range of dicotyledonous plants allowing the basic structure to be defined, but the presence of homologous proteins in monocots remains to be confirmed.

A typical 2S albumin (*e.g.* the napins of oilseed rape and other crucifers) comprises two subunits of about 30-40 residues and 60-90 residues, respectively, with two interchain disulphide bonds and two intrachain bonds within the large subunit. It is synthesised as a single precursor protein with posttranslational proteolysis resulting in the loss of an *N*-terminal prosequence and a linker peptide between the two subunits. However, there is considerable variation on this basic structure (see Fig. 3; Shewry and Pandya 1999).

1. Pro and/or linker sequences are absent from some albumins.
2. In castor bean two heterodimeric proteins are released by proteolysis of a single precursor protein.
3. In sunflower the mature protein consists of a single subunit (*i.e.* cleavage into large and small subunits does not occur) with either one or two albumins being released from a single precursor protein.
4. Most albumins have a conserved cysteine skeleton of eight residues, which form four disulphide bonds as discussed above. However, an additional unpaired cysteine is present in conglutin δ of lupin while the pea albumin subunits PA1a/PA1b contain a total of ten cysteine residues which do not apparently form interchain disulphide bonds.

Of particular interest is the presence in some species of methionine-rich components, the most widely studied being in Brazil nut (19 methionines out of 101 residues) and sunflower (16 methionines, 103 residues). The main interest in these proteins has been in relation to improving sulphur-poor forage and legume crops by genetic engineering. Although work on the Brazil nut protein was discontinued when it was shown to be allergenic, Molvig et al. (1997) reported that the expression of SFA8 in lupin seeds resulted in a 98% increase in seed methionine. However, the

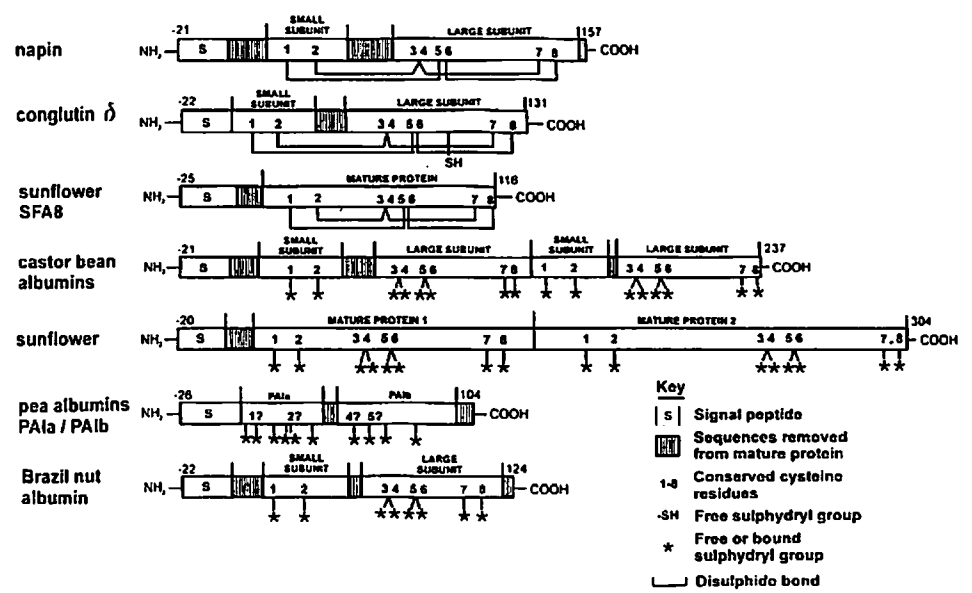

Fig. 3 : Schematic depictions of various types of 2S albumin, indicating their origins from precursor proteins and their disulphide structures. Conserved cysteine residues are numbered *1-8* and free sulphydryl groups shown as *-SH*. Cysteine residues whose behaviour is unknown are indicated by *. The precise correspondence between the cysteines in the pea albumins PA1a and PA1b and those in the other albumins is not known and potentially conserved residues are indicated by the *number* and ? The pea albumins PA1a and PA1b differ from the dimeric albumins in that the two subunits do not remain associated by interchain disulphide bonds. (Shewry and Pandya 1999)

total contents of sulphur and nitrogen in the seeds remained constant and the increase in methionine was at the expense of free sulphate and, to a lesser extent, cysteine (reduced by about 12%).

Although the small size of the 2S albumins would be expected to facilitate 3 D structure analysis, only one structure has so far been determined, for a 2S napin from oilseed rape by NMR spectroscopy (Rico et al. 1996). It shows five α-helices and a *C*-terminal loop in a right handed spiral (Fig. 4) with a global fold similar to other S-rich low M_r seed proteins (see below).

Although the major function of 2S albumins is undoubtedly storage, some components have been shown to exhibit biological activity. These include napins from kohlrabi (*Brassica napus* var *rapifera*), charlock (*Sinapis arverse*) and black mustard (*B. nigra*), all of which are inhibitors of serine proteinases (Svendsen et al. 1989, 1994; Genov et al. 1997) and may therefore play a role in defence. Similarly, napins from radish (*Raphanus sativus*) inhibit the growth of a range of plant pathogenic

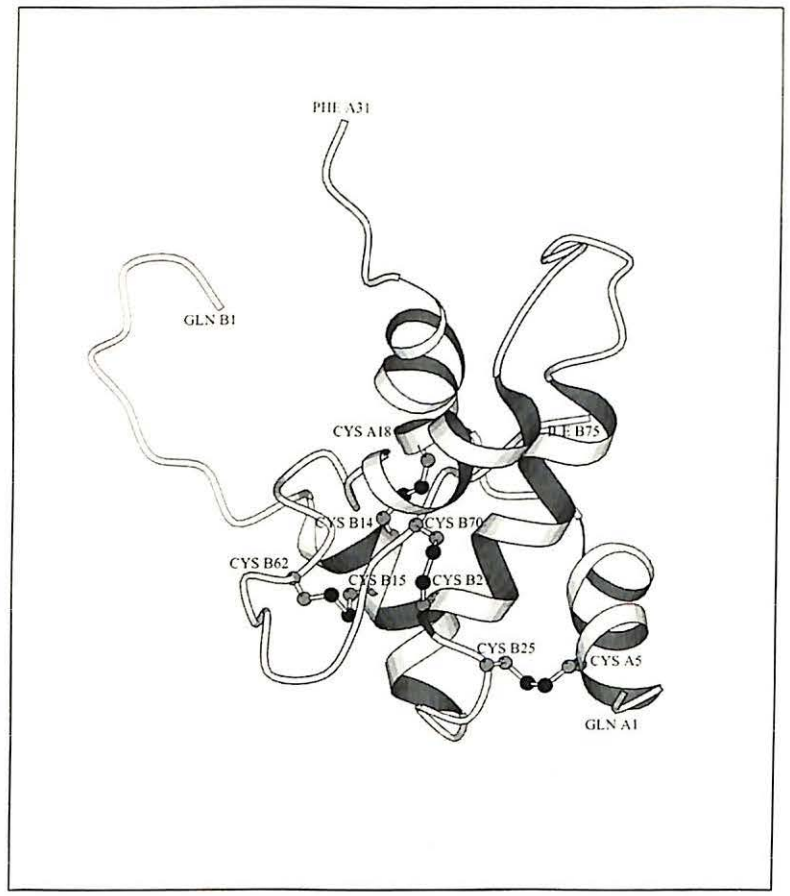

Fig. 4 : Ribbon diagram showing the 3-D structure of the *B. napus* napin Bn1b determined by NMR spectroscopy (Rico *et al.* 1996). The positions of cysteine residues and the *N*- and *C*-terminal residues of the polypeptide chains are indicated. (Rico et al. 1996)

fungi, particularly when in the presence of thionins (Terras et al. 1993). In this case, the mechanism probably involves membrane permeabilisation. A different type of biological activity, as allergens, has been referred to above in relation to the Brazil nut methionine-rich protein. This property is shared by albumins from other species such as castor bean, yellow mustard and oriental mustard.

Prolamins

The prolamin storage proteins differ from albumins and globulins in being restricted to the seeds of only one family of plants, the Gramineae (grasses,) which include the cereals. Because cereals form a major source of proteins for animal nutrition and food processing, their protein components have been studied in some detail. This has resulted in the availability of amino acid sequences and, in some cases, also structural data, for "typical" prolamins from all the major cereals, allowing their structural and evolutionary relationships to be determined.

The range of variation in the structures and properties of prolamins is vast, both within and between species. It is therefore not possible to provide a full account within the size limits of the present chapter. We will therefore focus on only two species, wheat and maize, referring the reader to recent review articles for more detailed accounts of these and other species (see, for example, Shewry (1995) and Shewry and Casey (1999)).

The prolamins of wheat correspond broadly to the gluten proteins (see p 317) and are classically divided into two groups on the basis of their solubility (gliadins) or insolubility (glutenins) in alcohol/water mixtures. The gliadins comprise monomeric proteins which interact in gluten by non-covalent forces (principally hydrogen bonds) and are further divided on the basis of their electrophoretic mobility at low pH into α-gliadins (fast), β-gliadins, γ-gliadins and ω-gliadins (slow) (Fig. 5).

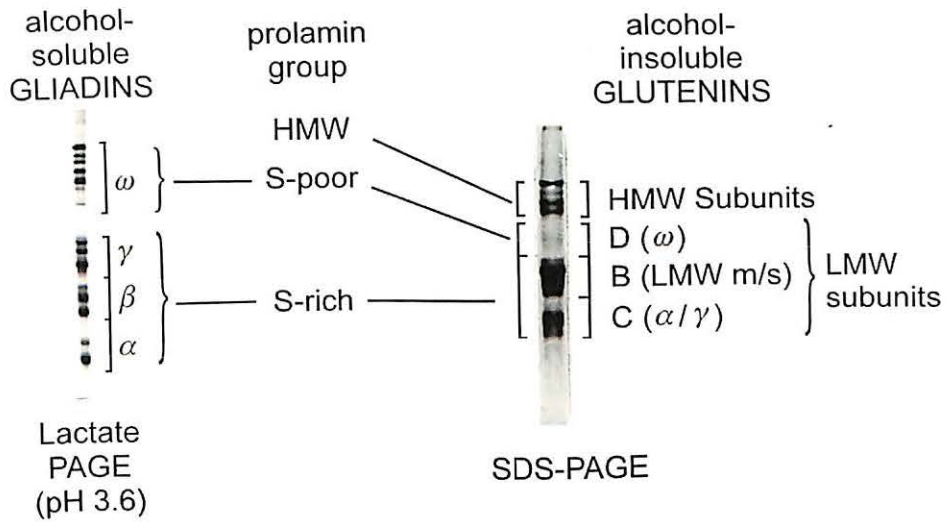

Fig. 5 : The classification and nomenclature of wheat gluten proteins separated by SDS-PAGE and electrophoresis at low pH. The D group of LMW subunits are only minor components and are not clearly resolved in the separation shown. (Shewry et al.. 1999)

In contrast, the glutenins consist of high M_r polymers stabilized by interchain disulphide bonds. Once these bonds are reduced the component subunits are soluble in alcohol/water mixtures and it is therefore usual to define both gliadins and glutenins as prolamins. The reduced glutenin subunits can be separated by SDS-PAGE into two major groups called high molecular weight (HMW) and LMW, the latter being further divided into B, C and D groups (Fig. 5).

Although the gliadin/glutenin classification is still routinely used, comparison of amino acid sequences shows that it is possible to divide the whole range of wheat prolamins into three broad groups and into subgroups within these (Fig. 5). Similar groups are also present in other members of the tribe Triticeae (barley, rye), confirming the validity of this classification (see Shewry 1995; Shewry et al. 1999).

The HMW prolamins comprise the HMW subunits of wheat glutenin. Either three, four or five individual HMW subunit proteins are present in cultivars of hexaploid bread wheat, each accounting for about 2% of the total grain protein (Halford et al. 1992). The availability of the complete amino acid sequences of a number of subunits, derived from genomic DNA sequences, shows that they have an organisation similar to that shown in Fig. 6. They comprise between 627 and 827 res-

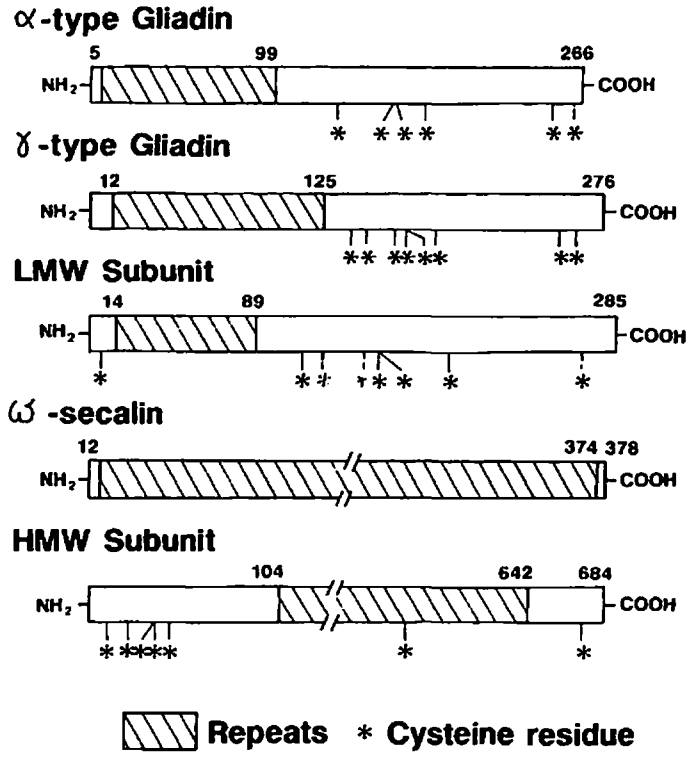

Fig. 6 : Schematic structures of typical wheat gluten proteins, based on published amino acid sequences. The repeated sequences are based on the following consensus sequences: α-type gliadin, PQPQPFP + PQQPY; γ-type gliadin, PQQPFPQ; LMW subunit, PQQPPFS + QQQQPCL; ω-secalin, PQQPFPQQ; HMW subunit, GYYPTSLQQ + PGQGQQ. Full references are given in Shewry and Tatham (1990) and Shewry et al. (1999). (Shewry et al. 1994a)

idues with M_r ranging from about 67 500 to 88 100. They also have a clear domain structure with an extensive repetitive domain flanked by shorter non-repetitive domains at the *N*- and *C*-termini (of 81-104 and 42 residues, respectively). The repeats are based on two (PGQGQQ + GYYPTSP or LQQ) or three (also GQQ) motifs which appear to form a loose spiral supersecondary structure resulting in an extended conformation for the whole molecule. Cysteine residues are largely restricted to *N*-terminal (three or five residues) and *C*-terminal domains (one residue), providing cross-linking sites for polymer formation.

The S-poor prolamins of wheat comprise the ω-gliadins and the D group of LMW subunits, which together account for about 10-20% of the total prolamin fraction. The ω-gliadins contain no cysteine residues but high contents of glutamine (40-50 mol%), proline (20-30 mol%) and phenylalanine (8-9 mol%) (Kasarda et al. 1983). No complete amino acid sequences of ω-gliadins have so far been reported but studies of the homologous ω-secalins of rye (Fig. 6) and C hordeins of barley show that they consist almost entirely of repeated sequences based on an octapeptide motif (consensus PQQPFPQQ). Although there is no homology between this motif and the repeated sequences present in the HMW subunits, the repeated sequences in the S-poor prolamins appear to form a similar loose spiral structure. The D group of LMW subunits are only minor components and appear to be derived from the ω-gliadins by the addition of a single cysteine residue, allowing polymer formation.

The final group of prolamins, called S-rich, forms about 60-80% of the total fraction and comprises the α-type gliadins (α/β-gliadins), the γ-type gliadins and the B and C groups of LMW subunits. These all have similar sequences, with repetitive *N*-terminal and non-repetitive *C*-terminal domains (Fig. 6). The *C*-terminal domains of the α-type and γ-type gliadins contain six and eight cysteine residues, which form three or four intra-chain bonds, respectively. The C-type LMW subunits appear to correspond to α-type and γ-type gliadins with the addition of unpaired cysteines, which allow the formation of polymers. In contrast, the B-type LMW subunits form a discrete group with three intra-chain disulphide bonds and one or two additional unpaired cysteines. The repetitive domains of the S-rich prolamins are rich in proline and glutamine and are based on either one or two short consensus motifs.

Despite the wide variation in sequence and structure within the prolamins of wheat and other members of the Triticeae, two lines of evidence indicate that they have evolved from a common ancestral gene (Kreis and Shewry 1989; Shewry and Tatham 1990). Firstly, it is possible to recognise three short conserved sequences, each comprising about 30 residues and labelled A, B and C in Fig. 7, within the *C*-terminal domains of S-rich prolamins. Related regions are also present in the HMW prolamins but in this case they are separated in the *N*-terminal (region A) and *C*-terminal (regions B, C) domains. These regions also show homology with each other, suggesting that they arose from the triplication of a short ancestral sequence. Insertion of further non-repetitive sequences between these regions and of repeated sequences may subsequently have given rise to the S-rich and HMW prolamins. The S-poor prolamins do not contain regions A, B and C but the repeated sequences, which comprise most of the proteins, are clearly related to those in the S-rich prolamins. It can therefore be proposed that they originated from the same ancestral protein as the S-rich prolamins by further amplification of the repeated sequences followed by loss of most of the non-repetitive sequences including regions A, B and C.

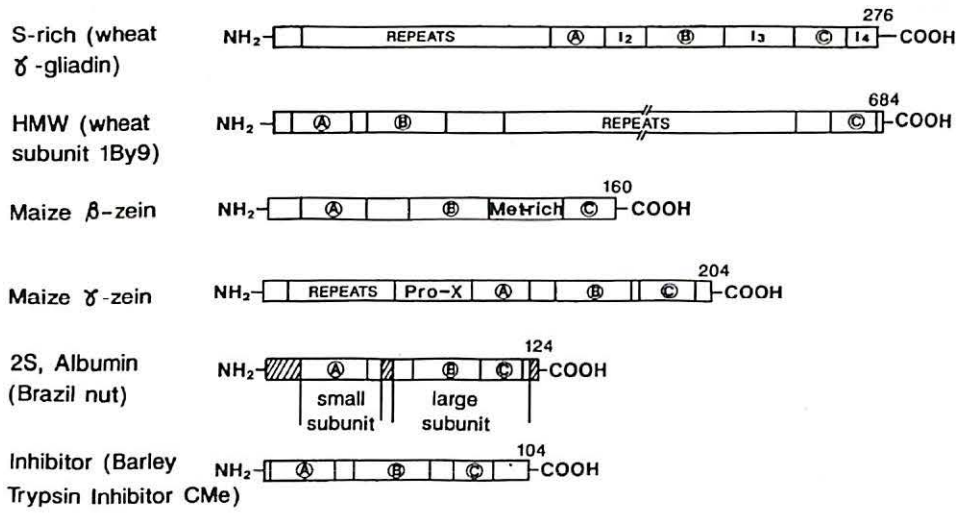

Fig. 7 : Schematic structures of various members of the prolamin superfamily. (After Shewry 1995)

Whereas the prolamins of wheat and other members of the Triticeae range in M_r from about 30 000-90 000, all of the maize prolamins (termed zeins) have M_r below 30 000. SDS-PAGE (Fig. 8) resolves two major bands of apparent M_r about 19 000 and 22 000 (α-zeins) with minor bands of apparent M_r about 27 000, 16 000 (γ-zeins), 14 000 (β-zeins) and 10 000 (δ-zeins). The Z19 and Z22 α-zeins actually comprise about 210-220 residues and 240-245 residues, with true molecular masses of about 23 000-24 000 and 26 500-27 000, respectively. They have similar structures (Fig. 7), with short *N*-terminal (36-37 residues) and *C*-terminal (10 residues) domains flanking either nine or ten blocks of degenerate repeats each comprising about 20 residues. These repeats are rich in non-polar amino acids (leucine, alanine) and each has been proposed to form an α-helix.

There is no evidence that the α-zeins are related to any of the prolamins of the Triticeae, or indeed to any other proteins except α-type prolamins in related panicoid cereals (sorghum, millets). In contrast, the other zein groups all appear to be related to the prolamins of the Triticeae (Fig. 7). The β-zeins and γ-zeins both contain regions corresponding to A, B and C in the Triticeae but have little other sequence homology with each other. In the γ-zeins either two (M_r 16 000) or eight (M_r 27 000) tandem repeats of the peptide PPPVHL are present, followed by a 22-residue Pro-X region in which almost every other residue is proline. In contrast, the β-zeins do not contain repeats but have a methionine-rich region close to the *C*-terminus. The δ-zeins are also methionine-rich but contain no sequence repeats or other distinctive features. Nevertheless, the identification of structural similarities to the methionine-rich 2S albumin of Brazil nut suggests that the δ-zeins may be related to other prolamins, as discussed below.

The α-zeins contain little or no methionine and only one or two cysteine residues, the latter resulting in their presence as monomers or oligomers. Higher levels of cysteine are present in the β-, γ- and δ-zeins, which is consistent with their pres-

ence in high M_r polymers. As in wheat, the reduced subunits are soluble in alcohol/water mixtures although the polymers may be soluble. The reduced γ-zein subunits are also readily soluble in water, being unique among prolamins in this respect.

The Prolamin Superfamily

Sequences related to those of regions A, B and C can also be identified in a range of other seed proteins, including prolamins from rice and oats, the 2S albumins and the cereal proteinase/amylase inhibitors (Kreis et al. 1985; Kreis and Shewry 1989). In the heterodimeric 2S albumins, such as the methionine-rich albumins of Brazil nut (Fig. 8), the site for proteolysis is located between regions A and B, with region A being present in the small subunit and regions B and C in the large subunit.

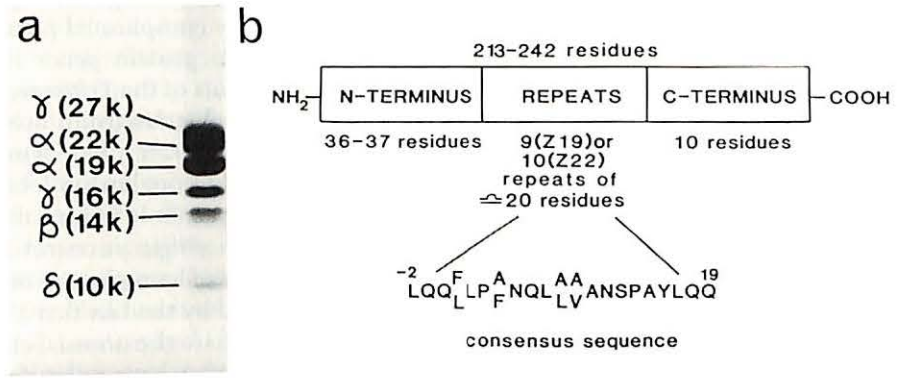

Fig. 8 : **a.** One-dimensional SDS-PAGE of total zeins of maize. **b.** Schematic structures of the M_r 19 000 (Z19) and M_r 22 000 (Z22) α-zeins of maize (After Shewry and Tatham 1990)

The cereal inhibitors have M_r ranging from about 12 000-16 000 and may be present as monomers or in dimeric or tetrameric complexes. The individual subunits may be inactive or exhibit activity against trypsin, exogenous α-amylases (notably from insect larvae), or both enzymes. They appear to be located in protein bodies within the mature cereal grain and are presumed to play a role in conferring broad-spectrum resistance to insect predators. They also have impacts on grain utilization, contributing to the development of respiratory allergy (baker's asthma) in workers in the milling and baking industry and affecting the stickness of pasta products made from durum wheat.

In addition to the 2S albumins and inhibitors, cereal grains contain several other groups of low M_r cysteine-rich proteins. These include the puroindolines, which are also rich in tryptophan and play a role in determining grain texture (*i.e.* hardness or softness) (Giroux and Morris 1998), and the non-specific lipid transfer proteins which are able to transport phospholipids *in vitro* (Breu et al. 1989). Both groups of proteins may also confer resistance to fungal pathogens *in vivo* (Dubreil et al. 1998; Terras et al. 1992). Both of these groups of proteins have highly conserved skeletons of cysteine residues, including Cys Cys and Cys Xaa Cys motifs, which are similar but not identical to those in the inhibitors (Gautier et al. 1994) and 2S albumins (Egorov

et al. 1996). Not surprisingly, all of these low-M_r S-rich proteins also appear to have similar 3-D structures, resembling the soybean hydrophobic protein in having four α-helices forming a right-handed superhelix (the α-type proteins) (Baud et al. 1993). Similar structures have been determined (by NMR or X-ray analysis) or predicted for 2S albumins from oilseed rape (Rico et al. 1996; Fig. 4) and sunflower (Shewry and Pandya 1999); a bifunctional α-amylase/trypsin inhibitor from ragi (Indian finger millet) (Strobl et al. 1995), non-specific lipid binding proteins from various species including wheat and maize (Gincel et al. 1994; Gomar et al. 1996) and wheat puroindolines (reviewed in Shewry et al. 2000).

Structure and Expression of Storage Protein Genes

Storage proteins of all the major crop species are encoded by multigene families clustered at different loci in the genome. This presents a very complicated picture, but there is considerable evidence to suggest that storage protein genes have evolved from a small number of ancestral genes. The prolamins of the *Triticeae*, for example, are encoded by genes in complex loci on the homoeologous group 1 chromosomes (*Glu-1, Sec3, Hor3; Gli-1, Sec1, Hor1; Glu-3, Hor2, Hor5*), minor remote loci on the same chromosomes (*Gli-3,4,5; Sec4, Hor4*) and additional major loci on chromosomes 6A, B and D of wheat and 2R of rye (see Shewry et al. 1999 for a more detailed review). This is consistent with their origin from single ancestral loci present on the group 1 chromosomes, and the translocation of some loci to other chromosomes in wheat and rye. This hypothesis is supported by the fact that all of the prolamin genes contain regions of some homology and share the unusual characteristic of containing no introns. The diversity seen today may have arisen from the multiplication of a single ancestral gene, the insertion of sequences and the duplication of repetitive regions (Kreis et al. 1985).

The α-zein genes of maize are encoded by a multigene family that is at least as complex as those in the Triticeae, with between 70 and 100 members (Wilson and Larkins 1984) present at several loci on chromosomes 1, 4, 7 and 10 (Shen et al. 1994). The β-δ- and γ-zeins, in contrast, are located at single loci (Das and Messing 1987). The β- and δ-zein genes are single copy, while there are either one or two γ-zein genes.

Legume globulin gene families are also quite large. In pea, for example, there are at least 18 genes encoding vicilin subunits (Casey et al.1986), divided into three small, related gene families (Casey et al. 1988), two genes encoding convicilin (Domoney and Casey 1985) and more than 10 genes encoding legumins (Domoney and Casey 1985; Domoney et al. 1986). As with the prolamin genes of the *Triticeae*, legume globulin genes show similarities in sequence, structure and organization that suggest that they may have evolved from a common ancestor. Sequence similarities *within* the 7S (vicilin-type) and 11S (legumin-type) globulin gene families from different species are greater than *between* the 7S and 11S globulins from any single species, but it is clear that the 7S and 11S families are related (Gibbs et al. 1989).

The expression of seed protein genes is subject to tissue-specific and developmental regulation. Prolamin genes of the *Triticeae* and maize are expressed exclusively in the starchy endosperm during mid- and late development, although there are some differences in spatial distribution of the different protein classes within the endosperm and in the exact timing of expression during development. They are also subject to nutritional regulation, responding to the availability of nitrogen and sulphur in the grain (Giese and Hopp 1984; Duffus and Cochrane 1992). Control of gene expression is exerted primarily at the transcriptional level (Bartels and Thompson 1986; Sörensen et al. 1989). Legume globulin genes are expressed in the parenchyma cells of the cotyledons and in the endosperm, where this is present.

Genes that share a common ancestry and show similar patterns of expression would be expected to have regulatory sequences in common, and several regulatory elements have been identified and characterized in seed storage protein gene promoters. The first to be identified was a sequence of approximately 30 bp positioned around 300 bp upstream of the transcription start site of several gliadin and hordein genes of wheat and barley (Forde et al. 1985). It was first termed the -300 element, subsequently the prolamin box or endosperm element, and one or more copies of it are present in the promoters of all the S-poor and S-rich storage protein genes of wheat, barley and rye characterized so far. The consensus sequence for the element is 5'- TGACATGTAA AGTGAATAAG ATGAGTCATG but it contains two, separate conserved motifs, TGTAAAGT and G(A/G)TGAGTCAT, with a more variable region in between. The former has been called the endosperm motif (Hammond-Kosack et al. 1993), the latter the GCN4-like motif (GLM), nitrogen element or N motif (Hammond-Kosack et al. 1993; Müller and Knudsen 1993). The N motif is similar to the binding site of the yeast transcription factor, GCN4, which is involved in nitrogen signalling. It is inverted with respect to the E motif in S-poor prolamin genes (Shewry et al. 1999).

Detailed functional analyses have been performed on versions of this element from C hordein and LMW subunit genes. Müller and Knudsen (1993) used particle bombardment of cultured barley endosperms with C hordein promoter/β-glucuronidase (GUS) constructs to confirm a regulatory role for the prolamin box and to show that the E and N motifs were separate elements. The N motif acted as a negative element at low nitrogen levels and interacted with the E motif and other upstream elements to give high expression when nitrogen levels were adequate. Hammond-Kosack et al.(1993) used *in vivo* footprinting and gel retardation assays to show that E motifs within the prolamin box and further upstream in the promoter of the wheat LMW subunit gene, *LMWG-1D1*, bound a putative transcription factor, ESBF-1, during early grain development, whereas no binding could be detected in other tissues. A second putative transcription factor, ESBF-II, bound the N motif prior to maximum expression of the gene. This study was followed by a functional analysis of the prolamin box of this gene in transgenic tobacco, which showed that both motifs were required for seed-specific expression, and the cloning of SPA, a bZIP transcriptional activator which recognized the N motif (Albani et al. 1997).

The E motif is also present in the promoters of zein genes and has been shown to act as a tissue-specific enhancer (Quayle and Faix 1992). It is not accompanied by an adjacent N motif, although motifs similar to the N motif are present elsewhere in the promoters of some zein genes (de Freitas et al. 1994). A transcription factor, encoded by the *Opaque2* gene, has been shown to bind a sequence close to the E motif and to regulate expression of α-zein genes (Schmidt et al. 1987, 1992; Lohmer et al. 1991). Another DNA-binding protein, OHP-1, has been shown to interact with the Opaque2 protein, but its exact role is not understood.

Perhaps surprisingly, the prolamin box is not present in the HMW prolamin gene promoters, despite the fact that HMW prolamin genes are subject to tissue-specific and developmental regulation similar to S-poor and S-rich prolamin genes. This may be because they are expressed at higher levels and require a more powerful enhancer element. The major regulatory element of the HMW prolamin promoters is a 38-bp sequence, GTTTTGCAAA GCTCCAATTG CTCCTTGCTT ATCCAGCT, first identified by Thomas and Flavell (1990) in the *Glu-1D-2* gene from the cultivar Chinese Spring at position -186 to -148. All of the HMW prolamin genes characterized so far contain this sequence and its position is tightly conserved (Shewry et al. 1999).

As with cereal genes, a regulatory element in legume storage protein genes was first identified by sequence comparisons. This revealed the presence of a conserved element of 28 bp, known as the legumin box, in the promoters of 11S globulin genes (Bäumlein et al. 1986). Functionality of this sequence has been tested in transgenic plants and the central core motif of CATGCAT shown to be required for high levels of gene expression (Bäumlein et al. 1992; Lelievre et al. 1992). A related sequence is present in the promoters of 7S globulins and has been termed the vicilin box. The legumin box is present in a β-phaseolin gene promoter within an upstream activator sequence (UAS1) that was identified by deletion analysis of the promoter using reporter gene expression in transgenic tobacco (Bustos et al. 1989). UAS1 was found to drive seed-specific expression at 80% of the level of the longest promoter sequence (795 bp) used in the study (Bustos et al. 1991). However, several other regulatory elements that do not resemble the legumin box were identified, including two other activating sequences, UAS2 and UAS3, and two elements, NRS1 and 2, that downregulated expression. Fine control of reporter gene expression in the transgenic plants required all of these elements (reviewed by Hall et al. 1999).

Storage Protein Synthesis and Deposition

Seed storage proteins are products of the secretory pathway which takes place within the endomembrane system of the cell. The proteins are initially synthesised on ribosomes attached to the rough endoplasmic reticulum (ER) and are directed into the lumen of the ER by an *N*-terminal signal peptide which is removed by proteolysis. The subsequent folding, assembly, processing and deposition of the proteins occur in one or more compartments of the endomembrane system: the ER lumen, the Golgi apparatus and the vacuole.

Protein folding and disulphide bond formation are thought to occur within the ER lumen. These processes may be assisted by proteins resident in the ER lumen, molecular chaperones such as a binding protein (BiP) and enzymes such as protein disulphide isomerase (PDI) and prolyl peptidyl *cis trans* isomerase (PPI). It is difficult to prove that such "helper" proteins are essential for storage protein folding and assembly, although the levels of BiP do increase in some seeds during the period of storage protein accumulation (Xia and Kermode 1999). In addition, the levels of BiP are increased in maize mutants in which storage protein synthesis is impaired and protein body morphology affected (Boston et al.1991; Fontes et al. 1991).

The assembly of proteins into oligomers and polymers stabilized by interchain disulphide bonds or non-covalent interactions is also considered to occur, or at least

to be initiated, within the ER lumen. In the case of cereal prolamins this may include the formation of extensive disulphide-stabilized polymers such as the glutenins, which are important in determining the functional properties of wheat. Assembly of 7S globulins into their mature trimeric form and of 11S globulins into an intermediate trimeric form of about 9S also occurs in the ER lumen.

In cereal prolamins no further modifications of the proteins occur but protein body formation occurs by two different routes. In maize and rice the prolamins accumulate within the lumen of the ER, giving rise to ER-derived protein bodies. In wheat and barley the situation is less clear but it is probable that two populations of protein bodies occur. Some prolamins, particularly the polymeric glutenins of wheat and polymeric hordeins of barley, appear to accumulate within the ER lumen as in rice and maize, but others (notably monomeric hordeins and gliadins) are transported via the Golgi apparatus to the vacuole, forming a second population of protein bodies (Levanony et al. 1992; Rubin et al. 1992). It is probable that the relative amounts of protein trafficking through these two pathways varies depending on the environmental conditions and developmental stage, while the two populations of protein body also fuse during the later stages of development to give a continuous proteinaceous matrix in the cells of the mature barley or wheat grain (Parker 1980).

The trafficking of proteins and their retention within specific compartments of the secretory pathway is usually determined by the presence of specific signals on the proteins, which may be cleavable or non-cleavable peptides at the *N*- or *C*-termini or surface patches on the folded protein, and their recognition by specific receptors. Proteins which do not contain such signals are thought to pass through the ER and Golgi to the cell membrane where they are secreted as a default destination. Retention within the ER is generally determined by the presence of *C*-terminal tetrapeptides (either KDEL or HDEL), which are present on the lumenal "helper proteins" discussed above, while vacuolar targeting signals vary considerably in their location and sequence.

No retention signals have yet been identified on prolamins and the retention of some components within the ER may be determined by their solubility properties which result in the formation of insoluble aggregates which are not readily transported through the endomembrane system. Similarly, no targeting signals have yet been identified on the prolamins destined for vacuolar storage.

The 2S albumins, 7S globulins and 11S globulins are all transported via the Golgi apparatus to the vacuole where proteolytic processing occurs to produce the mature proteins. Cleavage of the 11S globulin subunits into α/β chains also appears to be a prerequisite for assembly of the intermediate trimers formed in the ER into the mature hexamers in soybean but not in pea (Dickinson et al. 1989; Kermode and Bewley 1999).

The extent of glycosylation of 7S globulins varies between species and individual subunits whereas 11S globulin subunits are glycosylated only rarely (*e.g.* in lupin) (Casey 1999). Glycosylation of asparagine residues (*N* glycosylation) may occur during translocation across the ER membrane with further modifications occurring in the ER lumen and in the Golgi. In contrast, the glycosylation of serine and threonine residues (*O* glycosylation) occurs only in the Golgi. As with prolamins, the mechanisms that determine the targeting of 2S albumins and globulins to the vacuole are still unclear, but there is some evidence that specific sequences which are not subsequently cleaved are recognised. The vacuoles containing deposits of 2S albu-

min and globulin storage proteins then appear to divide to give discrete protein bodies.

Manipulation of Seed Storage Proteins

Classical and Mutation Breeding

Because the storage proteins are responsible for many aspects of the functional properties of seeds, their amount and composition have doubtless been manipulated by many generations of plant breeders whose aim has included the improvement of end-use properties. However, only rarely, and in recent years, have plant breeders been able to select for individual proteins with known quality attributes. The best-known example of this is the HMW subunits of wheat glutenin and breadmaking quality. The identification of specific allelic subunits associated with good or poor breadmaking performance (see Payne 1987, and below) and the ease with which these proteins can be followed in plant breeding programmes by SDS-PAGE of single grain extracts allowed breeders in the 1980s and a 90s to routinely select for quality associated subunits in their breeding programmes. There is no doubt that this has made an important contribution to quality improvement of European and other wheats. However, in most cases, it has not been possible to associate major quality traits with specific protein components.

Plant breeding has also been used to manipulate the total protein content of the seeds of some crops, exploiting variation which occurs between genetically determined limits for each species. Thus soybean varieties with between about 35 and 50% protein have been developed although there appear to be negative correlations between protein content and yield and between protein content and oil content (Salunkhe et al. 1992).

The ability to manipulate seed protein composition depends on the availability of genetic variation in the characters of interest, either within the crop itself or in related species with which it can be crossed. In some cases, such variation exists either within normal lines or spontaneous mutants. However, there is not sufficient variation available in many characters and further variation must be generated by mutagenesis or transformation.

Mutagenesis, using either chemical or physical mutagens, was extensively used in the 1960s and a 70s to generate variation in a range of characteristics. In barley, a major target was to generate high lysine mutants with improved nutritional quality, similar to the spontaneous mutants identified in maize (Misra et al. 1975; Bright and Shewry 1983). However, in both cereals there has been limited success in separating the high lysine character from deleterious associated effects to produce high-yielding lines with acceptability to farmers and end-users.

The high lysine mutants of maize were initially identified by visual inspection, due to associated changes in grain texture. About 12 mutant lines with lesions in different genes have been identified, with lysine contents ranging from about 110 to over 200% of those in the control non-mutant lines (Bright and Shewry 1983). The most widely studied, and most successful in terms of plant breeding, is the *opaque-2* mutation, which was also the first to be identified in 1964 (Mertz et al. 1964). The *Opaque-2* gene encodes a transcriptional activator which regulates the expression of

the M_r 22 000 α-zein genes. Thus in *opaque-2* mutants the amount of M_r 22 000 α-zeins is drastically reduced but there are also decreases in other zein classes, probably due to pleiotropic "knock on" effects (Coleman and Larkins 1999). Although *opaque-2* maize contains higher levels of lysine and tryptophan the kernal is also softer making it more liable to damage and infection. In addition, the yield may be lower than that of normal maize lines. Nevertheless, it has proved possible to produce hard-textured *opaque-2* lines suitable for commercial cultivation by using genetic modifiers. The resulting material, called quality protein maize (QPM), has only slightly lower quality than normal *opaque-2* lines. The mechanism of endosperm texture modification is still not completely understood but the lines contain elevated levels (two- to threefold) of γ-zein which result from increased steady state levels of transcripts (Coleman and Larkins 1999).

Several other high lysine mutants of maize have been studied in detail including *floury 2* and *opaque-15* which specifically affect M_r 19 000 α-zeins and M_r 27 000 γ-zeins, respectively. However, none of these mutants has been successfully used to produce commercial varieties.

In barley a number of induced high lysine mutants and one spontaneous mutant (Hiproly) are available, with increases in lysine of up to 40% (Bright and Shewry 1983). Despite intensive breeding efforts with Hiproly and Risø mutant 1508 (see Munck 1992), no commercially successful lines have been produced.

Genetic Transformation

Transformation provides an attractive opportunity to manipulate the protein composition of seeds as it allows single defined sequences to be inserted. It is, therefore, possible to add additional copies of endogenous genes and to regulate their levels and their temporal and spatial patterns of expression using specific promoters. It is also possible to use antisense or other technologies to downregulate the levels of expression of endogenous genes and to add completely new genes from other plants, microbes or animals. It is not surprising, therefore, that seed protein composition has been a major target for genetic engineering experiments.

Two strategies have been used in order to increase the levels of nutritionally limiting amino acids in seeds. The first is to increase the levels of free lysine based on manipulation of its biosynthetic pathway. Two key regulatory enzymes in the lysine biosynthetic pathway (aspartate kinase and dihydrodipicolinate synthase) are normally feedback-regulated by lysine, resulting in a low amount of free amino acid. Transformation with feedback-insensitive forms of these enzymes from bacteria resulted in two- and fivefold increases in total lysine in seeds of canola (oilseed rape) and soybean, respectively, (Falco et al. 1995) but no impact on total lysine was reported in transgenic barley (Brinch-Pedersen at al.1996).

The second approach is to transform plants with additional genes encoding proteins rich in essential amino acids. In legumes the focus has been on sulphur amino acid content, using genes for methionine-rich 2S albumins. The 2S albumin from Brazil nut contains about 26 mol% methionine and has been used to increase total seed methionine by up to 30% in tobacco and oilseed rape (Altenbach et al. 1989, 1992) and up to three-fold in narbon bean (*Vicia narbonensis*) (Saalbach et al. 1995). However, this protein is now known to be allergenic to humans limiting commercial development. The 2S albumin of sunflower contains 16 methionines

and 8 cysteines in a protein of 103 residues and does not appear to be allergenic. Expression in lupin resulted in a 94% increase in total seed methionine although there was a small decrease in cysteine and no increase in total seed sulphur (Molvig et al. 1997).

Although lysine-rich proteins have been characterized from plants (for example, the barley chymotrypsin inhibitors CI-1 and CI-2 which contain about 9.5 and 11.5 g% lysine, respectively) they have not so far been used for genetic engineering. Instead, Keeler et al. (1997) adopted a different approach, by designing and expressing a completely new protein containing 31% lysine and 20% methionine. This resulted in a significant increase in total lysine in seeds of transgenic tobacco but no significant effect on the methionine content.

Nutritional quality is readily defined in terms of the contents of nutritionally essential amino acids and is therefore also relatively easy to manipulate. In contrast, the functional properties that determine processing quality in food systems are more difficult to define in molecular terms and may involve multicomponent interactions (*i.e.* protein:protein, protein: starch, protein:lipid) which may change during the processing itself. Nevertheless, progress has been made in using genetic engineering to improve some aspects of seed protein functionality.

In wheat an obvious target for manipulation is the high molecular weight subunits of glutenin, due to the association between allelic variation in their number and composition and breadmaking quality (see above). Four laboratories have reported the transformation of bread wheat with additional genes for HMW subunits, in order to increase the number of expressed subunits and total amount of HMW subunit protein (Altpeter et al. 1996; Blechl and Anderson 1996; Barro et al. 1997; Alvarez et al., 2000), resulting in one case in increased dough strength (Barro et al. 1997). In one line the total amount of HMW subunit protein was increased from 12.7 to 20.5% of the total extractable protein, resulting in a highly elastic dough which appeared to be too strong to be mixed under normal conditions (Rooke et al. 1999).

Soybean proteins are extensively used in the food industry to confer functional properties, including gelation to form tofu (a traditional food in the Far East) and emulsification. Utsumi and colleagues have, therefore, focused on understanding and manipulating these properties using protein engineering of soybean 11S globulin (glycinin) subunits. They showed that several mutations resulted in increased gel hardness and improved emulsification properties, including deletion of short sequences at the *N*-terminus and the insertion of short methionine-rich sequences into variable domains (Utsumi 1992). The latter are of particular interest as they may allow the simultaneous improvement of functional properties and nutritional quality. Expression of one of the methionine-enhanced glycinins in transgenic tobacco resulted in accumulation up to about 4% of the total proteins but about half of the protein was partially degraded and not correctly assembled. Current work includes transformation of rice with the same constructs in order to improve its nutritional quality and confer novel functional properties.

These studies with wheat and soybean demonstrate therefore, that it should be possible to use genetic transformation to improve the functional properties of commercial crops, based on a detailed understanding of the molecular basis for protein functionality (see following section).

Biophysical Properties and Impact on Utilization

Seed storage proteins are used either as raw materials or ingredients for their functional properties. The latter are physicochemical properties that contribute to a specific function in the product, for example determining the desired processing and product characteristics (sensory, physical, textural, etc.). Functional properties are usually classified into three main groups:

- hydration properties depend on protein-water interactions and include water absorption, solubility, viscosity and adhesion,
- properties based on protein-protein interactions which include precipitation, gelatinisation, dough formation and shaping, viscoelastic properties,
- surface properties involve superficial strain and include emulsifying and foaming characteristics.

The objective of food scientists is to find a mechanistic explanation of functional behaviour. Obviously, elucidating molecular mechanisms should help us to optimize industrial processes and improve end-use qualities as well as to develop methods to screen for high quality genotypes in plant breeding programs or to test grain at harvest. However, our current state of knowledge and the complexity of the various food systems do not yet allow us to clearly understand how a given structure will determine, for example, the texture of a food product. The main difficulty lies in the fact that native structures are often modified during processing (*e.g.* mixing, extrusion, fermentation, heating, drying, cooking,) of raw materials or protein ingredients into the final complex food (Cheftel et al. 1985).

General Aspects of the Molecular Basis of Functionality

In order to understand the relationship between the structure of a protein and its functional properties, researchers have endeavoured to relate the latter to the protein structure as measured by a range of physicochemical parameters. For example, the size as given by the molecular weight, the charge and polarity as given by the electrophoretic mobility, isoelectric point and average hydrophobicity, and, finally, the intramolecular forces as given by the solubilization or dissociation behaviour in various solvents and stability to denaturation. At an intuitive level it is possible to select those physicochemical properties which are likely to contribute to specific functional properties. For example, a functional property such as solubility might be influenced by the molecular weight, hydrophobicity and net surface charge of the protein. Because, at a fundamental level, these properties can all be regarded as functions of the amino acid composition, primary sequence and structure of the protein, it follows that similarities in structural characteristics between proteins can be assumed to imply a corresponding similarity in both their physicochemical and functional properties (Wright 1983). To what extent this assumption is valid for seed storage proteins will be discussed below using cereals, mainly wheat, and legumes, mainly soybean, as examples.

Functionality of Cereal Proteins

Gluten proteins contribute unique technological characteristics to doughs, making wheat the most widely used cereal. Thus, in a developed wheat flour dough, the protein forms a continuous network which gives rise to the viscoelastic properties

essential for sheeting and gas retention. However, the molecular explanation for this uniqueness depends on differences which must be quite subtle, since cereals which are related genetically to wheat and show superficial similarities in protein composition do not form dough with similar viscoelastic properties. To understand the viscoelasticity of wheat gluten, a property which is inherent to polymers and associated with flexible thread-like molecules, it is necessary to know the structures of the individual proteins and how they interact in the developing and mature seed, as well as in food systems. Gaining this knowledge has required analyses at several levels. The amino acid sequences of individual monomers and subunits were originally determined by direct analysis of isolated proteins, but are now usually deduced from the nucleotide sequences of cloned cDNAs or genes encoding the proteins. This approach has been used to determine the sequences of a number of zeins, gliadins, glutenins and hordeins. Knowledge of the locations of intrachain disulphide bonds in monomers and interchain bonds in polymers determined by sequence analysis of peptides isolated from unreduced protein preparations has also proved to be extremely valuable (Müller et al. 1998).

The conformations and potential interactions of the proteins are important aspects to study. However, in contrast to legume proteins in which water or salt solvents extract almost all of the storage proteins, conformational analyses of cereal proteins are limited by the solubility properties of prolamins and studies of their structures using a range of spectroscopic or hydrodynamic procedures can be only performed in a limited range of solvents. Polymeric prolamins (*e.g.* the HMW subunits of glutenin or whole gluten) cannot be dissolved without reduction of disulphide bonds, which makes it difficult to relate the results to the native state, or restricts the choice of analytical procedures to those that can be used on material in the solid state (scanning probe microscopy and NMR and FT-IR spectroscopy) (Tatham et al. 1990) or to computer modelling approaches (Kasarda et al. 1994).

Functionality in Legume Proteins

Proteins of legume seeds are often refined using dry (air-classification) or wet (alkaline extraction followed by acid precipitation) procedures with selective removal or destruction of undesirable components. The resulting concentrates or isolates are then processed to make meat substitutes using texturization or functional agents for the food industry. Consequently, in legume proteins, functional properties mainly relate to an ability to stabilize emulsions or foams and impart textural attributes. Emulsions (*e.g.* margarine, salad cream) and foams (*e.g.* whipped desserts, toppings) represent disperse systems in which one phase (air or oil) is dispersed throughout a continuous phase (water). Proteins provide the stabilizing force that prevents these systems reverting to two separate phases. Proteins migrate to the air:water or to the oil:water interface and, on unfolding, form an interface layer with consequent alteration to the surface properties. It is generally accepted that the balance of hydrophobic to hydrophilic regions (and therefore the amino acid composition and sequence) of a given protein has a significant effect on its ability to stabilize such dispersions (Wright and Bumstead 1984). On the other hand, the textural properties of legume proteins originate in their capacity to heat-set and form a stable matrix or gel, *i.e.* in cakes. Gelling involves the denaturation of proteins, basically the rupture of intramolecular bonds and unfolding of polypeptide chains, followed by the formation of intermolecular crosslinks between newly

exposed residues of the denatured protein. The ability of a protein to form a gel depends upon its size, the structure and the nature of internal bonding, the previous history of the protein and also on extrinsic parameters such as solvent characteristics (Wright and Bumstead 1984).

Gluten Proteins and Breadmaking, Biscuitmaking and Pastamaking

Genetics of Protein Monomers and Protein Quality

Wheat is processed in the food industry into a range of products, including bread, other baked goods (cakes, biscuits), noodles and pasta. The ability to make these products is determined, to a great extent, by the gluten proteins which confer cohesive and viscoelastic properties to doughs. The concept of protein quality was born several decades ago, when it was realised that different wheat varieties had different baking performance. In the early 1970s it was demonstrated that the electrophoretic pattern of gliadins was a fingerprint of the wheat variety. At about the same time, wheat breeders in several countries began to develop varieties that had extremely high yield potential, but many of these varieties had unacceptable baking quality. Electrophoretic analysis of gliadins was quickly adopted for detecting the presence of admixtures in official grades of wheats or the presence of undesirable varieties in deliveries to the flour mill (Wrigley 1995). While widespread use has been made of the polymorphism of the gliadin proteins in wheat variety identification, it is research on glutenin subunits that has contributed significantly to a better assessment of protein quality and prediction in breeding programmes.

Glutenin subunits consist mainly of two types, the high molecular weight glutenin subunits (HMW subunits) and the low molecular weight glutenin subunits (LMW subunits). The HMW subunits (which account for about 6-12% of the total protein fraction) are so designated because they form a slower-moving group of components when reduced glutenin subunits are separated by SDS-PAGE with apparent M_r from about 80 000-120 000 (Fig. 9). Most wheat varieties have four or five HMW subunits, which frequently differ in their electrophoretic mobility among varieties. These are encoded on the long arms of chromosomes 1A, 1B and 1D and can be divided into x-type and y-type (Payne et al. 1984).

In the 1980s, researchers at the Plant Breeding Institute, Cambridge, UK, showed that the presence of certain subunits (*e.g.* 1Ax1, 1Ax2*, 1Dx5+1Dy10, or 1Bx7+1By9) were correlated with high breadmaking quality (Payne et al. 1987; Fig. 10). Scientists in several countries took advantage of this relationship between the genetics and quality of wheat proteins to develop new varieties more adapted to the modern baking technologies that require higher dough strength (*e.g.* the Chorleywood Bread Process, fast-food breads, rolls, buns, frozen doughs). Not all baking technologies, however, could benefit from this relationship. For example, for wholemeal bread, protein quantity is generally more important than protein quality. On the other hand, in southwestern Europe, breadmaking technologies are quite different from those commonly used in North America or in northern Europe. For example, in France, typical breads are made of essentially four ingredients, flour, water, yeast, and salt, with little or no additives, and they are normally baked on the oven hearth rather than in a pan. In this case, doughs with very high strength and tenacity are detrimental to baking score or loaf volume and a high extensibility of dough is required (Autran et al. 1997). To better understand the physicochemical bases of

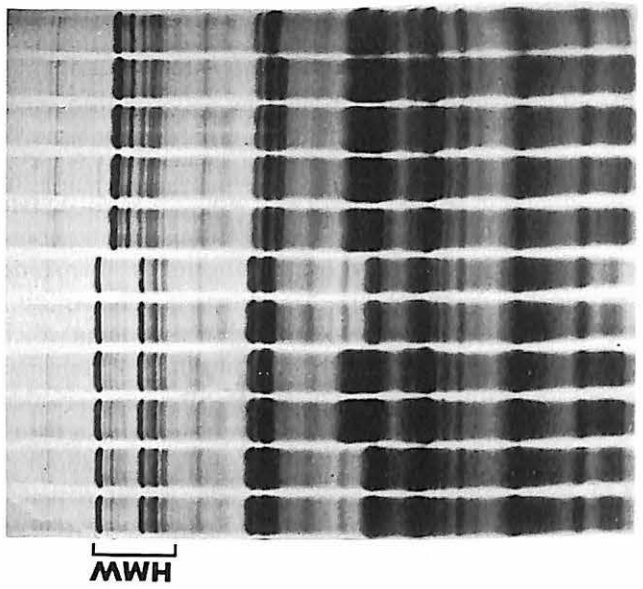

Fig. 9 : Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of wheat storage proteins, with indication of the high molecular weight (HMW) subunits of glutenin

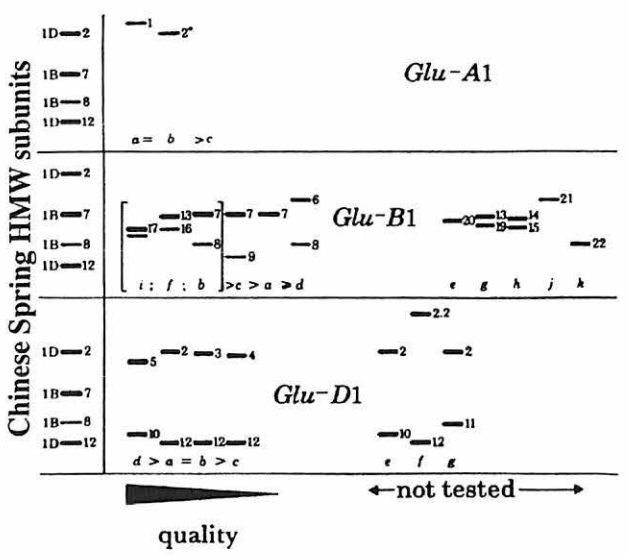

Fig. 10 : Allelic variation in the high molecular weight (HMW) subunits of glutenin and their relation to bread-making quality (Payne et al. 1984). On the *left hand side* are the standard HMW subunits from cultivar Chinese Spring. The subunits have been split into three groups according to whether their genes are controlled by chromosome 1A, 1B or 1D. The *letter beneath each allelic group* refers to the international HMW allele nomenclature. The subunits that associate most strongly with good quality have been placed on the *left hand side* of each group

dough extensibility and to facilitate breeding of new types of wheats with a satisfactory balance between dough strength and extensibility, it has been necessary to study protein fractions other than the HMW subunits of glutenin. Recent reports have emphasized the possible role of LMW subunits of glutenin, that are more abundant than the HMW subunits and are encoded by genes that are genetically linked to some of the gliadin genes on the short arms of chromosomes 1A and 1B. New protein markers in wheat breeding have been proposed. For example, combining the chromosome 1D-encoded HMW subunit (*Glu-D1*) alleles that impart a high dough tenacity (*e.g.* subunits 1Dx5+1Dy10) with the chromosome 1A-encoded (*Glu-A3*) LMW subunit alleles *o* and *n* that impart high extensibility should result in the development of new wheats adapted to the baking technologies of southwestern Europe. When aiming at breeding of biscuit-type wheats, it could be recommended to screen lines to select seeds containing both the *Glu-D1* allele 1Dx2+1Dy12 and the *Glu-B3* allele *III*. Moreover, because many of the food products made from soft milling wheats require little or no elasticity (*e.g.* to form doughs or batters) they can be made from a new type of wheat that has been produced by transferring null alleles at *Glu-D1* and *Glu-A1* loci into the soft wheat cultivar Galahad. One line, called Galahad-7 contains only one HMW subunit (subunit 1Bx7) and produces extremely extensible doughs (Payne and Seekings 1996).

A similar strategy had been used for improving pasta products before being used for breadmaking. Durum wheat (*Triticum durum* Desf.) is widely considered to be the best type of wheat for pasta making due to its excellent amber color and superior cooking quality. Differences in cooking quality (*i.e.* high firmness and good surface condition of cooked pasta) were attributed to the protein content and composition of the grain. A major breakthrough in our understanding of the biochemical and genetic basis of pasta quality was made by Damidaux et al. (1978), with the discovery of a clear-cut relationship between the electrophoretic pattern of γ-gliadins and gluten strength, an indicator of pasta firmness. The γ-45 gliadin was associated with strong gluten, whereas the allelic γ-42 gliadin was associated with weak gluten. In fact, the positive effect of the γ-45 gliadins (*Gli-B1*) locus arises from its genetic linkage with the *Glu-B1* locus encoding LMW subunits of glutenin (Pogna et al.1988) and probably results from differences in the amounts of expressed LMW subunits between the two allelic types.

However, new specifications for durum wheat proteins have recently been introduced as a result of the use of higher drying temperatures in the pasta industry. While the roles played by protein content and protein composition have almost the same importance when pasta is dried at low temperature (55 °C), when using 70-90 °C drying the protein content becomes most important. The question that is presently challenging researchers is, therefore, how to make it possible to increase of the protein content of the grain without simultaneous increases in the brownness and ash content of the semolina.

Functional Properties of Glutenin Polymers

To evaluate the functionality of wheat proteins and to manipulate them in breeding and during food processing, it is necessary to investigate the molecular basis of the above-mentioned correlations used by breeders. This shows that studies based only on electrophoresis are not adequate because functionality is primarily determined by the presence of large protein aggregates rather than protein monomers.

In fact, the gluten proteins include monomeric and polymeric prolamins. In the monomeric gliadins disulphide bonds are either intra-molecular (in α- β- and γ- gliadins) or are absent (ω-gliadins). Purified hydrated gliadins have little elasticity and cohesion and contribute mainly to the extensibility of the dough system. The aggregated glutenins, in contrast to the gliadins, have intermolecular disulphide bonds in addition to intramolecular ones (Kasarda 1989). Hydrated glutenin polymers, free of gliadins, are highly cohesive and elastic and contribute elasticity to the dough system. It seemed reasonable, therefore, to attribute this elasticity largely to the crosslinked nature of the glutenin subunits.

The aggregated prolamins comprise two groups of polypeptides that can be separated under reducing conditions: the HMW subunits (apparent M_r range 80 000-140 000) and LMW subunits (apparent M_r range 50 000-80 000). Although the HMW subunits account for about 6-12% of the total protein of wheat, they appear to be of particular importance in determining the viscoelastic properties of wheat gluten and in turn determine the functionality of wheat doughs in various food systems, specifically breadmaking.

Molecular and biophysical studies have revealed details of HMW subunit structure that may relate to their role in viscoelastic polymers. As reviewed by Shewry et al. (1998) and discussed above, analysis of genes encoding nine different HMW subunits (including allelic and homeoallelic forms derived from the A, B and D genomes) demonstrates that the proteins have similar structures. All consist of three domains, with short *N*-terminal and *C*-terminal domains (both containing one or several cysteine residues) flanking a central repetitive domain based on hexapeptide and nonapeptide motifs with tripeptide motifs present in x-type subunits only. Previous studies have suggested that these repeated sequences form an unusual supersecondary structure, a loose spiral based on β-reverse turns. This β-spiral is of special interest since it appears to be unique among proteins (although a similar structure has been demonstrated for a synthetic polypeptide based on a repeat motif present in elastin, an elastomeric protein of mammals). Such a structure may contribute to the mechanism of gluten viscoelasticity, via intrinsic elasticity and/or the formation of extensive hydrogen bonds with adjacent proteins, the latter being facilitated by the high content of glutamine residues (≅ 40 mol%). However, another important aspect of HMW subunit structure is the presence of cysteine residues near the ends of the polypeptide chain, a feature that can facilitate a linear extension of the glutenin polymer into huge molecules consisting of polypeptides attached to one another by disulphide bonds, with entanglement regions, a system that may give a rubber-like elasticity.

The other family of glutenin subunits, the LMW subunits, is one of the most abundant storage protein groups in wheat endosperm. However, they are much less well characterised than HMW subunits and the effects of individual components have not been studied (Sissons *et al.* 1998). Also, functional and structural studies of the LMW subunits have always been limited by the difficulty of preparing adequate amounts of single homogeneous polypeptides. This is because the LMW subunits are somewhat insoluble after reduction of intermolecular disulphide bonds, which is necessary for their purification, but which also causes the exposure of buried hydrophobic regions. The LMW subunits also derive from many more genes than the HMW subunits, and our knowledge is based on a limited number of complete DNA sequences, some of which may not encode major components.

The LMW subunits were initially classified into B and C types according to their mobility in SDS-PAGE. However, there is growing evidence of similarity of groupings of LMW polypeptides based on sequence studies. Lew et al. (1992) suggested that it would be more valid to classify LMW subunits into classes based on sequence rather than on their electrophoretic mobility. Six main sequence types were defined on the basis of *N*-terminal sequences. Three of these can be considered to form LMW subunits proper (*N*-terminal sequences: SHIPGLERPSGL-, METSHIPGL-, METS(R)CIPGL-) while the others closely resemble the α- γ-, and ω-type gliadins (Sissons et al. 1998). At first, the LMW subunits were assumed to be chain-terminators in the growth of the glutenin polymer as their cysteine residues were only present in the *C*-terminal part of the molecule (Kasarda 1989). More recently, LMW subunits were found with a cysteine residue near their *N* terminus, making such subunits also available as chain extenders. However, despite reported correlations between the allelic composition of LMW subunits and dough extensibility, there is still little evidence as to how individual LMW subunits affect dough quality.

In addition to the individual glutenin subunits, the structures of the polymeric glutenins that they form have also been studied. Pioneering studies relating the molecular weight distribution of such glutenins to breadmaking quality were based on solubility methods or on conventional chromatography and hence suffered from many disadvantages: they were tedious, lengthy and difficult to reproduce or to quantify. The advent of HPLC techniques that have capabilities of automation, reproducibility and quantitation has allowed studies of larger series of samples. In contrast with studies based on reversed-phase HPLC (RP-HPLC), which are generally aimed at fingerprinting varieties based on gliadins or reduced glutenin subunits, size-exclusion HPLC (SE-HPLC) has the potential to retain large aggregates in a quasi-native state, and to provide information on aggregate size. For example, SE-HPLC of unreduced phosphate-SDS extracts proved to be a powerful tool for studying the physico-chemical and structural basis of wheat quality, and one that is applicable to the rapid examination of size differences of glutenins from wheat flours and industrial glutens (Autran 1994; Fig. 11).

Recently, large-size protein aggregates have been investigated in a more dynamic way. For example, the amount and rheological characteristics of the SDS-insoluble gel protein fraction decreased during dough mixing, whilst the amount increased again during dough resting. In this way, differences in reactivity were shown for the various HMW glutenin subunits. In particular, subunits 1By9, 1Dy10 and 1Dy12 were incorporated into polymeric glutenin at a faster rate and to a higher level than the subunits 1Dx2, 1Dx5, 1Bx7, 1Ax1 and 1Ax2*, which can be of importance when blending flours of different subunit composition (Weegels et al. 1997). Also, the presence of subunits 1Dx5+1Dy10 appeared to be associated with higher polymer molecular weight distribution than the presence of subunits 1Dx2+1Dy12 (MacRitchie 1998). Popineau et al. (1993) used gluten extraction and fractionation by a sequential procedure that preserved functionality, as well as dynamic measurements in shear to investigate large glutenin polymers of various isogenic lines of the cultivar Sicco with different HMW subunit compositions, and found a strong correlation between the amount of large glutenin polymers and the viscoelasticity of gluten subfractions. They concluded that both the quantity of HMW subunits and subunit composition influenced gluten viscoelasticity by modifying the polymerisation state of gluten proteins. Also, glutens differing in their LMW and HMW subunit

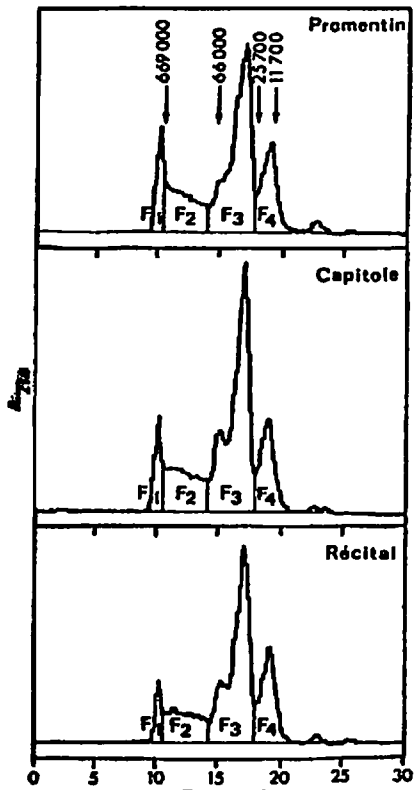

Fig. 11 : Elution profiles obtained by size-exclusion high performance liquid chromatography (SE-HPLC) of unreduced flour proteins extracted with phosphate-SDS from three bread wheat cultivars differing in baking strength (Alveograph W indices were 105, 110 and 200, respectively). Elution positions of molecular weight standards are indicated. The elution curve is divided into four regions: F_1 (high molecular size aggregates), F_2 (intermediate size aggregates), F_3 (gliadins), F_4 (albumins and globulins)

compositions were analysed to determine their size distribution and their rheology in the dynamic regime and by ESR spin-labelling, demonstrating a relationship between their aggregative properties, segmental flexibility and viscoelastic behaviour, with the proportion of rigid polypeptide segments related to the height of the elastic plateau (Hargreaves et al.1996).

Complementary to physicochemical techniques, the potential of immunochemical methods, especially those based on monoclonal antibodies, has been exploited to recognise protein conformation, to yield information on functionally important groups, and to quantify specific flour polypeptides. For example, Andrews and Skerritt (1994) reported positive correlations between antibody binding to chromosome 1D-encoded HMW glutenin subunits (*i.e.* 1Dx5+1Dy10 or 1Dx2+1Dy12) and dough strength.

With the advent of molecular genetics, it has become possible to produce whole wheat proteins and protein domains via heterologous expression in *E. coli* in suffi-

cient amounts for detailed spectroscopic analyses and to express wild-type and mutant subunits in transgenic wheat plants. These new developments, together with molecular modelling, should provide information on the relationship between the protein primary structure and functionality and on the molecular basis for gluten viscoelasticity and will aid the development of strategies to improve the functional properties of gluten and wheat for a range of end uses, as discussed by Ciaffi et al. (1998), Shewry (1998), Barro et al. (1997) and Rooke et al.(1999).

Effect of Environment on Functional Properties and End-Uses

The intrinsic processing quality of wheat cultivars is changed significantly by cultural practices (*e.g.* amount, type and application dates of nitrogen fertilisation; sulphur availability) and climate (grain filling duration, temperature and relative humidity during grain filling) via modification of flour protein content and composition. For example, a high amount of nitrogen fertiliser generally leads to a significant increase in the total protein content, but this increase affects mainly the gluten proteins with little effect on the albumins and globulins. Several reports have shown that a high amount of N fertiliser, especially when applied at late stages of plant growth, results in an increased gliadin to glutenin ratio (although the change in this ratio may not be consistent for certain genotypes) and an increased ratio of HMW to LMW subunits, but no change in the ratio of x-type: y-type HMW subunits (Pechanek et al. 1997). Whereas the proportions of the various groups of gluten protein may be affected by N fertiliser, the polypeptide composition within each group proved to be constant with respect to growing conditions, so that gliadin or glutenin patterns can be used for fingerprinting genotypes using techniques such as electrophoresis or RP-HPLC. The exception to this general rule relates to the increased synthesis of sulphur-poor prolamins (*i.e.* ω-gliadins) when sulphur is deficient (Wrigley et al. 1984; Fig. 12).

In recent years, interest in the effects of environmental factors has been stimulated by studies of heat stress during grain filling in wheat. High temperatures (>32 °C) for as few as 1 or 2 days during grain filling were shown to result in decreases in the protein quality and dough properties (Randall and Moss 1990), with a change in the protein composition generally resulting in a. decreased dough strength (Ciaffi et al. 1995). Blumenthal et al. (1998) recently reviewed the main hypotheses that have been advanced to account for the observed changes

1. A decrease in the ratio of glutenin: gliadin results from gliadin synthesis continuing during heat stress whereas glutenin synthesis is greatly decreased. This effect was explained by the presence of heat-stress elements (HSE) in the upstream regions of some gliadin genes, but not in the published sequences of glutenin genes.
2. A decrease of the size of glutenin polymers in the mature grain under heat stress, resulting in weakening of the resulting dough.
3. The synthesis of a M_r 70 000 heat-shock protein (HSP 70) as a reaction to heat stress, resulting (if still present in the mature grain) in a loss of dough quality.

In fact, the most recent investigations of Blumenthal et al. (1998) showed that the amount of HSP 70 in mature grain was not correlated with most indicators of dough strength, while incorporation of purified HSP 70 into dough showed no dramatic effect on dough properties. Also, sequencing the upstream regions of HMW subunit genes failed to show the presence of heat-shock promoters even in widely different genotypes. Consequently, research is presently focused on the degree of

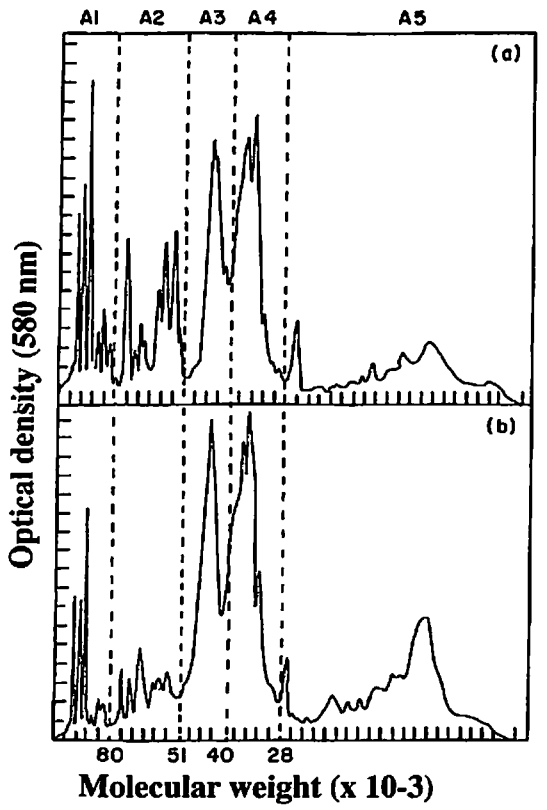

Fig. 12: Densitometer scans of SDS gel separations of polypeptides extracted with SDS + 2-mercaptoethanol from sulphur deficient flour (a 0.083% S, 1.84% N) and from normal flour (b 0.161 S, 1.96% N). The peaks in particular regions of the densitometer scans are divided into five groups: *A1* (HMW subunits of glutenin); *A2* (mainly ω-gliadins); *A3* (mainly LMW subunits of glutenin); *A4* (mainly α, β- and γ-gliadins); *A5* (albumins and globulins). (Wrigley et al. 1984)

polymerization of the glutenin chains and on the roles of HSP and chaperones in the developing grain (Blumenthal et al. 1998). This is because it is considered likely that HSP modifies the folding and aggregation of gluten proteins *in situ* during grain filling, especially during stress situations, thereby altering their dough-forming potential. HSPs have been implicated in such processes in other organisms.

Legume Seed Proteins

Legume seeds constitute the basic protein source in the diets of many developing countries. In developed countries they are used mainly as protein-rich food in intensive animal production, but they are also of importance (especially in the case of soybean products that are the most important in trade) as meat substitutes or functional agents in the food industry. Because it is not possible to grow soybean in many parts of the world, other leguminous species have been studied as vegetable protein sources (Gueguen 1983). For example, in Europe many studies have been

carried out on faba bean, pea and lupin in order to provide alternatives to meat by developing meat-like foods and to develop novel protein-rich foods as a complement to cereals. More recently, an interest has developed in using legume proteins for non-food markets. Various oilseeds have also been used to produce protein for processing. However, in oilseeds the yield and oil content rather than the protein properties have determined the choice of species.

This section will be concerned with the factors affecting the functional properties of legume storage proteins, with special reference to those of soybean, although most aspects may apply to proteins of other seeds such as pea, lupin or cruciferous oilseeds. Three principal aspects will be considered; processing, functionality and the possibilities available to manipulate functional properties.

Processing Legume Seed Proteins

Unlike many plant materials destined for manufacturing into foods, legume seeds are very rarely available in a form that is immediately usable by the food industry. Various types of processes are mandatory or desirable to purify the protein constituents and to transform them into ingredients suitable for the food processor. These processes may vary from one species to another, but all have implications with regard to the subsequent use of the material.

Refining Processes

An important constraint of refining is its cost, which favours simple processes with few steps, low energy consumption and a stable supply of raw material in large tonnages (the latter being readily fulfilled by seed proteins). However, processes may be mandatory, such as the removal of antinutritional factors (trypsin inhibitors, phytohaemagglutinin, goitrogen, saponin), desirable (removal of indigestible sugars) or optional (protein isolation, fractionation or specific modification) (Lillford 1981). Historically, a variety of procedures have been used to eliminate toxic substances and antinutrients. The processing steps generally included dehulling, boiling or cooking, grinding, toasting, puffing and fermentation (Deshpande and Damodaran 1990). Refinement processes can also be classified as dry (mechanical separation, air-classification) or wet (solvent extraction and washing, precipitation by pH adjustment, centrifugation). A typical dry process, pin-milling of legume seeds, leads to flours containing two populations of particles differentiated by both size and density. Using pin-milling, protein bodies can be detached from the surface of the starch granules so that, after air-classification, the heavy or coarse fraction (the starch fraction) can be separated from the light or fine fraction (the protein concentrate). However, only partial fractionation of protein and starch can be achieved, as even after repeated pin-milling and air-classification of pea flour, the lightest fraction still contains 8% starch in addition to 60% protein (this fraction is also enriched in lipids and ash) whereas the heavest fraction contains 5% protein in addition to 78% starch (Gueguen 1983).

Wet processes are recommended in order to prepare highly purified protein fractions (Fig. 13). For example, to prepare protein concentrates from defatted soy flour, the protein is generally immobilized by a choice of several treatments to enable removal of soluble sugars by washing with aqueous alcohol or dilute acid. However, these treatments are likely to leave the functionality of the proteins somewhat impaired (Wright and Bumstead 1984). On the other hand, isoelectric precipitation

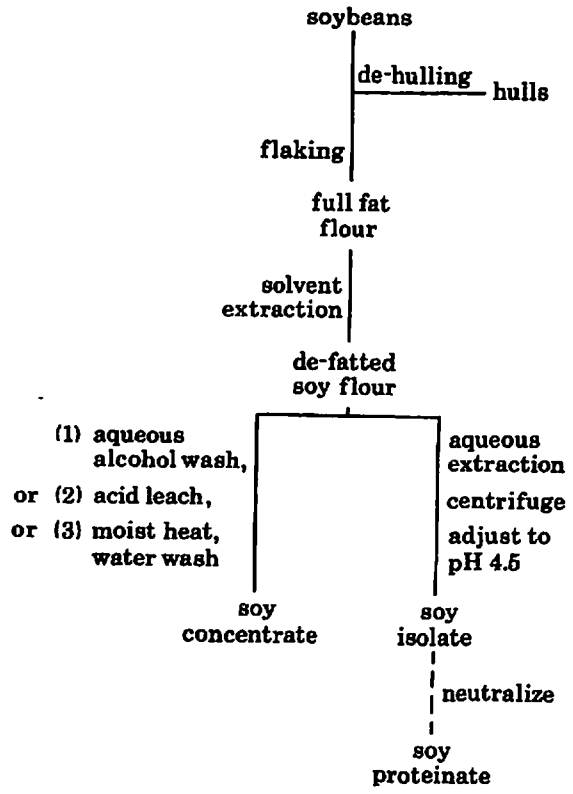

Fig. 13 : Processing of oilseeds (soybean). (Wright and Bumstead 1984)

at pH 4.5 is generally used to prepare protein isolates. However, even using mild conditions, precipitation can lead to loss of solubility and final neutralisation of the protein isolate before drying can pose a problem in subsequent processing and formulation because of residual salt. Following application of such processes to a defatted soy flour (containing 56% protein, 33% carbohydrates), typical protein concentrates can contain about 70% protein and 17% carbohydrates while protein isolates can contain as much as 95% protein with less than 1% carbohydrates.

Texturization of Seed Proteins:

As noted by Giddey (1983), the procedure used in the manufacture of concentrates or isolates is appropriate only if most of the protein remains in a native state. For example, concentrates obtained by a water/alcohol leaching process generally do not meet this requirement as partly denatured proteins cannot participate in the same way in stabilizing the artificial structures that are required for texturized end-products.

The primary task of texturization is to orientate the individual protein molecules so as to confer, after setting, a directional structure and thus an anisotropic resistance to the food. The texturization mechanisms therefore consist of protein insolubilization, including thermal (reversible or non-reversible) coagulation and denaturation.

Numerous patented processes of texturization have been described, which were divided by Giddey (1983) into nine classes: wet spinning, cooking extrusion, gel texturization, tear texturization, melt spinning, solvent texturization, texturization by surface deformation, texturization by freezing and biological texturization.

Processing and Functionality

Processes aim to confer desirable functional properties. These, in turn, relate mainly to conferring greater interfacial properties compared with most low-molecular-weight surfactants and emulsifiers, resulting in greater ability to stabilize emulsions or foams, and to impart textural attributes (Gueguen and Popineau 1998). Because of the complexity of food systems, information on how a protein will behave and on the relationships between microstructure and functional properties are extremely difficult to obtain, so that studies of functional properties generally start from simpler model systems before being extrapolated to mild processes and then to commercially available products. In addition, appropriate and relevant tools for functional characterization (water-binding capacity, hydrophobicity, charge and polarity, emulsifying and foaming characteristics, state of aggregation etc.) must be available and critically evaluated.

Functional properties are a manifestation of the inherent composition and structure of the protein, *i.e.* its amino acid composition, primary sequence and, finally, the organisation of the polypeptide chains and subunits in the native protein. For legume proteins, and specifically for soy proteins, the intrinsic properties of 7S and 11S globulins have been investigated in detail, including the amino acid compositions and sequences of subunits and their unfolding and association-dissociation upon heating, the latter using size separation and differential scanning calorimetry.

7S and 11S globulins have significantly different functional properties for emulsifying or gelation. Applying the same processing (*e.g.* to form tofu) to raw materials having different ratios of 7S and 11S globulins gives totally different product properties. For example, 7S globulins have higher emulsifying properties whereas gels made from 11S globulins are generally firmer with higher water holding capacities than their 7S analogues, the differences being ascribed to the contribution of disulphide bonds in the 11S globulins to the gel matrix. Thus, when a 7S protein solution is heated up to 100 °C at pH 7-8 and low ionic strength, the molecule dissociates into its three subunits, without further reaggregation, whereas reaggregation occurs if the protein solution is slightly concentrated (1%). As reviewed by Cheftel et al. (1985), the mechanism of denaturation during heating of 11S glycinin consists of the following steps:

1. The prevalence of hydrophobic interactions over electrostatic repulsion at the isoelectric point brings together the basic subunits,
2. Binding of the basic subunits together through disulphide bonds, leading to an oligomeric structure, followed by exchange of sulphydryl groups resulting in a release of the acid subunits which remain soluble,
3. Aggregation and precipitation of the basic oligomers as the result of hydrophobic interactions, leading to high M_r (10^7) through new exchanges of disulphide bonds, which may, in turn, initiate a three-dimensional protein network under specific conditions.

In 7S proteins (*e.g.* glycinin), the gelation mechanisms are different, as they involve more electrostatic interactions. The gelation mechanisms are also different

and more complex upon heating of a solution containing both 7S (emulsion-enhancing) and 11S (gel-enhancing) protein fractions (Fig. 14).

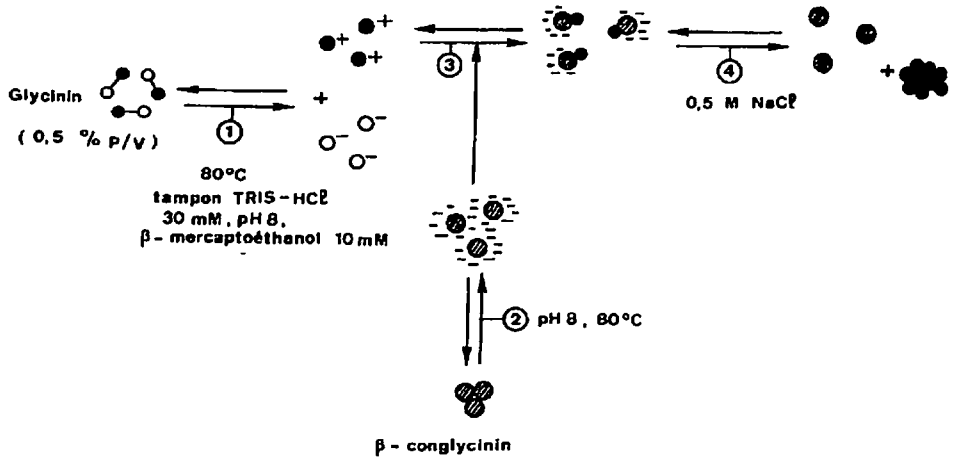

Fig. 14 : Hypothetical sequence of the dissociation-aggregation reactions during heating up to 80 ° C of a solution containing 11S glycinin + 7S β-conglycinin at pH 8.0.
○ acid (A) subunits of glycinin; ● basic (B) subunits of glycinin; ○ subunits of β-conglycinin. Glycinin ○–● is made of both A subunits ○ and B subunits ●. (Damodaran and Kinsella 1982)

In soybean the relationships between structural, chemical and functional properties have been largely substantiated. Because of the broad similarity with other seed proteins that have been studied, substitution of one protein for a related one generally presents few problems in food processing. However, large differences in the properties of 7S and 11S storage proteins will inevitably mean that the ratio of 7S and 11S proteins will play a significant role in determining the overall properties of the material (Wright 1983). In addition, other non-protein components present in the seeds may interact with proteins and thereby alter their basic properties. Finally, the structure/function relationships are based mainly on qualitative comparisons. According to Wright (1983), unless some quantitative relationships are established between functional and physicochemical properties, it will not be possible to predict with confidence the functional behaviour of any "protein X".

Manipulation of Functional Properties

Because the balance of 7S and 11S proteins is likely to play a significant role in determining the overall properties of the processed material, a number of ways to alter this ratio have been suggested. For example, this has been achieved during processing through the exploitation of intrinsic differences in the properties of the two protein types. For example, differences in the pH solubility profiles of the 7S and 11S globulins (Fig. 15) or differential effects of divalent cations on precipitation have been used in patented preparation procedures for 7S- and 11S-enriched fractions of legumes (Wright and Bumstead 1984). Technological approaches to quality improvement were also extended so as to include the elimination of undesirable components, including inactivation of trypsin inhibitors of soybean through heat treatments and high-shear extrusion to arrest lipoxygenase activity and thus eliminate "beany" flavours of soybean products that showed also have good shelf life against fat oxidation (Lusas 1998).

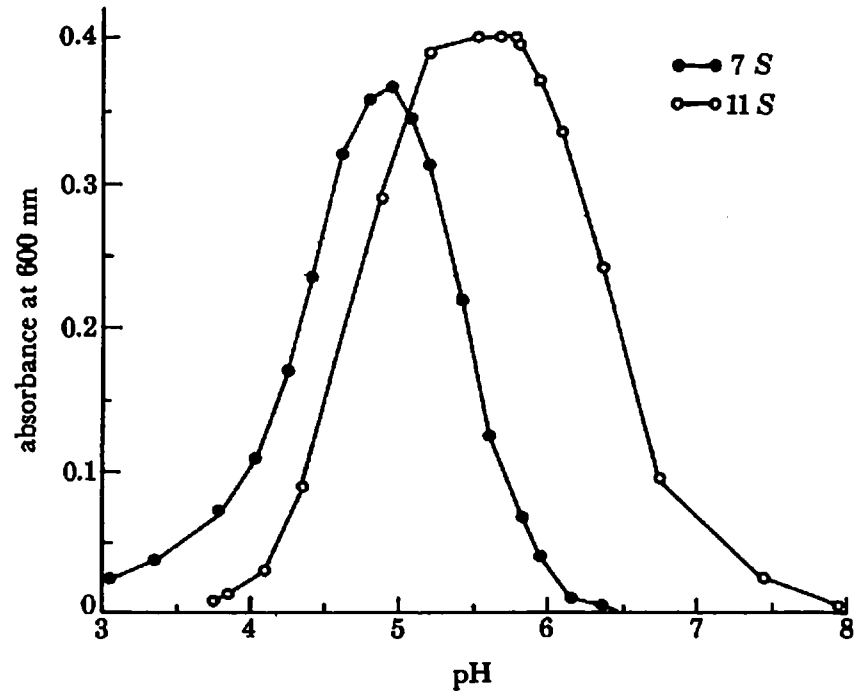

Fig. 15 : Effect of pH on precipitation of 7S and 11S soybean globulins in 0.06 M Tris-HCl buffer. (Wright and Bumstead 1984)

Another option for manipulating the ratio of 7S and 11S globulins is to exploit the natural variation in germplasm (*e.g.* the ratios vary from 0.6-2.0 in soybean, 0.7-5.0 in pea and 0.3-0.6 in faba bean) and, based on our knowledge of the inheritance of these proteins, to enhance the level of variation through breeding of high 7S or high 11S lines. Such a genetic approach has been successfully used to alter the ratios of 7S (emulsion-enhancing) and 11S (gel-enhancing) protein fractions. Soybeans with higher sulphur-amino acid contents (resulting from increased methionine) are also being developed (Lusas 1998).

A direct approach to modifying the functional properties is through the application of enzymic hydrolysis or chemical derivatization of specific residues to alter the structure or composition of the protein. Although the degree of hydrolysis is a highly critical parameter that must be carefully controlled to minimise the production of bitter-tasting peptides, peptide hydrolysates of soybean protein were obtained which had increased solubility and enhanced foaming and emulsifying properties (Wilde 1998).

Acknowledgement. IACR receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom.

References

Albani D, Hammond-Kosack MCU, Smith C, Conlan S, Colot V, Holdsworth M Bevan MW (1997) The wheat transcriptional activator SPA: seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes. Plant Cell 9: 171-184

Altenbach SB, Pearson KW, Meecker G, Staraci LC, Sun SSM (1989) Enhancement of the methionine content of seed proteins by the expression of a chimeric gene encoding a methionine-rich protein in transgenic plants. Plant Mol Biol 13: 513-522

Altenbach SB, Kuo C-C, Staraci LC, Pearson KW, Wainwright C, Georgescu A, Townsend J (1992) Accumulation of a Brazil nut albumin in seeds of transgenic canola results in enhanced levels of seed protein methionine. Plant Mol Biol 18: 235-245

Altpeter F, Vasil V, Srivastava V, Vasil K (1996) Integration and expression of the high-molecular-weight glutenin subunit *1Ax1* gene into wheat. Nat Biotechnoe 14 1155-1159

Alvarez ML, Guelman S, Halford NG, Lustig S, Reggiardo MI, Rybushkina N, Shewry PR, Stein J, Vallejos RH (2000) Silencing of HMW glutenins in transgenic wheat expressing extra HMW subunits. Theor Appl genet 100: 319-327

Andrews JL, Skerritt JH (1994) Quality-related epitopes of high Mr subunits of wheat glutenin. J Cell Sci 19: 219-230

Argos P, Narayana SVL, Nielsen NC (1985) Structural similarity between legumin and vicilin storage proteins from legumes. EMBO J 4: 1111-1117

Autran JC (1994) Size-exclusion high-performance liquid chromatography for rapid examination of size differences of wheat and cereal proteins. In: Kruger JE, Bietz JA (eds) HPLC of cereal and legume proteins. American Association of Cereal Chemists, St Paul, MN, pp 326-372

Autran JC, Hamer RJ, Plijter JJ Pogna NE (1997) To explore and improve the industrial use of wheats. Cereal Foods World 42: 216-227

Badley RA, Atkinson D, Hauser H, Oldani D, Green JP, Stubbs JM (1975) The structure, physical and chemical properties of the soybean protein glycinin. Biochim Biophys Acta 412: 214-228

Barro F, Rooke L, Békés F, Gras P, Tatham AS, Fido R, Lazzeri PA, Shewry PR, Barcelo P (1997) Transformation of wheat with high molecular weight subunit genes results in improved functional properties. Nat Biotechnol 15: 1295-1299

Bartels D,. Thompson RD (1986) Synthesis of messenger RNAs coding for abundant endosperm proteins during wheat grain development. Plant Sci 46: 117-125

Baud F, Pebay-Peyroula E, Cohen-Addad C, Odani S, Lehmann MS (1993) Crystal structure of a hydrophobic protein from soybean; a member of a new cysteine-rich family. J Mol Biol 231: 77-887

Bäumlein H, Wobus U, Pustell J, Kafatos FC (1986) The legumin gene family: structure of a B type gene of *Vicia faba* and a possible legumin gene specific regulatory element. Nucleic Acids Res 14: 2707-2720

Bäumlein H, Nagy I, Villaroel R, Inze D, Wobus U (1992) *Cis*-analysis of a seed protein gene promoter: The conservative RY repeat CATGCAT within the legumin box is essential for tissue-specific expression of a legumin gene. Plant J 2: 233-239

Blechl AE, Anderson OD (1996) Expression of a novel high-molecular-weight glutenin subunit gene in transgenic wheat. Nat Biotechnol 14: 875-879

Blumenthal C, Stone PJ, Gras PW, Békés F, Clarke B, Barlow EWR, Appels R, Wrigley CW (1998) Heat-shock protein 70 and dough-quality changes resulting from heat stress during grain filling in wheat. Cereal Chem 75: 43-50

Boston RS, Fontes EBP, Shank BB, Wrobel RL (1991) Increased expression of the maize immunoglobulin binding protein homologue b-70 in three zein regulatory mutants. Plant Cell 3: 497-505

Boulter D, Derbyshire E (1978) The general properties, classification and distribution of plant proteins. In: Norton G (ed) Plant proteins. Butterworths, London, pp 3-24

Breu V, Guerbette F, Kader JC, Kannangara CG, Svensson B, Von Wettstein-Knowles P. (1989) A 10 kD barley basic protein transfers phosphatidylcholine from liposomes to mitochondria. Carlsberg Res Commun 54: 81-84

Bright SWJ, Shewry PR (1983) Improvement of protein quality in cereals. CRC Crit Rev Plant Sci 1: 49-93

Brinch-Pedersen H, Galili G, Knudsen S, Holm PB (1996) Engineering of the aspartate family biosynthetic pathway in barley (*Hordeum vulgare* L.) by transformation with heterologous genes encoding feed-back-insensitive aspartate kinase and dihydrodipicolinate synthase. Plant Mol Biol 32: 611-620

Bushuk W, MacRitchie F (1989) Wheat proteins: aspects of structure that determine breadmaking quality. In: Dixon-Phillips R, Finley JW (eds) Protein quality and the effects of processing. Marcel Dekker, New York, pp 345-369

Bustos MM, Guiltinan MJ, Jordano J, Begum D, Kalkan FA, Hall TC (1989) Regulation of β-glucuronidase expression in transgenic tobacco by an A/T-rich, *cis*-acting sequence found upstream of a French bean β-phaseolin gene. Plant Cell 1: 839-853

Bustos MM, Begum D, Kalkan FA, Battraw MJ, Hall TC (1991) Positive and negative *cis*-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. EMBO J 10: 1469-1479

Casey R (1999) Distribution and some properties of seed globulins. In: Shewry PR, Casey R (eds) Seed proteins. Kluwer Dordrecht, The Netherlands, pp 159-169

Casey R and Domoney C (1999) Pea globulins. In: Shewry PR, Casey R (eds) Seed proteins. Kluwer Dordre cht, The Netherlands, pp 171-208

Casey R, Domoney C, Ellis N (1986) Legume storage proteins and their genes. Oxford Surv Plant Mol Cell Biol 3:1-95

Casey R, Domoney C, Ellis N, Turner S (1988) The structure, expression and arrangement of legumin genes in pea. Biochem Physiol Pflanz 183: 173-180

Cheftel J.-C, Cuq J-L., Lorient D (1985) Proteines alimentaires.. Technique et Documentation, Lavoisier-APRIA, Paris, 309 pp

Ciaffi M, Margiotta B, Colaprico G, de Stefanis E, Sgrulletta D, Lafiandra D (1995) Effect of high temperatures during grain filling on the amount of insoluble proteins in durum wheat J Genet Breed 49: 285-296

Ciaffi M, Lee YK, Tamas L, Gupta R, Appels R (1998) Molecular analysis of low Mr glutenin genes in *Triticum durum*. In: Gueguen J, Popineau (eds) Plant proteins from European crops – food and non-food applications. Springer, Berlin Heidelberg New York, pp 58-63

Coleman CE, Larkins BA (1999) Prolamins of maize. In: Shewry PR, Casey R (eds) Seed proteins. Kluwer, Dordrecht, The Netherlands, pp 109-139.

Damidaux R, Autran JC, Grignac P, Feillet P (1978) Mise en évidence de relations applicables en sélection entre l'électrophorégramme des gliadines et les propriétés viscoélastiques du gluten de *Triticum durum* Desf. C R Acad Sci Paris Sér D 287: 701-704

Damodaran S, Kinsella JE (1982) Effects of conglycinin on the thermal aggregation of glycinin. J Agric Food Chem 30: 812-817

Danielsson CE (1949) Seed globulins of the Gramineae and Leguminoseae. Biochem J 44: 387-400

Das OP, Messing JW (1987) Allelic variation and differential expression at the 27-kilodalton zein locus in maize. Mol Cell Biol 7: 4490-4497

de Freitas FA, Yunes JA, da Silva MJ, Arruda P, Leite A (1994) Structural characterization and promoter activity analysis of the γ-kafirin gene family from sorghum. Mol Gen Genet 245: 177-186

Deshpande SS, Damodaran S (1990) Food legumes: Chemistry and technology. In: Pomeranz Y (ed) Advances in cereal science and technology, vol 10. American Association of Cereal Chemists, St Paul, Minnesota, pp 147-241

Dickinson CD, Hussein EHA, Nielsen NC (1989) Role of posttranslational cleavage in glycinin assembly. Plant Cell 1:459-469

Doll H (1984) Nutritional aspects of cereal proteins and approaches to overcome their deficiences. Philos Trans R Soc Ser B 304: 373-380

Domoney C, Casey R (1985) Measurement of gene number for seed storage proteins in Pisum. Nucleic Acids Res 13: 687-699

Domoney C, Ellis THN, Davies DR (1986) Organization and mapping of legumin genes in *Pisum*. Mol Gen Genet 202: 280-285

Dubreil L, Gaborit T, Bouchet B, Gallant DJ, Broekaert WF, Quillien L, Marion D (1998) Spatial and temporal distibution of the major isoforms of puroindolines (puroindoline-a and puroindoline-b) and non-pecific lipid transfer protein (ns-LTPle$_1$) of *Triticum aestivum* seeds. Relationships with their in vitro antifungal properties. Plant Sci 138: 121-135

Duffus CM, Cochrane MP (1992) Grain structure and composition. In: Shewry PR (ed) Barley: genetics, biochemistry, molecular biology and biotechnology. CAB International, Wallingford, pp 291-317

Dunwell JM (1998) Cupins: A new superfamily of functionally diverse proteins that include germins and plant storage proteins. Biotechnol Genet Eng Rev 15: 1-32

Egorov TA, Odintsova TI, Musolyamov A Kh, Fido R., Tatham AS, Shewry PR (1996) Disulphide structure of a sunflower seed albumin: conserved and variant disulphide bonds in the cereal prolamin superfamily. FEBS Lett 396: 285-288

Falco SC, Guida T, Locke M, Mauvais J, Sanders C, Ward RT, Webber P (1995) Transgenic canola and soybean seeds with increased lysine. Bio/Technology 13: 577-582

Field JM, Shewry PR, Miflin BJ (1983) Solubilisation and characterization of wheat gluten proteins: correlations between the amount of aggregates proteins and baking quality. J Sci Food Agric 34: 370-377

Fontes EBP, Shank BB, Wrobel RL, Moose SP, O'Brian GR, Wurtzel ET, Boston RS (1991) Characterization of an immunoglobulin binding protein homologue in the maize floury-2 endosperm mutant. Plant Cell 3: 483-496

Forde BG, Heyworth A, Pywell J, Kreis M (1985) Nucleotide sequence of a B1 hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. Nucleic Acids Res 13: 7327-7339

Gautier M-F, Aleman ME, Guirao A, Marion D, Joudrier P (1994) *Triticum aestivum* puroindolines, two basic cysteine-rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Mol Biol 25: 43-57

Genov N, Goshev I, Nikolova D, Georgieva DN, Filippi B, Svendsen I (1997) A novel thermostable inhibitor of trypsin and subtilisin from the seeds of *Brassica nigra*: amino acid sequence, inhibitory and spectroscopic properties and thermostability. Biochim Biophys Acta 1341: 157-164

Gibbs PEM, Strongin KB, McPherson A (1989) Evolution of legume seed storage proteins – domain common to legumins and vicilins is duplicated in vicilins. Mol Biol Evol 6: 614-623

Giddey C (1983) Phenomena involved in the "texturization" of vegetable proteins and various technological processes used. In: Bodwell CE, Petit L (eds) Plant proteins for human foods. Martinus Nijhoff/Dr W. Junk, The Hague, pp 221-233

Giese H, Hopp E (1984) Influence of nitrogen nutrition on the amount of hordein, protein Z and β-amylase messenger RNA in developing endosperms of barley. Carlsberg Res Commun 49: 365-383

Gincel E, Simorre J-P, Caille A, Marion D, Ptak M, Vovelle F (1994) Three-dimensional structure in solution of a wheat lipid-transfer protein from multidimensional ^{1}H-NMR data. A new folding for lipid carriers. Eur J Biochem 226: 413-422

Giroux MJ, Morris CJ (1998) Wheat grain hardness results from highly conserved mutations in the friabilin components of puroindoline-a and b. Proc Natl Acad Sci USA 95: 6262-6266

Gomar J, Petit MC, Sodano P, Sy D, Marion D, Kader JC, Vovelle F, Ptak M (1996) Solution structure and lipid binding of a nonspecific lipid transfer protein extracted from maize seeds. Protein Sci 5: 565-577

Gueguen J (1983) Legume seed protein extraction, processing, and end product characteristics. In: Bodwell CE, Petit L (eds) Plant proteins for human foods. Martinus Nijhoff/Dr W Junk, The Hague, pp 63-99

Gueguen J, Popineau Y (1998) Plant proteins from European crops – food and non-food applications. Springer, Berlin Heidelberg New York, 339 pp

Halford NG, Field JM, Blair H, Urwin P, Moore K, Robert L, Thompson R, Flavell RB Tatham AS, Shewry PR (1992) Analysis of HMW glutenin subunits encoded by chromosome 1A of bread wheat (*Triticum aestivum* L.) indicates quantitative effects on grain quality. Theor Appl Genet 83: 373-378

Hall TC, Chandrasekharan MB Guofu L (1999) Phaseolin, its past, properties, regulation and future. In: Shewry PR, Casey R (eds) Seed proteins. Kluwer, Dordrecht, The Netherlands, pp 209-240

Hammond-Kosack MCU, Holdsworth MJ, Bevan MW (1993) *In vivo* footprinting of a low molecular weight glutenin gene (LMWG-1D1) in wheat endosperm. EMBO J 12: 545-554

Hargreaves J, Popineau Y, Cornec M, Lefebvre J (1996) Relationships between aggregative, viscoelastic and molecular properties in gluten from genetic variants of bread wheat. Int J Biol Macromol 18: 69-75

Kasarda DD (1989) Glutenin structure in relation to wheat quality. In: Pomeranz Y (ed) Wheat is unique: structure, composition, processing, end-use properties, and products. American Association of Cereal Chemists, St Paul, Minnesota, pp 277-302

Kasarda DD, Autran J-C, Lew E.J-L, Nimmo CC, Shewry PR (1983) N-terminal amino acid sequences of ω-gliadins and ω-secalins: implications for the evolution of prolamin genes. Biochim Biphys Acta 747: 138-150

Kasarda DD, King G, Kumosinski TF (1994) Computer molecular modeling of HMW-glutenin subunits. In: Molecular modeling. ACS Symposium Series No 576, American Chemical Society, Washington, DC, pp 209-220

Keeler SJ, Maloney CL, Webber PY, Patterson C, Hirata LT, Falco SC, Rice JA (1997) Expression of *de novo* high-lysine α-helical coiled-coil proteins may significantly increase the accumulated levels of lysine in mature seeds of transgenic tobacco plants. Plant Mol Biol 34: 15-29

Kermode AR, Bewley JD (1999) Synthesis, processing and deposition of seed proteins: the pathway of protein synthesis and deposition in the cell. In: Shewry PR, Casey R (eds) Seed proteins. Kluwer, Dordrecht, The Netherlands, pp 807-841

Ko T-P, Ng JD, McPherson A (1993) The three-dimenstional structure of canavalin from jack bean (*Canavalia ensiformis*). Plant Physiol 101: 729-744

Kreis M, Shewry PR (1989) Unusual features of seed protein structure and evolution. Bio-Essay 10: 201-207

Kreis M, Forde BG, Rahman S, Miflin BJ, Shewry PR (1985) Molecular evolution of the seed storage proteins of barley, rye and wheat. J Mol Biol 183: 499-502

Lawrence MC (1999) Structural relationships of 7S and 11S globulins. In: Shewry PR, Casey R (eds) Seed proteins. Kluwer, Dordrecht, The Netherlands, pp 517-541

Lawrence MC, Suzuki E, Varghese JN, Davis PC, Van Donkelaar A, Tuloch PA. Colman PM (1990) The three-dimensional structure of the seed storage protein phaseolin at 3 Å resolution. EMBO J 9: 9-15

Lawrence MC, Izard T, Beuchat RJ, Blagrove M, Coleman P (1994) Structure of phaseolin at 2.2 Å resolution. Implications for a common vicilin/legumin structure and the genetic engineering of seed storage proteins. J Mol Biol 238: 748-776

Lelievre JM, Oliveira LO, Nielsen NC (1992) 5'-CATGCAT-3' elements modulate the expression of glycinin genes. Plant Physiol 98: 387-391

Levanony H, Rubin R, Altschuler Y, Galili G (1992) Evidence for a novel route of wheat storage proteins to vacuoles. J Cell Biol 119: 1117-1128

Lew E J-L, Kuzmicky DD, Kasarda DD (1992) Characterization of low-molecular-weight glutenin subunits by reversed-phase high-performance liquid chromatography, sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and N-terminal amino acid sequencing. Cereal Chem 69: 508-515

Lillford PJ (1981) Extraction processes and their effect on protein functionality. In: Bodwell CE, Petit L (eds) Plant proteins for human food. Martinus Nijhoff/Dr W. Junk, The Hague, pp 199-205

Lohmer S, Maddaloni M, Motto M, Di Fonzo N, Hartings H, Salamini F, Thompson RD (1991) The maize regulatory locus *Opaque-2* encodes a DNA binding protein which activates the transcription of the b-32 gene. EMBO J 10: 617-624

Lusas E (1998) Achievements, status and challenges in food protein processing. In: Guegen J, Popineau Y (eds) Plant proteins from European crops – food and non-food applications. Springer, Berlin Heidelberg New York, pp 257-264

MacRitchie F (1998) Protein composition and physical properties of wheat flour doughs. In: Guegen J, Popineau Y (eds) Plant proteins from European crops – food and non-food applications. Springer, Berlin Heidelberg New York, pp 113-119

Mertz ET, Bates LS, Nelson OE (1964) Mutant gene that changes protein composition and increases lysine content of maize endosperm. Science 145: 279-280

Miflin BJ, Field JM, Shewry PR (1983) Cereal storage proteins and their effect on technological properties. In: Daussant J, Mossé J,. Vaughan J (eds) Seed proteins. Phytochemical Society of Europe Symposia, Series n° 20. Academic Press, London, pp 255-319

Misra PS, Mertz ET, Glover DV (1975) Studies on corn proteins. VI. Endosperm protein changes in single and double endosperm mutants of maize. Cereal Chem 52: 161

Molvig L, Tabe LM, Eggum BO, Moore AE, Craig S, Spencer D, Higgins TJV (1997) Enhanced methionine levels and increased nutritive value of seeds of transgenic lupins (*Lupinus angustifolius* L.) expressing a sunflower seed albumin gene. Proc Natl Acad Sci USA 94: 8393-8398

Müller M, Knudsen S (1993) The nitrogen response of a barley C-hordein promoter is controlled by positive and negative regulation of the GCN4 and endosperm box. Plant J 4: 343-355

Müller S, Vensel WH, Kasarda DD, Köhler P Wieser H (1998) Disulphide bonds of adjacent cysteine residues in low molecular weight subunits of wheat glutenin. J Cereal Sci 27: 109-116

Munck L (1992) The case of high-lysine barley breeding. In: Shewry PR (ed) Barley genetics, biochemistry, molecular biology and biotechnology. CAB International, Wallingford, pp 573-601

Osborne TB (1907) The proteins of the wheat kernel. Carnegie Inst. Washington, Washington DC, Publ No 84

Osborne TB (1924) The vegetable proteins. Longmans, Green, London

Parker ML (1980) Protein body inclusion in developing wheat endosperm. Ann Bot 46: 29-36

Payne PI (1987) Genetics of wheat storage proteins and the effect of allelic variation on breadmaking quality. Annu Rev Plant Physiol 38: 141-153

Payne PI, Seekings JA (1996) Characterisation of Galahad-6, Galahad-7 and Galahad-8, isogenic lines that contain only one HMW glutenin subunit. In: Wrigley CW (ed) Gluten '96, Proc 6th Int Gluten Workshop, 1-6 September, Sydney, Australia, pp 14-17

Payne PI, Jackson EA, Holt LM, Law CN (1984) Genetic linkage between endosperm storage protein genes on each of the short arms of chromosomes 1A and 1B in wheat. Theor Appl Genet 67: 235-243

Payne PI, Nightingale MA, Krattiger AF, Holt LM (1987) The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. J Sci Food Agric 40: 51-65

Pechanek U, Karger A, Groger S, Charvat B, Schöggl G, Lelley T (1997) Effect of nitrogen fertilization on quantity of flour protein components, dough properties, and breadmaking quality of wheat. Cereal Chem 74: 800-805

Pogna N, Lafiandra D, Feillet P and Autran J-C (1988) Evidence for a direct causal effect of low molecular weight subunits of glutenins on durum viscoelasticity in durum wheats. J Cereal Sci, 7: 211-214

Popineau Y, Cornec M, Lefebvre J, Marchylo B (1993) Influence of HMW glutenin subunits on glutenin polymers and rheological properties of glutens and gluten subfractions of near-isogenic lines of wheat Sicco. J Cereal Sci 19: 231-241

Quayle T, Feix G (1992) Functional analysis of the -300 region of maize zein genes. Mol Gen Genet 23: 369-374

Randall PJ, Moss HJ (1990) Somme effects of temperature regime during grain filling on wheat quality. Aust J Agric Res 41: 603-617

Rico M, Bruix M, González C, Monsalve RI, Rodríguez R. (1996) ^{1}H NMR assignment and global fold of napin BnIb, a representative 2S albumin seed protein. Biochemistry 35: 15672-15682

Rooke L, Békés F, Fido R, Barro F, Gras P, Tatham AS, Barcelo P, Lazzeri P, Shewry PR (1999) Overexpression of a gluten protein in transgenic wheat results in highly elastic dough. J Cereal Sci 30: 115-120

Rubin R, Levanony H, Galili G (1992) Evidence for the presence of two different types of protein bodies in wheat endosperm. Plant Physiol 99: 718-724

Saalbach I, Waddell D, Pickardt T, Schieder O, Muntz K (1995) Stable expression of the sulphur-rich 2S albumin gene in transgenic *Vicia narbonensis* increases the methionine content of seeds. J Plant Physiol 145: 674-681

Salunkhe DK, Chavan JK, Kadam SS (1992) World oilseeds – chemistry, technology, and utilisation. Van Nostrand Reinhold, New York, 554 pp

Schmidt RJ, Burr FA, Burr B (1987) Transposon tagging and molecular analysis of the maize regulatory locus *Opaque-2*. Science 238: 960-963

Schmidt RJ, Ketudat M, Aukerman M, Hoschek G (1992) *Opaque-2* is a transcriptional activator that recognizes a specific target site in 22-kDa zein genes. Plant Cell 4: 689-700

Shen B, Carneiro N, Torres-Jerez I, Stevenson B, McCreery T, Helentjaris T, Baysdorfer C, Almira E, Ferl RJ, Habben JE, Larkins BA (1994) Partial sequencing and mapping of clones from two maize cDNA libraries. Plant Mol Biol 26: 1085-1101

Shewry PR (1995) Plant storage proteins. Biol Rev 70: 375-426

Shewry PR (1998) The HMW subunits and their role in determining the functional of wheat gluten and dough. In: Workshop on Biopolymer Science – Food and Non Food Applications, 28-30 September, Montpellier (France)

Shewry PR, Casey R (eds) (1999) Seed proteins. Kluwer, Dordrecht, The Netherlands

Shewry PR, Miflin BJ (1985) Seed storage proteins of economically important cereals. In:Pomeranz Y (ed) Advances in cereal science and technology, vol 7. American Association of Cereal Chemists, St Paul, Minnesota, USA, pp 1-83

Shewry PR, Pandya M J (1999) The 2S albumins storage proteins. In Shewry PR, . Casey R (eds) Seed proteins. Kluwer Dordrecht, The Netherlands, pp 563-586

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

Shewry PR, Halford NG, Tatham AS (1994a) Analysis of wheat proteins that determine breadmaking quality. Food Sci Technol Today 8: 31-36

Shewry PR, Miles MJ, Tatham AS (1994b) The prolamin storage proteins of wheat and related species. Prog Biophys Mol Biol 61: 37-59

Shewry PR, Greenfield J, Buonocore F, Wellner N, Belton PS, Parchment O, Osguthorpe D, Tatham AS (1998) Conformational studies of the repetitive sequences of HMW subunits of wheat glutenin. In: Gueguen J, Popineau Y (eds) Plant proteins from European crops – food and non-food applications. Springer, Berlin Heidelberg New York, pp 120-125

Shewry PR, Tatham AS, Halford NG (1999) The prolamins of the Triticeae. In: Shewry PR, Casey R (eds) Seed proteins. Kluwer Dordrecht, The Netherlands, pp 35-78

Shewry PR, Tatham AS, Popineau Y (2000) The chemical basis of wheat grain quality. In: Autran JC, Hamer RJ (eds) Wheat Science and Technology in the European Union. IRTAC, Paris. In press

Shotwell MA, Larkins BA (1989) The biochemistry and molecular biology of seed storage proteins. In: Marcus A (ed) The biochemistry of plants, vol 5. Academic Press, San Diego, pp 297-345

Shutov AD, Kakhovskaya IA, Braun H, Bäumlein H, Muntz K (1995) Legumin-like and vicilin-like seed storage proteins: evidence for a common single-domain ancestral gene. J Mol Evol 41: 1057-1069

Sissons MJ, Békés F, Skerritt JH (1998) Isolation and functionality testing of low molecular weight glutenin subunits. Cereal Chem 75:30-36

Sörensen MB, Cameron-Mills V Brandt A (1989) Transcriptional and post-transcriptional regulation of gene expression in developing barley endosperm. Mol Gen Genet 217: 195-201

Spencer D (1984) The physiological role of storage proteins in seeds. Philos Trans R Soc Ser B 304: 275-285

Strobl S, Mühlhahn P, Bernstein R, Wilscheck R, Maskos K, Wunderlich M, Huber R, Glockhuber R, Holak TA (1995) Determination of the three-dimensional structure of the bifunctional α-amylase/trypsin inhibitor from ragi seeds by NMR spectroscopy. Biochemistry 34: 8281-8293

Svendsen I, Nicolova D, Goshev I, Genov N (1989) Isolation and characterization of a trypsin inhibitor from the seeds of kohlrabi (*Brassica napus* var. *Rapifera*) belonging to the napin family of storage proteins. Carlsberg Res Commun 54: 231-239

Svendsen I, Nicolova D, Goshev I, Genov N (1994) Primary structure, spectroscopic and inhibitory properties of a two-chain trypsin inhibitor from the seeds of charlock (*Sinapis arvensis* L), a member of the napin protein family. Int J Pept Protein Res 43: 425-430

Tatham AS, Shewry PR, Belton PS (1990) Structural studies of cereal prolamins, including wheat gluten. In: Pomeranz Y (ed) Advances in cereal science and technology, vol 10. American Association of Cereal Chemists, St Paul, Minn., USA, pp 1-78

Terras FRG, Goderis IJ, Van Leuven F, Vanderleyden J, Cammue BPA, Broekaert WF. (1992) *In vitro* antifungal activity of a radish (*Raphanus sativus* L.) seed protein homologous to nonspecific lipid transfer proteins. Plant Physiol 100: 1055-1058

Terras FRG, Schoofs HME, Thevissen K, Osborn RW, Vanderleyden J, Cammue BPA, Broekaert WF (1993) Synergistic enhancement of the antifungal activity of wheat and barley thionins by radish and oilseed rape 2S albumins and by barley trypsin inhibitors. Plant Physiol 103: 1311-1319

Thomas MS, Flavell RB (1990) Identification of an enhancer element for the endosperm-specific expression of high molecular weight glutenin. Plant Cell 2: 1171-1180

Utsumi S (1992) Plant food protein engineering. Adv Food Nutr Res 36: 82-208

Utsumi S, Gidamis AB, Mikami B, Kito M (1993) Crystallization and preliminary X-ray crystallographic analysis of the soybean proglycinin expressed in *Escherichia coli*. J Mol Biol 233: 177-178

Utsumi S, Gidamis AB, Takenaka Y, Maruyama N, Adachi M, Mikami B (1996) Crystallization and X-ray analysis of normal and modified recombinant soybean proglycinins: three-dimensional structure of normal proglycinin at 6 Å resolution. In: Parris N, Kato A, Creamer LK, Pearce J (eds) Macromolecular interactions in food technology. American Chemical Society, Washington, DC, pp 257-270

Weegels PL, Hamer RJ Schofield JD (1997) Depolymerisation and re-polymerisation of wheat glutenin during dough processing. II. Changes in composition. J Cereal Sci 25: 155-163

Wilde PI (1998) Modifying the interfacial behavior and functional characteristics of proteins. In: Guegen J, Popineau Y (eds) Plant proteins from European crops – food and non-food applications. Springer, Berlin Heidelberg New York, pp 105-112

Wilson DR, Larkins BA (1984). Zein gene organization in maize and related grasses. J Mol Evol 29: 330-340

Wright DJ (1983) Comparative physical and chemical aspects of vegetable protein functionality. Qual Plant Plant Foods Hum Nutr 32: 389-400

Wright DJ. Bumstead MR (1984) Legume proteins in food technology. Philos Trans R Soc B 304: 381-393

Wrigley CW (1995) Identification of food-grain varieties. American Association of Cereal Chemists, St Paul, Minnesota, 283 pp

Wrigley CW, du Cros DL, Fullington JG, Kasarda DD (1984) Changes in polypeptide composition and grain quality due to sulphur deficiency in wheat.J Cereal Sci 2:15-24

Xia J.-H, Kermode AR (1999) Analyses to determine the role of embryo immaturity in dormancy maintenance of yellow-cedar (*Chamaecyparis nootkatensis*) seeds: synthesis and accumulation of storage proteins and proteins implicated in desiccation tolerance. J Exp Bot 50: 107-118

Youle RJ, Huang AHC (1981) Occurrence of low molecular weight and high cysteine containing albumin storage proteins in oilseeds of diverse species. Am J Bot 68: 44-48