Seed storage proteins and approaches for improvement of their nutritional quality by genetic engineering

Surekha Mandal and R. K. Mandal*
Department of Biochemistry, Bose Institute, Calcutta 700 054, India

Seed storage proteins of grain crops meet the major dietary protein requirement of over half of the world population. However, seed proteins in general are deficient in some essential amino acids and hence are of poor nutritional quality. Therefore, intensive research is going on to isolate and characterize these proteins and their genes, and to produce transgenic crop plants with modified seed protein genes, with a view to improving their nutritive value as human food and animal feed. Many seed storage protein genes from cereals, legumes and oil seeds have been isolated, sequenced and their regulation has been studied by promoter deletion assay in transgenic plants. The amino acid composition of seed proteins of some transgenic crops has been marginally improved by modifying these genes for more methionine and lysine codons by site-directed mutagenesis or by introducing heterologous genes. The synthesis, processing and targeting of the introduced gene products and their stable expression in successive generations of transgenic plants have also been studied. The potentiality of such genetic engineering approaches has been amply demonstrated. But problems of control over copy number and site of integration of the foreign gene and the position effect of the changes in the amino acid in the polypeptide chain on the final level of expression and deposition of the correctly processed storage protein remains unsolved. More basic research is needed before large-scale application of this approach for improvement of nutritional quality of seed proteins becomes possible.

A major part of the human diet all over the world consists of cereals and legumes. According to FAO estimate, 70% of human food comprises cereals and legumes and the remaining 30% comes from animals, which again feed on these seed meals[1]. Animal proteins being more expensive, people in developing countries virtually depend on seed protein alone for their entire protein requirement. But unlike animal proteins like casein and egg albumin which are nutritionally more balanced in terms of all essential amino acids, plant proteins are generally deficient in some of them. As human beings cannot synthesize essential amino acids, these must be supplied from outside as food. Cereal proteins are generally deficient in lysine and tryptophan whereas legumes are deficient in sulphur-containing amino acids – methionine and cysteine. The presence of antinutritional compounds like glucosinolates (thioglucosides) in *Brassica* and proteinaceous toxic compounds like lectins, enzyme inhibitors, etc. in legumes adversely affects the nutritional quality of seed proteins. Thus higher or better quality with reference to seed proteins can have many different meanings, depending upon the crop and its utilization. But by convention, the 'quality' of seed protein refers to how far its amino acid composition matches the balanced amino acid composition recommended for human diet by the World Health Organization (WHO)[2].

Before the advent of recombinant DNA technology, studies on seeds mainly focused on the isolation, characterization and nutritive value of proteins they contained. In the last fifteen years, however, the ability to isolate, manipulate and reintroduce genes into other plant species has led to an enhanced interest in the means for qualitative and quantitative improvement of seed grains by genetic engineering.

Classification and isolation of seed proteins

Classification

Seed proteins can be broadly classified into two categories, viz. housekeeping proteins and storage proteins. The housekeeping proteins are responsible for maintaining normal cell metabolism[3]. A more recent classification divides these proteins into storage, structural and biologically active proteins[4]. The major biologically active proteins include lectins, enzymes and enzyme inhibitors. These are minor proteins and may have nutritionally more balanced amino acid composition than storage proteins. The seed storage proteins on the other hand, are non-enzymatic and have the sole purpose of providing proteins (nitrogen and sulphur source) required during germination and establishment of a new plant. Seed proteins were empirically classified many years ago by Osborne[5] based on their solubility as follows:

*For correspondence. (e-mail: rkmandal@bic.boseinst.ernet.in)

Albumins: Water extractable (1.6S–2S);
Globulins: Extractable in dilute salt solutions (7S–13S);
Prolamins: Extractable in aqueous alcohol;
Glutelins: Most difficult to solubilize; Extractable by weakly acidic or alkaline or dilute SDS solution.

Albumins and globulins comprise the storage proteins of dicots (e.g. pulses), whereas prolamins and glutelins are major proteins in monocots (e.g. cereals)[6]. The seed storage proteins can be distinguished from other proteins by some of their characteristics, e.g. (1) These accumulate in high amounts in seed during mid-maturation stage of seed development and are used up during germination. (2) These are synthesized only in the seed (in cotyledon or in endosperm) and not in other tissues. (3) They lack any other functional activity besides storage. (4) They are deposited mostly in special storage organelles called protein bodies.

Isolation

Seed proteins are generally isolated by extracting crushed whole seeds or seed meals left after oil extraction (defatted by solvent extraction, if necessary) with a low salt buffer. When desired, the different proteins are partially separated by successive differential extraction with water, salt solution, aqueous ethanol and then acid/alkali with detergent. The total or partially fractionated proteins are further purified by a combination of methods like heat and isoelectric precipitation, ammonium sulphate fractionation gel filtration and ion-exchange chromatography. For further separation of microheterogeneous forms/types, techniques like isoelectric focusing, two-dimensional electrophoresis and HPLC are used[7–10].

Characterization and structural studies

As expected of a reserve of nitrogen for the seedling, storage proteins are generally rich in asparagine, glutamine and arginine or proline[11]. The legume proteins (mostly 7–11S globulins) tend to be deficient in the sulphur-containing amino acid cysteine and methionine and also tryptophan while cereals generally lack lysine, threonine and tryptophan. The globulins are generally present as oligomers. The cereal prolamins are present as monomers or small aggregates, while glutelins form large disulphide-bonded aggregates[12]. Most of the seed proteins have high molecular weights and their water solubility is poor. There are many examples of small water soluble seed proteins in dicot seeds like castor bean[13], pea[14], mung bean[15], mustard[16–18], *Amaranthus*[19], *Chenopodium*[20], Alfalfa[21], many oil seeds[22], etc.

Storage proteins have also been studied from many seed grains of importance in India like chickpea[23], pigeon pea[24], black gram[25], mung bean[26] and several oilseeds[27,28]. All these studies revealed a close similarity in sizes, subunit nature and other physico-chemical properties of the same seed protein fraction from different legumes. For example, Uppal and Matta[25] working with black gram or urad bean (*Vigna mungo*) and Krishna and Bhatia[24] with pigeon pea (*Cajanus cajan*) found the major globulin fraction as 7S vicilin, a trimeric protein composed of one large (70–72 kDa) and two small (40–57 kDa) subunits not linked by disulphide bridges. This contrasts with the major 7S vicilin from pea which is a homotrimeric protein of 70 kDa subunits, also not linked by disulphide bridges[29]. These proteins are normally heterogeneous in terms of polypeptide size, charge and amino acid composition. Pea legumin or vicilin, a comparatively simple storage protein, contains up to 30 separable polypeptides[30] as observed in two-dimensional isoelectric focusing/SDS-PAGE analysis, whereas a complex fraction such as wheat storage protein contains more than 60 polypeptides[31]. In oilseed *Brassica* and other cruciferae, seed storage proteins are 2S albumin (napin) and 12S globulin (cruciferin) types. The 2S albumins are synthesized as large precursor molecules and after post-translational processing a signal peptide, an additional amino terminal fragment, an internal fragment and some carboxy-terminal residues are removed. The two subunits resulting from this cleavage process are linked by disulphide bridges[17].

Similar is the case with 2S albumin from *Arabidopsis thaliana* for which at least five isoforms have been characterized[32]. Castor bean 2S protein is unique in that two heterodimeric 11 kDa proteins of 97 and 99 amino acid residues are formed by processing of a single polypeptide synthesized by a single mRNA[33]. Eight members of the 2S albumin of sunflower are made of a single polypeptide chain, of which SFA 8 is rich in sulphur amino acids (16 Met and 8 Cys)[34]. In addition, Allen *et al.*[35] reported the deduced amino acid sequence of another napin-like sunflower 2S albumin, HaG 5, encoding heterodimeric protein with approximate molecular weight of 14 kDa. The 12S globulins are oligomeric complexes consisting of six subunit pairs, each subunit being composed of a heavy alpha chain and a light beta chain. Each pair is derived from a single precursor[36]. In the 12S globulins of legumes[6], cucurbits[37], rape seed[36] and oat[38], there are oligomeric structures of six subunits, in which each subunit is composed of disulphide-bonded large acidic and small basic polypeptides. The mechanisms of synthesis are also similar in different genera. Similar subunit nature of 2S albumins and 12S globulins from *Chenopodium album*[20,39] and *Sterculia foetida*[40] has also been described.

Amino acid composition

A large number of purified proteins have been characterized in terms of their amino acid composition[7] for their importance in human and animal nutrition. In general, it

has been found that cereal and oilseed proteins are nutritionally more balanced in amino acid composition compared to legume seed proteins. Table 1 shows the amino acid composition of some representative seed proteins compared with the WHO recommendation of essential amino acid content. A comprehensive early information on the isolation and composition of various seed proteins can be obtained in Prakash and Rao[7]. It can be seen that some mustard, *Chenopodium*[20] and *Amaranthus*[19] seed proteins are almost ideal proteins in terms of essential amino acids and meet or surpass WHO recommendation. The same holds true for 2S protein from cotton seeds[41] and buck wheat[42]. Also, most of the Asp and Glu residues in seed proteins are present as Asn and Gln, as was determined from nucleotide sequences later.

Structural studies

Understanding the structural features of seed storage protein is important for providing a basis for proposed engineered mutations at suitable sites of these proteins with the ultimate goal of enhancing nutritional quality, without affecting their functional role in seed metabolism and storage. In spite of large complexity in polypeptide subunits and amino acid sequences, diverse seed proteins may share considerable structural and sequence homology (see later in the article). This only indicates that notwithstanding evolutionary divergence, seed proteins have retained some essential structural features due to some functional constraint in seed metabolism and/or their function as storage proteins in specified organelles[43,44]. The 3D structure of seed proteins has been studied by spectroscopy, X-ray diffraction and other physicochemical techniques. One general feature of the proteins is their more or less compact structure with a hydrophobic core and some exposed loops or surface domains.

Published reports on direct structure determination of seed storage proteins by X-ray diffraction are scanty. Some protease inhibitors (many of which have amino acid sequence homology with seed storage proteins) have been well characterized both by X-ray diffraction[45,46] and by NMR studies[47]. De-Meester *et al.*[48] determined the 3D structure of soybean Kunitz-type trypsin inhibitor by molecular replacement using the coordinates of homologous *Erythrima* trypsin inhibitor. It was found that the conformation of the reactive site loop did not appreciably change by interaction with trypsin. The structure of a seed globulin having chymotrypsin inhibitory activity from winged bean *Psophocarpus tetragonolobus* has been determined by X-ray crystallography[49]. The protein structure revealed 12 antiparallel beta-strands with connecting loops arranged in a characteristic beta-trefoil fold common to other homologous serine protease inhibitors. The X-ray structure models of proCon A and Con A are available. This study confirmed that the residues involved in transpeptization reaction during the formation of mature Con A are neighbours[50]. Sainani *et al.*[51] studied the effect of chemical modification of nine different amino acids separately in pennisatin, a major storage protein of pearl millet, on the structural and functional properties[51]. As expected, changes in hydrodynamic properties indicated a change in the secondary structure for most substitutions. Thermostability remained unaffected except in the case of Glu and Asp modification. Some essential conserved features in the primary structure of napin-like 2S albumins have been described[52]. By proteolytic cleavage and subsequent peptide mapping by matrix-assisted laser desorp-

Table 1. Amino acid composition of some seed proteins (mole%) compared with WHO recommendation of essential amino acids

Amino acid	WHO[2]	Maize zein[8]	*B. camp.* (yellow sarson) 2S[52,†]	*Chenopodium* (goose foot) 2S protein[20]	*Amaranthus* AmA1[19,†]	Pea legumin[41a]
Ala (A)	–	13.7	4.2	4.3	5.3	6.0
Arg (R)	–	1.2	4.2	7.5	5.3	10.0
Asx (B)	–	4.7	1.6	5.0	16.6	13.1
Cys (C)	–	Trace	6.6	5.8	0.7	1.2
Glx (Z)	–	21.2	23.9	17.9	8.9	19.7
Gly (G)	–	2.1	6.6	5.8	12.0	6.9
His (H)	–	0.9	3.3	5.7	3.3	1.8
Ile (I)	4.0	4.1	4.2	3.8	5.0	4.0
Leu (L)	7.0	20.0	6.6	3.0	7.6	7.6
Lys (K)	5.5	0.2	6.6	4.9	6.6	4.2
Met (M)	3.5*	Trace	3.3	2.8	1.6	0.7
Phe (F)	6.0**	6.3	1.6	5.4	5.6	2.3
Pro (P)	–	9.7	8.2	6.1	3.6	5.5
Ser (S)	–	6.6	4.9	5.6	6.6	5.8
Thr (T)	4.0	2.4	3.3	5.1	5.3	3.1
Trp (W)	1.0	N.D.	0.82	N.D.	2.3	–
Tyr (Y)	**	3.1	0.82	3.0	4.8	3.6
Val (V)	5.0	3.6	4.9	3.4	5.6	5.0

†, Derived from DNA sequence; N.D., not determined; *, Met plus Cys; **, Phe plus Tyr.

tion ionization mass spectroscopy (MALDI-MS), the correct assignment of 8 cysteine residues (2 in the small and 6 in the large subunit) of napin involved in 4 disulphide linkages has been possible[53]. The high thermal stability of napin tertiary structure was corroborated by CD and high sensitivity differential scanning calorimetry[54]. Two consecutive transitions were attributed to two different folding domains.

Limited proteolysis with several endopeptidases, immunological comparison and CD spectrometry indicated that cleavable propeptide regions are exposed to the exterior of the mature napin. Moreover, the conformation of the mature napin was not significantly different from the pronapin expressed in insect cells[55]. Chemical, CD and fluorescence spectroscopic studies with the isolated small subunit of napin indicated that it has a strong tendency to form a non-native intramolecular disulphide bridge. Addition of the reduced large napin subunit inhibited this tendency[56]. This study indicates that interchain disulphide formation in napin may be an early step during maturation of this (and other) multisubunit seed protein. Using information from CD and NMR spectroscopy and sequence homology, Pandya *et al.*[57] proposed a model for the structure of sunflower 2S albumin SFA8, based on the known structure of ragi alfa-amylase/trypsin inhibitor[58]. The model is composed of a bundle of four alfa-helices tightly knit by 4 disulphide bridges. From CD and chemical denaturation studies, we proposed a similar structure for *Brassica* 2S albumin (Ghose *et al.* unpublished observation). From thermal denaturation, we observed no intermediate molten globule state for this protein.

Apart from the structure, other physicochemical and hydrodynamic properties of seed proteins are important determinants for their use in the food industry, e.g. viscosity, thermal-pH-emulsion stability, elastoextensibility, baking cooking quality, etc. Numerous studies have been carried out with different oilseeds, legumes and cereals[27,59,59a].

Homologies of seed storage proteins and their genes

With the increasing knowledge about the sequence of different proteins and efficient computer programs, homologies between drastically unrelated proteins are being detected. The homology studies not only give an insight into the structural and functional properties of different proteins but more importantly cast light on the evolutionary status of these proteins. The homologies of seed proteins can be studied by comparing their cross reactivity with specific antibody against one and amino acid sequence, either directly determined from protein or more easily derived from nucleotide sequence of cloned cDNA. The different classes of seed proteins exhibit considerable homologies among themselves[59a,59b]. Several diverse 2S proteins from different dicot seeds have shown sequence homology: the methionine-rich brazil nut seed protein[60], the allergenic seed protein in castor bean[61], the very basic 2S albumin in rape seed[16,52], a trypsin inhibitor from barley[62] and albumins of sunflower[35] and wheat[63]. The amino acid sequences of the coding region of 2S storage albumin of *Brassica* sp. showed 80% identity amongst themselves[52].

Homologies between seed proteins have also been studied with the help of antibody cross-reactivity studies and subsequently finding the homologous domains. Antigenically homologous seed proteins were found in different phylogenetically related species of *Cruciferae*[18] and of Chenopodiaceae and Amaranthaceae family[20]. Common antigenic determinants were in the prolamin fractions of wheat, barley and maize. N-terminal sequence analysis of wheat and barley prolamin shows homologous regions[64]. Prolamin c-DNAs bear strong homologies, 11S globulin cDNA of oat shares 74% homology with rice glutelins. It has been shown that parts of the non-repetitive domains A, B and C of one group of prolamins from barley, rye and wheat have considerable sequence similarity with sequences present in a large group of seed proteins from monocot and dicot plants[65,66], including Bowman–Birk protease inhibitors, cereal inhibitors of alfa amylase and trypsin, and 2S albumin storage proteins of castor bean, sunflower and oil-seed rape[70] (Figure 1). This implies an ancient origin of these domains from prolamin super family whose overall topology has been conserved in spite of large allowable changes in amino acid residues.

Storage proteins are encoded by families of polymorphic genes and gene copy number can be very high. The *Arabidopsis thaliana* family for example, contains four tandemly clustered members. The globulins of dicot seeds are also the products of multigene families. In monocots, one well-studied seed storage protein is zein of maize. It has been suggested that there may as many as ten families

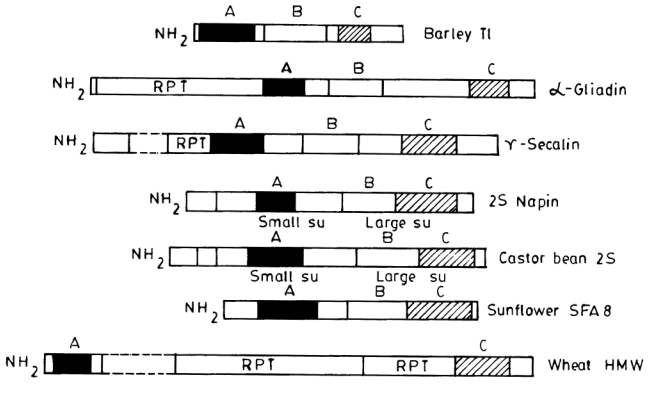


Figure 1. Schematic comparison of structures showing homologous regions of several representative seed proteins: barley trypsin inhibitor[62], alpha-gliadin of wheat[65,68], r-secalin of rye[65,66], 2S napin from oilseed rape[16], castor bean 2S protein[61], sunflower albumin SFA-8 (ref. 35) and wheat HMW subunit[70]. The regions homologous to domains A, B and C are present in all the proteins. Blocks of repeated amino acids (RPT) are present in cereal proteins but absent in dicot seed proteins.

of prolamin in maize, each having 3 to 10 genes. This means that there are 30 to 100 zein genes in maize. In spite of microheterogeneity in zein families, there is 90% homology between the nucleotide sequences. The microheterogeneity gives rise to mostly conservative mutations, so that in the amino acid level the replacement is mostly by homologous amino acid which does not alter the overall structure and solubility of the protein[67]. In a class of zein genes, there are 6 to 9 repetitive blocks. This type of block structures are also found in wheat prolamin and gliadin[68]. It has been suggested that homology which extends over such diverse plants, both monocots and dicots[69], must be due to constrained protein structure necessary to fulfil the basic physiological role of seed proteins as a nitrogen source for developing seedlings. Multigene families of seed proteins indicate that probably they are evolved by a complicated series of gene duplication. It could be implied that some non-reciprocal recombinational events occurred in the evolution of these genes[65]. Disposing diverse seed protein gene variants can be advantageous for the plants in some cases. Shewry *et al.*[70] have proposed a hypothetical evolutionary scheme of multigene families based on extensive sequence comparison. Both gene duplication and DNA exchange might have played a role in the evolution of 2S albumin genes of Brassicaceae[71,72].

Synthesis, processing and deposition of seed storage proteins

Proteins comprise one of the major and important storage materials which accumulate in high amounts during the second stage of seed development, i.e. the mid-maturation stage. This stage is preceded by the zygote development stage and followed by the desiccation stage in which seeds undergo preparation for dormancy[73]. During the first phase, cell division within the zygote gives rise to a globular embryo which differentiates into the embryonic axis and the cotyledons. The fate of the endosperm and cotyledon at this stage is different in monocots and dicots[74]; in monocots the endosperm becomes the major storage tissue while in most dicots the endosperm has only a transient presence and the cotyledons fulfil the major storage function.

During the second phase, cell expansion takes place, the embryo enlarges and there is rapid synthesis and accumulation of food reserves. In the pea seeds, the protein is stored in the protein bodies whereas starch is stored in plastids of cotyledon cells. In general, there is a sudden increase in storage protein which occurs within a few days during this period. Monocot plant species have specialized storage tissue, the endosperm (triploid) or perisperm (diploid) where storage reserves are laid down. This is followed by the third stage which consists of developmental arrest and desiccation which prepares the seed for dormancy. The seed loses more than 90% of its water content, RNA and protein synthesis terminate and embryonic dormancy begins. Although metabolic processes and *de novo* protein synthesis drops dramatically during this stage, late embryogenesis associated (LEA) proteins, which may be stress proteins produced to protect the seed tissue against desiccation, are synthesized at this stage. Dormancy ends in seed germination when seed storage proteins are used up for the seedling development. The duration of different stages of seed development varies with the species and environmental factors[73,75].

Seed proteins are usually synthesized as precursors having a leader pre-peptide sequence and pro-sequence which contain the signal for transport, processing and targeting from the site of synthesis to the storage organelles. The signal sequence is instrumental in directing the nascent chain from the rough side of the ER to the ER lumen. It is now well accepted that all storage proteins are synthesized on the rough ER. After cotranslational cleavage of the signal peptide, the polypeptides are subjected to glycosylation in specific cases and protein disulphide isomerase-catalysed disulphide formation, folded with the help of chaperones and targeted to protein bodies. Protein bodies are formed either by budding-off of storage proteins from ER (as in the case of zein and rice prolamin) or from protein storage vacuoles via the golgi apparatus (as in the case of legumins and rice glutelin; Figure 2). A more detailed discussion on processing and deposition of storage protein is presented in a recent review by Muntz[76].

Most seed storage proteins undergo limited proteolytic processing after initial removal of the signal peptide. The exact subcellular site of this processing and the enzymes involved may vary for different seed proteins and have not been fully elucidated yet. Some excellent reports, however, have appeared in recent years[76a–78]. Muntz and his group made a series of interesting studies on the processing, disulphide bridge formation and targeting of 11S globulins[79–81]. From *in vivo* (in transgenic tobacco seeds) and *in vitro* transcription/translation experiments, they showed that disulphide bridge formation probably occurs co-translationally and is necessary before the processed chains can form multimeric structures[79,80]. Prolegumins are thought to be processed into alfa and beta chains in the protein storage vacuoles by a cysteine endopeptidase (cysteine in the active site) with Asn or Asp cleavage specificity[81,82]. The napin-like 2S albumins of rapeseed, mustard and Brazil nut show more complicated processing. After cotranslational cleavage of the signal peptide in the ER, they undergo limited proteolytic processing in the vacuole whereby a propeptide, a short linker peptide between the small and large subunits of the mature protein and a short C-terminal peptide are cleaved off[83,84]. Controversy exists regarding the processing enzymes and recognized amino acid sequence motifs. The propeptide cleavage sites have an Asn (or Asp) at or near the p1 position and the detached propeptides are degraded by aspartate

protease(s), previously indicated to be the processing enzymes[55,85]. Asparaginyl endopeptidase (with cysteine at the active site) has strict substrate specificity towards carboxy side end of asparagine residues. A *Vigna mungo* processing enzyme termed VmPE-1 of this type is indicated to be involved in the vacuolar processing of yet another cysteine endopeptidase SH-EP which degrades seed storage proteins during germination[86].

From studies on 2S albumin of *Arabidopsis*, Krebbers *et al.*[87] indicated the occurrence of a Gly–Pro motif as a prerequisite for proteolytic processing, but subsequent sequence comparisons with *Brassica* napins and Brazil nut protein questioned the universality of this motif. It is speculated that some generally shared conformational features rather than amino acid sequences surrounding the cleavage sites are required for such processing[52]. Yamada *et al.*[88] recently reported an interesting case of the formation of multiple functional proteins by cleavage of Asn-Glu bonds of a single precursor by vacuolar processing enzymes in maturing pumpkin seeds. The processing of

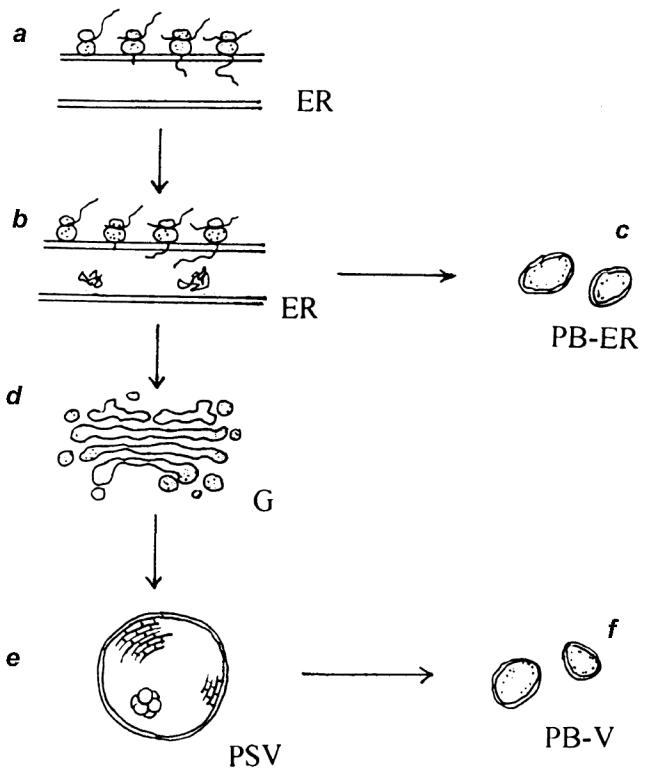


Figure 2. Storage protein processing and sorting. Glutelin and prolamin mRNA are translated at different areas of the rough ER: cER (*cis*-ternal R) and PB-ER (protein-body forming ER). ***a***, N-terminal signal peptide mediates their transport into the lumen of the ER. After cleavage of the signal peptide, further modifications like glycosylation, chaperone-aided folding, disulphide bridge formation and oligomerization take place before the protein leaves ER; ***b***, Prolamins of rice, maize, sorghum, coix, etc. also stored in protein body derived directly from ER (PB-ER); ***c***, Globulins of dicot (like legumins), glutelins and prolamins of some cereals are transported to the golgi apparatus G; ***d***, Further processing and cleavage of short N- and C-terminal targeting sequences takes place at G; ***e***, These get finally stored in protein bodies derived from vacuoles (PB-V); ***f***, see ref. 76.

proconcanavalin A (pro Con A) into Con A needs special mention here. Mature Con A from jack bean is a homotetrameric protein of unglycosylated 30.4 kDa subunits of 237 amino acids. After linker peptide deglycosylation, there occurs a circular permutation. In other words, after excision of internal 15 residues linker peptide, a head to tail transpeptidation (protein ligation) takes place which joins the N-terminal Ser of pro Con A via a peptide linkage to an Asn site, 9 residues upstream from the C-terminal of procon A. The C-terminal nona peptide is simultaneously detached[50].

Regulation of storage protein synthesis

The developing seed provides an attractive biological system for study of highly regulated gene cascade. Of the large number (about 20,000) of diverse genes expressed during embryogenesis, most encode rare mRNAs of unknown function. A small gene set, however, directs the synthesis of abundant mRNAs that encode seed storage proteins that are packed preferentially into protein bodies in the embryo, dicot cotyledons and monocot endosperms[73]. A large number of seed protein genes have been characterized at the molecular level[7,76,89]. These genes are excellent models for dissecting the different spatial, temporal and environmental regulatory processes. A number of reviews partly deal with some aspects of seed protein regulation[90,91].

Spatial regulation

There are several reports indicating qualitative as well as quantitative differences in the accumulation of seed storage proteins within different seed tissues. In cereals such as wheat, barley and maize, the major storage protein, prolamin, is deposited only in the starchy endosperm, whereas albumins and globulins tend to be largely concentrated in the embryo and aleurone tissue[92]. Although albumins and globulins are found throughout the legume seed, the outer cell layers of the abaxial cotyledon surfaces of the French bean and cowpea have been shown to contain higher levels of albumin protein[93]. In addition to storage proteins, several other proteins like lectins, proteinase/amylase inhibitors, LEA proteins are also synthesized in the embryo. Apart from possibly serving as storage proteins, these proteins have also been implicated as defence proteins against several biotic and abiotic stresses[94,95]. Such differential spatial (tissue-specific) accumulation of proteins may provide important clues to the function and regulation of these genes.

Developmental regulation

There are temporal controls of seed protein gene expression, as shown by numerous reports on a variety of

seeds[73,75,89]. This has been evident by detailed analysis of seed protein accumulation, measuring mRNA levels by cell-free translation and by nucleic acid hybridization using labelled cDNA sequences specific for each mRNA or a family of mRNA. These studies have revealed that each storage protein mRNA has a characteristic period during which it reaches a maximum level and then declines, although they all overlap[88]. For example, vicilin synthesis precedes legumin synthesis in *Vicia faba*[96]. In general, the proportion of low molecular weight proteins decreases along with concomitant increase in high molecular weight proteins as the seed matures. The low molecular weight proteins have a better amino acid composition. As a result, green seeds are nutritionally better than ripe ones[97]. Presumably, genes in each set respond to distinct regulatory signals, but the actual molecular interaction and mechanism of regulation could be effected at two levels. The short range control involves the sequences immediately upstream and downstream from 5′ and 3′ ends of the genes. The long range controls, although not identified like the former, may involve remote DNA sequences, genes or other structural features within chromosomes[98,99]. Such understanding is essential for practical application of genetic engineering tools for the improvement of protein quality and/or quantity.

Genetic regulation

Studies on the rates of accumulation and final levels of storage proteins in different genotypes and mutants of a plant provided clear evidence of genetic regulation of these proteins. In some cases, the defect is due to a mutation in the seed protein gene itself[73,99–101]; while in others, the defect is due to a mutation in unlinked regulatory sequences functioning as transacting loci[102]. The seed protein gene mutations are caused by several different events including deletion of an entire locus, insertional inactivation, point mutations that alter the protein reading frame, deletion of a 5′ flanking region and chromosomal rearrangement[103–105]. The mutations either eliminate or greatly reduce the levels of seed proteins or their mRNAs. For example, the mutant soybean lectin *le-1* gene is not transcribed detectably as a result of an insertion element present in the gene itself[106]. In maize, the opaque-2 mutation causes a reduction in the transcription of many unlinked zein storage protein genes[103]. Pysh *et al.*[104] demonstrated that *opaque-2* gene encodes a polypeptide belonging to the basic domain/leucine zipper class of transcription regulatory proteins and controls the expression of the 22 kDa zein proteins by binding to its promoter region. A defect in the seed protein gene may also be reflected in the final level of protein through its effect on the post-translational processing and targeting. Cotranslational cleavage of the signal prepeptide and removal of the prosequence of seed protein in the lumen of ER are necessary for correct processing, folding and targeting to vacuolar protein bodies. For example, fleury 2 mutant of maize having a mutant alfa-zein protein deficient in signal peptide cleavage has a much reduced level of alfa-zein and defective protein bodies[107]. Different regions of the N-terminal and C-terminal and even internal seed protein sequences have been shown to be involved in their correct targeting and deposition[76].

Environmental regulation

Environmental factors such as temperature, nutrients and hormones may alter seed protein composition and quantity. Seed protein composition is altered by temperature in wheat and pea[108,109]. Nutrient deficiency particularly of sulphur, selectively and markedly depresses sulphur-rich proteins compared to other seed proteins in rape, legumes and cereals[7,89,110,111]. A high osmotic pressure in tissue culture medium initiates legumin synthesis in cultured immature pea cotyledons[112] and 12S globulin in rapeseed[113]. Abscisic acid has also been shown to stimulate storage protein synthesis in cultured cotyledons of soybean and rapeseed[114,115]. It is assumed that the same conditions may regulate storage protein *in vivo* also.

Regulatory sequences in seed storage protein genes

The spatial and temporal controls of seed protein gene expression are exerted at the transcriptional level. They act like switch processes which may be turned 'on' and 'off' in the right tissue at the right time. Both *cis*-acting elements at the 5′ flanking regions of the protein genes and *trans*-acting binding protein factors are involved in this switching mechanism. Some conserved nucleotide sequences have been found in many well-characterized storage protein genes. Apart from 'TATA' and 'CAAT' boxes, some important regulatory elements have been identified by gel mobility shift on protein binding, and expression of a reporter gene product like CAT (chloramphenicol acetyl transferase) or GUS (beta-glucuronidase) linked to deleted promoters in transgenic plants. In the 11S legumin-like genes from pea, soybean, broad bean (*Vicia faba*) and sunflower, a conserved sequence of approximately 28 bp called the 'legumin box' (5′-TCCATAGCCATGCATGCTGAAGAATGTC-3′) is located within 250 bp of the transcription start site (TCS)[93,116]. The legumin box sequence has been implicated in the regulation of expression of alfa-subunit of beta-conglycinin in soybean at the transcriptional level[117]. Another sequence termed the 'Vicilin box' (5′-GCCCACCTCA/TTTTC/TGTTC/TAC/TTTCAACACNC-GTCAANNTNC AT-3′) located within 150 bp of the TCS in the 5′ region has been found in 7S protein genes in pea, soybean and *Phaseolus vulgaris* [93].

In the cereal prolamin genes of barley, wheat and maize, a sequence 300 bp upstream of TCS, termed the '-300 box' has been found[70]. The ATGAGTCA motif in the -300 element of rice glutelin gene has been found to be involved in nuclear protein binding[118]. The low mol. wt. glutelin gene *lmwg-1D1* promoter in wheat contains the 'endosperm box', a sequence motif highly conserved in the promoter region of a large number of monocot storage protein genes and is thought to confer endosperm-specific expression of prolamin genes[119].

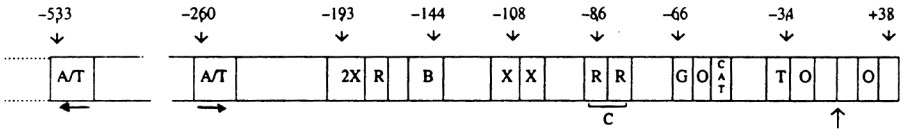
A class of elements of alternating purine–pyrimidine sequence called the RY repeats (5′-CATGCATGT-3′) is located in multiple copies in most legume seed protein genes and also in the promoter of 2S protein from *Brassica* species. The RY element plays a key role in tissue-specific and temporal expression of beta-conglycinin gene in coordination with other *cis*-acting elements[120,121]. Mutant promoter and deletion experiments suggest that RY element involves negative regulation in seeds[122]. A sequence motif CACGTG called 'G-box' was identified as a highly conserved protein-binding site upstream of many genes including a small subunit gene of ribulose-*bis*-phosphate carboxylase[123], *Brassica* 2S protein gene[120] and many other genes[124–126a]. Many positive and negative *cis*-regulatory elements were detected in 2S napin promoter of *B. juncea*[127] and *B. napus*[128] (Figure 3). Though these elements are recognized by transacting transcription factors both in transgenic tobacco and rape, some differences in tissue specificity of expression in these two plants were noted[129].

Expression of seed protein genes in transgenic plants

Though conventional breeding techniques will continue to be extensively employed, genetic engineering of crop plants becomes the only option when specific, modified genes are to be transferred between widely related species. The first transgenic plant was reported in 1983. Since then several reviews have been published on the introduction of heterologous/modified seed protein genes of agronomic importance[130–132].

There are several methods for transferring genes of interest into plants (Table 2). Of these methods, several techniques like macroinjection into floral stems or developing ovaries and direct uptake through pollen or germinating seed are not very reproducible. Particle bombardment is initially cost intensive. But the advantage is that it needs less trained manpower, i.e. molecular biologists, as technicians can try this method on a number of plants and tissues. On the other hand, *Agrobacterium* Ti (or Ri) vector-mediated gene delivery is more popular due to its being easy, reproducible and less expensive. The only disadvantage is that it is still generally applicable to dicot plants, although there are reports of successful *Agrobacterium* transformation of cereals using acetosyringone[133].

Plant transformation with numerous genes clearly established that dicot genes including seed protein genes, can be expressed and correctly processed in monocot plants and vice versa[134,135]. Thus, all the transcription/translation machinery, regulatory sequences, transacting factors and post-translational processing, modifying and targeting enzymes are homologous or similar in all plants. However, due to codon usage preference the level of expression may be different in different host plants and introduced gene combination. A number of seed proteins have been tissue-specifically expressed and correctly processed in transgenic plants like beta-phaseolin in tobacco, beta-conglycinine in *Petunia*, *Arabidopsis* 2S albumin in tobacco and *Brassica*[89]. In the author's laboratory, *Brassica* 2S protein gene has been expressed in



O = GATGAGTTT ; GATGAGTTA ; GATCGCCATG (Opaque–2 like)
T = TATAAAT (TATA Box)
CAT = TGCAAATC (CAAT Box)
G = TACACGTGAT (G–Box like)
C = CACATGCGTGCATGCA : R = CATGC (Ry Repeat)
2X = TCACACGACACACAT ; TACACGTTACACAT (X=TACACAT like repeat)
B = CTTGTCACTCCCTTCAAACACC
A/T = AATTTTATG (A/T rich sequence)
with arrows indicating the orientation.

Figure 3. Schematic representation of positions of different generally conserved and seed-specific consensus nucleotide motifs of a typical seed protein promoter (*B. juncea* 2S protein). Most of the protein binding regions have been identified by gel retardation assay. Vertical arrow at +1 nucleotide indicates the putative transcription initiation site (From ref. 166).

transgenic tobacco seeds, under its own promoter and CaMV 35S promoter. For ease of transformation, the most common plants to which seed protein genes have been transferred are the model plants, tobacco and *Petunia*, for studying the regulatory elements in the promoter, after tagging with a reporter gene like chloramphenicol acetyl transferase (CAT)[89] or beta-glucuronidase (GUS)[135–137] or firefly luciferase (LUX)[135].

Attempts to improve the nutritional quality of seed proteins

There can be two broad approaches for improving the nutritional quality of seed proteins by rDNA techniques: (1) Gene alteration, in which a seed protein gene already present and expressed in the plant seed is isolated and its coding sequence is altered by addition, substitution or deletion of nucleotides by site-directed mutagenesis, followed by reintroduction of the modified gene under its own promoter into the plant itself; or (2) Gene addition, in which a new capability or quality is conferred to the seed by introduction of new heterologous genes encoding higher levels of desired amino acids, under the seed's own promoter or more efficient heterologous promoter. Again, in the second approach by using a heterologous gene, an ideal animal protein gene like that of a milk protein introduced into a plant is poorly expressed due to difference in codon preference between the plant and animal in question. Therefore, heterologous plant protein genes with desired amino acid composition have been the choice of plant biotechnologists. After the first transgenic approach to improve Met content of seed proteins by Saalbach *et al.*[138] in 1988, there has been a large flow of literature on this approach. Of several reviews published in recent years, some comprehensive ones are mentioned here[139–142]. The promising strategy is to introduce genes encoding sulphur-rich and lysine-rich proteins under the control of an efficient promoter into legumes and cereals, respectively. The following illustrates some attempts to improve major crops.

Table 2. Methods of gene transfer to plants

Method	Nature of gene transferred
Agrobacterium tumefaciens *A. rhizogenes* (a) Vector-mediated (b) Virus-mediated (Agroinfection)	Any gene or DNA fragment between T-DNA borders
Somatic hybridization	Entire genome by nuclear fusion (hybrid) or cytoplasmic fusion (cybrid)
Direct transfer to protoplast (a) Electroporation (b) PEG-mediated (c) Liposome fusion (d) Microinjection	Segment of DNA or plasmid
Biolistic (particle gun)	Segment of DNA or plasmid
Other methods (a) Whole seed/seedling uptake (b) Vacuum infiltration (c) Macroinjection to floral tillers or ovary (d) Delivery through stigma/pollen	Segment of DNA or plasmid

T-DNA, Tumour-inducing DNA; PEG, Polyethylene glycol.

Improvement of legumes

An ideal candidate for enhancement of methionine content of legume seeds by this strategy is the seed-specific 2S albumin gene of Brazil nut (*Bertholletia excelsa*). Approximately 18% of the codons of this gene code for methionine and 8% for cysteine. This 2S albumin gene under cauliflower mosaic virus (CaMV) 35S promoter was transferred to tobacco and the grain legume *Vicia narbonensis*[143]. The 2S albumin was expressed in all parts of the plant both in tobacco and the legume and was found to be localized in the vacuoles of the leaf mesophyll cells of transgenic tobacco. In another attempt the 2S albumin gene, under a broad bean (*V. faba*) legumin B4 promoter for seed tissue-specific expression was transferred to *V. narbonensis*. This showed that up to 3% of soluble seed protein was 2S albumin. Some transformants had 3-fold increase in methionine content of the salt-soluble protein fraction in seed[139]. To improve the 2S albumin protein of Brazil nut which is sulphur-rich but lacks the essential amino acid tryptophan, residues of tryptophan were inserted at sites without altering the proper protein structure. Three constructs, one with 5 consecutive tryptophan codons and the other two modified genes encoding proteins carrying single tryptophan residues under 35S CaMV promoter were used and all three genes were correctly translated[144]. Several patents (e.g. Johnson, I. A., World patent no. 9741239, 1997; Wandelt, C. J., US patent no. 55633436, 1997) claim the enhancement of methionine content of soybean, alfalfa and lotus by the introduction of rice prolamin, sunflower sulphur-rich 2S albumin and maize zein genes. In another attempt, a maize gamma-zein gene encoding a sulphur amino acid-rich protein was used to transform alfalfa and trefoil (*Lotus corniculatus*) under CaMV 35S promoter and RUBISCO small subunit promoter. Expression level was rather low to the extent of 0.05% of alcohol soluble protein[145]. In yet another attempt, gamma-zein gene was modified by the introduction of $(\text{pro-lys})_n$ coding sequences at different positions. High levels of protein expression were observed when $(\text{pro-lys})_n$ sequence was contiguous or in substitution of pro-X region of the gamma-zein[146]. Muntz *et al.*[147] summarized the results of genetic engineering for high methionine grain legumes. The foreign protein

amounted to 5–10% of the seed proteins in the best transgenic lines. Transgenic soybean reached 100% and narbon bean and lupin reached 80% of the WHO standard for nutritionally balanced proteins. These results document that the Met problem of grain legumes can be solved by genetic engineering.

S-rich 2S albumin gene from sunflower has been introduced into the forage crop clover (*Trifolium subterraneum*) under CaMV 35S promoter along with a sequence encoding ER-retention-signal added to the 3′-end of the coding region. This prevented the translocation of the expressed protein to the vacuole. Under this condition, the protein was expressed to a maximum level of 0.3% of the total extractable protein in mature leaves[148,149].

Improvement of cereals

Most of the early work on maize and barley improvement was based on the characterization of 'high lysine' and 'Highproly' lines identified by screening of several thousand collections from wild germplasm. This Highproly mutant has reduced proportion of prolamin (which is deficient in lysine) with an accompanying increase in other more lysine-rich proteins. However, this mutation is highly linked to an agronomically harmful trait, namely, reduced starch content, thereby decreasing grain yield drastically. Attempts to disrupt this linkage by classical breeding did not seem promising[150]. An alternate approach is to introduce a gene for proteins having nutritionally enriched amino-acid composition under strong endosperm-specific promoter like that for high molecular weight glutelin. For example, beta-phaseolin (6% lysine) gene of common bean *Phaseolus vulgaris* under rice glutelin promoter was expressed to the level of 4% of total endosperm protein in transgenic rice seeds[151].

Mutant or less abundant seed storage protein zein from maize has been frequently used for quality improvement. To increase methionine level, a new methionine-rich zein, normally expressed at low levels was expressed at a high level using the 27 kDa zein promoter. This protein called the high sulphur zein (HS 7) was 21 kDa and contained 37% Met. The level of expression in the transgenic plant was not mentioned[152]. Soybean ferritin gene under rice glu-B-1 promoter has been introduced into rice to increase the iron content[153]. The glutelins together with gliadins are the most abundant proteins in cereal seeds. The high molecular weight glutelin subunit is important in determining the bread-making quality of wheat flour. To change the amount and composition of these proteins, a gene encoding a novel hybrid subunit under the control of native HMW-GS regulatory sequence was inserted into wheat. The hybrid subunit was accumulated to levels comparable to that of native HMW-GS[154]. These studies demonstrate the feasibility of constructing wheat and other cereals with altered levels and types of seed storage proteins.

Improvement of crucifers and other seeds

The small mol. wt. albumin (2S protein or napin) of rapeseeds/mustard is one of the most important targets for the improvement of its nutritional quality and also for 'Molecular Pharming', i.e. for producing medically important proteins in transgenic plants. There have been several promising attempts for the genetic engineering of 2S albumins. The 5′-upstream sequences of *Arabidopsis*[155], *B. napus*[156], and *B. juncea*[136] 2S protein confer seed specificity of expression of an introduced gene. These studies demonstrated that 2S protein is correctly expressed and targeted even in a heterologous host. Brazil nut 2S

Table 3. Some promising transgenic approaches for improving nutritional quality of seed proteins

Gene/source	Promoter/modification	Transformed plant	Improvement	Ref.
2S protein (Brazil nut)	CaMV 35S promoter	*V. narborensis*	Met	143
2S protein (Brazil nut)	*V. faba leg*B4 promoter	*V. narborensis*	Met	139
2S protein (Brazil nut)	CaMV 35S promoter, addl. Trp codons	Tobacco	Met	144
2S protein (Brazil nut)	Phaseolin promoter	Rapeseed	Met	158
S-rich γ-zein (maize)	CaMV 35S promoter	Alfalfa, trefoil	S-content	145
2S protein (sunflower)	CaMV 35S promoter, ER-retention signal	Clover	S-content rumen-protected	148, 149
2S protein (sunflower)	CaMV 35S promoter	Legumes	S-content	Prakash, C. S. (unpublished)
Different genes	Different promoters	Grain legumes	Met, Lys	141, 142
Beta-phaseolin (common bean)	Rice *glu* promoter	Rice	Lys	151
HS-7 zein (maize)	27 kDa zein promoter	Rice	Met	152
Ferritin (soybean)	Rice glu-B1 promoter	Rice	Fe-content	153
Ama 1 (amaranth)	Ama-1 promoter, Rice glu promoter	Rice, chickpea	Amino acids	Dutta, A. (unpublished)
Cruciferin (rapeseed)	Napin promoter, *cruA* antisense gene	Rapeseed	Met, Lys, Cys	159

protein gene[157] and a methionine-enriched modified napin were introduced into *B. napus* and were expressed to marginally improve the methionine content of total seed protein. Altenbach *et al.*[158] reported the highest level of expression of Brazil nut protein in transgenic *B. napus* ranging from 1.7% to 4% of total seed protein. This 4% methionine-rich protein contributed about 33% more methionine in the total seed protein. In all these studies, heterologous promoters like that of phaseoline, *Arabidopsis* AT2S1 and soybean lectin were used. The different levels of expression might be due to different promoter efficacy. Studies on *B. juncea* 2S promoter dissection indicated that a suitably modified (to exclude negative regulatory regions and by duplication of enhancer sequence) homologous promoter is more efficient in the expression of a reporter gene[95,128]. Also, the 3′-terminator sequences used have always been from a heterologous source. A homologous 3′-sequence might be better for expression of foreign/modified gene.

In another attempt, an indirect strategy was used in which the antisense gene for cruciferin was introduced into *B. napus*, which resulted in the suppression of cruciferin and increase in the relative proportion of napin, a more balanced protein. The antisense construct contained a part of *cruA* gene in reversed orientation, driven by 5′-flanking sequence of napin gene, such that the antisense cruciferin mRNA was expressed in a seed-specific manner[159]. The transgenic T1 generation seeds contained 10%, 8% and 32% higher Lys, Met and Cys, respectively, compared to their respective levels in control seeds. A summary of successful biotechnological approaches for improving nutritional quality is presented in Table 3. However, any transgenic seed with improved nutritional quality has not been commercialized yet.

Conclusion and future prospects

In the foregoing discussion, we have attempted to present an overview of the present status of seed protein quality improvement by genetic engineering. The basic understanding of regulation of seed protein gene expression and partial success in improving the nutritional quality make us optimistic about commercialization of some improved seed crops. Problems, however, remain with the low level of expression of the introduced gene, the existence of multiple copies of residing seed protein genes, and position effect on the insertion, including gene silencing[160,161]. But the last problem is not insurmountable since once a large number of stable transformed lines of a crop with different levels of expression of the desired gene are obtained, those with maximum level of expression and having the desired agronomic traits will be selected for commercialization. The Brazil nut 2S protein with its high methionine content has the drawback of being allergenic[162]. *Brassica* 2S protein is better in that sense. In this case, although allergenicity is reported[163], the issue is debatable[164].

Most of the genes used for crop improvement are isolated abroad and protected by patents. India has a vast resource of genetic diversity of minor cereals and legumes with better amino acid composition. Research is in progress for isolation of several such proteins and their genes. The seed protein genes of such economically minor plants like *Amaranthus*[19,164a], *Chenopodium*[20] and chickpea[59a] isolated in India will be prospective candidate genes for seed protein quality improvement. These genes can be introgressed into major crop plants through transgenic production followed by classical breeding for transfer to regionally established cultivars.

In the near future, high expression cassettes containing homologous 5′- and 3′-flanking sequences will be available for specific crops and/or for general crop plant transformation. Of the gene delivery techniques, the biolistic method is proving more and more promising for its species-independent general applicability, particularly for cereals. But even after all the technical problems of efficient expression of stably introduced gene with desired seed protein quality are solved, the commercial exploitation of such transgenic crops will depend largely on consumer acceptance. It is expected that the present public controversy and movement by green lobbyists against genetically modified crops (and all genetically modified organisms)[165] will fade out in time due to rigorous scientific evaluation and informed public opinion. In the present pace of research both by academic institutions and agribiotech industries, the benefits of transgenic grain crops with improved seed protein quality for food and feed are expected to be realized by the first decade of the 21st century.

1. Report No. 24, FAO, United Nations, 1970.
2. Technical Report Series, No. 522, World Health Organization, 1973.
3. Millerd, A., *Annu. Rev. Plant Physiol.*, 1975, **26**, 56–72.
4. Fukushima, D., *Food Rev. Int.*, 1991, **7**, 323–351.
5. Osborne, T. B., *The Vegetable Proteins*, Longmans Green, London, 1924.
6. Derbyshire, E., Wright, D. J. and Boulter, D., *Phytochemistry*, 1976, **15**, 3–24.
7. Prakash, V. and Narasinga Rao, M. S., *CRC Crit. Rev. Biochem.*, 1986, **20**, 265–363.
8. Righetti, P. G. and Gianazza, E., *Planta*, 1977, **136**, 115–123.
9. Shewry, P. R. and Casey, R., *Seed Proteins*, Chapman and Hall, London, 1998.
10. Gorinstein, S., Zemser, M., Fliess, A., Shnitman, I. and Paredes-Lopez, O., Kobayashi, S. and Taniguchi, H., *Biosci. Biotechnol Biochem.*, 1998, **62**, 1845–1851.
11. Higgins, T. J. V., *Annu. Rev. Plant Physiol.*, 1984, **35**, 191–221.
12. Larkins, B. A., *Genetic Engineering of Plants: An Agricultural Perspective*, Plenum Press, New York, 1983, pp. 93–118.
13. Youle, R. J. and Huang, A. H. C., *Plant Physiol.*, 1978, **61**, 13–16.
14. Murray, D. R., *Plant Cell. Environ.*, 1979, **2**, 221–226.
15. Manickam, A. and Carlier, A. R., *Planta*, 1980, **149**, 234–240.
16. Crouch, M. L., Tenbarge, K. M., Simon, A. E. and Ferl, R., *J. Mol. Appl. Genet.*, 1983, **2**, 273–283.

17. Ericson, M. L., Rodin, J., Lenman, M., Glimelus, K., Josefsson, L. G. and Rask, L., *J. Biol. Chem.*, 1986, **261**, 14576–14581.
18. Dasgupta, S. and Mandal, R. K., *Biochem. Int.*, 1991, **25**, 409–417.
19. Raina, A. and Dutta, A., *Proc. Natl. Acad. Sci. USA*, 1992, 89, 11774–11778.
20. Dey, N. and Mandal, R. K., *Biochem. Mol. Biol. Intl.*, 1993, **30**, 149–157.
21. Coulter, K. M. and Bewley, J. D., *J. Expt. Bot.*, 1990, **41**, 1541–1547.
22. Youle, R. J. and Hung, H. C., *Am. J. Bot.*, 1991, **68**, 44–48.
23. Singh, U., Rahu, S. M. and Jambunathan, S. M., *J. Food Sci. Technol.*, 1981, **18**, 86–88.
24. Krishna, T. G. and Bhatia, C. R., *Phytochemistry*, 1985, **24**, 2201–2203.
25. Uppal, K. and Matta, N. K., *Plant Physiol. Biochem.*, 1996, **23**, 27–32.
26. Dhillon, S. and Nainawatee, H. S., *Plant Physiol. Biochem.*, 1990, 17, 13–18.
27. Murty, B. S., Prakash, V. and Rao, M. S., *Indian J. Biochem. Biophys.*, 1990, **27**, 295–299.
28. Srinivas, H. and Rao, M. S. N., *J. Agric. Food Chem.*, 1987, **35**, 12–14.
29. Gatehouse, J. A., Croy, R. R. D., Marton, H., Tyler, M. and Boulter, D., *Eur. J. Biochem.*, 1981, **118**, 627–633.
30. Maltta, N., Gatehouse, J. A. and Boulter, D., *J. Exp. Bot.*, 1981, **32**, 183–197.
31. Payne, P. I., Holt, I. M., Jackson, E. A. and Law, C. N., *Philos. Trans. R. Soc. London*, 1984, **304**, 359–371.
32. Van der Klei, H., Van Damme, J., Casteels, P. and Krebbers, E., *Plant Physiol.*, 1993, **101**, 1415–1416.
33. Da Silva, J. G. *et al.*, *Arch. Biochem. Biophys.*, 1996, **336**, 10–18.
34. Kortt, A. A., Caldwell, J. B., Lilley, G. G. and Higgins, T. J., *Eur. J. Biochem.*, 1991, **195**, 329–334.
35. Allen, R. D. *et al.*, *Mol. Gen. Genet.*, 1987, **210**, 211–218.
36. Rodin, J. and Rask, L., *Plant Sci.*, 1990, **70**, 57–63.
37. Hara, I., Ohmiya, M. and Matsubara, H., *Plant Cell Physiol.*, 1978, **19**, 237–247.
38. Walburg, G. and Larkins, B. A., *Plant Physiol.*, 1983, **72**, 161–165.
39. Brinegar, C. and Goundan, S., *J. Agric. Food Chem.*, 1993, **41**, 182–185.
40. Dey, N., Ph D thesis, Calcutta Univ., 1998.
41. Galau, G. A., Wang, H. Y-C. and Hughes, D. W., *Plant Physiol.*, 1992, **99**, 779–782.
41a. Casey, R., *Biochem. J.*, 1979, **177**, 509–520.
42. Radovic, R. S., Maksimovic., R. V., Brikljacic, M. J. and Savic, P. A., *J. Agric. Food Chem.*, 1999, **47**, 1667–1670.
43. Byezynska, A. and Barciszewski, J., *J. Plant Physiol.*, 1999, **154**, 417–425.
44. Boutlier, K., Hatlori, J., Baum, B. R. and Miki, B. L., *Biochem. Syst. Ecol.*, 1999, **27**, 223–234.
45. Grutter, M. G., Priestle, J. D., Rahuel, J., Grossenbacher, H., Bode, W., Hofsteenge, J. and Stone S. R., *EMBO J.*, 1990, **9**, 2361–2365.
46. Li-de-la-Sierra, I., Quillien, I., Flecker, D., Gueguen, J. and Brunie, S., *J. Mol. Biol.*, 1999, **285**, 1195–1207.
47. Bode, W. and Huber, R., *Eur. J. Biochem.*, 1992, **204**, 433–451.
48. De-Meester, P., Lloyd, L. F., Blow, D. M. and Onesti, S., *Acta Crystollogr.*, 1998, **54**, 589–597.
49. Dattagupta, J. K. *et al.*, *Proteins*, 1999, **35**, 321–331.
50. Bowles, D. J. and Pappin, D. J., *Trends Biochem. Sci.*, 1988, **13**, 60–64.
51. Sainani, M. N., Gupta, V. S., Mishra, V. K., Lachke, A. H., Ranjekar, P. K. and Pillay, D. T. N., *Phytochemistry*, 1993, **34**, 919–925.
52. Dasgupta, J., Dasgupta, S., Ghosh, S., Roy, B. and Mandal, R. K., *Indian J. Biochem. Biophys.*, 1995, **32**, 378–384.
53. Gehrig, P. M. and Biemann, K., *Peptide Res.*, 1996, **9**, 308–314.
54. Kryzaniak, A., Burova, T., Haertle, T. and Barciszewski, J., *Nahrung Food*, 1998, **42**, 201–204.
55. Muren, E., Ek, B., Bjork, I. and Rask, L., *Eur. J. Biochem.*, 1996, **242**, 214–219.
56. Monslava, R. I., Villalba, M., Lopez-Otin, C. and Rodriguez, R., *Biochem. Biophys. Acta*, 1991, **1078**, 265–272.
57. Pandya, M. J., Sessions, R. B., Williams, P. B., Dempsey, C. E., Tatham, A. S., Shewry, P. R. and Clarke, A. R., *Protein Eng.*, 1997, **10**, 31–31.
58. Strobl, S., Muehlhahn, P., Berstein, R., Wiltscheck, R., Maskos, K., Wunderlich, M., Huber, R., Glockshuber, R. and Holak, T. A., *Biochemistry*, 1995, **34**, 8281–8293.
59. Halford, N. G. *et al.*, *Theor. Appl. Genet.*, 1992, **83**, 337–378.
59a. Pandya, M. J. *et al.*, *Proteins*, 2000, **38**, 341–349.
59b. Tai, S. S. K. *et al.*, *J. Agric. Food. Chem.*, 1999, **47**, 4932–4938.
60. Ampe, C., Van Damme, D., DeCastro, L., Sampalo, M. J., Van Montagu, M. and Vandekerckhove, J., *Eur. J. Biochem.*, 1986, **159**, 597–604.
61. Sharief, F. S. and Li, S. S. L., *J. Biol. Chem.*, 1982, **257**, 14753–14759.
62. Odani, S., Koide, T. and Ono, T. C., *Biochem. J.*, 1983, **213**, 543–545.
63. Lilley, G. G. and Inglis, A. S., *FEBS Lett.*, 1986, **195**, 235–241.
64. Shewry, P. R., Autran, J. C. and Ninmo, C. C. L., *Nature*, 1980, 286, 520–522.
65. Kreis, M., Shewry, P. R., Forde, B. G., Forde, J. and Miflin, B. J., *Ox. Sur. Plant Mol. Cell Biol.*, 1985, **2**, 253–317.
66. Hager, K. P., Braun, H., Czihal, A., Muller, B. and Baumlein, H., *J. Mol. Evol.*, 1995, **41**, 457–466.
67. Spena, A., Viotti, A. and Pirrota, V., *EMBO J.*, 1982, **1**, 1589–1594.
68. Bartels, D. and Thompson, R. D., *Nucleic Acids Res.*, 1983, **11**, 2961–2977.
69. Gorinstein, S., Zemser, M., Fliess, A., Shnitman, I., Paredes-Lopez, O., Yamamoto, K., Koyabashi, S. and Taniguchi, H., *Biosci. Biotechnol. Biochem.*, 1998, **62**, 1845–1851.
70. Shewry, P. R., Napier, J. A. and Tatham, A. S., *Plant Cell*, 1995, **7**, 945–956.
71. Boutlier, K., Hattori, J., Baum, B. R. and Miki, B. L., *Biochem. System. Ecol.*, 1999, **27**, 223–234.
72. Bhattacharya, S. and Mandal, R. K., *Indian J. Biochem. Biophys.*, 1999 (in press).
73. Goldberg, R. B., Barker, S. J. and Perez-Grald, *Cell*, 1989, **56**, 149–160.
74. Barrat, D. H. P., Whitford, P. N., Cook, S. K., Butcher, G. and Wang, T. L., *J. Exp. Bot.*, 1989, **40**, 1009–1014.
75. Dasgupta, S. and Mandal, R. K., *Seed Sci. Technol.*, 1993, **21**, 291–299.
76. Muntz, K., *Plant Mol. Biol.*, 1998, **38**, 77–99.
76a. Robinson, D. G. and Hinz, G., *Seed Sci. Res.*, 1999, **9**, 276–283.
77. Larkins, B. A., Vasil, I. K. (eds), *Cellular and Molecular Biology of Plant Seed Development*, Kluwer Academic, Dordrecht, 1997.
78. Muren, E. and Rask, L., *Planta*, 1996, **200**, 373–379.
79. Jung, R., Nam, Y–W., Saalbach, I., Muntz, K. and Nielsen N. C., *Plant Cell*, 1997, **9**, 2037–2050.
80. Jung, R., Saalbach, G., Nielsen, N. C. and Muntz, K., *J. Exp. Bot.*, 1993, **44**, 343–349.
81. Jung, R. *et al.*, *Plant Cell.*, 1998, **10**, 343–357.
82. Shutov, A. D., Kakhovskaya, T. A., Bastrygina, A. S., Bulmaga, V. P., Horstmann, C. and Muntz, K., *Eur. J. Biochem.*, 1996, **241**, 221–228.
83. Hara-Nishimura, I., Kinoshita, T., Hiraiwa, N. and Nishimura, M., *J. Plant Physiol.*, 1998, **152**, 668–674.
84. Saalbach, G., Rosso, M. and Schuman, U., *Plant Physiol.*, 1996, **112**, 975–985.

85. D'Hondt, K., Bosch, D., VanDamme, I., Goethals, M., Vandekerckhove, J. and Krebbers, E., *J. Biol. Chem.*, 1993, **268**, 20884–20899.
86. Okamoto, T. and Minamikawa, T., *Plant Mol. Biol.*, 1999, **39**, 63–73.
87. Krebbers, E., Herdies, L., De Clereq., A., Seurinck, J., Leemans, J., Van Damme, J., Sengura, M., Gheysen, G., Van Montagu, M. and Vandekerckhove, J., *Plant Physiol.*, 1988, **87**, 859–866.
88. Yamada, K., Shimada, T., Kondo, M., Nishimura, M. and Hara-Nishimura, I., *J. Biol. Chem.*, 1999, **274**, 2563–2570.
89. Shirsat, A. H., in *Developmental Regulation of Plant Gene Expression* (ed. Grierson, D.), Blackie, London, 1991, pp. 153–181.
90. Shewry, P. R., *Biol. Rev.*, 1995, **70**, 375–426.
91. Veena Rao, B. N., Manjunath, N. H. and Prasad, T., *J. Plant Biol.*, 1999, **26**, 119–129.
92. Payne, P. I., in *Seed Proteins* (eds Daussant, I., Mosse, I. and Vaughan, J.), Academic Press, New York, 1983, pp. 223–253.
93. Gatehouse, J. A., Evans, I. M., Croy, R. R. D. and Boulter, D., *Philos. Trans. R. Soc. London, Ser. B*, 1986, **314**, 342–348.
94. Dure, L., Crouch, M., Harada, J., David Ho, T. H., Mundy, J., Quatrano, R., Thomas, T. and Sung, Z. R., *Plant Mol. Biol.*, 1989, **12**, 475–484.
95. Gatehouse, J. A., Hilder, V. A. and Gatehouse, A. M. R., in *Plant Genetic Engineering* (ed. Grierson, D.), Blackie, London, 1991, pp. 105–135.
96. Adler, K. and Muntz, K., *Planta*, 1983, **157**, 401–404.
97. Chatterjee, S. R., Pokhriyal, T. C. and Abrol, Y. P., *Plant Biochem. J.*, 1977, **4**, 62–68.
98. Shewry, P. R., *Biol. Rev.*, 1995, **70**, 375–426.
99. Orellana, J., Fernandez-Calvin, B., Vazquez, J. F. and Carillio, J. M., *Theor. Appl. Genet.*, 1993, **85**, 639–643.
100. Voelker, T. A., Staswick, P. and Chirspeels, M. J., *UCLA Symp. Mol. Cell. Biol.*, 1989, **92**, 133–141.
101. Nambara, E., Naito, S. and Mecourt, P., *Plant J.*, 1992, **2**, 425–441.
102. Larkins, B. A., Pederson, K., Marks, M. D. and Wilson, D. R., *Trends Biotechnol. Sci.*, 1984, **7**, 306–308.
103. Kodrzycki, R., Boston, R. S. and Larkins, B. A., *Plant Cell*, 1989, **1**, 105–114.
104. Pysh, L. D., Aukerman, M. J. and Schmidt, R. J., *Plant Cell*, 1992, **5**, 227–236.
105. Dolfini, S. F., Landoni, M., Tonneli, C., Bernard, L. and Viotti, A., *Dev. Genet.*, 1992, **13**, 264–276.
106. Goldberg, R. B., Hoschck, G. and Vodkin, L. O., *Cell*, 1983, **33**, 465–475.
107. Coleman, C. E., Lopes, M. A., Gillikein, J. W., Boston, R. S. and Larkins, B. A., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 6828–6831.
108. Kolderup, F., *J. Sci. Food Agric.*, 1975, **26**, 583–592.
109. Thompson, R. and Casey, R. (eds), *Perspectives for Peas and Lupins as Protein Crops*, Martinus Nijhoff, The Hague, 1983.
110. Fujiwara, T., Hirai, M. Y., Chino, M., Komeda, Y. and Naito, S., *Plant Physiol.*, 1992, **99**, 263–268.
111. Pack, N. C., Imsande, J., Shoemaker, R. C. and Shibles, R., *Crop Sci.*, 1997, **37**, 498–503.
112. Domoney, C., Davies, P. R. and Casey, R., *Planta*, 1980, **149**, 454–460.
113. Mason, H. S., Dewald, D. B., Creelman, R. A. and Mullet, J. E., *Plant Physiol.*, 1992, **98**, 859–867.
114. Bray, F. A. and Beachy, R. N., *Plant Physiol.*, 1985, **79**, 746–750.
115. Finkelstein, R. R. and Crouch, M. L., *Plant Physiol.*, 1986, **81**, 907–912.
116. Baumelein, H., Wobus, U., Pustell, J. and Kafatas, F. C., *Nucleic Acids. Res.*, 1986, **14**, 2707.
117. Chamberland, S., Daigle, N. and Bernier, F., *Plant Mol. Biol.*, 1992, **19**, 937–949.
118. Kim, S. Y. and Wu, R., *Nucleic Acids Res.*, 1990, **18**, 6845–6852.
119. Wu, CC. Y., Suzuki, A., Washida, H. and Takaiwa, F., *Plant J.*, 1998, **14**, 673–683.
120. Dasgupta, S., Dasgupta, J. and Mandal, R. K., *Gene*, 1993, **133**, 301–302.
121. Fujiwara, T. and Beachy, R. N., *Plant Mol. Biol.*, 1994, **24**, 261–272.
122. Iida, A., Nagasawa, A. and Oeda, K., *Plant Cell. Rep.*, 1995, **14**, 539–544.
123. Giu lianno, G., Pichersky, E., Malik, V. S., Timko, M. P., Scolnik, P. A. and Cashmore, A. R., *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 7089–7093.
124. Williams, M. E., Foster, R. and Chua, N-H., *Plant Cell*, 1992, **4**, 485–496.
125. Menkens, A. E. and Cashmore, A. R., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 2522–2526.
126. Kawagoe, Y., Campell, B. and Murai, M., *Plant J.*, 1994, **5**, 885–890.
126a. Jain, R. K. *et al.*, *Crop Sci.*, 1999, **39**, 1696–1701.
127. Mandal, R. K., Dasgupta, J., Dasgupta, S. and Ghosh, S., in *Agric. Biotechnol.* (eds Chopra, V. L., Sharma, R. P. and Swaminathan, M. S.), Oxford–IBH, New Delhi, 1996, pp. 181–191.
128. Stalberg, K., Ellerstrom, M., Sjodahl, S., Ezcurra, I., Wycliffe, P. and Rask, L., *Transgen. Res.*, 1998, **7**, 165–172.
129. Herrera–Estrella, L., Depicker, A., Van Montague, M. and Schell, J., *Nature*, 1983, **303**, 209–213.
130. Kriz, A. L. and Larkins, B. A., *HortScience*, 1991, **26**, 1036–1041.
131. Shewry, P. R., Tatham, A. S., Halford, N. G., Barker, J. H. A., Hannapel, U., Gallois, P., Thomas, M. and Kreis, M., *Transgen. Res.*, 1994, **3**, 3–12.
132. Yamaguchi, D. and Minamikawa, T., *J. Plant Res.*, 1998, **111**, 1–6.
133. Nayak, P., Basu, D., Das, S., Basu, A., Ghosh, D., Ramakrishnan, N. A., Ghosh, M. and Sen, S. K., *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 2111–2116.
134. Sindhu, A. S., Zheng, Z. W. and Murai, N., *Plant Sci.*, 1997, **130**, 189–196.
135. Yoshihara, T. and Takaiwa, F., *Plant Cell Physiol.*, 1996, **37**, 107–111.
136. Ghosh, S. K., Dasgupta, J., Maiti, I. B., Hunt, A. and Mandal, R. K., *J. Plant Biochem. Biotechnol.*, 1995, **4**, 1–4.
137. Van-der-Geast, A. H. M. and Hall, T. C., *Plant Mol. Biol.*, 1996, **32**, 579–588.
138. Saalbach, G., Jung, E., Saalbach, I. and Muntz, K., *Biochem. Physiol. Pflanz.*, 1988, **183**, 211–218.
139. Hefford, M. A., *Biotechnol. Genet. Eng. Rev.*, 1996, **14**, 191–210.
140. Marcone, M. F., *Food Res. Int.*, 1999, **32**, 79–92.
141. Muntz, K., Cristov, V., Jung, R., Saalbach, G., Saalbach, I., Waddal, D., Pickardt, T. and Scheider, O., in *Sulphur Metabolism in Higher Plants: Molecular, Ecophysiological and Nutritional Aspects* (eds Cram, W. J. *et al.*), Backhuys, Leiden, 1997, pp. 71–86.
142. Tabe, L. and Higgins, T. V. J., *Trends Plant Sci.*, 1998, **3**, 282–286.
143. Saalbach, I., Pickardt, T., Waddel, D. R., Hilmer, S., Schieder, D. and Muentz, K., *Euphytica*, 1995, **85**, 181–192.
144. Marcellino, L. H., Neshich, G., Grassi-de-Sa, M. F., Krebbers, E. and Gander, E. S., *FEBS Lett.*, 1996, **385**, 154–158.
145. Bellucci, M., Lazzari. B., Viotti, A. and Arcioni, S., *Plant Sci.*, 1997, **127**, 161–169.
146. Torrent, M., Alvarez, I., Geli, M. I., Dalcol, I. and Ludevid, D., *Plant Mol. Biol.*, 1997, **34**, 139–149.
147. Muntz, K. *et al.*, *Nehrung*, 1998, **42**, 125–127.
148. Rafiqul, M., Khan, I., Ceriotti, A. T. L., Aryan, A., Mc. Nabb,

W., Moore, A., Craig, S., Spencer, D. and Higgins, T. J. V., *Transgen. Res.*, 1996, **5**, 179–185.
149. Sharma, S. B., Hancock, K. R., Ealing, P. M. and White, DWR., *Mol. Breed.*, 1998, **4**, 435–448.
150. Munck, L., in *Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology* (ed. Shewry, P. R.), CAB Internatiional, Wallingford, 1992, pp. 573–601.
151. Zhenweiz, Z., Sumi, K., Tanaka, K. and Murai, N., *Plant Physiol.*, 1995, **109**, 777–786.
152. Falco, S. C., Chui, C. F., Ward, T., Schreiner, R., McAdams, S., Smith, P., O'Brien, D. P. and Jones, T. J., *In Vitro: Plant*, 1995, **31**, 223–224.
153. Goto, F., Yoshihama, T., Shigemoto, N., Toki, S. and Takaiwa, F., *Nature Biotechnol.*, 1999, **17**, 282–286.
154. Blechi, A. E. and Anderson, O. D., *Nature Biotechnol.*, 1996, **14**, 875–879.
155. Declerq, A., Vanderwiele, M., VanDamme, J., Guerche, P., VanMontague, M., Vanderkerckove, J. and Krebbers, E., *Plant Physiol.*, 1990, **94**, 970–979.
156. Radke, S. E., Andrews, B. M., Moloney, M. M., Crouch, M. L., Kridl, J. L. and Knuf, V. C., *Theor. Appl. Genet.*, 1988, **75**, 685–694.
157. Guerche, P., DeAlmeida, E. R. P., Schwarztein, M. A., Gander, E., Krebbers, E. and Pelletier, G., *Mol. Gen. Genet.* 1990, **221**, 306–314.
158. Altenbach, S. B., Kuo, C-C., Staraci, L. C., Pearson, K. W., Wainwright, C., Georgesen, A. and Townsend, J., *Plant Mol. Biol.*, 1992, **18**, 235–245.
159. Kohno-Murase, J., Murase, M., Ichikawa, H. and Imamura, J., *Theor. Appl. Genet.*, 1995, **91**, 627–631.
160. Grierson, D., Lycett, G. W. and Tucker, G. A., *Mechanisms and Applications of Gene Silencing*, Nottingham University Press, Nottingham, 1996.
161. Stam, M., Mol., J. N. M. and Kooter, J. M., *Ann. Bot.*, 1997, **79**, 3–12.
162. Organization for Economic Cooperation and Development, *Food Safety Evaluation*, OECD Publications, Washington, DC, 1996, pp. 1–180.
163. Menendez-Arias, L., Dominguez, J., Moneo, I. and Rodriguez, R., *Mol. Immunol.*, 1990, 27, 143–150.
164. Murphy, D. J., *Immunol. Today*, 1999, **20**, 511–514.
164a. Chakraborty, S., Chakraborty, N. and Dutta, A., *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 3724–3729.
165. Rifkin, G., *New Sci.*, 1998, **160**, 34–37.
166. Mandal, R. K. and Ghosh, S. K., in *Recent Advances in Oilsed Brassicas* (eds Kalia, H. R. and Gupta, S. K.), Kalyani Publishers, New Delhi, 1997, pp. 135–161.

Received 23 July 1999; revised accepted 4 July 2000