PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Differential Gene Expression in the Developing Barley Endosperm

M. Kreis, M. S. Williamson, J. Forde, D. Schmutz, J. Clark, B. Buxton, J. Pywell, C. Marris, J. Henderson, N. Harris, P. R. Shewry, B. G. Forde and B. J. Miflin

Phil. Trans. R. Soc. Lond. B 1986 **314**, 355-365
doi: 10.1098/rstb.1986.0058

Phil. Trans. R. Soc. Lond. B **314**, 355–365 (1986)
Printed in Great Britain

Differential gene expression in the developing barley endosperm

By M. Kreis[1], M. S. Williamson[1], J. Forde[1], D. Schmutz[1], J. Clark[1], B. Buxton[1], J. Pywell[1], C. Marris[1], J. Henderson[2], N. Harris[2], P. R. Shewry[1], B. G. Forde[1] and B. J. Miflin[1]

[1] *Biochemistry Department, Rothamsted Experimental Station, Harpenden, Hertfordshire AL5 2JQ, U.K.*

[2] *Department of Botany, University of Durham, Science Laboratories, South Road, Durham DH1 3LE, U.K.*

Barley prolamin storage proteins account for 50% of the seed proteins. They are encoded by small multigene families that are only expressed in the developing endosperm. Previous work has shown that the major prolamins in barley are characterized by the presence of two or more unrelated structural domains, one of which contains repeated sequences. The non-repetitive domain is homologous with sequences present in other seed proteins found in the seed of mono- and dicotyledonous plants. Comparison of the 5′ flanking region of a B1 hordein storage protein gene of barley with those of other prolamin genes (zeins and α-gliadins) reveals short sequences within 600 base pairs (bp) of the translation initiation codon that are strongly conserved. A short sequence at −300 bp seems to be unique to the prolamin genes and is possibly involved in the control of gene expression in the developing cereal endosperm. Six DNA-binding proteins have been identified that might recognize and interact with the putative regulatory sequences identified in the B1 hordein gene.

Protease inhibitors account for a large proportion of the salt-soluble proteins of the barley seed, and contain up to 10% lysine. Cloned cDNAs for chymotrypsin inhibitors 1 and 2 have been isolated and characterized. All contain ochre stop condons in the sequences encoding a putative signal peptide. The two inhibitors are encoded by small multigene families that specify several subfamilies of mRNAs. The accumulation of chymotrypsin inhibitors in normal and mutant endosperms of barley is related to the abundances of their mRNAs.

Introduction

l proteins were early subjects for study in the development of protein chemistry (see Osborne (1924) for review). Similarly, in the development of plant molecular biology, research on seed protein genes has formed a high proportion of the total effort. In part this has been due to the amount of expression of these genes but the research has also been driven by the desire to understand a system that has important implications for the nutritional quality of the seed as a food for non-ruminants. In the execution of this aim, interest has focused on the prolamins, which are low in lysine, on mutations that confer a 'high-lysine' phenotype, and on proteins that contain relatively large proportions of lysine ('high-lysine proteins'). Our research group at Rothamsted has studied all of these topics, particularly with respect to barley. The initial phase of this work has concentrated on the description of the system, firstly, at the level of proteins (see Shewry & Miflin (1985) for review), and secondly, at the level of the genes.

The current phase is concerned with the factors controlling the expression of these genes. In this paper we shall summarize briefly the findings from the descriptive phase and then discuss in more detail work on the control of gene expression.

Barley seed proteins

Classification of seed proteins (and their genes)

A number of classification schemes have been used, based on different attributes of protein chemistry. These have been reviewed extensively and the brief scheme that follows is taken from Shewry & Miflin (1985) and Shewry *et al.* (1986). Firstly, the seed proteins may be roughly divided into three groups on the basis of function, the metabolic (enzymes, etc.), the structural (membrane or ribosomal proteins) and storage (those with a primary or major function in storing N, C and/or S for subsequent use in germination). We are concerned here chiefly with this last group. Secondly, the storage proteins can be divided into the prolamins and non-prolamin groups.

Prolamins are proteins that as individual polypeptides are soluble in aqueous alcohol solutions and found only in the seeds of cereals (Gramineae). They are rich in proline and glutamine and poor in basic amino acids such as lysine. They are usually present as a polymorphic series of related polypeptides coded by a number of related genes clustered in complex loci. Their amino acid sequences show the presence of structural domains and repeated blocks of residues. They can be subdivided on the basis of their amino acid composition (particulary of S-amino acids), the chromosomal location of their structural loci, and their amino acid sequence. For convenience, the prolamins of the Triticeae are often divided into the following categories: the S-rich, S-poor and high molecular mass (HMM) prolamins.

Non-prolamin proteins. There are a number of families of proteins present in the salt soluble fraction of cereals, some of which have strong homology to proteins present in the glutelin fraction. We have therefore proposed to classify them together. Among this diverse group are the 7 *S* and 11 *S* globulins, a wide range of inhibitor proteins and β-amylase. Although many of these proteins have defined functions – for example, β-amylase and the protease inhibitors such as chymotryspin inhibitors 1 and 2 (CI-1 and 2) – they also appear to act as storage proteins.

The prolamins of barley have been termed the B, C and D hordeins. These groups correspond to S-rich, S-poor and HMM prolamins. These are encoded by three complex loci on chromosome 5 of barley (*Hor* 2, *Hor* 1 and *Hor* 3 respectively). The hordeins make up about 50% of the protein in the endosperm and, because they contain less than 1% lysine, cause the overall nutritional quality of the protein to be low. In contrast, several of the non-prolamin proteins are rich in lysine. The chymotrypsin inhibitors 1 and 2 have lysine contents of 9.5 and 11.5% respectively (Boisen *et al.* 1981). They are small proteins of M_r about 9500 and are encoded by two loci (*Ica* 1 and 2) on chromosome 5 (Hejgaard *et al.* 1984). They normally make up about 0.35% of the soluble seed protein; however, they behave like storage proteins in that their relative accumulation increases significantly in response to nitrogen (Giese & Hejgaard 1984) fertilizer and they are rapidly degraded upon germination. Furthermore, they are greatly increased (CI-2 20-fold and CI-1 6-fold) (Boisen *et al.* 1981) in the 'high-lysine' barley Hiproly, which carries recessive *lys* gene on chromosome 7.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Evolution of seed proteins

Recent evidence derived from sequence comparisons (Kreis *et al.* 1985*a*, *b*; Casey *et al.* 1986) has shown that the seed proteins can be considered as a series of superfamilies. Consideration of all of the prolamin sequences obtained so far has shown that prolamins are characterized by having multiple domains which are broadly of two types, one consisting of a series of repeated sequences rich in proline and glutamine and a second that is non-repetitive (Kreis *et al.* 1985*a*; Forde *et al.* 1985*b*). Although prolamins are found only in grasses and have thus been considered to be of recent origin (Kasarda 1980), comparison of the non-repetitive domains with the sequences of other seed proteins has shown that the prolamins of the Triticeae, the 2 *S* globulins of castor bean and rape, the CM proteins of barley and wheat, and some protease and α-amylase inhibitors, constitute a superfamily of proteins with limited sequence homology as defined by Dayhoff (1978). The minor 15 kDa zein (Pedersen *et al.* 1986) and a protein known as RSP or glutelin-2 (Prat *et al.* 1985), but not the major 22 and 19 kDa family of zeins, are members of this family.

The prolamins of the Triticeae are considered to have arisen after the evolution of large repetitive domains within an ancient ancestral gene. Three classes of repeat can be recognized; the first, which is present in the S-rich and S-poor prolamins, varies in precise sequence but appears to have arisen from a common origin. The other two, which are more highly conserved, are found interspersed in the HMM prolamins (Forde *et al.* 1985*c*; Thompson *et al.* 1985; Sugijama *et al.* 1985). The origin of the repeats is unknown. They may have arisen from the insertion of pre-existing repeats into a progenitor gene or by internal amplification of sequences already present. In either case the event must have occurred more than once to give the different repeats in the different genes. Besides their evolutionary interest, these repeats also confer interesting structural properties on the proteins (Tatham *et al.* 1985).

Similarly, the chymotrypsin inhibitors of barley are part of a superfamily of inhibitors present in other plants, notably potatoes and tomatoes, and in animals (Ryan 1981).

Differential expression of seed protein genes in the barley endosperm

Accumulation of storage proteins during seed development

Some information about gene expression can be obtained by the study of protein and mRNA accumulation. Results of such studies (Rahman *et al.* 1982; Giese & Andersen 1982; Miflin *et al.* 1983) show that neither hordein protein nor RNA are found in tissues other than the endosperm and indicate that the expression of the hordein gene is under strict tissue specific controls. The hordeins also appear to be developmentally regulated as they and their mRNAs accumulate some days after endosperm development begins (Rahman *et al.* 1983). There is also some indication that not all of the hordeins accumulate at the same rate, the C hordeins being more predominant at the earlier and the B1 hordeins at the later stages (Rahman *et al.* 1984). Their tissue-specific expression also appears to be modulated by plant nutrition, in particular a deficiency of sulphur relative to nitrogen severely decreases the amount of the S-rich B-hordeins and their mRNAs relative to that of the S-poor C-hordeins (Rahman *et al.* 1983; Shewry *et al.* 1983; Giese & Hopp 1984). The chymotrypsin inhibitors are not endosperm specific because they are synthesized in the embryo and also, transiently, in the young root tip (Kirsi & Mikola 1971). However, their synthesis in the endosperm is under developmental control and they

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

appear at a relatively late stage (Rasmussen 1985). β-Amylase is also expressed at several stages of development of the barley plant although the endosperm β-amylase behaves as a seed storage protein in that it accumulates at about the same time as the hordeins and is also differentially affected by nitrogen nutrition (Giese & Hopp 1984). Since there appear to be multiple genes for β-amylase (authors' unpublished results) it may be that these are under different controls and that one or more members are truly seed specific. Evidence for seed-specific β-amylases has been found in soybean where mutants lacking the seed enzyme still express it in other tissues (Hymowitz 1983).

The amounts of barley seed proteins and their mRNAs are differentially affected by a number of 'high-lysine' mutations (Kreis *et al.* 1983*a*, *b*, 1984). Two of these occur at genes separate from those for the proteins under consideration. They may therefore be expected to act by *trans*-acting factors. The *lys* gene originally found in the line Hiproly (Munck *et al.* 1970) on chromosome 7 causes a slight decrease in hordein accumulation of all hordein groups and a marked increase in chymotrypsin inhibitors and β-amylase. In contrast the *lys3a* gene from Risø mutant 1508, which is also on chromosome 7, causes major decreases in B and C but not D hordein (Kreis *et al.* 1984), an increase in chymotrypsin inhibitors and a 20-fold decrease of β-amylase.

In summary, there is thus considerable evidence of a range of differential controls acting to regulate gene expression in the barley endosperm. To find out more how these operate we have isolated copy DNA (cDNA) and genomic clones for these proteins and are studying various aspects of their organization.

Analysis of the 5′ flanking region of the B1 hordein gene

A 2.9 kb *Eco*RI fragment containing a B1 hordein gene has been isolated by Forde *et al.* (1985*a*). The complete nucleotide sequence was determined including the 5′ and 3′ flanking regions. The 5′ and 3′ flanking regions were compared with those of three other cereal prolamin genes. We assumed that genes specifying proteins with the same function and pattern of expression might have conserved *cis*-acting regulatory elements. Using a graphic matrix homology plot, Forde *et al.* (1985*a*) found that within 600 bp of the ATG initiator codon there were three conserved sequences with more than 80% homology to sequences found in α-gliadin (Rafalski *et al.* 1984), zein-19 (Langridge & Feix 1983) and zein-21 (Pedersen *et al.* 1982) storage protein genes (figures 1 and 2). These each had a short core sequence common to all of the genes. The conserved sequence at about position −100 includes the conventional 'TATA' box, whereas the 11 bp sequence at around −150 has the core sequence 'CATC'. The 'TATA' box, as in other eukaryotic genes, has been found in almost all plant genes so far sequenced. The conserved sequence present at −150 has some of the characteristics of a 'CAAT' box but has been termed a 'CATC' box by virtue of the most strongly conserved tetranucleotide sequence. This sequence has been found in a variety of other cereal genes with the exception of the maize ADH-2 gene (Denis *et al.* 1984) but was not found in a manual survey of 15 dicotyledonous plant genes scanned. The −300 sequence was about 33 bp long, within which 29 nucleotides were identical in the B-hordein and α-gliadin genes. Sequences homologous to this element were found only in cereal storage protein genes and not in any of several other genes from a range of plants and tissues for which extensive upstream data was available (figure 2).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

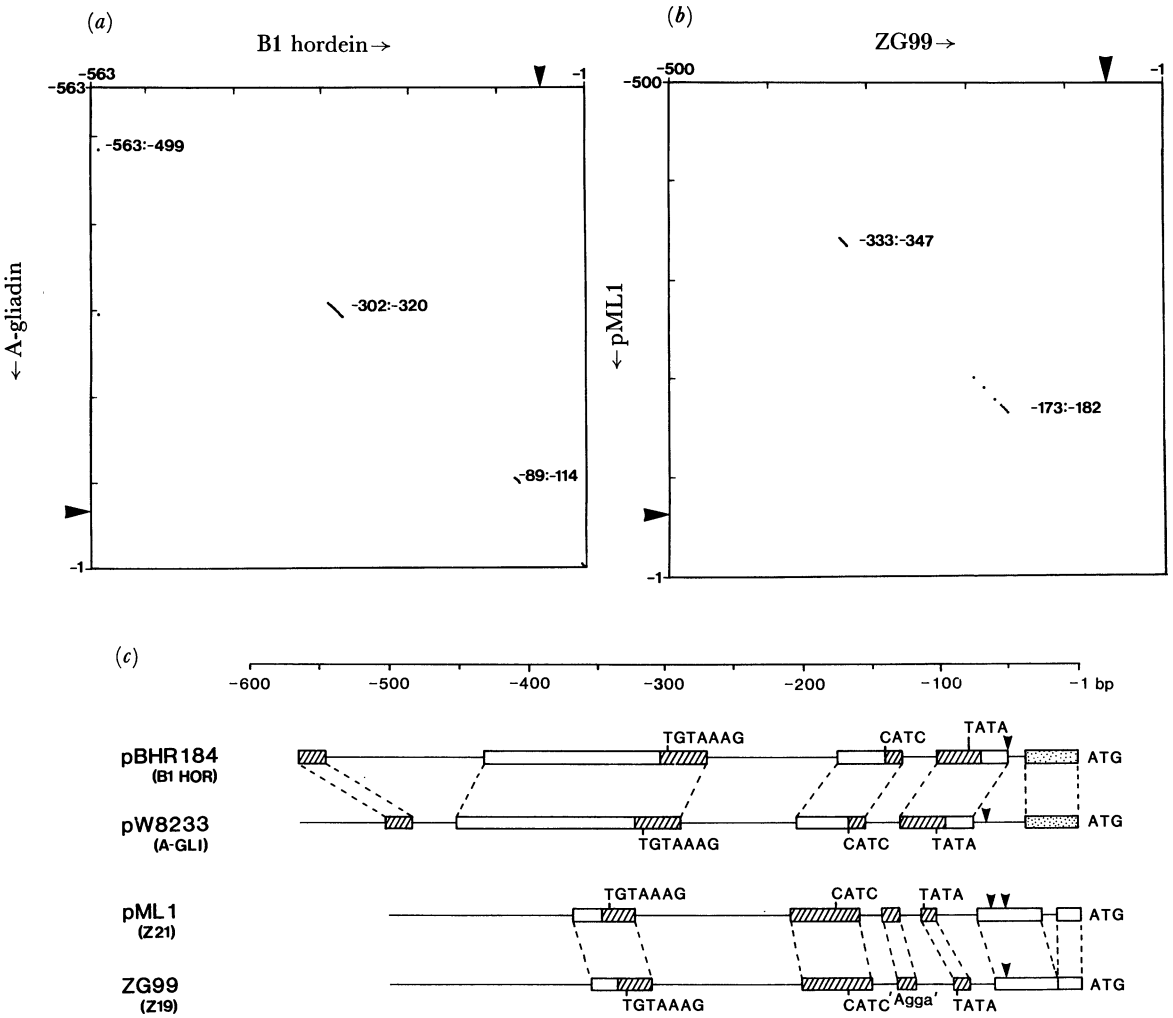

FIGURE 1. (*a*) DIAGON homology plot (Staden 1982) comparing the 5′-flanking regions of two S-rich prolamin genes: the B1 hordein gene (pBHR184) (Forde *et al.* 1985*a*) and a wheat A-gliadin gene (pW 8233 (Rafalski *et al.* 1984)). A span length of 17 was used and all blocks scoring 15 or more were plotted (double matching probability = 1.5×10^{-7}). (*b*) DIAGON plot comparing the upstream sequences of representatives of the two major families of zein genes in maize, pML1 (Langridge & Feix 1983), a Z21 gene, and ZG99 (Pedersen *et al.* 1982), a Z19 gene. All blocks scoring 14 or more in a span length of 17 were blotted (double matching probability = 0.68×10^{-7}). The arrows indicate the approximate locations of the RNA cap sites. The numbers on the diagonals refer to the coordinates of the first nucleotide in the adjacent region of homology, the location on the horizontal axis being given first. Numbering is from the first nucleotide upstream of the ATG codon. (*c*) Diagram showing the locations of the upstream sequences that are conserved in the S-rich prolamin genes (pBHR184 and pW8233) and in the zein genes pML1 and ZG99. The degree of homology is indicated by the shading within the rectangles: diagonal shading, 80–92% identity; stippled, 74% identity; open, 56–62% identity. Sequences outside the rectangles show little or no homology (less than 40%) (see Forde *et al.* 1985*a*).

We do not yet have any evidence as to whether these sequences have a function or what that function might be. However, it has been shown in other systems that conserved upstream regulatory systems do exist and are important for coordinating the induction of gene expression. Such sequences have been identified, mainly through generation of deletions or *in vitro* mutagenesis of the 5′ flanking region followed by testing of the expression of the altered genes by reintroduction into either an *in vitro* or an *in vivo* system. Subsequently an analysis of the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

A

Species	Clone	Gene Product	'CATC' Box	'TATA' Box	Cap Site(s)
Barley	pBHR184	B1 Hordein	-139 ACATCCAAACA.....-80	CTATAAATA.....	(-51)
Wheat	pW8233	A-Gliadin	-165 GCATCCAAGCA.....-105	CTATAAATA	(-67)
	gamb17	γ-Gliadin	-163 GCATTGAATCA.....-106	CTATAAAAA	(?)
	C11	HMW subunit	-160 ACATTTCTTCT.....-93	CTATAAAAG	(?)
	pTHO12	H3 Histone	-201 CCATCTCGACC.....-97	CTATTTAAC	(-62)
	pWS4.3	Rubisco SSu	-113 CCATCCCAACC.....-79	CTATATATA	(-42)
Maize	pML1	21kD Zein	-177 CCATCTATACC.....-110	GTATAAATA	(-65,-52)
	ZG99:	19kD Zein	-168 TCATCTCTACC.....-90	GTATAAGCA	(-57)
	p1S.1:	Alcohol Dehyd'ase	-167 CCATCTCTTCC.....-140	CTATATAAA	(-99)
	SS1	Sucrose Synthetase	-164 CCATCTGCACC.....-102	CTATTTATT	(-72)
		CONSENSUS:	CCATCTCNACC	CTATA$^{T}_{A}$ATA	

B

Clone	Gene Product	Position	Sequence	Score a	Score b
			a b		
pBHR184	B1 Hordein	-563	GAATTCGATGAGTCATG		-7.13
		-300	ACA..TGTAAA.GTGAATAAGGTGAGTCATG	-0.24	-7.42
pW8233	A-Gliadin	-510	ACA..TGTAAAAG.GAATTTGACGAGTCATG	-1.84	-9.62
		-318	AGC..TGTAAA.GTGAATAAGATGAGTCATG	-4.39	-5.63
gamb17	γ-Gliadin	-319	ACA..TGTAAA.GTGAATAAGATGAGTCAAT	-0.24	-8.40
pML1	21kD Zein	-347	ACATGTGTAAAGGTGAAG.CGATCATGCATG	-0.46	-8.63
Z7	21kD Zein	-323	ACATGTGTAAAGGTGAAG.AGATGATGCATG	-0.46	-6.83
ZG99	19kD Zein	-332	ACATGTGTAAAGGTATTGCATCACAC.CATT	-0.46	-15.43
ZE19	19kD Zein	-330	ACATGTGTAAAGGTATTGCATCACAC.TATT	-0.46	-17.51
		CONSENSUS:	ACA$^{::}_{TG}$TGTAAAGGTGAA$^{T}_{G}$NAGATGA$^{GT}_{TG}$CATG (a, b)		

FIGURE 2. Sequences common to the 5′-flanking regions of prolamin genes from barley, wheat and maize. Sequences that are common to the S-rich prolamins and the 19 kDa and 21 kDa zeins analysed in figure 1 have been aligned here for comparison. (A) Conserved sequences located within 200 bp upstream of the ATG codon. Similar sequences found at corresponding positions in a variety of other cereal genes are also shown: gamb17, wheat γ-gliadin (J. A. Rafalski, personal communication); C11, wheat HMW subunit (Thompson *et al.* 1985); pTHO12, wheat H3 histone gene (Tabata *et al.* 1984); pWS4.3: wheat Rubisco small subunit gene (Broglie *et al.* 1983); p1S.1, maize alcohol dehydrogenase (Dennis *et al.* 1984); SS1, maize sucrose synthetase gene (Weir *et al.* 1985). (B) Conserved sequences at around −300 that appear to be unique to the prolamin genes. The '−300 elements' from three S-rich prolamin genes and four zein genes are aligned for comparison (Z7 from Kridl *et al.* (1984)), and ZE19 from Spena *et al.* (1983)). The '−300 element' is divided into two segments, one of which is common to all occurrences of the element (*a*) and one that is absent or extemely divergent in the 19 kDa zein genes (*b*). The computer program GETFRQ (Staden 1984) was used to calculate a weight matrix for each set of sequences and then to derive a score indicating the degree to which each sequence fits the consensus. The scores confirm that segment (*b*) of the '−300 element' in the 21 kDa zeins is more closely related to segment (*b*) in the S-rich prolamins than to the corresponding sequence in the 19 kDa zeins.

transcriptional activity is made by measuring the accumulation of RNA or protein, and from this a deduction is made about the effect of the deletions or sequence changes on the expression of the gene. The chlorophyll *a*/*b* binding protein gene of wheat can be expressed in a heterologous system in a tissue and developmental specific manner (Lamppa *et al.* 1985); see Nagy *et al.*, this symposium). Generation of deletions of the 5′ flanking region of the gene showed that it contains *cis*-acting sequences that regulate the tissue and developmental specific expression.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Identification of DNA-binding proteins that recognize the 5′ flanking region of the B1 hordein gene

Purified eukaryotic polymerase II cannot initiate transcription selectively on defined promoters, but requires several distinct transcription factors which interact with the promoter and regulatory elements to direct gene expression (Dynan & Tjian 1985). The recognition and interaction of regulatory sequences by DNA binding proteins might be a fundamental process controlling gene activity. We are currently investigating whether there are specific interactions between the potential regulatory sequences indentified by Forde *et al.* (1985*a*) in the B hordein gene and DNA-binding proteins as reported for the *Drosophila* heat shock and alcohol dehydrogense genes (Heberlein *et al.* 1985; Dynan & Tjian 1985; Parker & Topol 1984). The existence of specific DNA-protein interactions in eukaryotes has been shown by different approaches. Miskimins *et al.* (1985) have recently described a procedure for the indentification of these proteins. This procedure requires the electroblotting of a sodium dodecyl sulphate polyacrylamide gel electrophoretically (SDS–PAGE) fractionated crude extract of nuclear proteins to a nitrocellulose filter, and incubation of the filters with ^{32}P-labelled DNA. J. Clark, D. Schmutz & M. Kreis (unpublished results) have used a similar technique to identify specific DNA–protein interactions in barley. A crude nuclear extract was prepared from transcriptionally active nuclei from barley shoots and endosperms by using a method modified from Willmitzer & Wagner (1981). The crude protein extract was prepared according to Siebenlist *et al.* (1984) and separated on a 13% SDS–PAGE and transferred electrophoretically to nitrocellulose. The filters were prehybridized in a 5% non-fat dry milk buffer (Johnson *et al.* 1984), which prevents any non-specific protein–DNA interactions. The filters were then incubated in binding buffer containing the ^{32}P-labelled probe derived from a 600 bp 5′ flanking region of the B-hordein gene. Six major protein bands (M_r 37000–200000) were detected with the nick-translated DNA fragment of 600 bp, containing promoter and possible upstream regulatory sequences. Some of these proteins were detected only in the endosperm nuclear extract. The labelled DNA fragment does not bind to a range of protein markers. Also, the six labelled protein band are not visualized upon staining of the gel, but a 600 bp coding region of the B hordein gene binds strongly to two proteins of 80 and 85 kDa respectively, suggesting that those are not involved in specific regulation. At present we do not know which sequences of the 5′ flanking region these proteins bind to and whether the protein–DNA interactions we observe *in vitro* have a physiological relevance.

Identification and characterization of chymotrypsin inhibitor (CI) cDNA clones

Poly(A)$^+$ mRNA derived from membrane-bound polysomes of Hiproly endosperms was used as a template for cDNA synthesis. A cDNA library was contructed in pUc9, following the Heidecker–Messing (1983) method. Several chymotrypsin inhibitor 2 (CI-2) and 1 (CI-1) related clones were indentified by using synthetic oligonucleotide probes. Restriction analysis and nucleotide sequence comparisons of about ten CI-2 cDNA inserts show that there are two subfamilies of mRNA, which are about 95% homologous. Southern blots show that there are from four to six CI-2 genes per haploid genome. The nucleotide sequence of one full-length cDNA is 450 bp long. Nucleotides 86–338 constitute an open reading frame that begins with an ATG and encodes a protein with a sequence identical to the amino acid sequence directly determined by Svendsen *et al.* (1980) for CI-2, except for one glutamine residue instead of

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

a glutamic acid. The deduced protein sequence would consist of 83 residues and have an M_r of 9120, which agrees well with the estimate from direct sequencing (Svendsen *et al.* 1980). It has been shown by Jonassen *et al.* (1981) and Williamson *et al.* (1986) that CI-2 is probably synthesized on membrane-bound polysomes. Recent *in situ* localization experiments by RNA–cDNA hybridization indicate that CT-2 mRNAs are localized on the rough endoplasmic reticulum (RER) of the barley endosperm (J. Henderson & N. Harris, unpublished results). Furthermore, immunogold labelling of barley endosperm tissue with antibodies to CI-2 shows that the protein is deposited mainly within membrane-bounded compartments, which is consistent with the above observations and with the localization of the chymotrypsin inhibitors in tomato and potato (Graham *et el.* 1985; Walker-Simmons & Ryan 1977). Therefore is was assumed that the CI-2 mRNA product has a short peptide that acts as a signal peptide directing the nascent protein to the membrane of the RER. This peptide could either be an internal sequence or an N-terminal extension of the nascent CI-2. The amino-terminal region of the deduced CI-2 protein has none of the characteristics of a signal peptide. The only internal region having the properties of a signal peptide (von Heijne 1985) found in the CI-2 is too close to the C-terminus and could not act as such (see Hortsch & Meyer 1984). However, the open reading frame of the full-length cDNA is preceded by a short nucleotide sequence beginning with another in frame ATG, this sequence could encode a peptide with the characteristics of a signal sequence if the ochre stop codon (TAA) in frame with the upstream ATG were suppressed (see von Heijne (1985) and Watson (1985)). It has been shown that suppression of termination codons can occur *in vivo* and *in vitro* (Heidsiak *et al.* 1982; Kohli *et al.* 1979; Kubli *et al.* 1982 for example by a suppressor tRNA, such as those found and purified from yeast and plants (e.g. tobacco and lupin) (Beier *et al.* 1984; Barciszewski et al. 1985). To examine the synthesis and processing of a single CI-2 polypeptide in more detail we have cloned the full-length cDNA clone, containing the stop codon, into an *in vitro* transcription vector. Synthetic mRNA has been made by using T7 polymerase and used to prime the wheatgerm translation system. A single translation product was obtained, on an SDS–urea–polyacrylamide gel, which co-migrated with the CI-2 synthesized by poly(A)$^+$ RNA from Hiproly. Since the wheatgerm system does not suppress nonsense codons, unless supplied with exogenous suppressor tRNAs (Kohli *et al.* 1979), we conclude that the most likely start site for protein synthesis *in vitro* is at the second ATG, which is downstream from the TAA codon. This result suggest that the poly(A)$^+$ RNA translation product does not posses an N-terminal extension. Furthermore the *in vitro* translation product has the same mobility as the mature protein on an SDS–urea–polyacrylamide gel. From these experiments, we conclude that all, or the vast majority, of CI-2 mRNAs are translated *in vitro* with no N-terminal extension.

However, if CI-2 is sequestered within a membrane, as mentioned above, then it must contain a functional signal sequence. It is possible that the TAA stop codon is read-through by an ochre suppressor tRNA. If this is so, then CI-2 would be the first example of a plant protein whose synthesis and subsequent deposition in protein bodies is dependent on a suppressor tRNA.

Conclusions

The seed proteins of plants are a complex collection of proteins of various natures and functions. Application of recombinant DNA techniques has allowed the sequence of many of these to be deduced and from this knowledge a clear picture of the relationships of the proteins,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

particulary the storage proteins, to one another has been obtained. Surprisingly this has shown that the prolamins, once thought to be present only in the Gramineae, belong to a wide ranging superfamily of seed proteins of ancient origin. The prolamin and several other seed protein genes under tissue-specific and developmental control and are also drastically affected by certain 'high-lysine' mutations. Besides the importance of these controls in regulating the amino acid composition of the seed and thus its nutritional quality, they also provide an interesting case study in plant systems. Although the regulation of expression is likely to occur during both transcription and translation, exact mechanisms are as yet unknown. The identification of conserved sequences in the 5′ untranscribed regions of hordein genes and of proteins that bind to them provide some leads towards such understanding.

This work was supported partly by contract no. GBI-4-023-UK(H) of the Biomolecular Engineering Programme of Commission of the European Communities.

References

Barciszewski, J., Barciszewska, M., Suter, B. & Kubli, E. 1985 Plant tRNA suppressors: *in vivo* read through properties and nucleotide sequence of yellow Lupin seeds tRNA (Tyr). *Pl. Sci.* **40**, 193–196.

Beier, H., Barciszewska, M., Krupp, G., Mitnacht, R. & Gross, H. J. 1984 UAG readthrough during TMV RNA translation: isolation and sequence of two tRNAs (TYR) with suppressor activity from tobacco plants. *EMBO J.* **3**, 351–356.

Boisen, C., Yding Andersen, C. & Hejgaard, J. 1981 Inhibitors of chymotrypsin and microbial serine proteases in barley grains. *Physiol. Pl.* **52**, 167–176.

Broglie, R., Coruzzi, G., Lamppa, G., Keith, B. & Chua, N.-H. 1983 Structure analysis of nuclear genes coding for the precursor of the small subunit of wheat ribulose-1,5-bisphosphate carboxylase. *Biotechnology* **1**, 55–61.

Casey, R., Domini, C. & Ellis, N. 1986 Legume storage proteins and their genes. *Oxf. Surv. Pl. molec. Cell Biol.* **3**. (In the press.)

Dayhoff, M. O. 1978 *Atlas of protein sequence and structure*, vol. 5, suppl. 3.

Dennis, E. S., Gerlach, W. L., Pryor, A. J., Bennetzen, J. L., Inglis, A., Llewellyn, D., Sachs, M. M., Ferl, R. J. & Peacock, W. J. 1984 Molecular analysis of the alcohol dehydrogenase 2 (Adh2) gene of maize. *Nucl. Acids Res.* **12**, 3983–4000.

Dynan, W. S. & Tjian, R. 1985 Control of eukaryotic messenger RNA synthesis by sequence-specific DNA binding proteins. *Nature, Lond.* **316**, 774–778.

Forde, B. G., Heyworth, A., Pywell, J. & Kreis, M. 1985*a* 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. *Nucl. Acids Res.* **13**, 7327–7339.

Forde, B. G., Kreis, M., Williamson, M. S., Fry, R. P., Pywell, J., Shewry, P. R., Bunce, N. & Miflin, B. J. 1985*b* Short tandem repeats shared by B- and C-hordein cDNAs suggest a common evolutionary origin for two groups of cereal storage protein genes. *EMBO. J.* **4**, 9–15.

Forde, J., Malpica, J. M., Halford, N. G., Shewry. P. R., Anderson, O. D., Greene, F. C. & Miflin. B. J. 1985*c* The nucleotide sequence of a HMW glutenin subunit gene located on chromosome 1A of wheat (*Triticum aestivum* L.). *Nucl. Acids Res.* **13**, 6817–6832.

Giese, H. & Andersen, B. 1982 The course of protein synthesis during grain filling in normal and high lysine barley. In *Proc. Second Res. Co-ord. Meet. on Cereal Protein Improvement*, pp. 217–226. Vienna: I.A.E.A.

Giese, H. & Hejgaard, J. 1984 Synthesis of salt-solubale proteins in barley. Pulse-labeling study of grain filling in liquid-cultured detached spikes. *Planta* **161**, 172–177.

Giese, H. & Hopp, E. H. 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.

Graham, J. S., Pearce, G., Merryweather, J., Titani, K., Ericsson, L. & Ryan, C. A. 1985 Wound-induced proteinase inhibitors from tomato leaves. *J. biol Chem.* **260**, 6555–6560.

Heberlein, U., England, B. & Tjian, R. 1985 Characterization of *Drosophila* transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. *Cell* **41**, 965–977.

Heidecker, G. & Messing, J. 1983 Sequence analysis of zein cDNAs obtained by an efficient mRNA cloning method. *Nucl. Acids Res.* **11**, 4891–4906.

Heidsiak, R. M., Laski, F. A., Raj Bhandary, U. L., Sharp, P. A. & Capecchi, M. R. 1982 Establishment of mammalian cell lines containing multiple nonsense mutations and functional suppressor tRNA genes. *Cell* **31**, 137–146.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Hejgaard, J., Bjørn, S. E. & Nielsen, G. 1984 Localization to chromosomes of structural genes for the major protease inhibitors of barley grains. *Theor. appl. Genet.* **68**, 137–130.

Hortsch, M. & Meyer, D. E. 1984 Pushing the signal hypothesis: What are the limits? *Biol. Cell* **52**, 1–8.

Hymowitz, H. 1983 Variation in and genetics of certain antinutritional and biologically active components of soybean seed. In *Ciba Foundation Symposium*, no. 97, pp. 49–60. London: Pitman.

Johnson, D. A., Gautsch, J. W., Sportsman, J. R. & Elder, J. H. 1984 Improved technique utilizing non-fat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose. *Gene analyt. Techn.* **1**, 3–8.

Jonassen, I., Ingversen, J. & Brandt, A. 1981 Synthesis of SPII albumin, β-amylase and chymotrypsin inhibitor CI-1 on polysomes from the endoplasmic reticulum of barley endosperm. *Carlsberg Res. Commun.* **46**, 175–181.

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

Kirsi, M. & Mikola, J. 1971 Occurrence of proteolytic inhibitors in various tissues of barley. *Planta* **96**, 281–291.

Kohli, J., Kwong, T., Altruda, F. & Soll, D. 1979 Characterization of a UGA-suppressing serine tRNA from *Schizosaccharomyces pombe* with the help of a new *in vitro* assay system for eukaryotic suppressor tRNAs. *J. biol. Chem.* **254**, 1546–1551.

Kreis, M., Forde, B. G., Rahman, S., Miflin, B. J. & Shewry, P. R. 1985*a* Molecular evolution of the seed storage proteins of barley, rye and wheat. *J. molec. Biol.* **183**, 499–502.

Kreis, M., Rahman, S., Forde, B. G., Pywell, J., Shewry, P. R. & Miflin, B. J. 1983*a* Sub-families of hordein mRNA encoded at the *Hor* 2 locus of barley. *Molec. gen. Genet.* **191**, 194–200.

Kreis, M., Shewry, P. R., Forde, B. G., Rahman, S. & Miflin, B. J. 1983*b* Molecular analysis of a mutation conferring the high-lysine phenotype on the grain of barley (*Hordeum vulgare*). *Cell* **34**, 161–167.

Kreis, M., Shewry, P. R., Forde, B. G., Rahman, S., Bahramian, M. B. & Miflin, B. J. 1984 Molecular analysis of the effects of the *1ys 3a* gene on the expression of *Hor* loci in developing endosperms of barley (*Hordeum vulgare*). *Biochem. Genet.* **22**, 231–255.

Kreis, M., Shewry, P. R., Forde, B. G., Forde, J. & Miflin, B. J. 1985*b* Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, barley and rye. *Oxf. Surv. Pl. molec. Cell Biol.* **2**, 253–317.

Kridl, J. C., Vieira, J., Rubenstein, I. & Messing, J. 1984 Nucleotide sequence analysis of a zein genomic clone with a short open reading frame. *Gene* **28**, 113–118.

Kubli, E., Schmidt, T., Martin, P. F. & Sofer, W. 1982 *In vitro* suppression of a nonsense mutant of *Drosophila melanagaster*. *Nucl. Acids Res.* **10**, 7144–7152.

Lamppa, G., Nagy, F. & Chua, N.-H. 1985 Light regulated and organ-specific expression of a wheat *cab* gene in transgenic tobacco. *Nature, Lond.* **316**, 750–752.

Langridge, P. & Feix, G. 1983 A zein gene of maize is transcribed from two widely separated promoter regions. *Cell* **34**, 1015–1022.

Miflin, B. J., Rahman, S., Kreis, M., Forde, B. G., Blanco, L. & Shewry, P. R. 1983 The hordeins of barley: developmentally and nutritionally regulated multigene families of storage proteins. In *Structure and Function of Plant Genomes* (ed. O. Ciferrri & L. S. Dure), pp. 85–92. New York: Plenum Press.

Miskimins, W. K., Roberts, M. P., McClelland, A. & Ruddle, F. H. 1985 Use of a protein-blotting procedure and specific DNA probe to identify nuclear proteins that recognize the promoter region of the transferon receptor gene. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6741–6744.

Munck, L., Karlson, K. E., Hagberg, A. & Eggum, B. O. 1970 Gene for improved nutritional value in barley seed protein. *Science, Wash.* **168**, 985–987.

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

Parker, C. S. & Topol, J. 1984 A *Drosophila* RNA polymerase II transcription factor specific for the heat-shock gene binds to the regulatory site of an *hsp* 70 gene. *Cell* **37**, 273–283.

Pedersen, K., Argos, P., Naravana, S. V. L. & Larkins, B. A. 1986 Sequence analysis and characterization of a maize gene encoding a high-sulfur zein protein of M_r 15,000. *J. biol. Chem.* (In the press.)

Pedersen, K., Devereaux, J., Wilson, D. R., Shelden, E. & Larkins, B. A. 1982 Cloning and sequence analysis reveal structural variation among related genes in maize. *Cell* **29**, 1015–1026.

Prat, S., Cortadas, J., Pingdomenech, P. & Palau, J. 1985 Nucleic acid (cDNA) and amino acid sequences of the maize endosperm protein glutelin-2. *Nucl. Acids Res.* **13**, 1493–1504.

Rafalski, T. A., Scheets, K., Metzler, M., Peterson, D. M., Hedgcoth, C. & Soll, D. G. 1984 Developmentally regulated plant genes: the nucleotide sequence of a wheat gliadin genomic clone. *EMBO J.* **3**, 1409–1415.

Rahman, S., Kreis, M., Forde, B. G., Shewry, P. R. & Miflin, B. J. 1984 Hordein gene expression during development of the barley (*Hordeum vulgare*) endosperm. *Biochem J.* **233**, 315–322.

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

Rahman, S., Shewry, P. R., Forde, B. G., Kreis, M. & Miflin, B. J. 1983 Nutritional control of storage protein synthesis in developing grain of barley (*Hordeum vulgare* L.). *Planta* **159**, 366–372.

Rasmussen, U. 1985 Immunological screening for specific protein content in barley seeds. *Carlsberg Res. Commun.* **50**, 83–93.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Ryan, C. A. 1981 Proteinase inhibitors. In *The biochemistry of plants* (ed. A. Marcus), vol. 6, pp. 351–370. London and New York: Academic Press.

Shewry, P. R., Franklin, J., Parmar, S., Smith, S. J. & Miflin, B. J. 1983 The effects of sulphur starvation on the amino acid and protein composition of cereal grain. *J. Cereal Sci.* **1**, 21–31.

Shewry, P. R. & Miflin, B. J. 1985 Seed storage proteins of the economically important cereals. In *Advances in cereal science and technology*, (ed. Y. Pomeranz), vol. 7, pp. 1–83. St Paul, Minnesota: AACC.

Shewry, P. R., Tatham, A. S., Forde, J., Kreis, M. & Miflin, B. J. 1986 The classification and nomenclature of wheat gluten proteins: a reassessment. *J. Cereal Sci.* **4**, 97–106.

Siebenlist, U., Hennigshausen, L., Batley, J. & Leder, P. 1984 Chromatin structure and protein binding in the putative regulatory region on the *c-myc* gene in Burkitt lymphona. *Cell* **37**, 381–391.

Spena, A., Viotli, A. & Pirotta, V. 1983 Two adjacent genomic zein sequences: structure, organisation and tissue-specific restriction pattern. *J. molec. Biol.* **169**, 799–811.

Staden, R. 1982 An interactive graphics programme for comparing and aligning nucleic acid and amino acid sequences. *Nucl. Acids Res.* **10**, 2951–2961.

Staden, R. 1984 Computer methods to locate signals in nucleic acid sequences. *Nucl. Acids Res.* **12,** 505–519.

Sugijama, T., Rafalski, A., Peterson, D. & Soll, D. 1985 A wheat HMW glutenin subunit gene reveals a highly repeated structure. *Nucl. Acids Res.* **13**, 8729–8737.

Svendsen, I., Martin H. B. & Jonassen, I. 1980 Characteristics of Hiproly barley. III. Amino acid sequences of two lysine-rich proteins. *Carlsberg Res. Commun.* **45**, 79–85.

Tabata, T., Fukosawa, M. & Iwabuchi, M. J. 1984 Nucleotide sequence and genomic organisation of a wheat histone H_3 gene. *Molec. gen. Genet.* **196**, 397–400.

Tatham, A. S., Miflin, B. J. & Shewry, P. R. 1985 The beta-turn conformation in wheat gluten proteins relationship to gluten elasticity. *Cereal Chem.* **62**, 405–412.

Thompson, R. D., Bartels, D. & Harberd, N. P. 1985 Nucleotide sequence of a gene from chromosome ID of wheat encoding a HMW-glutenin subunit. *Nucl. Acids Res.* **13**, 6833–6846.

Von Heijne, G. 1985 Signal sequences the limits of variation. *J. molec. Biol.* **184**, 99–105.

Walker-Simmons, M. & Ryan, C. A. 1977 Immunological identification of proteinase inhibitors I and II in isolated tomato leaf vacuoles. *Pl. Physiol.* **60**, 61–63.

Watson, M. E. E. 1985 Compilation of published signal sequences. *Nucl. Acids Res.* **12**

Weir, W., Frommer, W. B., Maas, C. & Starlinger, P. 1985 Structure of the sucrose synthase gene on chromosome 9 of *Zea mays* L. *EMBO J.* **4**, 1373–1380.

Williamson, M. S., Forde, J. & Kreis, M. 1985 Chymotrypsin inhibitor-2 of barley: characterization of cDNA clones and analysis of expression in normal and high-lysine barley. Manuscript in preparation.

Willmitzer, L. & Wagner, K. G. 1981 The isolation of nuclei from tissue-culture plant cells. *Expl Cell Res.* **135**, 69–77.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES