The EMBO Journal vol.4 no.5 pp.1103-1110, 1985

Each zein gene class can produce polypeptides of different sizes

A.Viotti, G.Cairo, A.Vitale and E.Sala

Istituto Biosintesi Vegetali, CNR, Via Bassini, 15, 20133 Milano, Italy

Communicated by S.Ottolenghi

The family of zein proteins in maize consists of several prototypes whose amino sequence is characterized by a repetitive block structure. We have now isolated and characterized a new series of zein cDNA clones from a cDNA library made from protein-body polysomal mRNA. The nucleotide sequences and the hybrid-selected translation results indicate that anomalous zein genes with termination codons within the coding region are transcribed, translated and possibly accumulated into zein protein bodies. The sequence analysis of these new clones and their comparison with already characterized zein sequences, all from the same maize line, indicate a large variability in the amino acid sequence of the blocks both among the prototypes and within members of each prototype. Northern blot analysis of total RNA from endosperms at different maturation stages shows a differential expression of each prototype sequence with the production, within each family, of zein transcripts coding for discrete size classes of polypeptides. The results allow a better definition of the zein system, clarify the relationship between sequence types and polypeptide size classes, and suggest the possibility of inserting lysine codons to raise the nutritional value of the seed storage protein.

Key words: zein prototypes/in-frame stop codons/truncated polypeptides/timing of expression

Introduction

The zein gene family in maize consists of >100 sequences, divided into several size classes and sequence prototypes each reiterated ~10–15 times per haploid genome (Viotti *et al.*, 1979, 1982; Hagen and Rubenstein, 1981). The various size classes and sequence prototypes are diversely distributed on three of the 10 chromosomes of the haploid maize complement (Viotti *et al.*, 1980, 1982; Soave and Salamini, 1984). Several loci, similarly scattered in the maize genome, differentially control the expression of the various genes of each subset during endosperm maturation (Di Fonzo *et al.*, 1979; Soave and Salamini, 1984).

The anatomy and organization of zein genes and their flanking sequences (Spena *et al.*, 1982, 1983; Pedersen *et al.*, 1982) suggest mechanisms of intergenic exchange, amplification and transposition that may explain the broad qualitative and quantitative differences in the restriction pattern of zein DNA fragments from different maize lines (Wienand and Feix, 1980; Viotti *et al.*, 1982). Sequence analysis of zein cDNA and genomic clones from different maize lines also indicates that zein proteins contain a central domain of seven to eight repeats of a block of 20 amino acids that constitutes the basic unit structure of the polypeptides (Spena *et al.*, 1982; Geraghty *et al.*, 1982; Pedersen *et al.*, 1982). Two other domains, the head of ~100 amino acids and the tail of eight amino acids, contribute to the determination of the length of the mature zein polypeptides (Spena *et al.*, 1982; Argos *et al.*, 1982). Discrepancies between the coding capacity of sequenced cDNA clones, the size of the products translated *in vitro* from selected mRNAs and the mol. wts. of zein polypeptides have been reported (Geraghty *et al.*, 1982; Viotti *et al.*, 1982). Moreover, the isolation of anomalous genomic clones with coding triplets mutated into stop codons suggests that the occurrence of truncated zein polypeptides might be a source of further size variability (Spena *et al.*, 1982, 1983).

Here we report the sequence of several cDNA clones from the same maize line, and provide evidence that anomalous zein genes are transcribed and their mRNAs are associated to the polysomal fraction. These data, together with the study on the differential expression of zein genes during endosperm development and the re-determination of the mol. wt. of zein polypeptides, explain the aforementioned discrepancies. Sequence analysis also shows a high frequency of amino acid substitutions in different positions of the repetitive blocks and of the head domain in genes belonging to each zein class, suggesting the existence of a degree of freedom in the structural model proposed for the zein polypeptides (Argos *et al.*, 1982) and indicating sites in which lysine codons might be introduced in this agronomically important seed storage protein.

Results

The zein components

Since the first characterization of zein proteins by electrophoresis on SDS-polyacrylamide gels different number of polypeptides and different mol. wt. determinations have been reported (Misra *et al.*, 1975; Burr and Burr, 1976; Gianazza *et al.*, 1976; Viotti *et al.*, 1979). Zein chains were classified as 'light' around 19 kd and 'heavy' around 22 kd. The recent data on the sequences of cDNA and genomic zein clones have, however, indicated higher mol. wt. values for both 'light' and 'heavy' chains (Pedersen *et al.*, 1982; Geraghty *et al.*, 1982; Spena *et al.*, 1982). These discrepancies have been tentatively explained on the basis of anomalous migration of zein polypeptides due to their high hydrophobicity (Geraghty *et al.*, 1982). Figure 1D shows that the electrophoresis of polypeptides from naked protein bodies (Viotti *et al.*, 1978a) of the W64A line resolves zein proteins into at least five discrete size classes (H1, H2, L1, L2 and L3). The number of polypeptides that can be resolved depends on the amount loaded and on the relative ratio of the various components. Figure 1D also shows that other maize lines have a similar pattern and, even though the relative abundance of the two size classes and of subspecies within each class is different, a constant pattern is obtained. This indicates the absence of artefacts during the run, as also confirmed by the experiment reported in Figure 1C, where the single bands were cut from a gel and rerun separately on a second gel. The separations reported in Figure 1 have been obtained in all the conditions tested (different acrylamide concentrations and acrylamide-bis-acrylamide ratio, see Materials and methods). However, a short run in high acrylamide concen-

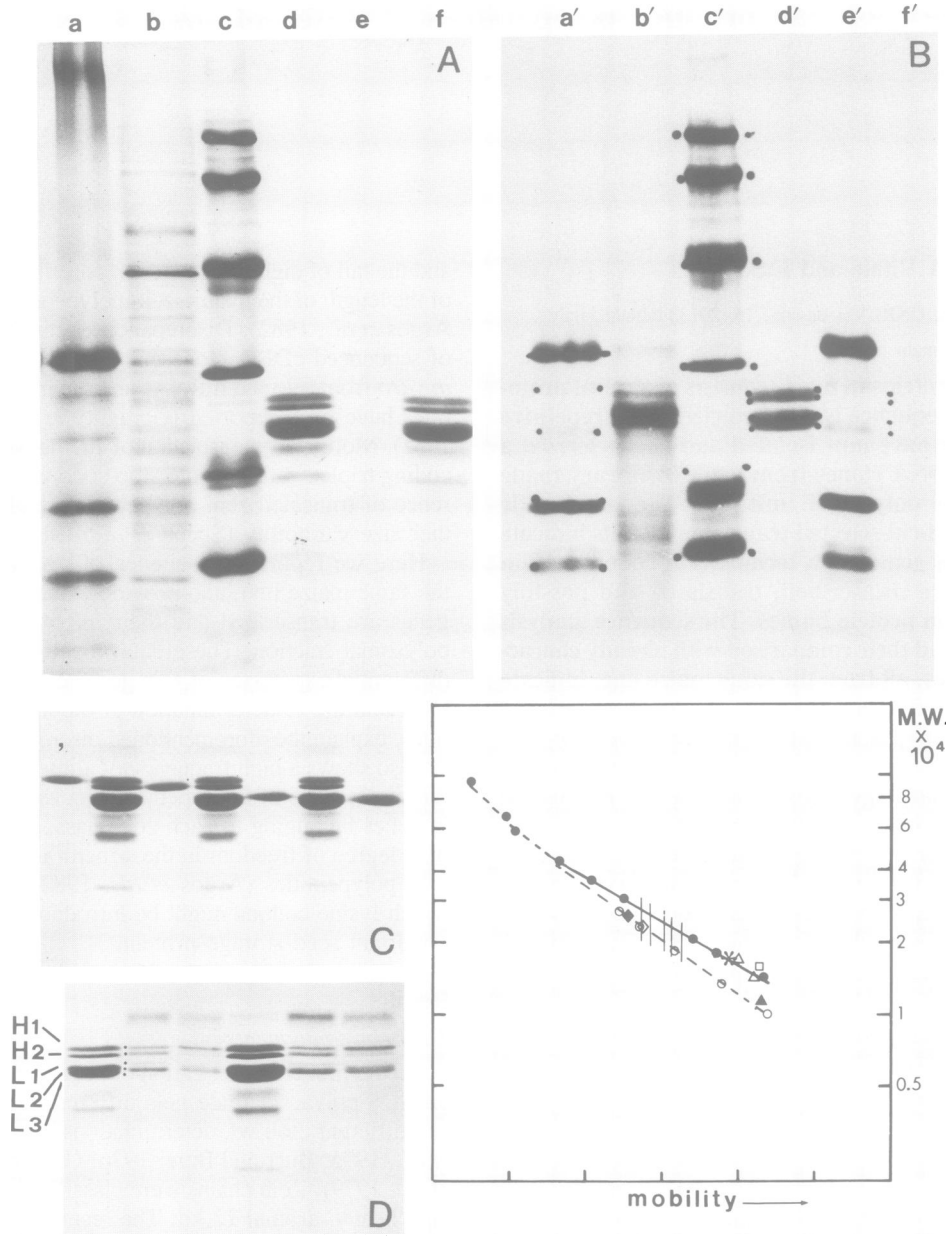

Fig. 1. SDS-PAGE analysis of zein proteins and mol. wt. estimation. (**A**) and (**B**) represent the same gel first stained and then fluorographed, the lanes were loaded with cold and radioactive proteins in the following combinations. (**a**) and (**a′**) cold and radioactive Con-A. (**b**) and (**b′**) *in vitro* products of protein body polysomal RNA translated in a wheat germ system. (**c**) and (**c′**) cold markers and radioactive markers. (**d**) and (**d′**) cold zein and [^{3}H]dansylated zein. (**e**) and (**e′**) only radioactive Con-A. (**f**) and (**f′**) only cold zein. The gel was first stained, then was photographed, processed for fluorography and dried. On the side of the markers and zeins, dots of radioactive ink were applied and then exposed to the film. Note the difference in mobility of α-lactalbumin (14 000) and lactoglobulin A (18 400). (**C**) Zeins H1, H2, L1 and L2 were cut from a first gel and then rerun in alternate loading with zeins from protein bodies. (**D**) From left to right zein from W64A, A69Y and Illinois high protein maize lines, in two different load series, are fractionated into the various components as indicated on the side. The estimation of mol. wts. reported in the plot of the log of mol. weight *versus* mobility, are from gels of different acrylamide concentrations and acrylamide-bisacrylamide ratios. Standard proteins from heavier to lighter, indicated by closed circles, are: phosphorylase B, BSA, catalase, ovalbumin, lactate dehyrogenase, carbonic anhydrase, trypsin inhibitor, ferritin, α-lactalbumin; the best fit of the continuous line also includes: (*) lactoglobulin A, (△) myoglobin and myoglobin I plus II, (□) human γ-globin. The open circles correspond to the mol. wts. of the five Con-A subunits according to Hauge (1975) and Wang *et al.* (1971). The best fit of the dashed line in the Con-A range also comprises: (◆) chymotrypsinogen A, (◇) trypsin, (▲) cytochrome C.

tration (Figure 2A) or a long run in low concentration (Figure 2B) separate poorly the various components which are clearly resolved in Figure 2C, D.

The molecular weight of zein polypeptides was originally determined using Con-A, trypsin, chymotrypsinogen A and cytochrome C as markers (Burr and Burr, 1976; Gianazza *et al.*, 1976). Figure 1 shows, however, that the mol. wt. of these proteins do not agree with their mobilities when compared with other available mol. wt. markers. It has been reported that some of these markers (α-chymotrypsinogen A and cytochrome c) migrate anomalously under similar conditions (Steele and Nielsen, 1978). We thought that this could explain the discrepancies reported in the mol. wt. of zein chains and we re-examined the problem using a battery of radioactive and non-radioactive markers (Figure 1A and B) including proteins whose mol. wt. has been accurately determined by sequencing. The plot reported in Figure 1 shows

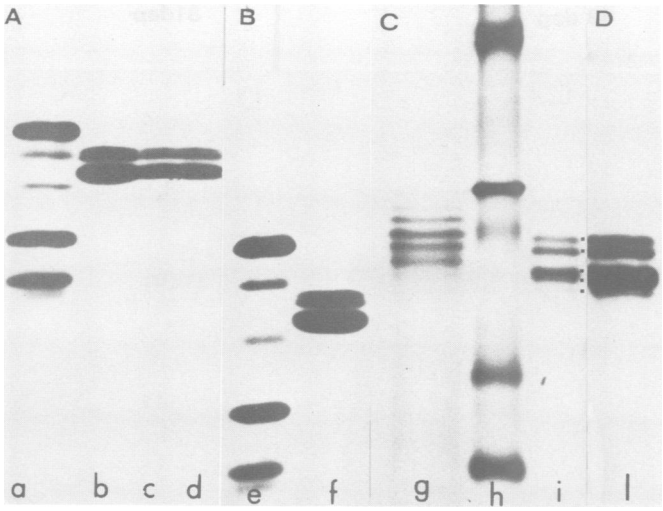

Fig. 2. Fluorography of [^{3}H]dansylated zein proteins fractionated by electrophoresis on SDS gel with different acrylamide concentrations. **(A)** 20%; **(B)** 12%; **(C)** and **(D)** 15%. **(a)** and **(e)** [^{3}H]concanavalin-A; **(b–d)**, **(i)**, and **(l)** [^{3}H]dansylated zeins; **(g)** *in vitro* products of protein body-polysomal RNA translated in a wheat germ system; **(h)** from top to bottom, radioactive markers: ovalbumin, carbonic anhydrase, lactoglobulin-A, cytochrome c.

that using the original (old) set of mol. wt. markers the mol. wt. of zein polypeptides is in the range 19 000 – 22 000 while using different, and probably more reliable, mol. wt. markers, the values obtained are 27 000, 26 500, 24 700, 23 500 and 21 600, for H1, H2, L1, L2 and L3, respectively. These values closely agree with those predicted by the sequence of zein genes (Geraghty *et al.*, 1982; Spena *et al.*, 1982). The L3 class, poorly resolved in stained gels, is clearly visible by fluorography and contains polypeptides that are coded at least in part by sequences homologous to the E19 gene (Figure 4).

The variable amounts of the five SDS-gel components in the lines investigated and the different qualitative/quantitative patterns obtained in Southern blot analysis of their DNA (Viotti *et al.*, 1982; Spena *et al.*, 1983) indicate that specific deletion or amplification of some zein sequences has occurred during the many generations that separate each line from the ancestral genotype. This also suggests that DNA fragments corresponding to the zein coding region might have undergone drastic changes possibly generating in each line a specific allele from each ancestral gene. This study has thus been undertaken on zein genes all belonging to the same maize line (W64A).

The light classes

The light zein genes found in the W64A line are subdivided on the basis of cross homology into two subclasses defined by the prototype clones M6 and MG1 (Viotti *et al.*, 1982). In Figure 3 we present the nucleotide sequence of M6 together with the sequence of M8, a member of this subclass, whose *in vitro* translation products are reported in Figure 4A. The two sequences are compared with that of E19, a genomic clone corresponding to the MG1-cDNA, and with that of My2, a new cDNA clone which has been selected by the criteria outlined in Materials and methods.

The M6 is a full length cDNA clone and like the A20 clone of Geraghty *et al.* (1982), it contains at the 5′ a sequence stretch also present in inverted orientations at the 3′ end. It codes for a mature zein polypeptide of 219 amino acids which, like other zein, can be divided into several domains: the signal, the head,

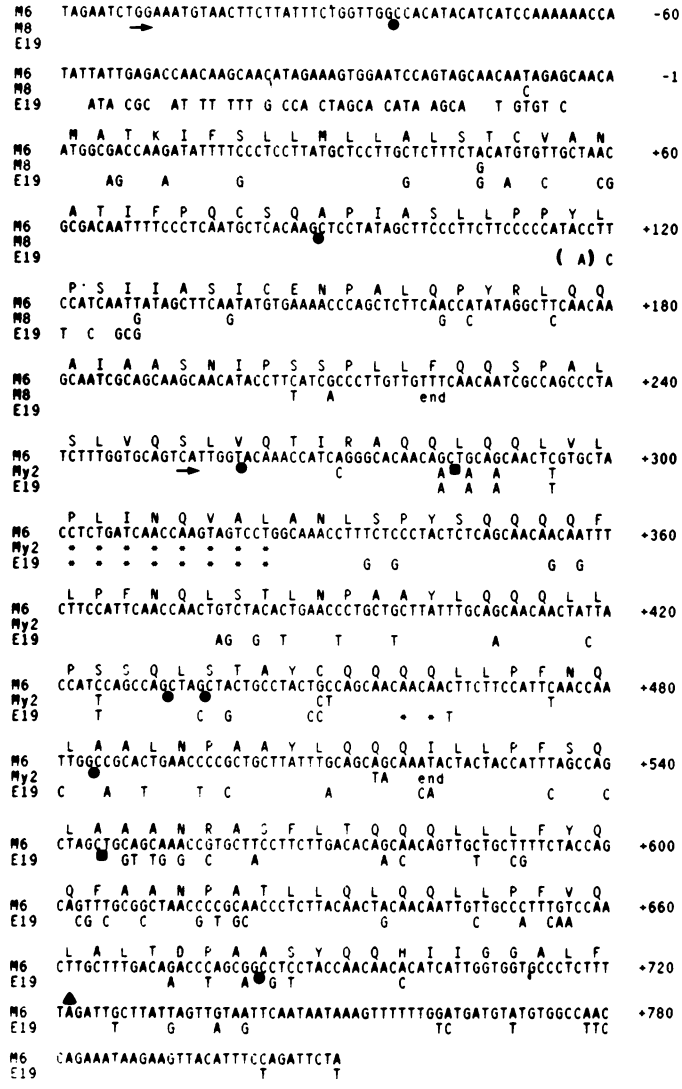

Fig. 3. Nucleotide sequence of light class clones. The sequence of M6 cDNA is reported, only sequence differences that occur in the M8 and My2 cDNA and in the E19 gene are listed. The nucleotides are numbered from the ATG initiation codon. Asterisks represent codon gaps introduced to optimize the homology. Physiological terminator codon is indicated by a triangle. In E19 sequence brackets indicate the in-frame stop codon. Closed circles and squares represent restriction sites used in the sequence strategy of M6 clone and labelled at the 5′ or 3′ ends, respectively.

the repetitive blocks and the tail. The M8 clone is highly homologous to M6, to which it strongly hybridizes, and selects mRNAs that direct the synthesis of zein polypeptides identical to those of M6 (Figure 4A). Homology and peculiarities between clones of the light size class can be also evaluated by comparison with the amino acid sequence reported in Figure 5. The sequences are aligned according to the block structure and are compared with the sequence of the A20, a cDNA clone from IHP maize line (Geraghty *et al.*, 1982) that is 100% homologous to M8 cDNA.

The My2 clone contains a cDNA sequence of 240 bp, that is partially homologous to both the M6 and to the E19 sequences. At the 5′ side My2 is highly homologous to E19, with the eight amino acid gap characteristic of the L2 zein type, while further downstream it resembles the M6 sequence; My2 in fact hybridizes in stringent conditions to both M6 and E19 sequences while M6 and E19 in the same condition do not cross-hybridize (data not

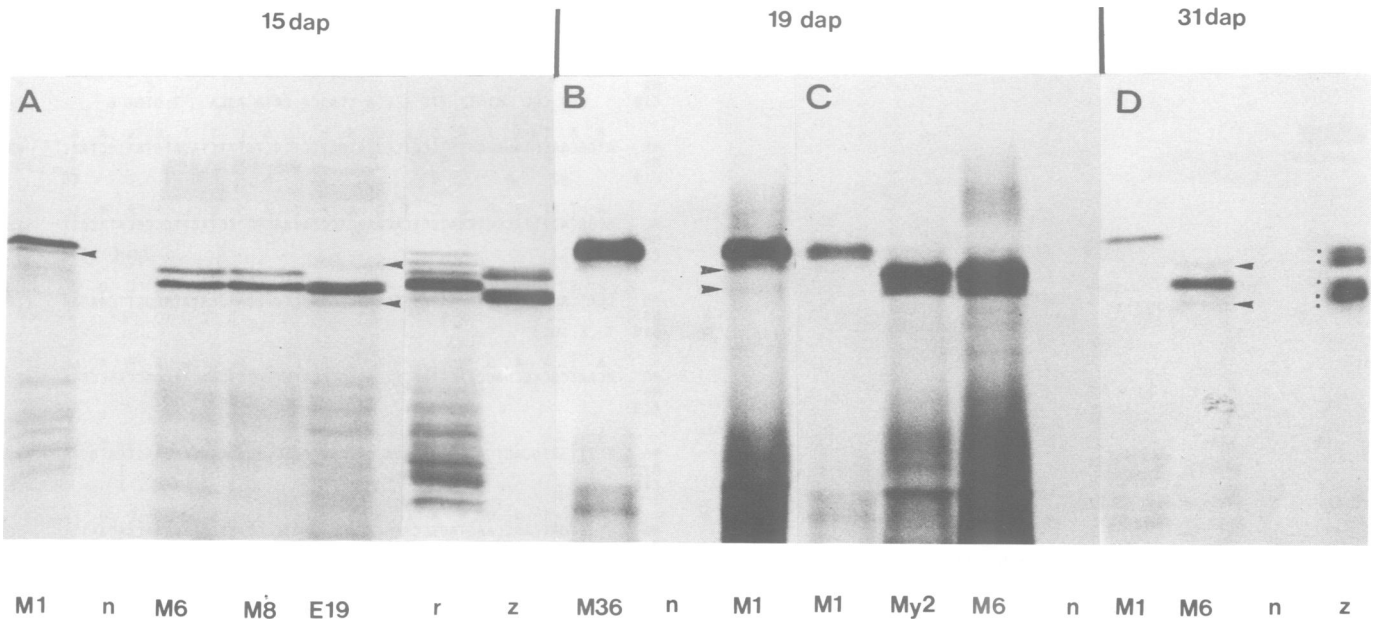

Fig. 4. Fluorography of *in vitro* translated products from mRNAs selected by zein clones and fractionated by SDS-PAGE (15%). The RNAs used in the hybridization were from endosperms at different developmental stages (days after pollination, dap) as indicated. Hybridization was carried out at high stringency (T_m −10°C, 65% formamide, NaCl 0.4 M, Pipes 10 mM pH 6.8, Viotti *et al.*, 1979). Selection by each clone is indicated in the corresponding lane by upper case leters. Lower case letters represent: (n) no RNA added; (r) total RNA; (z) mature zein labelled with [^{3}H]dansyl chloride. Arrow heads indicate minor products selected by E19, M6 and M1 clones.

```
M6   M A T K I F S L L M L L A L S T C V A N A   21
M8                                 A
A20                                A
E19      A       C           G     A S A   T

M6   T I F P Q C S Q A P I A S
M8
A20
E19

M6   L L P P Y L P S I I A S I C E N P A L Q P Y R L Q Q
M8                   M       V
A20                  M       V
E19          ▲   S P A V S   V         I           I

M6   A I A A S N I P S S P L L F Q Q
M8                   L         end
A20                  L
E19          G I L   L       F L

M6   S P A L = = = = S L V Q S L V Q T I R A Q Q L Q Q L
A20
My2
E19    S     L Q Q L P     H L   A   N

M6   V L P L I N Q V A L A N L S P Y S Q Q Q Q   119
A20       V
My2       * * * * * * * *
E19       * * * * * * * *         A A
```

```
M6   F L P F = N Q L S T L N P A A Y L Q Q Q L   139
A20
My2
E19                  A A     S   S           Q

M6   = L P S = S Q L S T = = = = A Y C Q Q Q Q
A20        F           A                 S
My2        F           A                 S
E19        F           P A               P     * *

M6   L L P F = N Q L A A L N P A A Y L Q Q Q I
A20
My2                                        L     end
E19  F                       S P             Q

M6   L L P F = S Q L A A A N R A S F L T Q Q Q
A20
E19                H     G V S P   T         P

M6   L L L F = Y Q Q F A A N P A T L L Q L Q Q
A20       P
E19       P           H A   P   A G

M6   L L P F = V Q L A L T D P A A S Y Q Q H I I G G A L F
A20                                        P
E19             N             N L     F    P
```

Fig. 5. Comparison of the amino acid sequences of clones of the light classes. The complete sequence of M6 is reported, for the other clones only variable amino acids are given at the site of their occurrence. The first three blocks after the 21 amino acids of the signal peptide represent the head domain. The following blocks are the central repetitive part; the arrow indicates the start of the tail piece. Dashes represent gaps introduced to maximize homology among blocks of the same clone. Asterisks represent gaps introduced to maximize homology among blocks of different clones. Filled triangle represents the terminator codon in E19 clone.

shown). My2 represents a bridge between the prototypes M6 and E19 (Viotti *et al.*, 1982; Geraghty *et al.*, 1982) suggesting that diversification from one to another occurred through recombination and small changes.

The amino acid coding sequence of the two prototypes (M6-E19) differs by six amino acids and partly explains the difference in mol. wts. of the L1 and L2 classes, However, sharp bands in the L1 and L2 region of the SDS-gels have never been

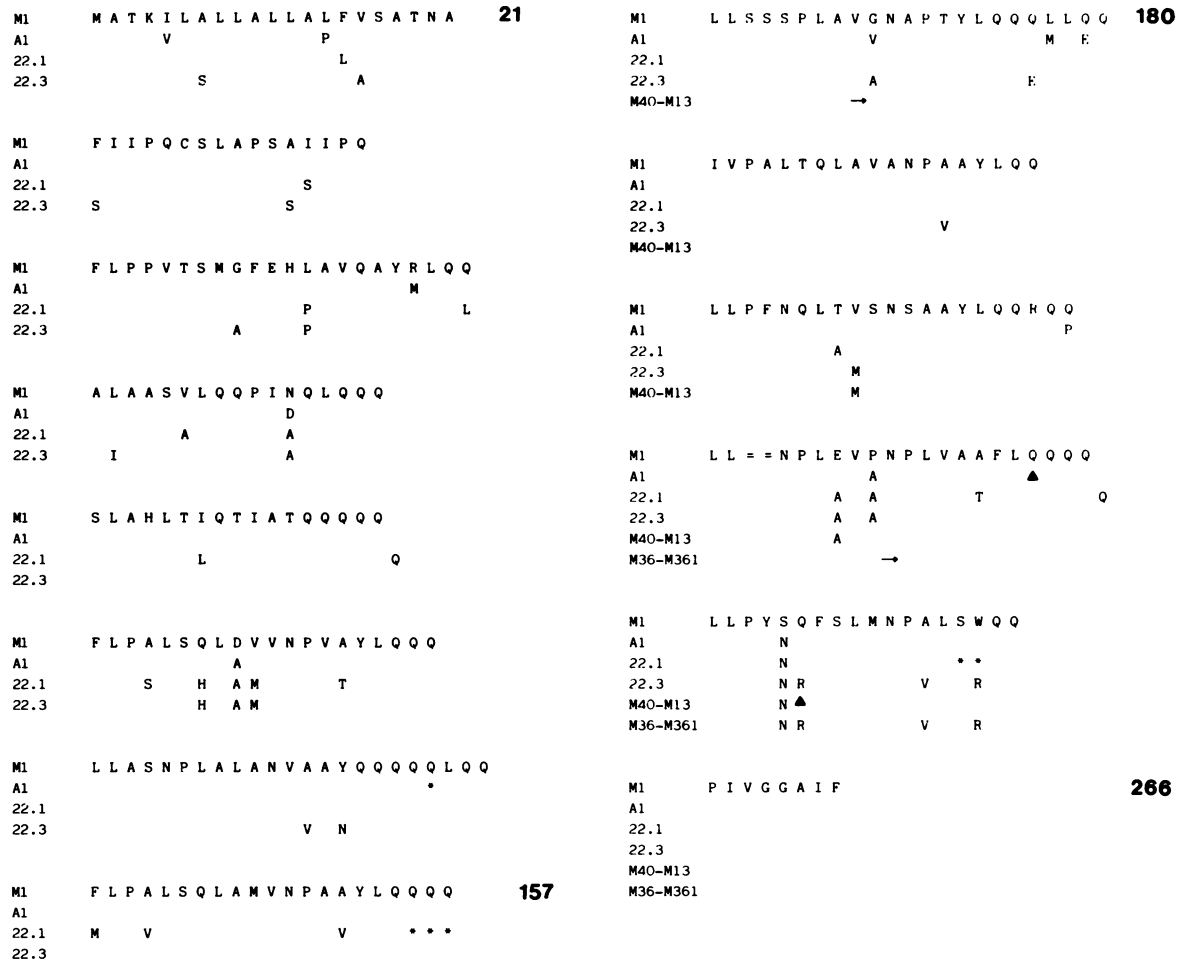

Fig. 6. Comparison of the amino acid sequences of clones of the M1 family. The complete sequence of M1 is presented, only variable amino acids at corresponding positions are listed for the other clones. The residues are divided according to the block structure described in the legend to Figure 5. Dashes and asterisks have the same meaning as those in Figure 5.

obtained, indicating that polypeptides differing in length by few amino acids are present in each class. The hybrid-selected translation reported in Figure 4A shows that E19 prototype selects messengers whose products range in at least four discrete size classes, for instance two of these with a difference of three amino acids, as estimated from the mobility of the SDS-PAGE bands of Figure 4A. Similarly M6 and M8 select mRNAs coding for more than one discrete size class but with mol. wts. different from those of the E19 type (Figure 4A).

The M1 family

The heavy class of zein polypeptides is resolved in SDS-PAGE into two main bands H1 and H2, both of which are preferentially suppressed in maize lines carrying the *opaque-2* mutation. We have previously characterized a clone coding for the H1 zein chain (Viotti *et al.*, 1982; Spena *et al.*, 1982). We have not yet been able to identify cDNA clones corresponding to the H2 polypeptides. However, we have selected four cDNA clones that do not hybridize to *opaque-2* cDNA (Viotti *et al.*, 1982) and hybridize more weakly to M1. These clones still hybridize principally mRNAs coding for H1 polypeptides.

The nucleotide sequences (data not shown) of these clones (M40, M13, M36 and M361) compared with that of M1 and with that of the genomic clone A1 (Spena *et al.*, 1982), show some base changes that account for the different extent of their cross-hybridization to M1 prototype and for the different amino acid sequences reported in Figure 6. The M36 and M361 are completely homologous in the coding region, but differ by a few nucleotides in the trailer part and thus represent two different genes of the M1 family. Interestingly, M40 and M13, have a stop codon at the amino acids minus 21 from the carboxy terminus. The codons CAG of M1 and CGG of M36 and M361 have mutated in M40 and M13 into TAG. Figure 6 also reports the sequence of two cDNA clones isolated by Marks and Larkins (1982) from the same maize line (W64A). Their sequence is highly homologous to M1 with some insertions or deletions that reduce the polypeptide encoded by 22.1 clone by four amino acids.

The family thus consists of a set of sequences more or less homologous to each other but with a coding capacity that may differ by a few to up to 20 or 30 amino acids. The latter case, exemplified by the M40 and A1 genes, suggests that the H2 polypeptides might derive from genes of the M1 type with proper deletions or stop mutations. In fact the M1 and M36 clones, in addition to messenger(s) for the H1 size class, also poorly select some mRNAs coding for the H2 class (Figure 4B). On the other hand polypeptides of the H2 size class are also coded by mRNAs selected by E19, a genomic clone of L2 size (213 amino acids). This is supported by the data reported in Figure 4 and by the existence of a genomic clone (Z4 of Geraghty *et al.*, 1982) homologous to our E19 but with the additional insertion of one and a half repetitive blocks that brings its coding capacity to 246

amino acids (see also below). In summary, our interpretation of these data is that the H2 size class is constituted by some shortened sequences of M1 type and by some E19-like sequences lengthened by insertion. This could explain our failure to identify clones for the H2 class by means of differential hybridization.

Zein gene expression and mRNA selection

In a preliminary study on zein gene expression we observed, by Northern blot analysis of total endosperm RNA, that zein genes corresponding to each prototype are differentially transcribed during endosperm development. In the total RNA extracted from endosperm at three different stages of maturation (15, 21 and 31 days after pollination), the amount of RNA that hybridizes to M1 and M6 is drastically different (Figure 7). Sequences homologous to M1 are highly expressed at the early stage, but are poorly represented later on. Only long exposure of the filter reveals some transcripts of the M1 type at 21 and 31 days (Figure 7A, B). The M6-like sequences seem, however, to be expressed in alternate fashion (Figure 7C). The decrease in the middle stage is followed at 31 days by an increase in the amount of transcripts that are also slightly slower in migration, suggesting the activation of a new subset of genes and a partial suppression of the previous ones. Translation in a wheat germ system of mRNAs selected by M6 from total RNAs of these three stages results in different polypeptide patterns specific for each stage (Figure 4A compared with D). The E19-like sequences have a timing of expression and type of transcripts similar to that of M6 sequences (data not shown). The same results were obtained if the RNA was extracted with the phenol-chloroform or CsCl methods and were independent of the year of harvest. Moreover, selection or loss of specific zein sequences during extraction are ruled out as the RNAs corresponding to the three stages are the same for the filters hybridized to M1, M6 and E19 probes. These results make it clear that a proper classification of zein genes should be based on sequence analysis, common control and time of expression.

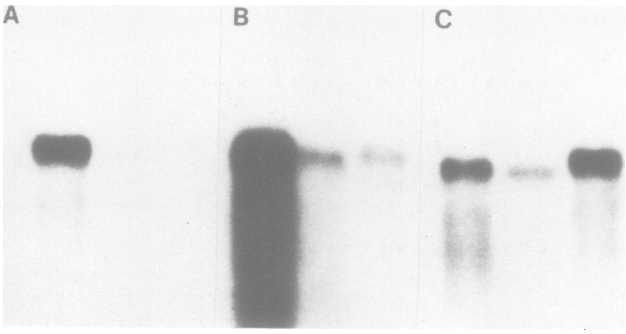

Fig. 7. Autoradiography of Northern blot hybridization analysis of zein genes expressed during endosperm development. Total endosperm RNA was electrophoresed in denaturing agarose-formaldehyde gels, blotted onto nitrocellulose filters and hybridized to the M1 (**A** and **B**) cDNA clones radioactively labeled by nick translation. For each filter, the developmental stages of the endosperm (days after pollination) from left to right were 15, 21 and 31. **Panel B** represents a longer exposure of **A**. Filters were washed at high stringency (0.2 x SSC for 2 h at 68°C) that detects sequences with homologies >96%.

Discussion

This study of the zein system points out an extreme variability at the genic level and consequently a complex situation amongst the zein proteins. Each size class comprises the transcripts of different prototype families that are differentially expressed during endosperm development.

Within each prototype the nucleotide sequence analysis reveals

Table I. Comparison of the amino acid sequences of the zein domains: signal peptide, head and tail

Clones	Maize lines	Zein domains		
		Signal peptide	Head	Tail
A1	W64A	M A T K V L A L L A L L A P F V S A T N A	F I I P Q C S L	S W Q Q P I V G G A I F
M1	W64A	I L		
M40	W64A	----------	----------	
M36	W64A	----------	----------	R
Z22.1[a]	W64A	I L L		R
Z22.3[a]	W64A	I S L A	S	* *
B49[b]	IHP			
M6	W64A	M A T K I F S L L M L L A L S T C V A N A	T I F P Q C S Q	S Y Q Q H I I G G A L F
M8	W64A	A		----------
A20[b]	IHP	A		
E19	W64A	M A A K I F C L L M L L G L S A S A A T A	T I F P Q C S Q	F Y Q Q P I I G G A L F
Z19.1[c]	W64A	I	S	
A30[b]	IHP			
Z4[b]	W22	I	S	
ZG7[d]	W22			
ZG19[d]	W22	----------	----------	
ZG31[d]	W22			
ZG124[d]	W22			
ZG99[c]	Bl;Mex	I	S	
E25	W64A	L A V K I F C L L M L L A L S A S A A N A	T N F L Q C S Q	----------

Sequences were derived from cDNA and genomic clones from different maize lines. Genomic clones are underlined. The sequence of each zein prototype is reported; otherwise only variable amino acids are given at the site of their occurrence. Lines represent sequences not present in the clone considered. Asterisks represent gaps introduced to maximize homology.

[a]Marks and Larkins, 1982.

[b]Geraghty *et al.*, 1982.

[c]Pedersen *et al.*, 1982.

[d]Heidecker and Messing, 1983.

a greater variability in the repetitive blocks than in the other regions with insertions/deletions and substitutions resulting, at times, in drastic changes in the amino acid sequence (polar amino acids into hydrophobic or *vice versa*). However, the comparison of the amino acid sequence between members of each prototype indicates that three domains, signal, head and tail, are strongly conserved and characteristic of each class (Table I); with the clones, M1, M6 and E19 as representatives of three prototype sequences. In this respect the E25 gene (Spena *et al.*, 1983) on the basis of both the nucleotide and amino acid sequences (Table I) may be considered a fourth prototype. In consequence a possible classification of the zein genes might be based on the amino acid sequence of the three conserved regions, even though the corresponding polypeptides might be variable in length and charge.

It has been shown that the recessive mutations at the O2 and O7 loci reduce respectively the synthesis of the 'heavy' and of the 'light' size classes (Di Fonzo *et al.*, 1979; Soave and Salamini, 1984). However, this reduction is only partially specific depending on the maize line investigated; for example the O2 mutation may reduce the light class by 10–20% but the heavy class by 60–90% and never cause the complete disappearance of the heavy component (Gentinetta *et al.*, 1975; Salamini, 1980). This apparent non-specific effect of the O2 and O7 mutations can be explained by the above observations: the heavy and light chains are constituted by polypeptides belonging to different prototypes that are differentially regulated on the basis of sequence type rather than of size class.

Our data on the sequences of the cDNA clones constructed from a polysomal mRNA fraction indicate that zein transcripts with premature stop codons are translated. The resulting protein would be shorter than the normal one and thus, according to Argos *et al.* (1982), might be defective in some functions necessary for its folding and packaging in the protein body. However the amino acid sequences of several cDNA clones and genes indicate that in the W64A line, as also in other lines (Geraghty *et al.*, 1982; Pedersen *et al.*, 1982), a great variability in total length, number of repetitive blocks and block length exists both among prototypes and within members of each family. Moreover, the drastic amino acid changes (e.g., basic to hydrophobic) in several positions, which are supposed by the model of Argos *et al.* (1982) to contribute to the inter and intramolecular interactions, indicate either a high degree of freedom or that alternative interactions regulate the folding and aggregation of zein proteins. The differential expression of zein genes during endosperm maturation suggests that at various developmental stages the protein bodies contain members of each family in different proportions. Alternatively, or additionally, the protein bodies may contain layers of different composition which enlarge asynchronously during development; accordingly, specific interactions among differentially expressed 'heavy' and 'light' polypeptides should occur. However, recent data on zein proteins extracted from protein bodies (Pogna and Viotti, unpublished) indicate preferential aggregation of heavy with heavy or light with light chains rather than between them, suggesting that expression should be regulated in such a way as to maintain a proper ratio.

Others (Langridge *et al.*, 1982) have reported a pattern of developmental expression, different from that presented in this study. The discrepancy may be due to the different maize lines and experimental procedures utilized, or more probably to the structure of their zein clones. One of them, recently sequenced (Langridge and Feix, 1983), shows an amino acid sequence almost completely different from that of several zeins sequenced so far. This clone also contains deletions of the terminal part of the signal peptide and the beginning of the head domain and probably constitutes another example of the anomalies found in the zein gene family. This rearrangement causes the occurrence of a lysine residue never before observed in zein proteins.

In general, none of the many codon changes observed in the sequences of the various zein clones results in a lysine, even though glutamines are frequently substituted by the two other basic amino acids, histidine and arginine. Such a constraint is unexpected and certainly not due to the lack of a corresponding tRNA or aminoacyl-tRNA synthetase activity in the endosperm (Viotti *et al.*, 1978b). The bias is suggestive of constraints in structural conformation of the polypeptide. However, it may be possible to substitute lysine by site-directed mutagenesis into sites which already contain basic amino acid residues and to test the modified zein gene in an heterologous system.

Materials and methods

Screening of cDNA libraries

The cDNA libraries constructed from protein body polysomal mRNA or total endosperm poly(A) mRNA in λ-641 vector (Viotti *et al.*, 1982) were plated on Q358 *Escherichia coli* at a low number of plaques per plate. Parallel filters taken from each plate were hybridized to the previously characterized zein sequences and to cDNAs made from wild-type or *opaque 2* protein body polysomal mRNA (Viotti *et al.*, 1982). Recombinant clones differentially hybridizing to cDNA but not to previously characterized zein sequences M1, M6, MG1 (Viotti *et al.*, 1982) were selected, purified and subcloned in the *Eco*RI site of the pUC8 plasmid (Vieira and Messing, 1982).

Hybrid-selected translation

Purified DNA of plasmid clones was bound to Biodyne A (Pall) filters as described by Viotti *et al.* (1982) and hybridized to RNAs of normal seeds of the W64A line extracted from endosperm at 15 and 31 days after pollination or from protein-body polysomal RNA of 19-day-old endosperms. Hybridization was carried out at high stringency (T_m −10°C). Translation of recovered RNA in a wheat germ cell-free system and SDS-PAGE of *in vitro* translated products were performed as previously described (Viotti *et al.*, 1982).

DNA sequencing

Restriction fragments were labeled at the phosphatase-treated 5′ end or 3′ end with polynucleotide kinase or DNA polymerase (Klenow), respectively. Fragments labeled at a single end were obtained by strand separation (Szalay *et al.*, 1977) and sequenced using the chemical degradation technique of Maxam and Gilbert (1980). All sequences were determined on both strands and the relevant regions were sequenced several times using the G, G+A, T+C, C and A>C reactions.

RNA preparation, Northern transfer and hybridization

Total RNA from endosperms of the W64A line was prepared by homogenization of the frozen tissue in 100 mM Tris-HCl (pH 9), 25 mM EDTA, 0.5% sodium lauroyl sulphate, 0.2% diethyl pyrocarbonate. The homogenate was cleared by low-speed centrifugation and made 0.9 g/ml by CsCl concentration. The solution was vigorously shaken and then centrifuged at 20 000 *g* for 20 min. The clear phase was centrifuged on a 0.8 ml cushion of 5.7 M CsCl in 150 mM EDTA (pH 8) in a SW56 rotor at 35 000 r.p.m. for 12 h according to Glisin *et al.* (1974). Alternatively the homogenate was immediately extracted twice with one volume of phenol-chloroform (1:1, v:v) and the nucleic acids were precipitated by ethanol. The pellet was resuspended in the homogenization buffer and processed as above. In both cases the pelleted RNA was resuspended in homogenization buffer, made 70% in ethanol and stored in aliquots at −80°C. RNA samples were centrifuged and resuspended in 20 mM MOPS (pH 7.0), 2.2 M formaldehyde, 50% formamide, 5 mM sodium acetate, 1 mM EDTA (loading buffer) at a concentration of ~2 μg/μl. Samples from these solutions were carefully diluted 100–200 times and read on a spectrophotometer. After estimation of RNA concentration, the original samples were diluted to 1 μg/μl and after denaturation at 65°C for 5 min, 20 μl of each were loaded, to avoid differential migration of RNA due to different concentration or loading volume. The RNA was fractionated by electrophoresis at 2 V/cm for 10 or 15 h on a 1.2% or 1.5% agarose gel containing 2.2 M formaldehyde. Ribosomal RNAs from *E. coli* and rabbit globin mRNA were used as mol. wt. markers. The gel was soaked for 60–90 min in 20 x SSC (3 M NaCl, 0.3 M sodium citrate) and then the RNA was blotted with 10 x SSC on GENE SCREEN filter (NEN) for 18 h. After blotting, the gel was stained to control the transfer. The baked filters were pre-hybridized and hybridized

at 42°C in 50% formamide, 5 x SET (0.6 M NaCl, 8 mM EDTA, 0.15 M Tris-HCl pH 7.8), 5 x Denhardt (Denhardt, 1966), 10 μg/ml of sonicated and denatured salmon sperm DNA and 1 μg/ml of yeast soluble RNA. The filters were washed at 68°C twice in 2 x SSC for 3 h and then, for higher stringency, in 0.1 or 0.2 x SSC for 2 h.

SDS-polyacrylamide gel electrophoresis

The analysis and mol. wt. determinations of zein polypeptides by SDS-PAGE were carried out according to Laemmli (1970). Acrylamide concentrations of separating gel (13 cm long) were 12%, 15% and 18% with two acrylamide-bis acrylamide ratios (75:1 or 150:1, w/w). Samples were denatured at 90°C for 5 min in 60 mM Tris-HCl pH 6.8, 1.5% SDS, 2.5% 2-mercaptoethanol, 15% glycerol and 0.08% bromophenol blue. Electrophoresis was carried out at constant 12 mA for 15–22 h depending on acrylamide concentration. Gels were stained with Coomassie blue (G-250) or fluorographed according to Laskey and Mills (1976). Mol. wt. markers (from Pharmacia) were 94 000 phosphorylase b (rabbit muscle); 67 000 bovine serum albumin; 43 000 ovalbumin (egg white); 30 000 carbonic anhydrase (bovine erythrocyte); 20 100 trypsin inhibitor (soybean); 14 000 α-lactalbumin (bovine milk): from BDH, low mol. wt. marker kit, 16 949 myoglobin; 14 404 myoglobin I and II; 8159 myoglobin I; 6214 myoglobin II; 2512 myoglobin III; from Boehringer, 68 000 bovine serum albumin; 60 000 catalase (beef liver); 25 000 chymotrypsinogen A (Bovine pancreas); 21 500 trypsin inhibitor (soybean): [^{14}C]methylated radioactive markers were from NEN, 97 000 phopshorylase B; 69 000 bovine serum albumin; 46 000 ovalbumin; 30 000 carbonic anhydrase; 18 400 lactoglobulin A; 12 300 cytochrome C. The mol. wts. reported above are those indicated by the manufacturers. 15 980 human γ-globin (from S.Ottolenghi); cold concanavalin-A was from jack bean (Sigma); radioactive concanavalin-A ([^{3}H]acetylated) was from Amersham or (^{3}H-generally labelled) was from NEN. Mature zein was labelled with [^{3}H]dansyl chloride according to Talbot and Yphantis (1971).

Acknowledgements

We are very grateful to Dr. V.Pirrotta for helpful suggestions and for reading the manuscript.

References

Argos,P., Pedersen,K., Marks,D. and Larkins,B.A. (1982) *J. Biol. Chem.*, **257**, 9984-9990.

Burr,B. and Burr,F.A. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 515-519.

Denhardt,D.A. (1966) *Biochem. Biophys. Res. Commun.*, **23**, 641-646.

Di Fonzo,N., Fornasari,E., Salamini,F., Reggiani,R. and Soave,C. (1979) *J. Hered.*, **71**, 397-402.

Gentinetta,E., Maggiore,T., Salamini,F., Lorenzoni,C., Pioli,F. and Soave,C. (1975) *Maydica*, **20**, 145-164.

Gianazza,E., Righetti,P.G., Pioli,F., Galante,E. and Soave,C. (1976) *Maydica*, **21**, 1-17.

Geraghty,D.E., Messing,J. and Rubenstein,I. (1982) *EMBO J.*, **1**, 1329-1335.

Glisin,V., Crkvenjakov,R. and Byus,C. (1974) *Biochemistry (Wash.)*, **13**, 2633-2637.

Hagen,G. and Rubenstein,I. (1981) *Gene*, **13**, 239-249.

Hague,D.R. (1975) *Plant Physiol.*, **55**, 636-642.

Heidecker,G. and Messing,J. (1983) *Nucleic Acids Res.*, **11**, 4891-4906.

Laemmli,U.K. (1970) *Nature*, **227**, 680-684.

Langridge,P. and Feix,G. (1983) *Cell*, **34**, 1015-1022.

Langridge,P., Pintor-Toro,J.A. and Feix,G. (1982) *Planta*, **156**, 166-170.

Laskey,R.A. and Mills,A.D. (1975) *Eur. J. Biochem.*, **56**, 335-341.

Marks,M.D. and Larkins,B.A. (1982) *J. Biol. Chem.*, **257**, 9976-9983.

Maxam,A.M. and Gilbert,W. (1980) in *Methods Enzymol.*, **65**, 449-560.

Misra,P.S., Mertz,E.T. and Glover,D.V. (1975) Dowden, Hutchinson and Ross, Stroudsburg, pp. 291-305.

Pedersen,K., Devereux,J., Wilson,D.R., Sheldon,E. and Larkins,B.A. (1982) *Cell*, **29**, 1015-1026.

Salamini,F. (1980) *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 467-476.

Soave,C. and Salamini,F. (1984) *Phil. Trans. R. Soc. Lond. Ser. B*, **304**, 341-347.

Spena,A., Viotti,A. and Pirrotta,V. (1982) *EMBO J.*, **1**, 1589-1594.

Spena,A., Viotti,A. and Pirrotta,V. (1983) *J. Mol. Biol.*, **169**, 799-811.

Steele,J.C.H. and Nielsen,T.B. (1978) *Anal. Biochem.*, **84**, 218-224.

Szalay,A.A., Grohmann,K. and Sinsheimer,R.L. (1977) *Nucleic Acids Res.*, **4**, 1569-1578.

Talbot,D.N. and Yphantis,D.A. (1971) *Anal. Biochem.*, **44**, 246-253.

Viera,J. and Messing,J. (1982) *Gene*, **19**, 259-268.

Viotti,A., Sala,E., Alberi,P. and Soave,C. (1978a) *Plant Sci. Lett.*, **13**, 365-375.

Viotti,A., Balducci,C. and Weil,J.H. (1978b) *Biochim. Biophys. Acta*, **517**, 125-132.

Viotti,A., Sala,E., Marotta,R., Alberi,P., Balducci,C. and Soave,C. (1979) *Eur. J. Biochem.*, **102**, 211-222.

Viotti,A., Pogna,N., Balducci,C. and Durante,M. (1980) *Mol. Gen. Genet.*, **178**, 35-41.

Viotti,A., Abildsten,D., Pogna,N., Sala,E. and Pirrotta,V. (1982) *EMBO J.*, **1**, 53-58.

Wang,J.L., Cunningam,B.A. and Edelman,G.M. (1971) *Proc. Natl. Acad. Sci. USA*, **68**, 1130-1134.

Wienand,U. and Feix,G. (1980) *FEBS Lett.*, **116**, 14-16.

Received on 12 February 1985