Sequence analysis of zein cDNAs obtained by an efficient mRNA cloning method

Gisela Heidecker*

Department of Bacteriology, University of California, Davis, CA 95616, USA, and

Joachim MessingDepartment of Biochemistry, University of Minnesota, St. Paul, MN 55108, USA

Received 21 March 1983; Revised and Accepted 24 June 1983

ABSTRACT

A cDNA library was generated from mRNA isolated from the developing endosperm of W22 maize inbred. cDNA clones for zein, the maize storage protein family, were isolated and analyzed by DNA sequencing. The DNA sequences of four clones containing cDNA copies of mRNAs belonging to one zein subfamily were determined. The data support the following conclusions: a) genes encoding the larger of the two zein species contain eleven instead of nine repeat units within the coding sequence of the gene; b) transcription can be terminated at either of the two polyadenylation signals and c) transcription starts 31 basepairs downstream from the first T in the TATA box. To facilitate this analysis a new method for the construction of cDNA libraries was developed. The mRNA was annealed to linearized and oligo-dT tailed pUC9 plasmid DNA, which then primed synthesis of the first strand of the cDNA. Oligo-dG tails were added to the cDNA-plasmid molecules, which were then centrifuged through an alkaline sucrose gradient. The gradient step removed small molecules and separated the two cDNAs which were formerly attached to the same double stranded plasmid molecule. An excess of oligo-dC tailed denatured pUC9 DNA was added and the DNA was renatured under conditions that favor the circularization of monomers by the oligo-dC and oligo-dG tails. The oligo-dC tail served as primer for the synthesis of the second strand of the cDNA. The library was screened by colony hybridization using ³²P-labelled cDNA and DNA from genomic zein clones as probes. We obtained 20,000 clones hybridizing total cDNA starting with 1 µg of plasmid DNA and 1 µg of mRNA.

INTRODUCTION

Zeins, the major storage proteins in the corn kernel, consist of a group of alcohol soluble proteins (1). They account for up to 60% of the total protein in the mature seed. Previous protein studies showed that they share a common amino acid composition (2,3) and that the amino terminus is conserved at 22 of 33 positions (4). Electrophoretically, zeins can be separated into two major (21 and 19 kd) and several minor bands (10 to 15 kd) on polyacrylamide/dodecylsulfate gels (2, 5). On two dimensional gels the proteins can be separated into at least 25 different species (6). These data suggest that the proteins are encoded by a multigene family similar to the chorion multigene family in *Bombyx mori* (7).

Similar to the chorion proteins, zeins are synthesized in high amounts and only in one specific organ of the organism, i.e. the endosperm of the developing kernel 18 to 32 days after pollination. During this time span the membrane bound mRNA isolated from the endosperm directs the synthesis of zeins in an in vitro translation system (8, 9, 10). Using cDNA clones, mRNAs were divided into 5 subfamilies by hybrid arrest translation experiments (11, 12, 13) and dot blot analysis under high stringency conditions (14, 15). Southern blot hybridization experiments with individual cDNA clones as probes showed heterogeneity within the subfamilies (16). Lowering the stringency of hybridization in a stepwise manner demonstrated that all members of the zein family show at least 60% homology (12).

To investigate the sequence divergence between and within the subfamilies we generated a cDNA library from the mRNA of the developing endosperm of W22 maize inbred. In this report we describe the generation of the cDNA library and the sequence data for four cDNA clones belonging to the same subfamily. The cDNA library was constructed by a method which combined efficiency of yield and completeness of the cDNA copies with the simplicity of previously described methods (17, 18, 19). The DNA sequence analysis of the four cDNA clones reveals that genes within one subfamily can encode both the 19 and 21 kd proteins. The length of the peptide chain depends upon the number of repetitive units which have been previously found to account for most of the structural part of zein genes (20, 21, 22, 23). Furthermore, the sequence data suggests that the start of transcription occurs 31 basepairs downstream from the first T in the TATA box and that there is heterogeneity in respect to which of the two polyadenylation signals, frequently found in plant genes (24), is utilized.

MATERIALS AND METHODS

Strains

The pUC plasmids, the M13mp phage vectors and the strains JM103 and JM83 have been described elsewhere (25, 26).

Media, Transformation, Histochemical assay, Chemicals and Enzymes

Media and plates were as described by J. Miller (27). Transformation of competent cells was as described by Cohen et al. (28) except that the CaCl_2 concentration was raised to 50 mM. Cells were transfected with M13mp phage and plated in soft agar with 5×10^8 fresh cells in the presence of 1 mM IPTG and 0.004% X-gal. Cells transfected with pUC plasmids were allowed to grow for 40 min in the absence of selection before they were streaked on YT plates

containing 100 $\mu\text{g/ml}$ ampicillin, 1 mM IPTG and 0.004% X-gal. IPTG was omitted when JM83 was used.

IPTG (Isopropyl- β -D-galactopyranoside) was obtained from Sigma, X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) from Bachem, dNTPs and ddNTPs from PL biochemicals, radiolabelled dNTPs from Amersham and RNasin from biotec. All enzymes were either purchased from Bethesda Research Laboratories or New England Biolabs except for terminal deoxynucleotidyl transferase (terminal transferase) and reverse transcriptase, which were obtained from Ratilff Biochemicals and Dr. J. Beard, respectively.

Tailing Reactions

20 μg of pUC9 DNA were digested with PstI and the degree of digestion was monitored by agarose gel electrophoresis. The DNA was extracted with phenol and phenol/chloroform and ethanol precipitated twice to ensure the removal of residual phenol. The precipitate was washed with 70% ethanol and dried under vacuum for 15 min. The DNA was resuspended in 20 μl ddH₂O and divided into two aliquots. For the oligo-dT-tailing reaction 10 μl 1 M K-cacodylate pH 7.0, 23 μl H₂O, 0.5 μl 0.1 M DTT, 2 μl 1 mM ^{32}P - α -dTTP, 1.5 μl terminal transferase (16U/ μl) and 5 μl 10 mM CoCl₂ were added in the listed order. The components for the oligo-dC-tailing reaction were the same except that 2 μl 0.5 mM ^{32}P - α -dCTP were used instead of the dTTP. The mixtures were incubated at 37 °C for 30 min. The DNA was phenol extracted and ethanol precipitated three times to remove the Co²⁺ which would interfere in the reverse transcriptase reaction. The tailed plasmid DNAs were finally resuspended in 10 μl of low Tris buffer (10 mM Tris pH 7.6, 10 mM NaCl, 1 mM EDTA).

To tail the cDNA-plasmid conjugates the DNA was resuspended in 10 μl ddH₂O, 4 μl 1 M K-cacodylate, 1 μl 1 mM dGTP, 1 μl 0.05 M DTT, 2 μl 20 mM MnCl₂ (29) and 1 μl terminal transferase (16U/ μl). The reaction was incubated for 15 min at 37 °C. The DNA was phenol extracted, ethanol precipitated and dissolved in 50 μl low Tris buffer.

cDNA Synthesis

The cDNA synthesis reaction was done in a final volume of 15 μl and incubated 90 min at 37 °C. The components were: 800 μM dATP, dCTP, dGTP, dTTP, 70 mM KCl, 50 mM Tris pH 8.2, 10 mM MgCl₂, 2 mM DTT, 1 U/ μl RNasin, 25 $\mu\text{g/ml}$ actinomycin D, 40 nM oligo-dT-tailed pUC9 DNA (1 μg), 250 nM RNA, 100 U/ml reverse transcriptase. The poly A-mRNA and the oligo-dT-tailed plasmid DNA anneal under these conditions during the reaction. After the reaction the DNA was phenol extracted once and ethanol precipitated three times. To prevent the precipitation of unincorporated nucleotides the ethanol precipi-

tates were warmed to room temperature before centrifugation.

Alkaline Sucrose Gradient Centrifugation

A 5 ml linear sucrose gradient was used (5-20% sucrose w/v in 0.2 M NaOH, 0.8 M NaCl, 1 mM EDTA with a 0.5 ml 60% sucrose cushion). The sample was diluted with 50 μ l of the 5% sucrose solution and layered on the gradient. Centrifugation was carried out in a SW 50.1 rotor at 36k rpm for 17 hr at 4°C. The gradient was collected from the bottom in 0.3 ml fractions. The profile of the gradient was established based on the Cerenkov radiation of the fractions.

Reannealing and Circularization of the cDNA-Plasmids

The amount of plasmid in the pooled fractions was calculated based on the relative amount of radioactivity and oligo-dC-tailed pUC9 DNA was added in 5 to 10 fold excess. The solution was then dialyzed against low Tris buffer in the cold to remove the NaOH and NaCl. The DNA was concentrated in the presence of 25 μ g/ml carrier RNA and resuspended in 50 μ l low Tris buffer. Concentrated NaCl, Tris pH 8., formamide and ddH₂O were added to give final concentrations of 1-5 μ g/ml plasmid DNA, 32% (v/v) formamide, 50 mM NaCl, 10 mM Tris (30, 31). The annealing mix was incubated for 24 hr at 37° C, dialyzed against 100 mM NaCl, 10 mM Tris pH 8.0, 1 mM EDTA in the cold and concentrated by ethanol precipitation.

Fill-in Reaction

The annealed DNA was taken up in 50 μ l of cold 50 mM NaCl, 20 mM Tris pH 7.6, 10 mM MgCl₂, 1 mM DTT, 100 μ M dATP, dCTP, dGTP, dTTP. 3 units of DNA polymerase I-large fragment were added and the mixture was incubated for 60 min at 15 C and 60 min at room temperature. The DNA was phenol extracted, ethanol precipitated and resuspended in 50 μ l low Tris buffer.

Isolation of Maize Endosperm mRNA

Maize kernels were harvested 22 days after pollination and mRNA was extracted as described by Burr *et al.* (9). The mRNA was size fractionated by centrifugation through a 5-20% sucrose DMSO gradient and purified subsequently by passage through an oligo-dT cellulose column. The RNA was assayed for biological function in a cell free translation system as described by Park *et al.* (12).

DNA Sequencing

The cDNA inserts were sequenced using the chain termination method (32) and the M13 subcloning procedure (33) except that a synthetic universal primer was used. A detailed protocol for the subcloning strategy and sequencing procedure is presented elsewhere (34). Storage and processing of

the sequencing data was conducted with the Apple II Microcomputer (35, 36).

Electrophoresis

Sequencing gels were essentially prepared as described elsewhere (37), except that for long runs 5% gels (acrylamide/bisacrylamide 40:1) were used. Agarose gel electrophoresis and visualization of nucleic acids were carried out as described elsewhere (33). Denaturing agarose gels contained 5 mM CH_3HgOH and were prepared in 50 mM boric acid, 5 mM $\text{Na}_2\text{B}_4\text{O}_7$, 10 mM Na_2SO_4 , 1 mM EDTA Na_4 pH 8.2.

RESULTS AND DISCUSSION

Construction of the cDNA Library

The major steps of the procedure are outlined in Fig. 1. The mRNA is annealed to oligo-dT-tails which had been added at the PstI site of pUC9 plasmid DNA. The oligo-dT-tails prime the cDNA synthesis along the mRNA. The plasmid-cDNA conjugates are in turn extended with oligo-dG-tails, denatured and sized by centrifugation through an alkaline sucrose gradient. Molecules of the appropriate length are renatured under dilute conditions in the presence of an excess of oligo-dC-tailed, single-stranded pUC9 DNA. The low DNA concentration favors the formation of monomers circularized by hybridization between the dC- and dG-tails. The second strand of the cDNA is primed by the oligo-dC-tail and synthesized by DNA polymerase I-large fragment.

Step 1: Preparation of Primer Vector and Second Strand Vector

pUC9 plasmid DNA was prepared as previously described (25) and cleaved with restriction endonuclease PstI. The tailing reactions were carried out as described in Materials and Methods. The reactions were monitored in three ways. The total number of nucleotides incorporated per molecule of plasmid was calculated from the proportion of TCA precipitable radioactivity. To determine the length of the individual tails an aliquot of the tailing reaction was digested with EcoRI. This releases one tail per plasmid molecule which can be visualized by electrophoresis through a denaturing polyacrylamide gel followed by autoradiography. The resulting picture showed a ladder (data not shown). Based on a sequencing reaction of a known template which was run in parallel the steps of the ladder corresponded to chains of 65 to 85 nucleotides for the dT-tailing reaction and 45 to 65 for the dC-tailing reaction. The actual tails are shorter by the distance between the EcoRI and PstI site which is 26 bp. The integrity of the plasmid DNA was tested by digesting another aliquot with HaeII. The fragments were

Figure 1: Flowsheet for the cDNA cloning procedure. Explanations are given in the text. E,B,S, and H stand for the EcoRI, BamHI, SalI and HindIII sites in the multicloning site of pUC9.

separated into four bands on a 1.5% agarose gel. The smallest two bands had shifted up in comparison to a HaeIII digest of PstI cleaved pUC9 DNA (Fig. 2). The autoradiograph of this gel showed that greater than 90% migrated with the lower two bands. Higher labelling of the other bands would have indicated that the plasmid DNA had been nicked before or during the tailing reaction to

Figure 2: Analysis of oligo-dT-tailed pUC9 DNA. Homopolymer tails of deoxythymidilate were added to PstI cleaved pUC9 DNA. An aliquot of this DNA was digested with HaeIII and the resulting fragments were separated on a 1.5% agarose gel (lane 3). The autoradiograph (lane 4) of the gel reveals that >90% of the radiolabel resides in the smallest two fragments. These two fragments show a decreased mobility when compared to non-tailed pUC9 DNA cleaved with PstI and HaeIII (lane 2). Lane 1 shows M13mp2 RF DNA digested with HaeIII, the sizes of some fragments are given.

a degree that would interfere with the following steps in the procedure.

Step 2: cDNA Synthesis

mRNA from developing seeds of corn Inbred W22 was isolated as described in Materials and Methods. The mRNA was converted into cDNA using the oligo-dT-tailed plasmid DNA as primer (see Materials and Methods). A three fold molar excess of mRNA had previously been determined to saturate the system. Under these conditions about 60% of the tails primed cDNA synthesis as judged by comparing equimolar amounts of oligo-dT-tailed plasmid and plasmid-cDNA conjugates on a denaturing agarose gel (data not shown). The length of the cDNA transcripts can be estimated from the same gel.

Step 3: Addition of oligo-dG Tails

The cDNA-plasmid molecules were tailed in the presence of dGTP and $MnCl_2$.

This combination results in optimal (though not equally efficient) tailing of any kind of end (29). There is no easy way of monitoring this reaction. As both the molar concentration (see below) and the kind of ends in the reaction are unknown the incorporation of labelled dGTP does not allow a calculation of the average tail length. It was, therefore, necessary to perform this reaction under conditions where both dGTP and terminal transferase were present in excess. Under these conditions the length of the tails depended on the incubation time and the conditions given in Materials and Methods resulted in the addition of 10 to 25 residues of dGMP. This had been established in a control experiment previously and was verified by sequencing data.

Step 4: Sizing and Strand Separation of cDNA-Plasmids

After the dG-tailing reaction the molecules were fractionated on an alkaline sucrose gradient (see Materials and Methods). The vector and cDNA-vector molecules banded in a well defined peak separated by four fractions from a gradually rising slope of small molecules (Fig. 3). The gradient served four purposes:

- Enrichment of molecules longer than the vector itself. This will increase the proportion of clones that contain cDNA inserts over those that just contain "tails".
- Elimination of the short DNA molecules. These small molecules seem to be generated during the cDNA synthesis step and may be primed by RNA fragments or by the poly-A tail of the mRNA hybridized to the

Figure 3: Profile of cDNA-plasmid conjugates on an alkaline sucrose gradient. Oligo-dT-tailed pUC9 DNA was used to prime cDNA synthesis with maize endosperm mRNA serving as template. The cDNA-plasmid molecules were extended with terminal transferase in the presence of dGTP and subjected to alkaline sucrose gradient centrifugation as described in Materials and Methods. Sedimentation is from right to left.

primer vector. As these molecules also carry oligo-dG tails they would interfere in the following step, especially as their molar concentration is difficult to determine.

- c) Removal of the RNA by base hydrolysis.
- d) Separation of the two strands of plasmid DNA and, thus, the two cDNAs which were formerly attached to the same double-stranded vector molecule. This is necessary in order to obtain plasmid molecules with a single cDNA insert.

Step 5: Reconstitution and Circularization of the Plasmid

To reconstitute a double-stranded circular plasmid molecule with only one cDNA insert, 3 μ g of oligo-dC tailed pUC9 DNA were added to the pooled fractions 4 to 6. This represented a seven to ten fold excess of dC-tailed plasmid over dG-tailed cDNA-plasmid. Base and salt were removed and the DNA was reannealed in the presence of formamide. The conditions given in Materials and Methods result in greater than 90% reannealing of the DNA as judged from an agarose gel on which the position of the reannealed material was compared to native and denatured pUC9 DNA (data not shown). The formamide was gradually removed by dialysis to allow circularization of molecules with dC- and dG-tails. As the DNA concentration was low, circularization was more likely to occur than the formation of concatemers. Finally, the DNA was concentrated by ethanol precipitation and the second strand of the cDNA was filled in by large fragment of DNA polymerase I.

Step 6: Screening of the cDNA Library

Transformation of 2.5% of the final product of the cloning procedure gave about 500 white and 100 blue colonies. The blue colonies most likely arose from uncleaved vector DNA present in the oligo-dC tailed preparation. The white colonies were screened for the presence of cDNA inserts by colony hybridization using ³²P-labelled, randomly primed cDNA or nick translated DNA from two genomic Zein clones representing two subfamilies (38). About 80% of the clones hybridized to the cDNA probe, and about 20% hybridized to the two combined zein probes (data not shown). Colonies that only hybridized weakly to the Zein probes were later shown to contain cDNAs belonging to a different subfamily. Extrapolating the number of clones that hybridized to the various probes, we estimate that 30 to 40% of our clones contained Zein cDNAs. This proportion is surprisingly low as Zein proteins were by far the most prevalent products when the mRNA used in this experiment was translated in vitro (8, 9, 10, 11). To determine the size of the inserts plasmid DNA from 60 white colonies was isolated and sized by agarose gel electrophoresis.

Twelve had inserts of about 1200 bp, 7 carried short inserts of 100 to 200 bp and the remaining 41 had inserts between 400 and 900 bp (data not shown).

COMMENTS

The method presented here describes a scheme that allows the synthesis and cloning of cDNA copies of mRNA in only 6 steps. The three enzymes that are used in the procedure are all commercially available. Except for the oligo-dG-tailing reaction of the cDNA-plasmid conjugates every reaction in the scheme can be easily monitored and thus controlled. The length of the cDNA inserts primarily depends on the length of the mRNA. In the experiment reported here 400 to 1200 nucleotide long mRNAs were used and both the length of the first strand cDNA copies, as well as the length of the inserts in the final clones, reflected this size distribution. In a pilot experiment using 9S RNA from rabbit reticulocytes a large proportion of the cDNA-plasmid conjugates were of uniform length and indicated that first strand cDNA transcripts of 550 nucleotides were the major product. On the bases of length and restriction pattern, about 50% of the cDNA clones isolated in this experiment had inserts that were full length cDNA copies of globin mRNAs (data not shown). First strand cDNA copies of mouse mammary tumor virus RNA prepared under these conditions again showed the same size distribution as the RNA (1000 to 9000 nucleotides) (data not shown). The only step in the procedure that might reduce the length of the cDNA and still result in its being cloned is the oligo-dG-extension of the cDNA-plasmid conjugates. This can be prevented by using terminal transferase that is free of nuclease activity.

Even though any plasmid that contains a cloning site with 3' protruding ends could be used with this protocol, the pUC plasmids are particularly suitable. Since the PstI site is flanked by several other restriction sites the cDNA insert can be released by one (pUC7) or two (pUC8, 9, 12, 13) restriction enzymes. Furthermore, by choosing the right pUC plasmid the procedure can be adapted to experimental goals that go beyond the simple cloning of cDNA. In the scheme presented here cDNA inserts in both orientations are obtained. By removing one of the oligo-dT-tails with the appropriate restriction enzyme before the first strand cDNA synthesis all the resulting clones should contain cDNA copies in either the sense or nonsense orientation in respect to the lac promoter. This allows the introduction of defined deletions from either the 5' or 3' end of the insert in subsequent experiments. Insertion of the cDNA in the sense orientation should also result in the expression of fusion proteins between the aminoterminal end of

beta-galactosidase and cDNA encoded peptides in some of the clones. This would allow one to screen the library by immunological assays and make it possible to isolate genes for which only the protein has been isolated so far (40).

Since the M13mp vectors contain the same array of restriction sites as the pUC plasmids, sequences cloned in pUC plasmids can be shuttled between these two vector systems and the cloned DNA can be easily isolated as single stranded DNA. This is of use for the preparation of strand specific hybridization probes (41), site specific mutagenesis with oligonucleotides (42) and for DNA sequencing by the chain termination method (32). If the cDNA is cloned directly into the M13mp vectors for sequencing, then the universal primer directs DNA synthesis through a run of dA or dG (33). Depending on the length of these homopolymers, the Sanger sequencing reaction (32) is inhibited (not shown). Most sequencing studies require at least one subcloning step and, thus, we do not feel that it is necessary to use the M13mp vectors as the primary cloning vehicle in the cDNA-plasmid procedure. There are several important reasons for using the pUC plasmids instead of the M13mp vectors. First, pUC is a smaller vector which allows a better separation of the cDNA-plasmid conjugates from the vector molecules. Second, pUC plasmids are not restricted to a male host. Thus, bacterial strains that exhibit high transformation efficiencies can be used (43). Although strains with the M15 deletion either in the F- traD36 or in the *E. coli* chromosome, are host strains (44) which allow pUC plasmids with inserts to be monitored by a simple color reaction (25), this test is usually superfluous when used with recombinant DNA libraries. If the library has to be amplified before screening, M13 phage containing inserts have a severe growth disadvantage when compared to "empty" phage, which is in contrast to lambda and plasmid libraries (45).

Sequence Analysis of cDNA Clones

Four Zein cDNA clones that hybridized labelled Z4 DNA (38) were isolated. Based on the restriction maps of these clones, Sau3A, AluI, RsaI and EcoRI* were used to fragment the cDNA inserts. The fragments were subcloned into M13mp vectors both in a shotgun fashion and by forced cloning (46, 26). The M13mp subclones were sequenced by the chain termination method (32).

Figure 4 shows a comparison of the four cDNA sequences ZG7 and ZG19, ZG31 and ZG124 to the sequence of the genomic clone Z4 (21). Also included in the comparison is the sequence of cDNA clone A30 (20, 47), which is the

Figure 4: Comparison of the nucleotide sequences of cDNA clones belonging to the Z4 family of zein genes. The sequences are numbered starting with the first transcribed nucleotide. The translated part of the sequence is written in triplet form. Dashes indicate nucleotides not present in cDNA clones ZG124, A30, ZG31 and ZG19. The complete sequence of genomic clone Z4 is presented, but only the variable nucleotides for the cDNAs are given at the site of their occurrence. The sequences of Z4, A30 and ZG31 were previously published (21).

copy of a mRNA isolated from a different inbred line (Illinois High Protein).

ZG7 and ZG124 are probably full length cDNA copies as they have a common start which is located 31 bp downstream from a consensus TATA sequence in the genomic clone Z4 (24). Two other cDNA clones (ZG14 and ZG15; Heldecker and Messing, in preparation) which belong to a different subfamily start at this position, whereas all other cDNAs sequenced vary at their 5' ends. Both ZG124 and ZG7 start with a Guanidyl residue which does not match the genomic sequence. This G-residue could be due to contaminating nucleotides either in

the dG-tailing reaction of the cDNA-plasmids or in the dC-tailing reaction of the plasmid. Alternatively, it could be the beginning of a loop back at the end of the cDNA. The latter explanation seems more likely as none of the tails show any evidence of contaminating nucleotides. Moreover, we observed the same GATC palindrome at the start of the other two full length cDNA sequences (ZG14 and ZG15), which were mentioned above.

The comparison shows that overall the sequences are indeed very highly conserved, the lowest degree of homology between any two clones being 95%. However, even this closely related subfamily can again be divided into two subgroups. cDNA clones ZG19, ZG31, ZG124 and A30 share most of the 40 differences in comparison to the genomic clone Z4. Any pairwise comparison shows at least 99% homology. cDNA clone ZG7 and Z4 share an internal duplication of 96 basepairs and an extra codon in comparison to the other four clones. These two sequences differ by only 10 single nucleotide exchanges. The extensive colinearity between Z4 and ZG7 identifies Z4 as an active gene and supports our previous data (21) showing that our genomic clone does not contain any intervening sequences but represents a member of the larger size class of zein genes. Because of the lack of matching cDNA and genomic clones a more indirect approach, including electron microscopy and S1 mapping, has been used in other studies to document the absence of intervening sequences in zein genes (22, 48).

ZG7 and ZG19 share a remarkable feature. Both clones originated from mRNAs that terminated after the first of the two polyadenylation signals frequently found in plant genes (24). The distance between the signal and the site of polyadenylation is about the same for all five cDNA clones. Thus, plant genes may contain either one or two polyadenylation signals and in at least some cases where two signals are present, either signal can be recognized. This situation has been found for only a few animal genes (49, 50, 51). It remains to be elucidated whether this feature is used to regulate or control gene expression at the mRNA transcription and/or processing level, or whether it is just a random event without any particular consequences.

The proteins encoded by the 6 closely related genes described in this report represent both major zein size classes. The molecular weights predicted from the nucleotide sequence all are higher by about 4 kd than those determined by SDS/polyacrylamide electrophoresis. The latter values most likely are underestimates due to the hydrophobicity of the proteins. The molecular weights of the proteins deduced from the sequences of Z4 and

ZG7 are 29.2 kd for the precursor and 27 kd for the mature protein. The sequences of ZG124 and A30, and probably of ZG31 and ZG19, can likewise be translated into proteins of 25.4 kd and 23.3 kd for the two forms. These data do not support a report by Marks and Larkins (15) who found that each subfamily, as defined by cross-hybridization under stringent conditions, directed the synthesis of proteins of only size class. These conflicting results may be caused by differences in the inbred line, mRNA isolation, and/or peculiarities of in vitro translation experiments. Park et al. (12) reported translation arrest experiments which suggested that clone A30 hybridized with mRNA that coded for both the higher and the lower mol. wt. species of proteins in most of the inbred lines used in these studies. Taking into account the variables described above, the hybridization data of the latter report agree very well with our sequencing data. Although 2 out of 6 cloned sequences of this subfamily code for the larger protein, a quantitative representation of both size classes cannot be calculated, because these numbers are too small to allow a valid statistical analysis. On the other hand, sequencing 5 clones from the same subfamily in the same inbred line did not result in any identical sequences, which gives us an indication that this subfamily alone must be far more complex than 5 members.

ACKNOWLEDGEMENTS

We would like to thank Dave Pratt, John Ingraham, Bob Cardiff, Irwin Rubenstein and Thomas Fanning for their help and discussion throughout the work and Ida Fierro and Kris Kohn for their aid in preparing this manuscript. This research was supported by the Department of Energy, DE-AC02-81ER 10901, the National Institutes of Health, St32 GM07467-05, and the Minnesota Exp. Station, MN-15-030.

*Permanent address: Department of Pathology, School of Medicine, University of California, Davis, CA 95616, USA

REFERENCES

1. Wall, J. S. and Paulis, J. W., in Pomeranz, Y (ed.). *Adv. Cereal Science and Technol.*, Vol. II. Am. Assoc. Cereal Chem. St. Paul, Minn, pp. 135-219, 1975.
2. Lee, K. H., Jones, R. A., Dalby, A. and Tsal, C. (1976) *Biochem. Genet.* 14, 641-650.
3. Gianazza, E., Viglenghi, V., Rhigetti, P. G., Salamini, F. and Soave, C. (1977) *Phytochemistry* 16, 315-317.
4. Bietz, J. A., Paulis, J. W. and Wall, J. S. (1979) *Cereal Chem.* 56, 327-332.

5. Gianazza, E., Rhigetti, P. G., Pioli, F., Galante, E. and Soave, C. (1976) *Maydica* 21, 1-17.
6. Hagen, G. and Rubenstein, I. (1980) *Plant Sci. Lett.* 19, 217-223.
7. Eickbush, T. H., Jones, C. W. and Kafatos, F. C. In Brown, D. D. (ed.). *ICN-UCLA Symposia on Molecular and Cellular Biology*, Vol. XXIII. Academic Press, New York, pp. 135-153, 1981.
8. Larkins, B. A., Jones, R. A. and Tsai, C. Y. (1976) *Biochemistry* 15, 5506-5511.
9. Burr, B., Burr, F. A., Rubenstein, I. and Simon, M. N. (1978) *Plant Sci. Lett.* 13, 365-375.
11. Wienand, U., Bruschke, C. and Felix, G. (1979) *Nucl. Acids Res.* 6, 2707-2715.
12. Park, W. D., Lewis, E. D. and Rubenstein, I. (1980) *Plant Physiol.* 65, 98-106.
13. Viotti, A., Abildsten, D., Pogna, N., Sala, E. and Pirrotta, V. (1982) *EMBO J.* 1, 53-58.
14. Burr, B., Burr, F. A., St. John, T. P., Thomas, M. and Davis, R. W. (1982) *J. Mol. Biol.* 154, 33-49.
15. Marks, M. D. and Larkins, B. J. (1982) *J. Biol. Chem.* 257, 9976-9983.
16. Hagen, G. and Rubenstein, I. (1981) *Gene* 13, 239-249.
17. Maniatis, T., Kee, S. G., Efstratiadis, K. and Kafatos, F. C. (1976) *Cell* 8, 163-182.
18. de Martynoff, G., Pays, E. and Vassart, G. (1980) *Biochem. biophys. Res. Comm.* 93, 645-653.
19. Okayama, H. and Berg, P. (1982) *Mol. Cell. Biol.* 2, 161-170.
20. Geraghty, D., Pelfer, M. A., Rubenstein, I. and Messing, J. (1981) *J. Nucl. Acids Res.* 9, 5163-5174.
21. Hu, N.-T., Pelfer, M. A., Heldecker, G., Messing, J. and Rubenstein, I. (1982) *EMBO J.* 1, 1337-1342.
22. Pedersen, K., Devereux, J., Wilson, D. R., Sheldon, E. and Larkins, B. A. (1982) *Cell* 29, 1015-1026.
23. Argos, P., Pedersen, K., Marks, M. D. and Larkins, B. (1982) *J. Biol. Chem.* 257, 9984-9990.
24. Messing, J., Geraghty, D., Heldecker, G., Hu, N.-T., Kridl, J. and Rubenstein, I., In Hollaender, A., Kosuge, T. and Meredith, (eds.). *Genetic Engineering of Plants*. Plenum Press, New York, pp 211-228, 1983.
25. Vieira, J. and Messing, J. (1982) *Gene* 19, 259-268.
26. Messing, J. and Vieira, J. (1982) *Gene* 19, 269-276.
27. Miller, J. *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, Cold Spring harbor, NY, 1972.
28. Cohen, S. N., Chang, A. C. Y. and Hsu, L. (1972) *Proc. Natl. Acad. Sci. USA* 69, 2110-2114.
29. Deng, G. and Wu, R. (1982) *Nucl. Acids Res.* 9, 4172-4188.
30. Fanning, T. G., Schreier, P. F. and Davies, R. W. (1976) *Eur. J. Biochem.* 62, 173-170.
31. Dugalczyk, A., Boyer, H. W. and Goodman, H. (1975) *J. Mol. Biol.* 96, 171-184.
32. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-4567.
33. Heldecker, G., Messing, J. and Gronenborn, B. (1980) *Gene* 10, 69-73.
34. Messing, J., in Wu, R. (ed.). *Methods in Enzymology*, Vol. 101. Academic Press, New York, pp. 29-78, 1983.
35. Larson, R. and Messing, J. (1982) *Nucl. Acids Res.* 10, 39-49.
36. Larson, R. and Messing, J. (1983) *DNA* 2, 31-35.
37. Sanger, F. and Coulson, A. R. (1978) *FEBS Lett.* 87, 107-110.

38. Lewis, E. D., Hagen, G., Mullins, J. I., Mascia, P., Park, W. D., Benton, W. D. and Rubenstein, I. (1981) *Gene* 14, 205-215.
39. Holmes, D. S. and Quigley, M. (1981) *Anal. Biochem.* 114, 193-197.
40. Hellman, D. M., Feramisco, J. R., Fiddes, J. C., Thomas, G. P. and Hughes, S. H. (1983) *Proc. Natl. Acad. Sci. USA* 80, 31-35.
41. Hu, N.-T. and Messing, J. (1982) *Gene* 17, 271-277.
42. Zoller, M. J. and Smith, M., in Wu, R. (ed.). *Methods in Enzymology* Vol. 101. Academic Press, New York, in press.
43. Morrison, D. A., in Wu, R. (ed.). *Methods in Enzymology*, Vol. 68. Academic Press, New York, pp. 326-331, 1979.
44. Messing, J. *Recombinant DNA Technical Bulletin*, NIH Publication No. 79-99, 2, No. 2 (1979) 43-48.
45. Messing, J., in Setlow, J. and Hollaender, A. *Genetic Engineering*, Vol. 4, Plenum Publishing Company, New York, pp. 19-35, 1982.
46. Messing, J., Crea, R. and Seeburg, P. H. (1981) *Nucl. Acids Res.* 9, 309-321.
47. Geraghty, D., Messing, J. and Rubenstein, I. (1982) *EMBO J.* 1, 1329-1335.
48. Wienand, U., Langridge, P. and Felix, G. (1981) *Mol. Gen. Genet.* 182, 440-444.
49. Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980) *Cell* 20, 313-319.
50. Setzer, D. R., McGrogan, M., Nunberg, J. H. and Schimke, R. T. (1980) *Cell* 22, 361-370.
51. Tosi, M., Young, R. A., Hagenbuchle, O. and Schibler, V. (1981) *Nucl. Acids Res.* 9, 2313-2323.
52. Grunstein, M. and Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* 72, 3961-3965.