Primary Structure of a Proline-Rich Zein and Its cDNA

SHU-ZHEN WANG AND ASIM ESEN*
Department of Biology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

ABSTRACT

Eighty-five cDNA clones for γ-zein (proline-rich zein) from a cDNA expression library were isolated using specific antibody and cDNA probes. Nucleotide sequences of seven independent clones were determined and found to be identical in regions where they overlapped. The primary structure of the mature protein, determined from the sequence of one near full-length clone, consists of 204 amino acids. It has a molecular weight of 21,824 daltons, about 5 kilodaltons less than that estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The N-terminal one-half of the sequence contained eight essentially identical tandem repeats of the hexapeptide Pro-Pro-Pro-Val-His-Leu and two of the octapeptide Gln-Pro-His-Pro-Cys-Pro-Cys-Gln. The codon specifying the third proline in the hexapeptide repeating unit is identical (CCG) in all of the eight repeats. The coding region has a very high G-C-C content (69.8%). The multiple charge components of γ-zein detected by isoelectric focusing do not seem to be encoded by members of a multigene family. Moreover, it was found that the codon preference in γ-zein is, in fact, the base preference in the wobble position. A codon usage value was devised to express this phenomenon.

Storage proteins in maize endosperm are classified into four solubility groups: albumin, globulin, prolamin, and glutelin (17). The prolamins of maize, zein, is a group of highly hydrophobic proteins insoluble in aqueous buffers but readily soluble in organic solvents, alcohol in particular. According to a nomenclature system used by the junior author, there are three classes of zein: α, β, and γ. These classes differ from each other with respect to solubility, mol wt, amino acid composition, and sequence and immunological properties. α-Zein (80% of total zein) (9) consists of 20 to 24 kD polypeptides and has been studied extensively at both protein and nucleic acid levels (4, 11, 16, 18). Viotti et al. (21) estimated that there were 120 copies of α-zein gene per haploid genome. β-Zein (15% of total zein) (9) consists of 17 and 18 kD polypeptides and has also been studied at both protein and nucleic acid levels.

γ-Zein is the least abundant zein (5–10% of total zein) (9) and is soluble in alcohol as well as in water in the presence of a reducing agent. When separated by SDS-PAGE, γ-zein gives a broad (diffuse) band with an estimated mol wt of 27 kD (6). The analysis of γ-zein fraction isolated by differential solubility at low pH indicated the presence of at least two size classes (14). As for the charge heterogeneity, one can readily detect 14 to 16 components focusing between pH 6.0 and 9.5 after IEF1 in polyacrylamide gels containing 6 M urea (6). Based on these observations, it was postulated that γ-zein was encoded by a multigene family.

Esen et al. (7) determined the sequence covering the first 58 residues of the N-terminal region by amino acid sequence analysis. They found that tandem repeats of a hexapeptide unit, Pro-Pro-Pro-Val-His-Leu, occurred at least six times in the N-terminal sequence. In the present paper we report additional data on γ-zein including the nucleotide sequences of seven independent γ-zein cDNA clones and the complete deduced primary structure of the mature protein.

MATERIALS AND METHODS

Polysome Isolation and in Vitro Translation. Polysomes from developing maize kernels of inbred line SSA419 were isolated 26 d after pollination according to the procedure of Vodkin (22). Polysome bound mRNAs were in vitro translated in a Wheat Germ System (BRL) according to the procedure of the supplier. Translation products were centrifuged for 10 min in a microfuge at 4°C to remove insoluble zeins (α and β) which otherwise coprecipitate with antibody-γ-zein complexes later. Antibody against γ-zein was added to the supernatant fluid and the mixture was incubated at 4°C for 4 h. Then goat antirabbit IgG (GAR) was added, incubated at 4°C overnight, and centrifuged. The pellet was dissolved in SDS gel sample buffer (13) and analyzed by SDS-PAGE. The gel was treated with EN3 HANCE (New England Nuclear) as described by the supplier, dried, and autoradiographed for 2 to 4 d at −70°C to visualize the radioisotope-labeled protein. Antiserum against γ-zein was prepared as described by Esen et al. (8).

cDNA Synthesis and Isolation of γ-Zein Clones from an Expression Library. cDNA synthesis, construction of cDNA expression library, immunoscreening, and hybridization screening of the library with antibody, and labeled γ-zein cDNA probes were according to the procedure of Helfman et al. (12) and its modification (23). cDNA inserts were cleaved out from a total of 85 positive clones by digestion of the recombinant plasmid with Eco RI and Sal I, their lengths were compared after agarose gel electrophoresis, and seven clones with different lengths of inserts were selected for final characterization.

Subcloning of γ-Zein cDNA for Sequencing. γ-Zein cDNA inserts from six clones (γZM5X, 8-1, 8-3, 8-7, 11-6, and 11-7) isolated by immunoscreening of cDNA expression library, and one clone (γZGS71) obtained by hybridization screening were subjected to DNA sequence analysis. They were double digested out of the pUC8 plasmid by Eco RI and Sal I and subcloned into M13 mp10 and mp1 in order to sequence both the 5' and 3' ends. Clone γZM5, the longest insert among the ones selected by antibody probe, was partially and completely digested with Pst I. The digested fragments were purified by electrophoresis and subcloned into M13 mp10.

DNA sequencing was performed following the Sanger's dideoxy method using 32PdATP as label (2). Single stranded DNA preparation, sequencing reactions, and gel electrophoresis were performed following the procedures (1) provided by the supplier of reagents (Amersham).

___

1 Abbreviation: IEF, isoelectric focusing.
RESULTS

In Vitro Translation and Immunoprecipitation of the Product. Polysome bound mRNAs of inbred line SSA419 were in vitro translated and then the in vitro translation product was immunoprecipitated. The precipitate produced a narrow band (Fig. 1A) on SDS-PAGE, contrary to that observed for mature protein which usually produces a broad and diffuse band on stained SDS-PAGE gel (6). In addition, to rule out the possibility that antibody might not react with all the ‘forms’ or ‘bands’ of γ-zein, they were transferred from an IEF gel (Fig. 1B) (Western-blot) onto nitrocellulose paper and incubated with antibody. The results (Fig. 1C) indicated that the antibody reacts with all the bands detectable on an IEF gel.

Sequence Determination. γ-Zein cDNA inserts subjected to DNA sequencing varied in length from 250 to about 900 base pairs as shown in Figure 2. Clone γZM5 having the longest insert among those selected by immunoscreening was chosen for complete sequencing while others were sequenced both from their 5' and 3' ends except γZM8-1 whose 5' end was sequenced only (Fig. 2). The length and position of each clone and its sequenced regions are indicated in Figure 2 with reference to the first base of the codon specifying the N-terminal residue of the mature polypeptide. Clone γZM8-3 was sequenced from nucleotide number -15 to 186 and from 92 to 236, clone γZM8-7, from 157 to 438 and from 530 to 620, clone γZM8-1, from 157 to 359, clone γZM11-7, from 151 to 402 and from 521 to 730, clone γZM11-6, from 13 to 165 and from 496 to 746, clone γZGS71 from 243 to 316 and 150 nucleotides at 3' end. When the sequence data from all seven clones were compared, it was found that they were identical in all the regions where they overlapped. While this manuscript was in preparation, the sequence of a γ-zein cDNA clone with full length coding sequence but shorter 3' nontranslated region than γZM5 was published by Prat et al. (19). Their sequence data and ours are in full agreement.

The sequences of both DNA and the deduced protein are shown in Figure 3. When the deduced N-terminal amino acid sequence was compared to the determined sequence (7), the two sequences showed perfect agreement. The amino acid composition calculated from the sequence showed that γ-zein contained 25% proline, 15% glutamine, 8% histidine, and does not have aspartic acid, asparagine, lysine, and tryptophan. There was also very close agreement between the amino acid composition calculated from the sequence and the one determined by amino acid analysis (6). The deduced mature protein has 204 amino acid residues with a mol wt of 21.824 kD which is about 5 kD smaller than that estimated (27 kD) by SDS-PAGE.

Tandem Repeats in γ-Zein Sequence. The existence of tandem repeats of the hexapeptide Pro-Pro-Pro-Val-His-Leu in the N-terminal sequences of gamma-zein was reported earlier (7). The intriguing question was how far the hexapeptide repeats extended. As can be seen in Figure 3, the sequence Pro-Pro-Pro-
Val-Val-His-Leu repeats tandemly and perfectly eight times except the No. 7 repeat in which Leu is replaced by Val. It is interesting to note that this is the only substitution within the eight repeats and is the result of a single nucleotide substitution, C to G. There are also four successive proline residues flanking the stretch of eight tandem repeats at both ends. Another intriguing point is that when the repeats are lined up as in Figure 4, the codon for the third proline in each is invariably CCG. Moreover, codons for the first prolines are either CCG or CCA while those for the second prolines are either CCG or CCT. At other positions of the coding sequence, however, proline codons could be any one of the four possible codons, CCC.

In addition, there are two tandem repeats of 24 nucleotides between positions 232 and 279 (CAG CCA CAC CCA TGC CCG TGC CAA and CAG CCG CAT CAG AGC CCG TGC CAG) coding for the octapeptides Gln Pro His Pro Cys Pro Cys Gln and Gln Pro His Pro Ser Pro Cys Gln. When these two repeats are compared, they differ from each other by four base substitutions, only one of which is in the first position of a codon and results in amino acid substitution (missense) while the rest are in the wobble position and thus do not change the amino acid sequence. The only amino acid substitution is from cysteine to serine resulting from a single base change from T to A. The first unit of the octapeptide tandem repeat is preceded by the sequence Gin-Pro-His-Pro which is the same as the first half of the octapeptide repeating unit itself (Fig. 3).

γ-Zein primary structure also contains such dipeptide repeats as Ala-Ala, Gin-Gln, Gly-Gly, Leu-Leu in its second one-half. Similar repeats were also observed by Geraghty et al. (11) in α-zein.

**Codon Usage and GC Ratio.** The γ-zein cDNA and its deduced amino acid sequence are divided into two distinct halves based on their nucleotide and amino acid composition, respectively. The first part is the N-terminal one-half of the mature protein (102 amino acids) which is rich in proline, low in glutamine, but devoid of alanine (0 Ala; 41 Pro; 8 Gin) and includes all the tandem repeats. The second part includes the C-terminal one-half (102 amino acids) and is rich in glutamine and alanine (10 Ala; 10 Pro; 22 Gin). The nucleotide sequence coding for the eight tandem hexapeptide repeats plus the four flanking prolines is extremely GC-rich, 73.6%. Similarly, the whole coding region is also GC-rich, 69.8%. In contrast, the 3' nontranslated region has
Amino Acid Pro Pro Pro Val His Leu
Codon CCN CCN CCN GTN CAT CTN CAC TTA(G)
Repeating Unit
1 CCG CCG CCG GTT CAT CTA
2 CCG CCG CCG GTG CAT CTG
3 CCA CCT CCG GTT CAC CTG
4 CCA CCT CCG GTG CAT CTC
5 CCA CGG CCG GTC CAC CTG
6 CCA CGG CCG GTC CAC CTG
7 CCA CGG CCG GTC CAT GTG
8 CCG CGG CGG GTT CAT CTG

Fig. 4. Tandem repeats (18-base long) in γ-zein DNA. The conserved CCG for all the third position prolines are blocked.

a 53.9% GC content. The base composition study of the antisense strand indicated that A=T (92:93) whereas C>G (253:174).

Like other genes, γ-zein gene has preference in codon usage as shown in Table I. In fact, the codon preference was dictated by the base preference in the wobble position. To reflect this phenomenon, a codon usage value, U=O/T, was calculated for each of the four bases by dividing its observed number of occurrence (O) by the theoretical one (T). As shown in Table II, there are significant differences in base usage (U value) in the wobble position of codons. For example, G is a consistently favored nucleotide for the wobble position. It was used in this position twice as frequently as would be expected on a random basis. T and A are not favored for the third position of codons especially in the last part of protein. For example, out of 100 theoretical ‘A’ positions, only 14 was actually used in codons specifying the last 102 amino acids.

DISCUSSION

The sequence data revealed that γ-zein has an unusual primary structure. The existence of eight essentially identical hexapeptide and two octapeptide repeats suggest that they play an important role in the structure and function of γ-zein. The codon for the third proline in the hexapeptide repeating unit is invariably CCG and some degree of preference, i.e. CCG/CCA and CCG/CCT, is also apparent for the first two prolines. Some degree of conservation at the nucleotide level, at other positions of the repeats, is also apparent. Whether or not the high degree of conservation at the nucleotide sequence level has significance or is just fortuitous is presently not known. The fact that there occurred only one amino acid substitution, a conservative one, in the sixth position of the seventh repeat and otherwise all eight repeats of the hexapeptide unit are identical suggests that this sequence is extremely critical to the function of γ-zein. Although the biological function of γ-zein is presently not known, it is expected to be the focus of future research efforts. γ-Zein was found to be located in the periphery of the protein bodies and upon its removal the protein body changed its general shape (15). This may suggest that γ-zein not only serves as nutrient reserve, but also has some structural function in the protein body. This idea is favored by the fact that there are regular repeats at the N-terminal region of the protein. One hypothesis is that γ-zein may serve as a skeleton or lining on the inner surface of protein body membrane where it can form multimeric networks through disulfide linkages with itself or other cysteine-rich proteins like glutelins and some zeins facilitating the packaging of protein into the protein body.

The mol wt of γ-zein calculated from deduced amino acid sequence is about 5 kD less than that estimated by SDS-PAGE. High histidine plus arginine content of γ-zein would reduce the net negative charge of the SDS-protein complexes and slow their rate of migration in electric field towards the anode in comparison to that of an average protein. In fact, mobilities slower than expected from their mol wt have been observed with other basic proteins such as histones and Cyt c. Alternatively, the preponderance of prolines in the primary structure of γ-zein may cause it to convert to a fully extended polypeptide chain as a result of destruction of its secondary and tertiary structure when subjected to a denaturant. This possibility is a strong one because a fully extended polypeptide chain would be retarded to a greater extent than a partially denatured one of the same size. Overestimation of mol wt by SDS-PAGE has also been reported for proline-rich proteins from other cereals as well as from animals (3) although the reason for this anomaly is not fully understood. It is also conceivable that addition of prosthetic groups to γ-zein during or after its synthesis can cause the increase in the observed mol wt.

The sequence information obtained from seven independent clones suggest that there is probably only a single coding sequence for γ-zein although the existence of a small gene family with two

Table 1. Codon Usage in γ-Zein cDNA

<table>
<thead>
<tr>
<th>Phe</th>
<th>TTT</th>
<th>0</th>
<th>Ser</th>
<th>TCT</th>
<th>0</th>
<th>His</th>
<th>CAT</th>
<th>9</th>
<th>Pro</th>
<th>CCT</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTC</td>
<td>2</td>
<td></td>
<td>TCC</td>
<td>1</td>
<td></td>
<td>CAC</td>
<td>7</td>
<td></td>
<td>CCC</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>TTA</td>
<td>0</td>
<td>TCA</td>
<td>0</td>
<td>Gin</td>
<td>CAA</td>
<td>4</td>
<td></td>
<td>CCA</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGT</td>
<td>3</td>
<td>TCG</td>
<td>2</td>
<td>Asn</td>
<td>AAT</td>
<td>0</td>
<td></td>
<td>Thr</td>
<td>ACT</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>CTT</td>
<td>0</td>
<td>AGT</td>
<td>0</td>
<td>Arg</td>
<td>AAG</td>
<td>5</td>
<td></td>
<td>AAC</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTC</td>
<td>3</td>
<td>AGG</td>
<td>5</td>
<td>Lys</td>
<td>AAA</td>
<td>0</td>
<td></td>
<td>ACA</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CTA</td>
<td>1</td>
<td>Arg</td>
<td>AGA</td>
<td>0</td>
<td>Ala</td>
<td>GAT</td>
<td>0</td>
<td>Ala</td>
<td>GCT</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>CCG</td>
<td>3</td>
<td>Arg</td>
<td>GAC</td>
<td>0</td>
<td>Ala</td>
<td>GAG</td>
<td>0</td>
<td>GCA</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCG</td>
<td>3</td>
<td>Arg</td>
<td>GAC</td>
<td>0</td>
<td>Ala</td>
<td>GAG</td>
<td>0</td>
<td>GCA</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>ATG</td>
<td>1</td>
<td>CCG</td>
<td>3</td>
<td>Glu</td>
<td>GA</td>
<td>0</td>
<td></td>
<td>GCA</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GTT</td>
<td>4</td>
<td>Tyr</td>
<td>TAT</td>
<td>0</td>
<td>Gly</td>
<td>GGT</td>
<td>1</td>
<td>Trp</td>
<td>TGG</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>GTC</td>
<td>7</td>
<td>TAC</td>
<td>4</td>
<td>Gly</td>
<td>GGT</td>
<td>1</td>
<td></td>
<td>Trp</td>
<td>TGG</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>GTA</td>
<td>0</td>
<td>Cys</td>
<td>TGT</td>
<td>1</td>
<td>GGA</td>
<td>1</td>
<td></td>
<td>TAG</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GTG</td>
<td>4</td>
<td>GAC</td>
<td>3</td>
<td>GGG</td>
<td>2</td>
<td></td>
<td>TGA</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Downloaded from on July 22, 2017 - Published by www.plantphysiol.org
Copyright © 1986 American Society of Plant Biologists. All rights reserved.
Table II. Evaluation of Base Usage at the Third Position of γ-Zein Codons by U Value (U=O/T)∗

<table>
<thead>
<tr>
<th>Base</th>
<th>Codons 1–102</th>
<th>Codons 103–204</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>63.72%</td>
<td>30.77%</td>
</tr>
<tr>
<td>C</td>
<td>99.12%</td>
<td>153.85%</td>
</tr>
<tr>
<td>A</td>
<td>65.92%</td>
<td>14.08%</td>
</tr>
<tr>
<td>G</td>
<td>180.22%</td>
<td>199.43%</td>
</tr>
</tbody>
</table>

∗ U, utilization; O, observed occurrence of a base in the third position; T, theoretical occurrence of a base in the third position.

or three members cannot be ruled out at this time. The perfect agreement between our sequence data and that of Prat et al. (19), who obtained their mRNA for reverse transcription from a different maize genetic source and employed different cloning and sequencing procedures from ours, support the hypothesis of single gene or small gene family for γ-zein. These results do not provide any support for the earlier hypothesis that multiple γ-zein polypeptides resolved by IEF were encoded by members of a large multigene family. The basis for extensive charge heterogeneity within γ-zein as revealed by IEF may be an artifact of the method of IEF or due to posttranslational processing and modifications. Either way it deserves a thorough investigation. When γ-zein bands in an IEF gel were transferred onto nitrocellulose, all bands reacted with specific antibody against γ-zein (Fig. 1C) indicating that the antibody does react with all the different pl forms of γ-zein. The immunoprecipitation of in vitro translation product, on the other hand, gave a narrow single band on SDS gel (Fig. 1A) as opposed to the broad, diffuse band that the mature protein yields (6). It is possible that the in vitro product is subjected to some posttranslational modifications, e.g. glycosylation and deamidation (20) which may not occur in the in vitro translation system used. Our attempts to detect glycosylation by periodic acid-Schiff staining gave no definitive results to conclude that γ-zein was a glycoprotein. It should be borne in mind that the periodic acid-Schiff stain reacts only with certain types of sugar residues, and the sensitivity of the assay is not high enough (10) to detect low amounts of carbohydrate in a protein. In addition, we cut out individual bands of γ-zein from an IEF gel and subjected them to IEF. It was found that each band could give multiple bands although not as many as 14 to 16 bands as detected for the mature γ-zein (data not shown). These results suggest that multiple bands on an IEF gel may be, at least in part, an artifact resulting from interaction of γ-zein protein either with ampholyte, urea, and other chemicals being used, or with itself to give aggregation through covalent and noncovalent linkages.

The sequence data also indicated that most of the clones covered the middle portion of the gene. This was obviously the result of breakage of DNA (or RNA/DNA duplex) during the first strand synthesis (most likely by a contaminating nuclease). The shortest insert, clone γZMB-3, corresponded to the 5' end of mRNA. It encodes 84 amino acids, 5 of which were part of the signal peptide and 79 were the N-terminal region of the mature protein. Because this clone reacted with antibody against γ-zein as strongly as others, it was concluded that at least one antigenic determinant resides in this 79 amino acid region. This presumptive determinant might reside in the region containing tandem repeats. The antigenity of tandem repeats of an oligopeptide was also observed by Enea et al. (5).

LITERATURE CITED

2. BIGNIN MD, TJ GIBSON, GF HONG 1983 Buffer gradient gels and 32P label as an aid to rapid DNA sequence determination. Proc Natl Acad Sci USA 80: 3963–3965
4. BURR B, FA BURK 1976 Zein synthesis in maize endosperm by polyosomes attached to protein bodies. Proc Natl Acad Sci USA 73: 515–519
11. GERAGHTY DE, J MESSING, J RUBENSTEIN 1982 Sequence analysis and comparison of cDNAs of the zein multigene family. EMBO J 1: 1329–1336