Volume 15 Number 8 1987 Nucleic Acids Research

The nucleotide sequence of a cDNA clone encoding the wheat E_m protein

James C.Litts, Greggory W.Colwell, Randy L.Chakerian and Ralph S.Quatrano

Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331-2902, USA

Received July 22, 1986; Revised and Accepted January 27, 1987 Accession no. Y00123

ABSTRACT

The nucleotide sequence of a cDNA clone isolated from developing wheat embryos and encoding the E_m protein is reported. The entire coding region for E_m and the 3' non-translated flank are contained within this clone. The amino acid sequence deduced for E_m is very rich in glycine (18 mol%) as well as both basic and acidic residues. The molecular weight of the protein is ca. 9,900 daltons. The deduced sequence is supported by direct amino acid sequencing of cyanogen bromide cleavage fragments obtained from purified E_m protein. E_m is shown by Southern blots to be a product of a gene family of approximately ten members.

INTRODUCTION

The E_m protein was first identified as a major protein being synthesized by wheat embryos during the very early stages of germination by Cuming and Lane (1). When poly A+ RNA isolated from dry wheat embryos is translated *in vitro* in the presence of ^{35}S-methionine, E_m is a dominant product which is labeled to high specific activity, hence the name "early methionine labeled" (E_m) protein (2,3). As the embryo germinates the levels of E_m and its translatable mRNA decline rapidly (2). This decline in the mRNA is also evident when total RNA is hybridized to a cDNA probe for E_m using a northern blot (3). *In vivo* labeling studies indicate that detectable synthesis of the protein is absent five hours after imbibition (4), and by 10 hours E_m is not detectable in cell free translations (2,3). Williamson et al. (5) studied the expression of this protein during the development of the wheat embryo from anthesis to maturity. Accumulation of E_m commences in Stage IV (6), late in embryo development (30 days post-anthesis). However, precocious accumulation of E_m occurs when Stage II/III embryos are cultured in the presence of the phytohormone abscisic acid (ABA). A number of proteins which are synthesized late in embryogenesis in other seeds also appear to be under the regulation of ABA (7).

Cuming (4) isolated a short cDNA clone which hybrid-selected an mRNA for

the E_m protein from poly A+ RNA prepared from mature wheat embryos. Williamson et al. (5) selected several clones from a cDNA library constructed from poly A+ RNA isolated from ABA-treated stage III embryos that exhibited differential responses to ABA. One cDNA clone (p1015) from those that had an increased expression in the presence of ABA, was estimated to contain about 87% of the sequence of the mature mRNA for the E_m protein (5). Using p1015 to probe RNA slot-blots, Williamson et al. (5) found that the levels of E_m mRNA varied in the same manner as E_m protein during embryo development, both *in vivo* and in response to ABA *in vitro*. They suggested that modulation of E_m expression by ABA involved a transcriptional control point, but that ABA was probably not involved with the embryo specificity of E_m expression.

The amino acid composition of E_m was reported by Grzelczak et al. (3). Its high content of Glx and Gly, coupled with its rapid disappearance upon germination, has led to its tentative identification as an embryo storage protein (3). More recently, however, it was suggested that E_m may provide protection for the cytoplasm during the desiccation stage of embryo development (8). This latter suggestion was supported by hydrodynamic and optical data and calculations of Gibbs energies of hydration which indicate E_m exists in solution largely as a random coil and has a high affinity for water. The proposed functions of the E_m protein and the regulation of its expression by ABA make it an interesting gene product for further structural and functional analyses.

In this paper we report the complete amino acid sequence of the E_m protein as deduced from the nucleotide sequence of cDNA clone p1015 and supported by limited amino acid sequencing of peptides. We also report preliminary results describing the gene family which encodes this protein, including copy number estimates and limited restriction analysis.

MATERIALS AND METHODS

Materials

Restriction enzymes were obtained from Bethesda Research Labs, New England Biolabs, and Pharmacia-P-L Biochemicals; T4 Ligase from New England Biolabs; DNA Polymerase I (Klenow fragment) from Pharmacia-P-L Biochemicals; radioisotopes from New England Nuclear; and "NYTRAN" blotting membrane from Schleicher and Schuell. All chemicals were reagent grade or ultrapure.

DNA Isolation

Nuclei were prepared from wheat embryos essentially as described by Luthe and Quatrano (9). DNA was isolated from these nuclei using ethidium

bromide, as suggested by Kislev and Rubenstein (10) and by the method described by Blin and Stafford (11).

Plasmid DNA and M13 RF DNA were isolated by the alkaline-SDS method of Birnboim and Doly (12) followed by CsCl-Ethidium Bromide equilibrium centrifugation. Single-stranded M13 DNA was obtained according to the methods of Messing and Vieira (13).

DNA Sequencing and Subcloning

The Pst I insert from the cDNA clone p1015 described in (5) was transferred from pBR322 to the high copy number plasmid pSP64 obtained from Promega Biotec, Inc. The insert was isolated from this vector by extracting the DNA from low melting temperature agarose with phenol as described by Weislander (14). The isolated insert was digested separately with Msp I, Hae III, Alu I, and Sau 3A. These digests were shotgun cloned into the Acc I, Sma I, Sma I, and Bam HI sites, respectively, of M13 mp19 by standard methods. Sequencing was done by the dideoxy chain termination method of Sanger et al. (15), using ^{35}S-dATP as the radioactive label. Sequence analysis was conducted with an Apple II personal computer using the sequence analysis software obtained from Larson and Messing (16) and Fristensky et al. (17).

Southern Blot Analysis

Hybridization of probes to wheat embryo nuclear DNA was performed essentially as described by Southern (18). DNA was digested with various restriction endonucleases and electrophoresed in 0.9% agarose (SeaKem LE, FMC) gels (3 mm thick) for 16 hours in a "Protean" apparatus (Hoefer). Twelve μg of DNA was loaded in each sample lane. Reconstructions were made by linearizing pBR322 and pSP64 plasmids containing the insert. Plasmids used in this application were purified through two rounds of CsCl-Ethidium Bromide centrifugation and quantitated by their absorbance at 260 nm. The reconstructions were done assuming a 1C DNA content of the wheat genome of 17.5 pg (19). Linearized plasmid was diluted appropriately, mixed with 12 μg of salmon sperm DNA digested with HinD III or Eco RI and 500 ng of HinD III digested bacteriophage lambda DNA, heated to 65^{o}C for 2 min and run in lanes adjacent to the wheat DNA.

Following electrophoresis the gels were stained with 0.5 μg/ml ethidium bromide, briefly destained, and photographed on a UV light box. The DNA was denatured with 0.5 M NaOH, 1.5 M NaCl, a 5 min rinse with agitation, and a 20 min soak with fresh solution. The gel was then neutralized in a like manner with 0.5 M Tris-Cl, 3.0 M NaCl. The DNA was transferred to a derivatized nylon membrane by capillary action with 5 X SSPE. The transfer was allowed to

proceed for 16 hrs. The blot was then baked under vacuum at 80°C for 1 hour. The membrane was wetted briefly with distilled water, and then prehybridized for 15 hrs at 42°C with 50% formamide containing 5 X SSPE, 5 X Denhardts, 2 mM EDTA, 0.5% SDS, and 200 μg/ml sheared denatured salmon sperm DNA. The prehybridization solution was then removed and the membrane was hybridized in fresh prehybridization solution except the salmon sperm DNA was reduced to 100 μg/ml and heat denatured radioactive probe was included at 5 X 10^6 cpm/ml. The hybridization was allowed to proceed for 36 hrs.

Probe was prepared from insert isolated as above. Radioactive probe was prepared by the random hexamer labeling procedure of Feinberg and Vogelstein (20) using ^{32}P-dCTP (3000 Ci/mmol); specific activities were near 10^9 cpm/μg. Following hybridization the membrane was washed at room temperature for 1 hr with 50% formamide containing 5 X SSPE, 0.1% SDS. The membrane was then washed for 30 min each as follows: twice with 2 X SSPE, 0.1% SDS at 50°C, once with 0.25 X SSPE, 0.1% SDS at 60°C, and twice with the same solution at 63°C. The washes were all done in the heat-sealable bag. The membrane was then removed from the bag, dried, and exposed to Kodak XAR5 film for 3 days with a DuPont Lightning-Plus intensifying screen at -70°C.

Protein Sequencing

The products of a cyanogen bromide cleavage reaction were chromatographed on a Bio-gel P-6 (Bio-Rad Laboratories, Richmond, CA) column and the excluded material was subjected to automated Edman degradation in an Applied Biosystems, Inc. 470A gas phase protein sequencer. Following conversion with trifluoroacetic acid, the PTH-derivatives were resolved on an IBM cyanocolumn with a Beckman microflow HPLC system.

RESULTS AND DISCUSSION

A partial restriction map of the insert from p1015 is shown in Figure 1 with the M13 subclones used in sequencing shown below the map. With the exception of 30 base pairs at the 5' end and 79 base pairs at the 3' end, the entire region was sequenced on both strands. The poly G, C and A tails present at the ends interfered with the sequencing reactions, so these regions were sequenced only from the inside toward the ends. Figure 2 shows the nucleotide sequence of p1015 with the deduced amino acid sequence of the E_m protein directly under the only long open reading frame.

The length of the mature RNA for the E_m protein was determined by Northern blot analysis to be approximately 780 base pairs (5). Clone p1015 contained the entire coding region of 282 base pairs, 21 base pairs of 5'flank,

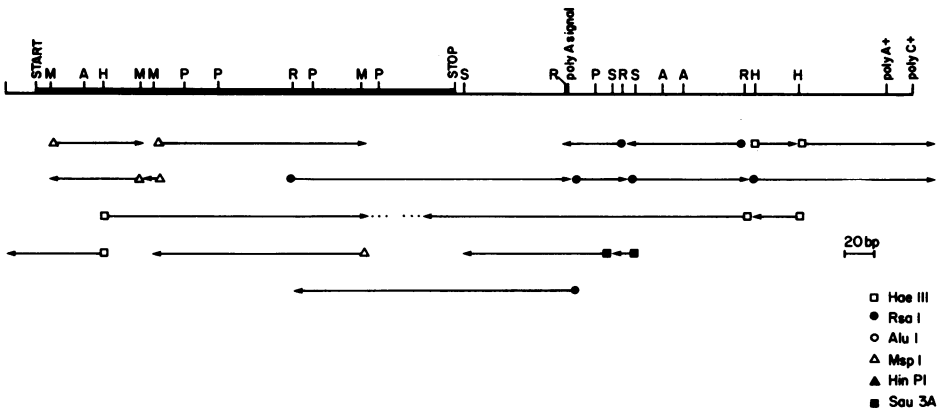

Figure 1. A schematic representation of the clone p1015 including restriction enzyme cut sites used in obtaining subclones for sequencing, positions of important structural features, and the actual subclones used to obtain the nucleic acid sequence. Also included is the start position of the Kozak sequence (START), the coding region (heavy line), position of the termination codon (STOP), site of the first and least divergent polyadenylation site (poly A signal), beginning of the poly A tail (poly A+) and start of the poly C tail added in the cloning procedure (poly C+). Restriction enzyme cut sites: (M) Msp I, (A) Alu I, (H) Hae III, (P) Hin P I, (S) Sau 3A, (R) Rsa I.

288 base pairs of 3' flank and a poly-A tail of 17 base pairs. Thus, we have sequenced a total of 608 base pairs of the mature mRNA. Assuming 100 base pairs as a rough estimate of the minimum length of the poly-A tract of the mature mRNA, this implies that the sequence reported here represents over 88% of the mature mRNA which may extend up to 100 base pairs further 5' than we have obtained.

The translation start codon, which begins at position 22 in Figure 2, is consistent with the consensus sequence, (A/G)NNATG(G/A), proposed by Kozak (21).

It is common for the 3' flanking regions of plant genes to contain two, or sometimes more, putative polyadenylation signals (22). The first signal is usually closer to the stop codon than the polyadenylation site, and usually conforms closely to the polyadenylation signal recognized in non-plant, eukaryotic genes, AATAAA (23). The second tends to be within 10 or 20 bases of the polyadenylation site and is more likely to diverge from that consensus found in other eukaryotes, (A/G)ATAA(A/T) (22). In the case of p1015 the positions (marked with overlines in Figure 2) of these sequences are consistent with

```
                                        30                  45                  60
    5         15        MET ALA SER GLY GLN GLN GLU ARG SER GLN LEU ASP ARG
CAAGG CGAGGCAGAG AGAGCA ATG GCT TCC GGT CAG CAG GAG AGG TCG CAG CTG GAC CGC

                   75                  90                  105
LYS ALA ARG GLU GLY GLU THR VAL VAL PRO GLY GLY THR GLY GLY LYS SER LEU GLU
AAG GCC CGC GAG GGC GAG ACC GTC GTC CCC GGT GGC ACC GGC GGC AAG TCC CTC GAG

120                 135                 150                 165
ALA GLN GLU ASN LEU ALA GLU GLY ARG SER ARG GLY GLY GLN THR ARG ARG GLU GLN
GCG CAG GAG AAC CTC GCC GAA GGG CGC AGC CGT GGC GGG CAG ACT CGC AGG GAG CAG

    180                 195                 210                 225
MET GLY GLU GLU GLY TYR SER GLN MET GLY ARG LYS GLY GLY LEU SER THR ASN ASP
ATG GGG GAG GAG GGG TAC AGC CAG ATG GGG CGC AAG GGC GGG CTG AGC ACC AAC GAC

        240                 255                 270                 285
GLU SER GLY GLY ASP ARG ALA ALA ARG GLU GLY ILE ASP ILE ASP GLU SER LYS PHE
GAG TCC GGC GGC GAC CGC GCC GCC AGG GAG GGC ATC GAC ATC GAC GAG TCC AAG TTC

            300
LYS THR LYS SER ***          315             330             345
AAG ACC AAG TCC TAG ATGATCTGTTAC CCTAGCAAGCAAGAC GAATCTCCTTAGGTC GGCTGTTGGT

  360             375             390             405             420
TTGCC CTAGTGTCTACGTAC CGAATAATGTAGTTC AGGCGCACGTAGGCG ATCGTACGTAGTATG ATCAT

       435             450             465             480             495
GTGTGCTTCG TAGCTGGTAGTCTGTAGCTACTCTACGGTGAGTAGCGTCAGGTTAAGTCGGAAGCGTACT GTG
                <-^^---^^--------^^++^^--------^^---^^->
          510             525             540             555
GCCTGGTCAGGT GGCTGGTCTGCCTCT GGGCCCAGGAGTGTC GCAGTGTCCACTGTA TCCTTGCTTCTT

570             585   591
CAGTAATGCATGTCGTCT TGCTCA(n)
```

Figure 2. Nucleotide sequence of p1015 with the deduced amino acid sequence of the E_m protein. Kozak's sequence is underlined. The sites cleaved by cyanogen bromide to produce the peptides used for amino acid sequencing are indicated with (↓). The stop codon is noted by (***). The two putative polyadenylation signals are overlined, one beginning at nucleotide 379, the other at 570. The inverted repeats in the 3' flank are marked with arrows (+--^--->), with (-) where perfect matches occur and (^) where mismatches occur.

this pattern; however, the sequences are more divergent in both cases, AATAAT and AGTAAT, respectively,

One feature of particular interest in the 3' sequence of p1015 is an inverted repeat of 19 bases centered around position 461 (marked with opposing arrows in Figure 2). In this structure 13 of the 19 bases would be able to pair to form a stem structure in the mRNA. A much smaller, but perfect, 5 base inverted repeat is found at position 528. Also of interest is the high content of G(40%), and C(30%) in the coding strand within the boundaries of the coding region. The high G+C content is particularly apparent in the third positions of the codons. At this position there is a G or a C in 87 of the 93 codons (93.5%). We do not know the significance of this observation, but it is interesting to note that a similar pattern is developing in a globulin gene

Table I. Comparison of amino acid residues recovered from automated Edman degradation of E_m protein to those predicted from p1015 nucleotide sequence.

Cycle no.:	1	2	3	4	5	6	7	8
			Amino Acid Recovered					
Peptide 1:	Gly	Glu	Glu	Gly	Tyr	Ser	Glu	Met
Peptide 2:	Gly	?	Lys	Gly	Gly	Leu	Ser	Thr
			Predicted Amino Acid					
Peptide 1:	Gly	Glu	Glu	Gly	Tyr	Ser	Gln	Met
Peptide 2:	Gly	Arg	Lys	Gly	Gly	Leu	Ser	Thr

which we are currently sequencing that is also expressed during embryogenesis under the control of ABA (data not shown). GC enrichment in the third codon-position has also been observed for another wheat embryo protein (Byron Lane, unpublished result), and in two prokaryotic genes (24, 25). There is no such enrichment apparent in the gliadin (26) or glutenin (27) genes of wheat which are expressed slightly earlier in the course of development in the endosperm of the seed. Outside the coding region the G and C contents more closely approach 25% each.

The molecular weight of the protein encoded by the open reading frame is 9,965 daltons. If the N-terminal methionine is removed, and the resulting terminal alanine is acetylated to produce the N-terminal blockage which we have observed and which was reported by Hofmann et al. (28), a molecular weight of 9,894 daltons is calculated. This is in close agreement with a value of 9,700 daltons determined by equilibrium centrifugation if allowances are made in the calculation of this value for the abnormally low partial specific volume of E_m (0.68 cc/gm) reported by McCubbin et al. (29).

We have conducted sequence analysis of the cyanogen bromide cleavage products obtained from the purified protein. These results are summarized in Table I. The results are consistent with the deduced amino acid sequence presented here, with two exceptions. In the second cycle only Glu was obtained, at about 50% of the yield expected based upon the first and subsequent cycles. We interpret this as simply being due to a poor recovery of Arg. The second exception was in the seventh cycle when we obtained a Glu where our cDNA sequence predicted a Gln. We interpret this as either a deamidation of our E_m protein preparation during the process of purification or a point mutation (e.g. GAG:Glu to CAG:Gln) in the particular gene family member that our cDNA represents. As discussed below, Figure 3 indicates that E_m is the product of a small multigene family. With such families, microheterogenity among

Table II. Comparison of the amino acid composition of embryo albumins from wheat (E_m) and mung bean as determined by chemical analysis and sequence.

	E_m PROTEIN			MUNG BEAN PROTEIN
Amino Acid[a]	Sequence[b] Translation (#)	(mol%)	Chemical[c] Analysis (mol%)	Chemical[c] Analysis (mol%)
LYSINE	6	6.5	7.1	9.7
HISTIDINE	0	0.0	0.9	1.6
ARGININE	10	10.9	8.0	7.0
THREONINE	5	5.4	5.7	6.0
SERINE	9	9.8	11.4	10.8
PROLINE	1	1.1	1.5	1.5
GLYCINE	17	18.5	18.0	14.0
ALANINE	6	6.5	7.1	6.2
VALINE	2	2.2	2.5	3.0
METHIONINE[b]	3(2)	2.2	1.7	1.9
ISOLEUCINE	2	2.2	2.1	4.3
LEUCINE	4	4.3	3.8	3.9
TYROSINE	1	1.1	1.1	1.2
PHENYLALANINE	1	1.1	1.2	2.8
ASPARTATE	5	5.4	-	-
ASPARAGINE	2	2.2	-	-
ASX	7	7.6	7.5	6.9
GLUTAMATE	12	13.0	-	-
GLUTAMINE	7	7.6	-	-
GLX	19	20.6	20.4	14.5
TOTAL AMINO ACIDS[c]	93 (92)			
CALCULATED M.W.	9,965 (9,894)			

[a]Cysteine and tryptophan were not detected.
[b]The N-terminal amino acid is methionine in the deduced E_m sequence. It is likely that this methionine is not in the mature protein and that the N-terminal Ala is acetylated. This is reflected in the numbers in parenthesis at the bottom of the table. The mol% for the deduced sequence is calculated assuming the methionine is removed and there are 92 residues in the protein. See (7) for further discussion.
[c]From (3) for E_m and (30) for mung bean.

the members of the family at certain positions is not unusual. In support of this, preliminary data that we have obtained recently from the sequence of a genomic clone indicates a slightly different amino terminal for another member of the gene family (data not shown).

The amino acid compositions deduced from the cDNA sequence and obtained by chemical analysis (3, 30) are shown in Table II. The very good agreement between these two determinations provides further support for our sequence. Glycine is the most prevalent amino acid (18.5 mol%) in E_m, and of particular interest are the high ratios of Glu/Gln and Asp/Asn. Our data indicate that

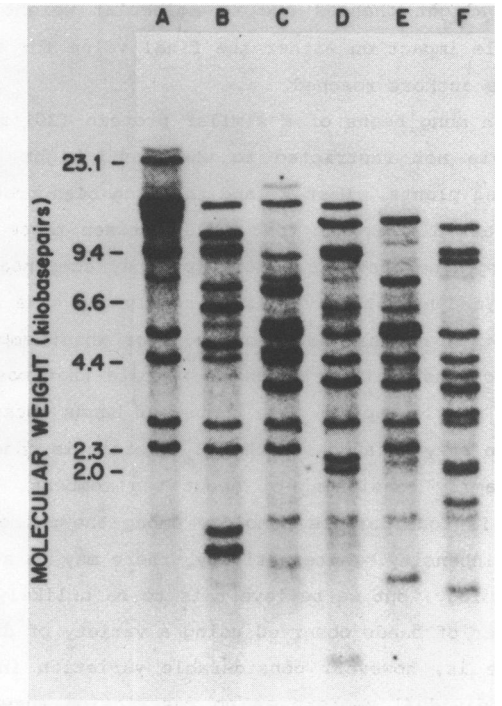

Figure 3. Southern blots of wheat genomic DNA digested with various restriction endonucleases and probed with ^{32}P-labelled p1015. All lanes contained 12.5 µg of wheat DNA first digested with Eco R I, followed by a second digestion with (A) no further enzyme, (B) HinD III, (C) Xba I, (D) Bgl I, (E) Bgl II, or (F) Sst I. The positions of molecular weight markers obtained by HinD III digestion of lambda DNA are indicated on the left.

in both cases the respective acids are present in this protein about twice as often as the amides. Also of interest is the very high content of basic residues, 10.9 mol% arginine and 6.5 mol% lysine. This high content of both acidic and basic residues results in a protein that is very highly charged in solutions significantly removed from its pI. A pI of 5.9 was predicted for the E_m protein by a calculation of the proton binding isotherm as described by Tanford (31). There is substantial disagreement between this calculated value and that reported by Cuming (4). The reason for this discrepancy is not clear.

We have recalculated the hydration potential of E_m based on our amino acid sequence and compared it with several other proteins as initially done by McCubbin et al. (8). The substitution of the proper values for the acids and

the amides and the slight changes due to molecular weight and composition differences had little impact on either the final value for this parameter or the conclusions those authors reached.

The presence in mung beans of a similar protein (30) may indicate that an Em-like protein is not restricted to wheat and might be quite broadly distributed among seed plants. When E_m and the mung bean protein are compared in Table II the largest differences that occur between these two proteins are in the lysine and arginine contents. However, they compensate each other to maintain both charge and hydration values at nearly the same levels.

A Southern blot of nuclear DNA isolated from wheat embryos is shown in Figure 3. Reconstructions (data not shown) indicate that most of the visible bands are present as single copies. The number of bands obtained with several different restriction enzymes suggests that E_m protein is coded for by a relatively small gene family consisting of about ten members. These data also indicate that there is considerable homology among the members as most bands are of fairly equal intensity. Alternatively, there may be a larger number of members of lower homology, but we believe this to be unlikely due to the consistency in the number of bands observed using a variety of different restriction enzymes. There is, however, considerable variation in the restriction sites surrounding individual members as no restriction enzyme tested gave a single or a few bands even though some fairly low molecular weight bands were observed. We are currently investigating the extent of this variation by including enzymes which cut more frequently than the six-base recognition restriction enzymes employed here.

Several authors have suggested that the E_m protein functions in the embryo primarily as a storage protein for nutrients necessary early in the germination process (3, 28, 30). This is supported by its high level of abundance early in germination, its rapid disappearance upon the onset of germination, and a high content of Glx (20.6 mol%). McCubbin et al. (8) suggested that E_m might function as a cryptobiotic protein analogous to that found in Artemia (32). As such, it would supply a form of desiccation tolerance to the cytoplasm while it is dehydrated at seed maturation. This suggestion was made based upon the high degree of hydophilicity predicted by the amino acid composition and the proposal that it exists largely as a random coil as indicated by hydrodynamic and optical measurements (8).

These two hypotheses are not necessarily mutually exclusive; however, the E_m protein is an atypical storage protein for two reasons. First, most storage proteins, including storage albumins (33), tend to be found in protein

bodies. The E_m protein, however, shows no sign of a signal peptide, which would probably indicate that it is localized in the cytoplasm. Certainly, a cytoplasmic locale would facilitate a role in protecting cytoplasmic components. A second common characteristic of storage proteins is their high content of glutamine residues. This is consistent with the role of this amino acid in transport and metabolism as a source of reduced nitrogen. E_m is, again, atypical in this respect; however, its solubility in the cytoplasm at high concentration might be facilitated by its having a relatively high charge at neutral or slightly basic pH, which would in turn be enhanced by a higher ratio of glutamate to glutamine.

ACKNOWLEDGEMENTS

We gratefully acknowledge the contributions of Dr. Reg McParland in conducting the protein sequencing, and Roswitha Hopkins for her isolation of the E_m protein. Research support to R. Q. was given by the U.S.D.A. Competitive Grant Program (84-CRCR-1-1380), and the M. J. Murdock Charitable Trust through the Center for Gene Research and Biotechnology, Oregon State University.

BIBLIOGRAPHY

1. Cuming, A. C. and Lane, B. G. (1979) Eur. J. Biochem. 99, 217-224.
2. Thompson, E. W. and Lane, B. G. (1980) J. Biol. Chem. 255, 5965-5970.
3. Grzelczak, Z. F., Sattolo, M. H., Hanely-Bowdoin, L. K., Kennedy, T. D. and Lane, B. G. (1982) Can. J. Biochem. 60, 389-397.
4. Cuming, A. C. (1984) Eur. J. Biochem. 145, 351-357.
5. Williamson, J. D., Quatrano, R. S. and Cuming, A. C. (1985) Eur. J. Biochem. 152, 501-507.
6. Rogers, S. O. and Quatrano, R. S. (1983) Amer. J. Bot. 70(2), 308-311.
7. Quatrano, R. S. (1986) In Oxford Surveys of Plant Molecular and Cell Biology (ed. B. J. Miflin), Oxford Press, in press.
8. McCubbin, W. D. and Kay, C. M. (1985) Can. J. Biochem. 63, 803-811.
9. Luthe, D. S. and Quatrano, R. S. (1980) Plant Physiol. 65, 305-308.
10. Kislev, N. and Rubenstein, I. (1980) Plant Physiol. 66, 1140-1143.
11. Blin, N. and Stafford, D. W. (1976) Nuc. Acids Res. 3, 2302-2308.
12. Birnboim, H. C. and Doly, J. (1979) Nuc. Acids Res. 7, 1513-1523.
13. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
14. Weislander, L. (1979) Anal. Biochem. 98, 305-309.
15. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
16. Larson, R. and Messing, J. (1982) Nuc. Acids Res. 10, 39-49.
17. Fristensky, B., Lis, J. and Wu, R. (1982) Nuc. Acids Res. 10, 6451-6563.
18. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
19. Bennett, M. D. and Smith, J. B. (1976) Phil. Trans. R. Soc. London, Ser. B, 274(933), 227-274.
20. Feinberg, A. P. and Vogelstein, B. (1984) Anal. Biochem. 132, 6-13.
21. Kozak, M. (1981) Nuc. Acids Res. 9(20), 5233-5252.
22. Messing, J., Geraghty, D., Heidecker, G., Hu, N.-T., Kridl, J. and Rubenstein, I. (1983) In Genetic Engineering of Plants: An Agricultural Perspective, Kosuge, T., Meredith, C. P. and Hollaender, A. (eds.), Plenum Press, New York, pp. 211-227.

23. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
24. Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T. and Oshima, T. (1984) J. Biol. Chem. 259, 2956-2960.
25. Dunn, R., McCoy, J., Simsek, M., Majumdar, A., Chang, S. H., RajBhandary, U. L. and Khorana, H. G. (1981) Proc. Natl. Acad. Sci. USA 78, 6744-6748.
26. Anderson, O. D., Litts, J. C., Gautier, M.-F. and Greene, F. C. (1984) Nuc. Acids Res. 12, 8129-8144.
27. Sugiyama, T., Rafalski, A., Peterson, D. and Soll, D. (1985) Nuc. Acids Res. 13, 8729-8737.
28. Hofmann, T., Kells, D. I. C. and Lane, B. G. (1984) Can. J. Biochem. Cell Biol. 62, 908-913.
29. McCubbin, W. D., Kay, C. M. and Lane, B. G. (1986) Biochemistry and Cell Biol. 64, 264.
30. Manickam, A. and Carlier, A. R. (1980) Planta 149, 234-240.
31. Tanford, C. (1961) Physical Chemistry of Macromolecules, pp. 554-573. John Wiley & Sons, Inc., New York.
32. DeHerdt, E., De Voeght, F., Clauwaert, J., Kondo, M. and Slegers, H. (1981) Biochem. J. 194, 9-17.
33. Higgins, T. J. V., Blagrove, R. J., Chandler, P. M., Beach, L. R., Randal, P. J., Craig, S. C. and Spencer, D. (1985) First Int. Cong. Pl. Mol. Biol. (ed. G. A. Galau) p. 57.