The primary structure of a plant storage protein: zein

Dan Geraghty, Mark A. Peifer, Irwin Rubenstein and Joachim Messing*

Department of Genetics and Cell Biology, and *Department of Biochemistry, University of Minnesota, St. Paul, MN 55108, USA

Received 10 July 1981

ABSTRACT

The protein sequence of a representative of the zeins, the major storage proteins of maize, has been derived from the nucleotide sequence of a zein cDNA clone. This cDNA was sequenced both by the Maxam and Gilbert and the M13-dideoxy techniques. The nucleotide sequence encompasses the non-translated 3' terminus of the mRNA, the entire coding sequence specifying both the mature zein protein and a small signal peptide, and a portion of the non-translated 5' region. The deduced amino acid composition and the amino-terminal amino acid sequence closely resemble those derived from chemical analysis of the zein protein fraction. The data presented represent the first complete amino acid sequence of a plant storage protein.

INTRODUCTION

Storage proteins form one of the most abundant protein classes in the developing seed. This abundance and the resultant nutritional importance of these proteins has led to their characterization in the seeds of many legumes and cereals. Since they have no known enzymatic function, storage proteins are classified on the basis of their physical properties. In maize, the ethanol-soluble or prolamin fraction, the zeins, constitutes as much as half of the total protein of the endosperm (1). These proteins are synthesized on polysomes bound to the surface of protein bodies, the membrane-bound subcellular structures within which they are stored (2-4). This synthesis is developmentally regulated, beginning about 14 days after pollination (5).

Analysis of the overall amino acid composition of the zein fraction reveals a group of hydrophobic proteins, rich in glutamine, leucine, proline, and alanine, but low in the essential amino acids lysine and tryptophan (6). SDS gel electrophoresis of the fraction indicates two predominant components of apparent molecular weights 19,000 and 22,500 (7), but minor components having higher and lower molecular weights may be present. Other means of analysis, however, including isoelectric focusing (7,8), partial amino acid sequence analysis (9,10), and cDNA hybrid-selection translation (11) reveal

extensive microheterogeneity, involving twenty or more different component proteins. This combination of near-homogeneity according to size with microheterogeneity limits the success of direct protein sequencing of unfractionated zeins. Bietz and colleagues (10) have succeeded in obtaining the sequence of the 33 N-terminal amino acids of a mixture of zeins, but microheterogeneity restricts extension of this effort.

These difficulties led us to explore another approach to the determination of a zein amino acid sequence. Isolation of protein bodies and their attached polysomes allows one to obtain zein mRNAs (12), which may then be reverse transcribed into cDNA (13). Cloned cDNAs, derived from individual zein mRNAs, allow the sequencing of a single message and consequently the unambiguous derivation of a zein protein sequence. We selected one such cDNA clone, A30, for sequence analysis. This clone had been previously characterized by restriction analysis and hybrid-selection translation (11). These studies indicated that A30 represents a zein cDNA clone containing about 80% of the length of the mature mRNA, a length we expected would include the entire protein coding sequence.

Using a combination of the Maxam and Gilbert (14-16) and M13-dideoxy sequencing techniques (17,18), we determined the entire nucleotide sequence of A30. Within this sequence we found one long open reading frame, coding for a protein of 225 amino acids. This sequence was compared to published data concerning the size, amino acid composition, and amino acid sequence of the zein fraction; we conclude that the deduced protein is a zein and that a short signal peptide is attached at its N-terminus. We also observed that the carboxy-terminal two-thirds of the protein appears to consist of 7 or 8 tandem repetitions of a highly conserved 20 amino acid repeating unit.

MATERIALS AND METHODS

Reagents

The restriction endonucleases Rsa I and Pvu II were purchased from New England Biolabs. T4 polynucleotide kinase, DNA polymerase (Klenow subfragment), and the other restriction endonucleases were purchased from Bethesda Research Labs. Bovine alkaline phosphatase was obtained from Worthington Biochemicals. T4 DNA ligase was isolated, stored, and used as detailed by Tait and colleagues (19). All other enzymes were used as recommended by their supplier. ($\gamma^{32}\text{P}$)dATP (5000Ci/mmol) was obtained from ICN and ($\gamma^{32}\text{P}$)dATP (400Ci/mmol) was obtained from Amersham.

Cloning vectors and their hosts

The plasmid A30 and its growth conditions (11), and the phage vector M13mp7, its *E. coli* host JM103, and its growth conditions have been previously described (18).

Sequencing by chemical modification

Selected restriction fragments of A30 were isolated, end labelled, processed and sequenced as described by Maxam and Gilbert (14,15); a modification of the A+G reaction, described by Gray et al. (16), was employed. Fragments labelled with ^{32}P on only one end were obtained by cleavage with a second, asymmetrically cleaving restriction enzyme, followed by preparative polyacrylamide gel electrophoresis and electro-elution.

Subcloning and dideoxy sequencing

A30 was restricted by Taq I, releasing the insert flanked by some pMB9 sequences. This fragment was isolated by electro-elution from an agarose gel and cleaved to completion with Alu I. The resulting subfragments were cloned into the Hind II site of M13mp7 by blunt end ligation. To obtain the 3'-terminal sequence, an additional 250 base pair (bp) Alu I-Rsa I fragment was isolated by preparative polyacrylamide gel electrophoresis and separately cloned into the same site. Conditions of cloning, transformation, propagation of M13mp7, isolation of both replicative form and single-stranded DNA, and dideoxy sequencing were as previously described (17,18).

All experiments have been performed in accordance with NIH guidelines.

RESULTS AND DISCUSSION

Sequencing strategy

In sequencing the zein cDNA A30 we employed a two-stage procedure. First, restriction mapping begun previously (11) was extended. These data were used to facilitate easy and unambiguous cleavage, isolation, end-labelling, recleavage, and reisolation of singly labelled fragments which could then be sequenced by the method of chemical modification (14-16). This approach yielded the sequence of at least one strand of the entire clone, as illustrated in Fig. 1.

This information was used to determine all the endonuclease restriction sites of the cDNA insert. Based on this examination we selected the restriction enzymes Alu I and Rsa I for subcloning. They were used to cleave an isolated Taq I fragment of A30 that contained the cDNA insert. These subfragments were of a size to be conveniently cloned into M13mp7 and subsequently sequenced by the dideoxy technique (17). The sequence

Fig. 1. A restriction map of the zein cDNA A30 and the sequencing strategy employed.

- (a) Selected restriction sites. Numbers are from the 5' end of the noncoding strand. *AluI* (A) sites are located at positions 90, 99, 192, 369, 422, 524 and 528, *RsaI* (R) sites at positions 117 and 779, and single *PvuII* (P), *BamHI* (B), *HhaI* (H), *DdeI* (D), and *XbaI* (X) sites at positions 191, 173, 592, 335 and 786 respectively.
- (b) Sequences obtained by the Maxam and Gilbert technique. Lines indicate extent of sequences determined from particle restriction sites. Arrow indicates direction and thus the strand from which sequence was read. The dotted line indicates sequences read outside the insert.
- (c) *AluI* and *RsaI* subclones of A30 in M13mp7 sequenced using the dideoxy-chain terminator technique. A number is given each clone for future reference. Dashed lines indicate regions present in a clone but not sequenced. Other conventions are as in (b).

information derived from these subclones was fitted to the other DNA sequencing data (Fig. 1), obtained from the Maxam-Gilbert procedure. This combination of the two sequencing techniques provided independent confirmation of the entire sequence, usually by sequences obtained from both strands.

Nucleotide sequence yields complete zein protein sequence

The complete nucleotide sequence of the non-coding strand of the zein cDNA A30, consisting of 795 nucleotides between the homopolymer A-T tails, is presented in Fig. 2. As expected, its length is about 80% of that of the mature messenger (12). Within this sequence there is only one long open reading frame. Beginning with the initiation codon ATG near one end, it extends to a termination codon not far from the other terminus, defining a region coding for a polypeptide of 225 amino acid residues, flanked by a 90 bp 3'-nontranslated region and a 30 bp fragment of the nontranslated region at the 5'-terminus of the mRNA.

Previous work had presented evidence that the 19,000 dalton zein protein is initially synthesized as a protein precursor, bearing a 1100 dalton signal peptide on its N-terminus which is cleaved off during maturation (12). This

(POLY T) CAA TGG CAG

```

GCC AAA ATA TTT TGC CTC CTT MET LEU LEU GLY LEU SER ALA SER ALA ALA
ATG CTC CTT GGT CTT TCT GCA AGT GCT GCT

THR ALA THR ILE PHE PRO GLN CYS SER GLN ALA PRO ILE ALA SER LEU LEU
ACG GCG ACC ATT TTC CCG CAA TGC TCA CAA GCT CCT ATA GCT TGC CTT CTT

PRO PRO TYR LEU SER PRO ALA VAL SER SER VAL CYS GLU ASN PRO ILE LEU
CCG CCG TAC CTC TCA CCA GCG GTG TGT TCG GTA TGT GAA AAC CCA ATT CTT

GLN PRO TYR ARG ILE GLN GLN ALA ILE ALA ALA GLY ILE LEU PRO LEU SER
CAA CCC TAC ACG ATC CAA CAG GCA ATC ACA GCT GGC ATC TTA CCT TTA TCA

PRO LEU PHE LEU GLN GLN SER SER ALA LEU LEU GLN GLN LEU PRO LEU VAL
GCC TTG TTC CTC CAA CAA TCA TCA TCA CTA TTA CAG CAG TTA CCT TTG GTG

HIS LEU LEU ALA GLN ASN ILE ARG ALA GLN GLN LEU GLN GLN LEU VAL LEU
CAT TTA TTG GCA CAA AAC ATC ACG ACA CAA CAA CTA CAA CAA CTT GTG CTA

ALA ASN LEU ALA ALA TYR SER GLN GLN GLN GLN PHE LEU PRO PHE ASN GLN
GCA AAC CTT GCT GCG TAC TGT CAG CAA CAG CAG CAG TTT CTT CCA TTC AAC GAA

LEU GCT ALA LEU ASN SER ALA SER TYR LEU GLN GLN GLN GLN LEU PRO PHE
CTA ACA GCA TTG AAC TCT GCT TCT TAT TTG CAA CAA CAA CAA CTA CCA TTC

SER GLN LEU PRO ALA ALA TYR PRO GLN GLN PHE LEU PRO PHE ASN GLN LEU
AGC CAG CTA CCT GCT GCG TAC CCC CAG CAA TTT CTT CCA TTC AAC CAA CAG

ALA ALA LEU ASN SER PRO ALA TYR LEU GLN GLN GLN GLN LEU LEU PRO PHE
GCA GCA TTG AAC TCT CCT GCT TAT TTA CAG CAG CAA CAA CTA CTA CCA TTC

SER GLN LEU ALA GLY VAL SER PRO ALA THR PHE LEU THR GLN PRO GLN LEU
AGC CAG CTA GCT GGT GTG ACG CCT GCT ACC TTG ACA CAA CCA CCA CAG TTC

LEU PRO PHE TYR GLN HIS ALA ALA PRO ASN ALA GLY THR LEU LEU GLN LEU
TTG CCG TTC TAC CAG CAC GCT GCG CCT AAC GCT GGC ACC CTC TTA CAA CAG

GLN GLN LEU LEU PRO PHE ASN GLN LEU ALA LEU THR ASN LEU ALA ALA PHE
CAA CAA TTG CTG CCA TTC AAC CAA CTT GCT TTG ACA AAC CTA GCA GCG TTC

TYR GLN GLN PRO ILE ILE GLY GLY ALA LEU PHE stop
TAC CAA CAA CCC ATC ATT GGT GGT GCC CTC TTT TAG ATT TGT TAT GAG TTA

stop
TAG TTC AAT AAT AAA GTT TTT TGT CTG ATG TTT GTG GCT TCG CAG AAA TAA

GAA AGT ACA TTT CTA GAT TCT (POLY A)

```

Fig. 2. The nucleotide sequence of a zein cDNA and the deduced amino acid sequence.

The lower line (NTDS) gives the nucleotide sequence of the noncoding strand of the cDNA while the upper line presents the amino acid sequence deduced from the only long open reading frame, in the three letter code.

was confirmed by comparison of the derived amino acid sequence presented in Fig. 2 to the partial amino acid sequences obtained by chemical analysis of mixtures of zein proteins (9,10). This comparison revealed a 12 amino acid signal peptide, not found on mature zeins, joined to the N-terminus of a zein protein whose first 33 amino acids agree quite well with those previously determined (Fig. 3). Further, the deduced amino acid composition (minus the signal peptide) closely parallels that of the zein fraction as a whole (21, Table 1). In summary, these comparisons confirm that A30 represents a member of the zein mRNA family.

Fig. 3. Comparison of the deduced amino acid sequence with previously published sequences of the NH₂-terminal amino acids of mixtures of zein proteins. The figures below the experimental sequence indicate the relative amounts of the two amino acid residues. The * indicates that the relative amounts of the two residues in these positions cannot be specified.

The major ambiguity remaining after comparison to published data involves the molecular weight of the protein. As stated above, zeins fall into two apparent molecular weight classes, 19,000 and 22,500 daltons as measured by SDS gel electrophoresis, while the deduced weight of the zein protein minus the signal peptide is 23,329 daltons. Since SDS gel electrophoresis does not reflect the absolute molecular weight of a protein, the classification of this zein protein into one of the two major size classes depends on additional data. Handa and colleagues (9) have obtained partial amino acid sequence data from the two zein classes, demonstrating specific differences in the amino acid residues in the N-terminal region. Comparison of the deduced amino acid sequence to these data reveals that the sequence coded by A30 resembles more closely that of the lower molecular weight classes (Fig. 3). In addition, the amino acid composition of the A30 protein more closely parallels the properties of the 19,000 dalton class (21, Table 1). Finally, in hybrid-selection translation the A30 DNA clone selects predominantly messengers coding for the lower molecular weight zeins (11). It will be interesting to see how other members of the zein family correspond to these findings.

Novel features of the zein protein

Sequencing the zein cDNA clone A30 has allowed us to determine the first complete amino acid sequence of a plant storage protein. The result has revealed a number of novel features which may furnish a model for plant storage proteins. Although full appreciation of the importance of such features must await comparison to other sequence data, a description of some

TABLE 1

Comparison of the amino acid composition of A30 zein with the amino acid composition of the 19,000 and 22,500d zeins

Residue	Number in A30	<u>g amino acid</u> <u>100g protein</u>		
		A30	19,000d ^b	22,500d ^b
Nonpolar				
Ala	29	8.8	8.8	7.7
Ile	9	4.4	3.5	3.5
Leu	43	20.9	18.5	14.5
Met	0	0	0.4	1.9
Phe	13	8.2	7.4	6.7
Pro	23	9.6	11.2	13.7
Trp	0	0	n.d.	n.d.
Val	5	2.1	3.4	4.2
SUM	122	54.0	53.2	52.2
Polar				
Asn	10	4.9	5.0 ^a	4.1 ^a
Cys	2	0.9	---	---
Gln	41	22.5	21.2 ^a	18.0 ^a
Gly	5	1.2	2.1	4.2
Ser	15	5.6	5.3	4.5
Thr	5	2.2	2.5	2.4
Tyr	8	5.6	4.8	6.4
SUM	86	42.9	40.9	39.6
Basic				
Arg	2	1.3	2.4	4.1
His	2	1.2	1.2	2.2
Lys	0	0	---	---
SUM	4	2.5	3.6	6.3
Acidic				
Asp	0	0	a	a
Glu	1	0.6	a	a
SUM	1	0.6	a	a

a. We assumed that the aspartic and glutamin acid residues were in the form of asparagine and glutamine.

b. Reference 21.

of these features is worth noting.

As mentioned above, the precursor to the zein protein coded by A30 apparently possesses a 12 amino acid signal peptide, with a highly hydrophobic composition. Numerous examples of such peptides preceding animal, viral, and prokaryotic proteins have been characterized (22). The zein signal peptide is similar to these in that the final residue is an alanine and the sequence is hydrophobic. In contrast to these other examples, which range in length from 16 to 29 residues, this signal peptide is apparently only 12 amino acids long. This size is in agreement with previous published estimates (12). Of additional interest is the presence of an ATG codon one nucleotide out of phase at what would be residue position 21 from the N-terminal threonine. The significance of this reduced size and the second ATG sequence is difficult to gauge; zein is, however, a very hydrophobic protein and thus its hydrophobic leader may differ in structure or function from those of other secreted proteins.

Examination of the sequence of the mature zein protein reveals other interesting features. Dipeptide repeats are particularly common, comprising about 30% of the residues in the sequence: the dipeptide Ala-Ala, present 7 times, the dipeptide Leu-Leu, present 8 times, and especially the dipeptide Gln-Gln, present 14 times, are particularly prevalent. Most distinctive, however, is the unusual organization of the carboxy-terminal two-thirds of the protein, where one can find evidence of seven or eight tandem repetitions of a highly conserved 20 amino acid repeating unit. Alignment of the individual repeats to maximize homology (Fig. 4) allowed us to construct a consensus amino acid and nucleotide sequence. Subsequent comparison of each repeat to the consensus revealed striking homology, ranging from 55 to 90% at the level of the amino acids and from 70 to 98% at the level of the nucleotides. Such intragenic duplications have been previously noted in other genes, where they have been viewed both as relics of evolution and also as functionally important units (23,24,30). Speculation about the functional and evolutionary significance of this repeating unit must await comparison to other members of the zein family and studies of other plant storage proteins.

Interesting elements of the nucleotide sequence

Two features of the nucleotide sequence are also worthy of mention. The first involves codon usage, displayed in Table 2. As has proved characteristic of other mRNAs sequenced (25,26), codon usage is nonrandom. For example, note the dichotomy between the use of the asparagine codons AAC (used 10 times) and AAT (not used). In addition, the dinucleotide CpG, known to be

Fig. 4. Comparison between several directly repeated units within the coding sequence.

The nucleotide sequence of the noncoding strand of A30 is presented, beginning at residual position 76 and ending at position 210. The alignment serves to maximize homology between repeats. Above each line is the deduced amino acid sequence in the one letter code. A consensus nucleotide and amino acid sequence for the repeating unit is presented beneath.

underrepresented in animal DNA (26), is also underrepresented here. It constitutes only 1.5% of the dinucleotides, compared to an expected frequency of 4.7%. In contrast, the dinucleotide GpC represents 7.8% of the dinucleotides in A30, while its expected frequency is also 4.7%.

Second, most eukaryotic messengers previously examined have near their 3'-terminus the sequence 5'-AATAAA-3', thought to be a signal for polyadenylation or processing (27,28). This sequence may also be found about 60 nucleotides from the 3' end of the A30 insert; this position, however, is much farther from the site of polyadenylation than has been previously observed. Nearer the noted location for this putative recognition sequence one finds the sequences 5'-AATAAG-3', a possible variant of the usual signal. This arrangement might represent a situation similar to that seen in the mouse α amylase mRNAs where multiple polyadenylation sites are present (31).

Future use of the M13-A30 recombinant clones

As a byproduct of the process of dideoxy sequencing, we have generated a number of M13-A30 recombinant clones, containing well defined regions of the zein messenger sequence. One such clone, for example, contains the carboxy

TABLE 2

Use of Codons in the Zein cDNA Clone A30

SECOND LETTER

		T		C		A		G			
F I R S T L E T T E R	T	PHE	3	SER	5 (6)	TYR	2	CYS	1	T	T H I R D
		PHE	10	SER	1	TYR	6	CYS	1	C	
		LEU	7	SER	5	OCHRE	0	OPAL	0	A	
		LEU	11	SER	1	AMBER	1	TRP	0	G	
	C	LEU	8 (10)	PRO	7	HIS	1	ARG	0	T	L E T T E R
		LEU	4 (5)	PRO	5	HIS	1	ARG	0	C	
		LEU	10	PRO	8	GLN	28	ARG	0	A	
		LEU	3	PRO	3	GLN	13	ARG	0	G	
	A	ILE	3	THR	0	ASN	0	SER	0 (1)	T	T H I R D
		ILE	5	THR	3	ASN	10	SER	3	C	
		ILE	1	THR	2	LYS	0	ARG	0	A	
		MET	0 (1)	THR	0 (1)	LYS	0	ARG	2	G	
	G	VAL	0	ALA	13 (15)	ASP	0	GLY	3 (4)	T	L E T T E R
		VAL	0	ALA	4	ASP	0	GLY	2	C	
		VAL	1	ALA	9 (10)	GLU	1	GLY	0	A	
		VAL	4	ALA	3 (4)	GLU	0	GLY	0	G	

The numbers indicate the codon usage in the zein protein. The numbers in parentheses indicate codon usage in the zein protein including the signal peptide.

terminus and the 3' nontranslated region, another carries the 5' leader fragment and the signal peptide, while a third contains a single copy of the amino acid repeating unit. Such clones will prove useful in future studies, for they can be easily made into highly specific and extremely sensitive probes which may be used in hybridization studies of both the messenger population and the genome (29,32).

CONCLUSION

The approach utilized by this work has allowed us to obtain the amino acid sequence of a zein protein. Additional protein sequences derived by this procedure will be needed, however, before we can be sure that artifacts have not been introduced during the formation of the cDNA or the growth of the plasmid in *E. coli*. Comparison of cDNA clones from a given zein mRNA family (11) will be particularly instructive since the amino acid sequences should be

closely related. The cloning procedure offers what may be the most effective way to obtain this type of amino acid sequence data for the group of closely related polypeptides that represent the zein storage protein fraction in maize endosperm.

ACKNOWLEDGEMENTS

We thank Janet McCallister and Kris Kohn for their aid in preparing this manuscript. This work was supported by a grant from the USDA/SEA Competitive Grant Program--Genetic Mechanisms for Crop Improvement; 59-2271-0-1401-0; from the Minnesota Agricultural Experiment Station: MN 15-030; from the Department of Energy: DE-AC02-81 ER 10901; and from the National Institutes of Health: 5T32 GM07094.

REFERENCES

1. Nelson, O.E. (1966) Fed. Proc. Fed. Am. Soc. Exp. Biol. 25, 1676-1678.
2. Duvick, D.M. (1961) Cereal Chem. 38, 374-385.
3. Wolf, M.J., Khoo, U. and Seckinger, H.L. (1967)
4. Burr, B. and Burr, F.A. (1976) Proc. Natl. Acad. Sci. USA 73, 515-519.
5. Misra, P.S. and Mertz, E.T. (1975) Cereal Chem. 52, 734-739.
6. Wall, J.S. (1964) in Proteins and Their Reactions, Symposium on Foods, Shutz, H.W. and Anglemier, A.F., Eds., pp. 315-341, Avi Publ. Co., Westport.
7. Gianazza, E., Righetti, P.G., Pioli, F., Galante, E. and Soave, C. (1976) Maydica 21, 1-17.
8. Gianazza, E., Viglienghi, V., Richetti, P.G., Salamini, F., and Soave, C. (1977) Phytochemistry 16, 315-317.
9. Handa, A.K., Harmodson, M., Tsai, C.Y., and Larkins, B.A. (1979) Plant Physiol. 63 Supplement, 94.
10. Bietz, J.A., Paulis, J.W., and Wall, J.S. (1979) Cereal Chem. 56, 327-332.
11. Park, W.D., Lewis, E.D. and Rubenstein, I. (1980) Plant Physiol. 65, 98-106.
12. Burr, B., Burr, F.A., Rubenstein, I., and Simon, M.N. (1978) Proc. Natl. Acad. Sci. USA 75, 696-700.
13. Burr, B. and Burr, F.A. (1980) in Genome Organization and Expression in Plants, Leaver, C.J. Ed., pp. 227-231, Plenum, New York.
14. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
15. Maxam, A.M. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K. Eds., Vol. 65 I, pp. 499-560, Academic Press, New York.
16. Gray, C.P., Sommer, R., Polke, C., Beck, E., and Schaller, H. (1978) Proc. Natl. Acad. Sci. USA 75, 50-53.
17. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
18. Messing, J., Crea, R., and Seeburg, P.H. (1981) Nucl. Acids Res. 9, 309-321.
19. Tait, R.C., Rodriguez, R.L., and West, Jr., R.W. (1980) J. Biol. Chem. 255, 813-815.
20. Larkins, B.A., and Harkman, W.J. (1978) Plant Physiol. 62, 256-263.

21. Lee, K.H., Jones, R.A., Dalby, A., and Tsai, C.Y. (1976) *Biochem. Genet.* 14, 641-650.
22. Inouye, M., and Halegoua, S. (1980) *CRC Critical Reviews in Biochemistry* 7, 339-371.
23. Barker, W.C., Ketcham, L.K., and Dayhoff, M.O. (1978) *J. Mol. Evol.* 10, 265-281.
24. Villafranca, J.E., and Robertus, J.D. (1981) *J. Biol. Chem.* 256, 554-556.
25. Grantham, R., Gautier, C., and Gouy, M. (1980) *Nucl. Acids Res.* 8, 1893-1912.
26. Bird, A.O. (1980) *Nucl. Acids Res.* 8, 1499-1504.
27. Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* 263, 211-214.
28. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) *Nucl. Acids Res.* 8, 127-142.
29. Hu, N.-T., and Messing, J., in preparation.
30. Yamada, Y., Avredimento, E., Mudryi, M., Ohkubo, H., Vogeli, G., Irani, M., Pastan, I. and deCrombrughe, B. (1980) *Cell* 22, 887-892.
31. Tosi, M., Young, R.A., Hagenbuchle, O. and Schibler, V. (1981) *Nucl. Acids Res.* 9, 2313-2323.
32. Messing, J. (1981) in the *Third Cleveland Symposium on Macromolecules: Recombinant DNA*. Elsevier, Amsterdam, in press.