Plant Science, 47 (1986) 95—102
Elsevier Scientific Publishers Ireland Ltd.

ISOLATION AND SEQUENCING OF A 28 kD GLUTELIN-2 GENE FROM MAIZE. COMMON ELEMENTS IN THE 5′ FLANKING REGIONS AMONG ZEIN AND GLUTELIN GENES

A. BORONAT*, M.C. MARTÍNEZ**, M. REINA, P. PUIGDOMÈNECH and J. PALAU***

Institut de Biologia de Barcelona, C.S.I.C. C/Jordi Girona Salgado, No. 18-26, 08034 — Barcelona (Spain)

(Received May 23rd, 1986)
(Revision received July 30th, 1986)
(Accepted August 26th, 1986)

Four 28 kD glutelin-2 genomic clones were selected from a partial *EcoRI* lambda *sep* 6 *lac* 5 maize genomic library. The complete nucleotide sequence of a 1.9 kilo base pairs (kb) *EcoRI-HindIII* fragment containing the coding region of a glutelin-2 gene, as well as its 5′ and 3′ flanking regions, was worked out. The most relevant characteristics of the gene are: (a) it has no introns; (b) putative TATA and CAAT boxes are found at positions −104 and −141, respectively; (c) three polyadenilation signals are present in the 3′ region at position +743, +797 and +910. In comparison with the best fitting zein gene, more than 50% homology can be detected in long stretches of both 5′ and 3′ flanking regions, as well as in the signal peptide coding region. A number of common short conserved nucleotide sequences are found in glutelin-2 and zein genes upstream the CAAT box. The possibility for these sequences to be signals for the coordinated expression of these genes is discussed.

Key words: glutelin-2; storage protein genes; maize; regulatory sequences; coordinated expression

Introduction

The best studied storage proteins of maize are zeins. They are located in vesicles called protein bodies, which are derived from the rough endoplasmic reticulum of endosperm cells [1,2]. During the last years, great advances have taken place in the study of cDNA and genomic DNA clones obtained for high molecular weight zeins [3—15].

Our laboratory has demonstrated that another maize storage protein, the 28 kD glutelin-2, is accumulated in the perifery of protein bodies and that the process of biosynthesis of 28 kD glutelin-2 and zeins occurs in a parallel way [16]. In maize endosperm, glutelin-2 fraction accounts for about 15% of the total protein and shows little heterogenity (about 6 components) as determined by two-dimensional gel electrophoresis and immunodetection (M. Torrent et al., manuscript in preparation).

In a previous work [17], a cDNA coding for the 28 kD glutelin-2 protein from maize endosperm was cloned and the complete amino acid sequence of the derived protein was worked out. Several domains in the protein sequence can be defined in terms of characteristic sequences (a repeated region formed by eight units of Pro-Pro-Pro-Val-His-Leu, an alternating Pro-X stretch, a Cys rich domain and a C-terminal region rich in Gln). Such motives are drastically different in comparison with those present in high molecular weight zeins.

*Present address: Department de Bioquímica. Facultat de Farmàcia. Universitat de Barcelona. Pedralbes. 08028 - Barcelona.
**Present address: Departament de Bioquìmica i Biologia Molecular. Universitat Autònoma de Barcelona. Bellaterra. Barcelona.
***To whom correspondence should be sent.
Abbreviation: kb, kilo base pair.

0618-9452/86/$03.50

In maize endosperm, zeins and glutelins, the two major groups of storage proteins, accumulate during the maturation process of the seed. Although the amino acid composition and the structure of these two groups of proteins have revealed to be very different, they show a coordinated timing of synthesis during the seed maturation [16].

The amino acid sequence of zeins shows no homology with the major storage proteins found in other cereals, thus suggesting that zeins have evolved independently [18,19]. Nevertheless, the homologies found between glutelin-2 and gliadin or hordein [17] indicate that glutelin-2 may have the same evolutionary origin as the major storage proteins of wheat and barley.

In this paper, we report the isolation and sequencing of a 28 kD gene from maize and an analysis of its flanking regions. The comparison with data obtained for zein genomic clones [8,9,11,14,20] may give some clues to the general functioning of the gene flanking regions of the maize storage proteins taken as a whole family.

Methods

Materials

Restriction enzymes were from Amersham, Boehringer and New England Biolabs. T4 DNA ligase was from New England Nuclear. DNA polymerase I was for Boehringer. Nitrocellulose was from Schleicher and Schull. Hybridization primer from sequencing was from New England Biolabs. ^{32}P-dATP and ^{32}P-dCTP (400 Ci/mmol) were from Amersham.

Maize DNA isolation

Kernels of inbred line W64A (21 days after pollination) were frozen in liquid nitrogen, ground to a fine powder and thawed in 300 mM NaCl, 100 mM EDTA (pH 8.0), 1% SDS and 100 μg/ml proteinase K. After incubation at 50°C for 4 h the solution was clarified by centrifugation at 1500 × *g*, at room temperature for 10 min. The supernatant was extracted twice with an equal volume of phenol/chloroform and the DNA precipitated by the addition of two volumes of ethanol. The DNA was dissolved in 10 mM Tris, 1 mM EDTA (pH 8.0) and purified on cesium chloride/ethidium bromide gradients by centrifugation at 43 000 × rev./min in a 50 Ti rotor at 20°C for 60 h. The DNA band was removed from the gradient, the ethidium bromide extracted with isopropanol and the DNA precipitated with ethanol [21].

Construction of a maize genomic library

An *EcoRI* partial genomic library was constructed using lambda *sep* 6 *lac* 5 (described in [22]) as a vector. Maize DNA, prepared as described above, was partially digested with *EcoRI* and sedimented through a 10—40% sucrose gradient according to the method described by Maniatis et al. [22]. Fractions containing fragments in the 12—20 kb range were pooled and the DNA recovered by precipitation with ethanol. The arms of lambda *sep* 6 *lac* 5 were obtained by digestion with *EcoRI*, annealing at 42°C and centrifugation on a 10—40% sucrose gradient as described by Maniatis et al. [22]. The DNA was ligated at a concentration of 130 μg/ml vector DNA and 50 μg/ml insert DNA. The ligated recombinant phages were packaged in vitro according to protocol II of Maniatis et al. [22].

Selection of clones containing the glutelin-2 gene

The phages from the library (approx. 200 000) were plated out on a *recA* derivative of *E. coli* LE392. Nitrocellulose blots were prepared according to the procedure of Benton and Davis [23] and hybridized against cDNA probe pME119 corresponding to a glutelin-2 mRNA [17] labelled with ^{32}P by nick translation. After autoradiography, positive plaques were purified by replating and hybridization.

Filter hybridization

Filters were prehybridized in 6 × SSC, 2 × Denhardt, and 100 μg/ml denatured salmon sperm DNA at 68°C for 4—6 h. Hybridization with the labelled probe (2-3 × 10^8 cpm/μg

DNA) was carried out under the same conditions but containing 50 μg/ml of salmon sperm DNA, at 68°C for 20 h. Filters were washed out in 6 × SSC, 2 × Denhardt at 68°C for 30 min; 2 × SSC, 0.1% SDS at 45°C for 90 min and 0.1 × SSC, 0.1% SDS at 45°C for 90 min.

DNA techniques

Subcloning of specific fragments into plasmids pUC18 and pUC19 or into the replicative form of phages M13mp18 and M13mp19 [24] was performed as described by Maniatis et al. [22]. Phage and plasmid DNA were prepared according to standard procedures described basically by Maniatis et al. [22].

DNA sequence analysis

DNA sequencing was performed by the dideoxy chain termination method described by Sanger et al. [25], using a synthetic dodecadeoxy-nucleotide as a primer. Sequencing reaction products were electrophoresed on 6% polyacrilamide-urea gels.

S1 nuclease mapping

S1 nuclease mapping analysis was performed basically as described by Maniatis et al. [22]. A 652 bp *HaeIII-HaeIII* fragment spanning the 5′ end of the glutelin-2 gene (see Fig. 1) was used. Total RNA (200 μg) (obtained from kernels 20 days after pollination) was mixed with 0.2 μg of the purified DNA fragment and coprecipitated with ethanol. After centrifugation, the pellet was washed with 70% ethanol and resuspended in 30 μl of hybridization solution (80% formamide, 0.4 M NaCl, 1 mM EDTA and 20 mM PIPES, pH 6.4). After heating at 85°C for 15 min, hybrids were allowed to form at 65°C for 3 h. The hybridization mixture was added of 300 μl of a solution containing 280 mM NaCl, 50 mM sodium acetate (pH 4.6), 4.5 mM $ZnSO_4$, 20 μg/ml calf thymus DNA, and 200 units nuclease S1/ml, and incubated at 37°C for 45 min. The reaction was stopped by the addition of 50 μl of a solution containing 4 M ammonium acetate and 100 mM EDTA and extracted once with phenol-chloroform. After addition of 20 μg tRNA (calf liver) nucleic acids were precipitated with isopropanol. The pellet obtained after centrifugation was washed with 70% ethanol and resuspended in 40 μl of TE (pH 8.0). The protected fragment was detected after gel electrophoresis (2% agarose), blotting to a nylon membrane (Zeta-Probe, BioRad) and hybridization to a ^{32}P-labelled homologous probe. pBR322 digested with Hinf I was used as a marker.

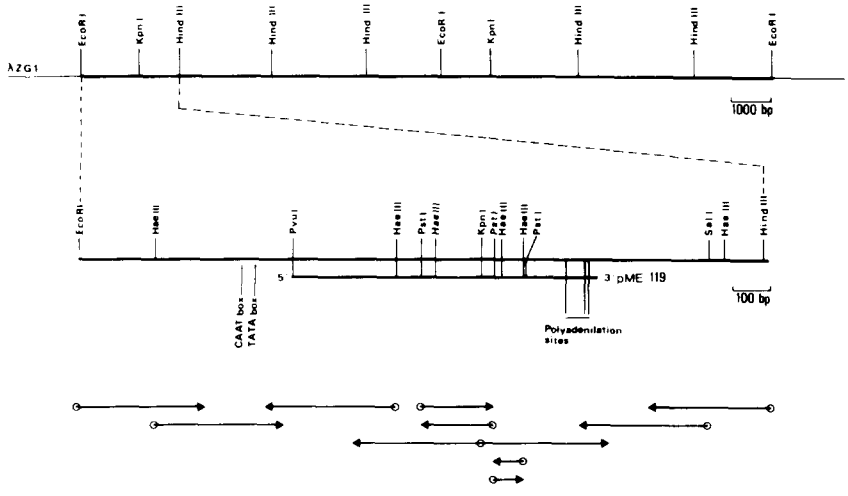

Fig. 1. Restriction enzyme maps of the genomic fragment cloned in lambda ZG1 phage and the 1.9 kb *EcoRI-HindIII* fragment containing the glutelin-2 gene. Sequencing strategy is also indicated.

Results

Identification of glutelin-2 genomic clones

cDNA clone pME119 corresponding to glutelin-2 [17] was used as a probe to screen a partial *EcoRI*, lambda *sep* 6 *lac* 5 maize genomic library (see Methods). After repeated re-screenings, 4 positive clones, out of 200 000, were selected.

The inserted DNA from the 4 recombinant phages was analyzed by restriction endonuclease mapping showing the same restriction pattern after digestion with *EcoRI*, *BamHI* and *HindIII*. One of these phages, coded as ZG1, was used for further studies.

Figure 1 (upper part) shows the restriction map of the insert of 12 kb of genomic DNA cloned in phage lambda ZG1. A 6 kb *EcoRI* fragment containing sequences homologous to the cDNA clone pME119 was subcloned into plasmid pUC18 and analyzed in more detail. By Southern blot analysis, the gene was located in a 1.9 kb *HindIII-EcoRI* fragment and the sense of transcription determined using a 5′ fragment from the cDNA clone as a probe. The 1.9 kb *HindIII-EcoRI* fragment was subcloned into plasmid pUC18 and a fine restriction map was established.

Nucleotide sequence of the glutelin-2 gene

Figure 1 also shows the restriction map of the 1.9 kb *EcoRI-HindIII* fragment and the sequencing strategy. Suitable restriction fragments were subcloned into phages M13mp18 and M13mp19 and sequenced by the dideoxy chain termination method of Sanger et al. [25]. The complete nucleotide sequence of the glutelin-2 gene and its 5′ and 3′ flanking regions is represented in Fig. 2. Some parts of the nucleotide sequence were obtained from a single strand. In these cases the sequencing determinations were performed at least twice and no apparent ambiguity in the reading was observed.

By comparison with the sequence of the cDNA clone pME119, no introns were found in the genomic clone. Furthermore, the sequence of the cDNA and the genomic clones

Fig. 2. Nucleotide sequence of the 28 kD glutelin-2 gene. CAAT and TATA boxes, as well as polyadenilation signals, are underlined. Dots indicate the start and the end of the cDNA clone pME119. The different bases between the genomic clone and the cDNA clone are indicated. Other putative regulatory sequences discussed in the text are also underlined.

can be perfectly aligned with the only exception of 3 nucleotides at postions +31, +393 and +582 (Fig. 2). These nucleotide substitutions are silent since they do not change the amino acid coded by the corresponding triplet. The complete homology spans also to the untranslated regions of the transcript represented in the cDNA clone.

Several sequences related to the putative control elements discussed by Breathnach and Chambon [26] are found in the 5′ flanking region. A typical TATA box (TATAAAT) is found at position −104. Thirty-seven bases further upstream from the TATA box, at position −141, there is the sequence CATAAGA that resembles the CAAT or AGGA box consensus sequences described for other plant genes [27]. The transcription start

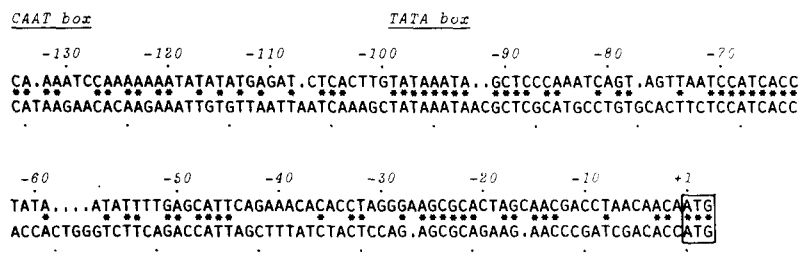

Fig. 3. Comparison of the 5′ flanking sequences of the 23 kD zein gene (PML1) and the 28 kD glutelin-2 gene. Nucleotides are numbered from the ATG initiation codon (boxed) and presented as negative values. Dots indicate gaps introduced to emphasize homology.

point has been located around nucleotide −70 by S1 mapping analysis (see Methods).

As it has been described in the corresponding cDNA clone [17], three conventional polyadenilation signals are found in the 3′ region of the genomic clone: AATAAT (+743), AATGAA (+797) and AATGAA (+810). Like other plant genes, the glutelin-2 gene does not have conventional polyadenilation signals, although the AATGAA sequence has also been described as polyadenilation signals in pea and soybean [28,29].

Comparison of glutelin-2 and zein genes

Comparing the nucleotide sequence of the glutelin-2 gene with that of zein genes reported by other authors, homologies can be observed both at the 5′ and 3′ flanking regions but not at the coding region (with the only exception of sequences coding for the signal peptide).

Figure 3 shows the alignment of the region extending from the ATG initiation codon to the CAAT box of the glutelin-2 and the zein gene pML1 [20]. In this region the homology between both genes is 52%. No further clear homology is found upstream from this region.

A higher level of homology (54%) is also observed in the region coding for the signal peptide, increasing the level of homology to 62% when the zones coding for the hydrophobic part of the signal peptide are compared (not shown).

At the 3′ end of the gene a level of homology similar to that described for the 5′ region is found (Fig. 4). The region of homology spans over more than 200 bases downstream from the stop codon.

Discussion

The results presented above clearly suggest that the cloned glutelin-2 gene corresponds to

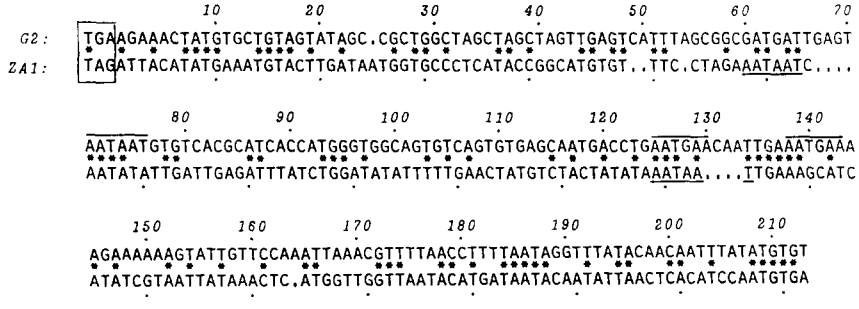

Fig. 4. Comparison of the 3′ flanking sequences of a 23 kD zein gene (ZA1) and the 28 kD glutelin-2 gene. Nucleotides are numbered from the terminator colon (boxed) and presented as positive values. Dots indicate gaps introduced to emphasize homology. Putative polyadenilation signals are marked.

a functional gene. On one hand the coding region contains the information for a protein giving the same amino acid sequence than that deduced from a previously isolated cDNA clone [17]. The fact that the cDNA and the genomic clones were obtained from different maize varieties (the cDNA was from the double hybrid inbred line E10 and the genomic clone from the inbred line W64A) indicates that glutelin-2 proteins do not show the high polymorfism found in zein proteins [30,31]. On the other hand, the gene contains consensus sequences at the 5′ flanking region that can act as promoters (TATA and CAAT boxes) and polyadenilation sequences at the 3′ region.

The coordinated expression of glutelin-2 and zein genes in maize endosperm, two structurally unrelated genes, prompted us to study the 5′ and 3′ untranslated regions of both groups of genes in order to find possible common regulatory elements. The comparative studies of the nucleotide sequence of the flanking regions of the glutelin-2 gene reported in this paper and some of the zein sequences described by other authors, clearly indicate homologies at both the 5′ and 3′ ends of the genes.

The comparison of the nucleotide sequence of the regions covering the 140 bases upstream from the ATG initiation codon of the glutelin-2 gene and the pML1 zein gene [20] shows that the level of homology increases to 60% when the sequences between nucleotides −65 to −140 are compared (Fig. 3). It is interesting to point out that this region of maximal homology includes the transcription start point, the TATA box and the CAAT box. When the −1 to −140 region of the glutelin-2 gene is compared to other zein genes, like ZA1 [11] and ZG99 [9], a long gap of 36 bases has to be introduced in the corresponding zein gene to obtain the same level of homology than that found for the pML1 gene (data not shown). Nevertheless, the same gap has also to be introduced to match the sequences of pML1 and ZG99 genes.

It is interesting to point out that the homology between glutelin-2 and zein genes at the 5′ end extends also over the translated region corresponding to the signal peptide. The conservation of this part of the protein can be related to the fact that both type of proteins, although being structurally different, have to be transported into the same cell vesicles. At present we do not know if the differences found in the hydrophilic part of the signal peptide may be responsible for the differential deposition of glutelin-2 and zeins into the protein bodies [16].

Similar levels of homology to those in the 5′ region are found when the 3′ ends of the genes are compared. The region of homology clearly extends over more than 200 bases downstream the stop codon. Interestingly the putative polyadenilation signals are placed in similar positions (Fig. 4).

There are increasing evidences suggesting that the transcriptional regulation of coordinately induced genes can be mediated through the interaction of specific factors with short repeated sequences located in the 5′ region [32]. In the specific cases of the *Drosophila* heat shock genes [33] and four genes involved in amino acid synthesis in yeast [34], short conserved sequences at the 5′ region have been shown to be required for the coordinated induction of the corresponding sets of genes. For the heat shock genes of *Drosophila* a defined protein factor that is able to bind to specific 5′ sequences has been described [35]. Up to now a number of such conserved sequences has also been reported upstream of some coordinately regulated plant genes [36–39].

Since in maize endosperm zein and glutelin genes show a coordinated timing of expression, we searched for such conserved sequences in the 5′ region. The sequence GTGTAAAG located at position −354 in the glutelin-2 gene is also found in a similar position in all zein genes sequenced so far: −342 in PML1 [20], −319 in Z7 [15], −328 in ZG99 [9] and Z4 [7] and −326 in ZE19 and ZE25 [14]. It is relevant to point out that the same conserved sequence, similar to the core sequence of the SV40 enhancer: GTGGAAAG

[40], is found in other storage protein genes as also reported by Kreis et al. [19] and Forde et al. [39]. The fact that this sequence is present in similar positions in gliadin, hordein, zein and glutelin genes suggests that it can be of relevance in the expression of these genes during the seed maturation process.

Another set of short conserved sequences has been found in zeins and glutelins but not in other cereal storage protein genes. The two glutelin-2 sequences CAACAAA (position − and ATGTCAA (position −320) are also present in the 19 kD zein genes (at position −211 and −193 in ZG99; −210 and −192 in Z4, ZE19 and ZE25). It is worth to point out that in all cases the two 7 bases conserved sequences are separated by an 11 bases spacer of non conserved sequence. We call this arrangement the 7-11-7 sequence. In addition, for the 23 kD zein genes a slightly modified version of the 7-11-7 sequence is found (GATGAAA at −211 and ATGTCGA at −193 for the Z4 gene). The fact that these conserved sequences seem to be specific of maize storage protein genes, suggests that they may have a role in their coordinated induction during the kernel maturation. Nevertheless, until a functional analysis of these sequences could be made, it will not be possible to make definite statements about their importance in regulating gene expression.

Acknowledgments

We wish to thank M. Meunier (Institut Jacques Monod, Paris) and R. Cortese (EMBO Lab., Heidelberg) for their help in setting up the chain termination sequencing method and M. Caballero (CMB, Madrid) for his help in the computer sequence analysis. A.B. was supported by a postdoctoral fellowship from the Ministerio de Educacion y Ciencia M.C.M. was supported by a postdoctoral fellowship from the Generalitat de Catalunya. M.R. held a graduate fellowship from the Fundación Ramón Areces. The work was supported by grants from CAICYT, CIRIT and Fundación Ramón Areces.

References

1 B. Burr and F.A. Burr, Proc. Natl. Acad. Sci. U.S.A., 73 (1976) 515.
2 B.A. Larkins and W.J. Hurkman, Plant Physiol., 62 (1978) 256.
3 D. Geraghty, M.A. Peifer, I. Rubenstein and J. Messing, Nucl. Acid. Res., 9 (1981) 5163.
4 E.D. Lewis, G. Hagen, J.I. Mullins, P.N. Mascia, W.D. Park, W.D. Benton and I. Rubenstein, Gene, 14 (1981) 205.
5 U. Wienand, P. Langridge and G. Feix, Mol. Gen. Genet., 182 (1981) 440.
6 D. Geraghty, J. Messing and I. Rubenstein, EMBO J., 1 (1982) 1329.
7 N.T. Hu, M.A. Peifer, G. Heidecker, J. Messing and I. Rubenstein, EMBO J., 1 (1982) 1337.
8 M.D. Marks, B.A. Larkins, J. Biol. Chem., 257 (1982) 9976.
9 K. Pedersen, J. Devereux, D.R. Wilson, E. Sheldon and B.A. Larkins, Cell, 29 (1982) 1015.
10 J.A. Pintor-Toro, P. Langridge and G. Feix, Nucl. Acids Res., 10 (1982) 3845.
11 A. Spena, A. Viotti and V. Pirrotta, EMBO J., 1 (1982) 1589.
12 A. Viotti, D. Abildstein, N. Pogna, E. Sala and V. Pirrotta, EMBO J., 1 (1982) 53.
13 G. Heidecker and J. Messing, Nucl. Acid Res., 11 (1983) 4891.
14 A. Spena, A. Viotti and V. Pirrotta, J. Mol. Biol., 169 (1983) 799.
15 J.C. Kridl, J. Vieira, I. Rubenstein and J. Messing, Gene, 28 (1984) 113.
16 M.D. Ludevid, M. Torrent, J.A. Martínez-Izquierdo, P. Puigdomènech and J. Palau, Plant Mol. Biol., 3 (1984) 227.
17 S. Prat, J. Cortadas, P. Puigdomènech and J. Palau, Nucl. Acid Res., 13 (1985) 1493.
18 P.R. Shewry, B.J. Miflin and D.D. Kasarda, Phil. Trans. R. Soc. London, B-304 (1984) 297.
19 M. Kreiss, P.R. Shewry, B. Forde, J. Forde and B.J. Miflin, Oxf. Surv. Plant Mol. Cell Biol., 2 (1985) 253.
20 J.W.S. Brown, C. Wandelt, G. Feix, G. Neuhaus and H.G. Schweiger, Eur. J. Cell Biol., 42 (1986) 161.
21 R.W. Davis, D. Botstein and J.R. Roth, Advanced Bacterial Genetics. A Manual for Genetic Engineering, Cold Spring Harbor Laboratory, New York, 1980.
22 T. Maniatis, E.F. Fritsch and J. Sambrook, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory, New York, 1982.
23 W.D. Benton and R.W. Davis, Science, 196 (1977) 180.

24 J. Norrander, T. Kempe and J. Messing, Gene, 26 (1983) 101.
25 F. Sanger, S. Nicklen and A.R. Coulson, Proc. Natl. Acad. Sci. U.S.A. 74 (1977) 5463.
26 R. Breatnach and P. Chambon, Annu. Rev. Biochem., 50 (1981) 349.
27 J. Messing, D. Geraghty, G. Heidecker, N. Hu, J. Kridl and I. Rubenstein, Plant gene structure, in: T. Kosuga, C.P. Meredith and A. Hollaender (Eds.), Genetic Engineering of Plants, Plenum Press, New York, 1983, p. 221.
28 S.L. Berry-Lowe, T.D. McKingth, D.M. Shah and R.B.. Meagher, J. Mol. Appl. Genet., 1 (1982) 483.
29 A.R. Cashmore, Nuclear genes encoding the small subunit of Ribulose 1,5 bisphosphate carboxylase, in: T. Kosuge, C.P. Meredith and A. Hollaender (Eds.), Genetic Engineering of Plants, Plenum Press, New York, 1983, p. 29.
30 A. Viotti, E. Sala, R. Marotta, P. Alberi, C. Balducci and C. Soave, Eur. J. Biochem., 102 (1979) 211.
31 G. Hagen and I. Rubenstein, Gene, 13 (1981) 239.
32 E.H. Davidson, H.T. Jacobs and R.J. Britten, Nature, 301 (1983) 468.
33 H.R.B. Pelham, Cell, 30 (1982) 517.
34 T.F. Donahue, R.S. Davis, G. Lucchini and G.R. Fink, Cell, 32 (1983) 89.
35 C. Wu, Nature, 311 (1984) 81.
36 F. Schoffl, E. Raschke and R.T. Nagao, EMBO J., 3 (1984) 2491.
37 V.P. Mauro, T. Nguyen, P. Katinakis and D.P.S. Verma, Nucl. Acids Res., 13 (1985) 727.
38 E.S. Dennis, M.M. Sachs, W.L. Gerlach, E.J. Finnegan and W.J. Peacock, Nucl. Acids. Res., 13 (1985) 727.
39 B. Forde, A. Hayworth, J. Pywell and M. Kreiss, Nucl. Acids Res., 13 (1985) 7327.
40 H. Weiher, M. Konig and P. Gruss, Science, 219 (1983) 626.