The EMBO Journal vol.8 no.10 pp.2795 - 2801, 1989

The *O2* gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators

Hans Hartings, Massimo Maddaloni, Nadia Lazzaroni, Natale Di Fonzo, Mario Motto, Francesco Salamini[1] and Richard Thompson[1]

Istituto Sperimentale per la Cerealicoltura, Sezione di Bergamo, Via Stezzano 24, 24100 Bergamo, Italy and [1]Max-Planck Institut für Züchtungsforschung, Egelspfad, D-5000 Köln 30, FRG

Communicated by J.Schell

The structure of the zein regulatory gene *Opaque 2* of *Zea mays* has been determined by sequence analysis of genomic and cDNA clones. The size of *O2* mRNA is 1751 bp [poly(A) tail not included] containing a major open reading frame (ORF) of 1380 bp preceded by three short ORFs of 3, 21 and 20 amino acid residues. The main ORF comprises 1362 bp and is composed of six exons ranging in size from 465 to 61 bp and five introns of 678 bp to 83 bp. A putative protein 454 amino acids long was derived by the theoretical translation of the genomic sequences corresponding to exons. The opaque 2 protein contains a domain similar to the leucine zipper motif identified in DNA binding proteins of animal proto-oncogenes such as *fos, jun* and *myc*, and in the transcriptional activators *GCN4* and *C/EBP*. The region of 30 amino acid residues next to the leucine repeats towards the N terminus is rich in basic amino acids and is also homologous to a domain present in *fos, jun* and *GCN4*. Moreover, in the carboxy terminal region an amino acid motif closely resembling a metal binding domain is present.

Key words: O2 locus/DNA binding protein/*myc*, *jun*, *fos* proto-oncogenes/transcriptional activators/*Zea mays*

Introduction

The storage proteins of maize seed consist of a group of alcohol soluble polypeptides collectively known as zeins. These proteins are synthesized in the endosperm tissue, between 15 and 40 days after pollination, on the rough endoplasmic reticulum and on the surface of protein bodies (Burr and Burr, 1976; Larkins and Hurkman, 1978). Zeins account for ~50% of the total endosperm protein in the mature seed.

The zein polypeptides, although similar in sequence and size, can be resolved into two major mol. wt classes of 20 and 22 kd, each composed of many isoelectric point variants (Lee *et al.*, 1976; Gianazza *et al.*, 1976; Righetti *et al.*, 1977). The 22 and 20 kd zein components are encoded by large multigene families with high levels of sequence similarity among them (Hagen and Rubenstein, 1981; Heidecker and Messing, 1986). The expression of zein genes is coordinately regulated and zein mRNAs accumulate to high concentration during early stages of endosperm development (Boston *et al.*, 1986; Marks *et al.*, 1985).

Several loci have been identified which exert a regulatory effect on the production of zein proteins during endosperm development (for review see Motto *et al.*, 1989). One of these, the *Opaque 2* locus (*O2*), has been particularly studied because kernels of its recessive mutants are better suited in amino acid composition to human and monogastric animal nutrition (Mertz *et al.*, 1964). In fact, in these mutants the synthesis of the nutritionally unbalanced zein proteins is reduced by 60% compared to normal genotypes, and the 22 kd zein class component is nearly absent (Jones *et al.*, 1977).

The reduction of the 22 kd zeins in *o2* mutants is correlated with reduced amounts of the corresponding mRNAs suggesting that this mutation alters transcription (Pedersen *et al.*, 1980; Burr and Burr, 1982; Marks *et al.*, 1985). More recently, Kodrzycki *et al.* (1989), provided more direct evidence that zein gene expression is regulated transcriptionally. Furthermore, genetic evidence indicated that the *O2* locus is located on the short arm of chromosome 7, while zein genes encoding the 22 zein kd class are mainly located on chromosome 4 (Soave and Salamini, 1984). Taken together, all available data suggest that *O2* is a *trans*-acting transcriptional activator of zein gene expression.

The *O2* locus was recently cloned using a transposon tagging strategy with the help of the mobile elements *Spm* (Schmidt *et al.*, 1987) and *Ac* (Motto *et al.*, 1988). We present here the main structural characteristics of this gene as evident from sequence data from genomic and cDNA clones. The opaque 2 protein contains a putative domain similar to the leucine zipper motif identified in DNA binding proteins of animal proto-oncogenes and in transcriptional regulators of yeast (Vogt *et al.*, 1987; Struhl, 1987; Landschulz *et al.*, 1988a,b). Adjacent to this sequence, towards the amino terminus, a conserved cluster of basic residues is present similar to those adjacent to the leucine zipper of transcriptional activators such as c-*myc*, v-*jun*, v-*fos* and *GCN4* (Landschulz *et al.*, 1988a). Moreover, in the carboxy terminal region of the 50 kd *O2* protein a motif closely resembling a metal binding domain is present.

Results

Identification of O2 cDNA and genomic clones

A maize endosperm cDNA library was prepared in λ NM1149 using a poly(A) rich RNA fraction from wild-type endosperm of the line A69Y as the template for cDNA synthesis. The primary library consisted of 2×10^6 clones. Approximately 0.002% of these phages were positives when screened with the 0.9 kb *Xho*I genomic fragment specific of the *O2* gene (Motto *et al.*, 1988). Phage DNA was isolated from the positive clones and the cDNA insert size

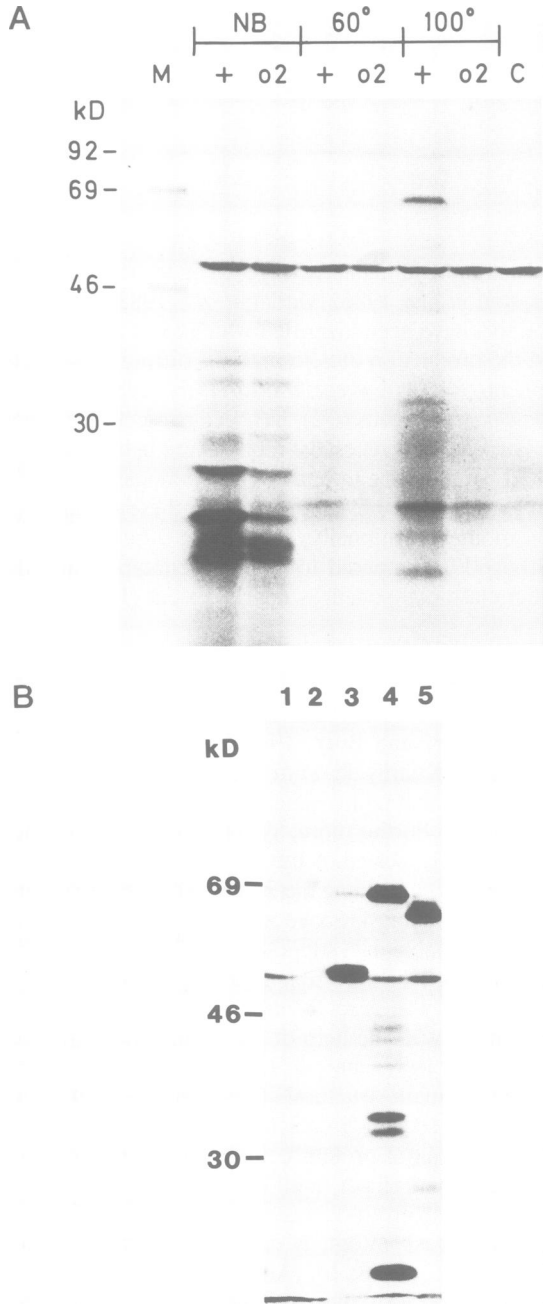

Fig. 1. (**A**) Hybrid-selected translation of *O2* gene product. The cDNA clone insert from pOp3 was used as a probe for hybrid selection. **Lane M** = ^{14}C-methionine-labelled mol. wt markers. 5 μg of pOp3 insert was bound to nitrocellulose and hybridized to 30 μg poly(A)$^+$ RNA from A69Y wild-type endosperm (+) and A69Y mutant (*o2*). NB = *in vitro* translation products of non-bound RNA. *In vitro* translation products of RNA eluted from the filters at 60°C or at 100°C are correspondingly indicated. C = *in vitro* translation products without added mRNA. (**B**) *In vitro* transcription and translation of full-length *O2* cDNA insert. Translation products were prepared as described in Materials and methods. (1) Endogenous translation products of rabbit reticulocyte lysate. (2) Standard set of mol. wt markers. (3) Transcription product of plasmid pOp3 after digestion with *Sal*I. (4) *Pvu*I cuts of coding region in pOp3. (5) *Bam*HI cuts of coding region in pOp3.

determined by agarose gel electrophoresis. The largest clone, pOp1, showed a cDNA insert of ~1200 bp and was further used to rescreen the same library. Out of a total of 40 clones isolated in the two screens the clone pOp2, with a length of 700 bp was chosen because it overlapped pOp1 and extended it. Both cDNA clones were subcloned in the pGEM3Zf(+) cloning vector and sequenced. A complete in-frame cDNA sequence was obtained by overlapping the two cDNA clones and obtaining the pOp3 plasmid.

By hybrid-selected translation in a rabbit reticulocyte system, the *O2* cDNA insert is able to select and tightly bind a mRNA from a population of 20 day old maize wild-type endosperm RNAs which directs *in vitro* the synthesis of a polypeptide with an apparent mol. wt of 58 000. This polypeptide is not detected by *in vitro* translation of hybrid-selected mRNA extracted from *o2* mutant endosperms (Figure 1A). These results estabish that the cloned cDNAs are homologous to a mRNA present only in wild-type extracts. Moreover the pOp3 plasmid was transcribed *in vitro* using SP6 RNA polymerase and the resulting RNA translated in the rabbit reticulocyte lysate system which showed that the polypeptide coded had a size identical to that of the *in vivo* product (Figure 1B).

Upon screening of a maize genomic library with the 0.9 kb *Xho*I fragment of the *o2-m5* allele, two clones containing overlapping inserts were selected and analyzed by restriction mapping. Suitable restriction fragments were subcloned into the phagemids pGEM3Zf(+) and pGEM3Zf(−) and sequenced according to the chain termination method of Sanger *et al.* (1977).

The structure of the O2 coding region

The genomic sequence corresponding to the region of DNA encoding the *O2* gene was obtained. The sequence of the genomic *O2* coding unit and 5′ and 3′ flanking regions from −1548 to +3213 will be made available to a gene data bank (Maddaloni *et al.*, submitted). In Figure 2 we reproduce only the nucleotide sequences from −350 to −291, from −40 to +20 and from +2480 to +2599 (coordinate +1 is considered the first base of the ATG codon which opens the main ORF; see later). The region with the highest homology to a TATA box consensus sequence (5′-CTATTTG-3′) starts at −339. In addition, no obvious CAAT or AGGA boxes are evident in the sequenced 1548 bp which precede the cited ATG start site of translation, even on the opposite strand, but a CAT-rich region occurs between nucleotide −375 and −398 at an appropriate distance from the cap site. The transcription start point of the gene can be putatively located at around 300 bp upstream of the start codon. Here a putative sequence, ATGCAT, is in agreement with the consensus sequence for the transcription start reported for plant genes (Joshi, 1987). The leader sequence preceding the ATG at position +1 is rather long because it extends for 289 nucleotides downstream of the 5′ end of the longest cDNA clone. Three in line termination signals located at positions −162, −153 and −138 respectively, that would invalidate the translation of any longer polypeptide than that coded by the main ORF, are presented in the 5′ untranslated region of the cDNA and genomic clones, leaving no doubt about the start position of the major polypeptide coded by *O2*. The sequence surrounding the first ATG codon, 5′-GGCATGG, is consistent with the proposed eukaryotic translation initiation sequence of (A/G)NNATGG (Kozak, 1984).

Translation terminates at a TAG stop codon (position 2490; Figure 3). A polyadenylation signal is present 38 bp downstream of the TAG stop codon. The poly(A) tail begins 34 bases past the polyadenylation signal, as is typical of

```
-350 ATCACTTGTC CTATTTGCTG CCCTGCAGGT TCACATTGAG TGCAAGGCCG ATGCATTTTT

 - 40 TTGCTTGGAA CCATTGATTG ATAGTTACTT ATTATTGGGC ATGGAGCACG TCATCTCAAT

+2480 AGGCGTCGCT GAATAAGGCT GGTTGTCTCG ATCTCCCTTG ACATGAAATC CAAATAACTC
```

Fig. 2. Segments of the *O2* genomic clone sequence relevant for the characterization of the gene. Underlined are TATA box, ATG, stop codon and the polyadenylation signal.

eukaryotic genes, resulting in a 3′ transcribed, non-translated region of 79 bases. In addition to the polyadenylation signal AATAAA, two other sequences have been shown to play a role in the 3′ processing of mRNA. The consensus YGTGTTTYY (Y = pyrimidine; McLauchlan *et al.*, 1985) is located about 24 to 30 bp downstream of polyadenylation signals, i.e. 3′ of the poly(A) site, in 67% of the mammalian sequences. Good homology to this consensus is only found 16 bp (AGTGTTCC, position 2553) downstream of the polyadenylation signal for the *O2* mRNA.

The other conserved sequence (CAYTG), located adjacent to the polyadenylation site of vertebrate genes is thought to direct the 3′ cleavage point (Berget, 1984). Accordingly, there is a similar signal in the *O2* gene (CCTTG) located 11 bp 5′ to the corresponding *O2* poly(A) site (+2516 in Figure 2).

To analyze the structure of the *O2* gene in the region coding for the main *O2* protein, DNA sequences of both genomic and cDNA clones were compared. As shown in Figure 3 the coding region present in the genomic sequence is composed of six exons, ranging in size from 465 bp to 61 bp, and five introns of 678 bp to 83 bp. It is worth noting that the introns are located approximately in the central part of the gene. Moreover, these introns (not shown) conform to the GT−AG rule of exon−intron borders (Breathnach and Chambon, 1981). The translatable cDNA sequences and the genomic ones can be perfectly aligned with the exception of the presence of an in-frame deletion of 18 bp in the genomic clone located at position 67 and a base substitution (A to C) at position +449. This nucleotide substitution implies an amino acid change from Ala to Asp. A complete homology was also evident between the 5′ and 3′ untranslated regions of the transcript present in the cDNA clone and the corresponding genomic sequences, with the exception of two nucleotides (one insertion and one base substitution).

The O2 transcripts

The minimum size of the *O2* mRNA as evaluated by considering the two cDNA clones with longest 5′ and 3′ stretches is 1751 bp [poly(A) tail not included]. The search for putative ORFs in this sequence revealed four regions of interest. The first three code for three putative short peptides 3, 21 and 20 amino acids long. Their ATG codons are respectively positioned at −147, −131 and −106 bp from the ATG where the main ORF starts. All three ORFs end in a stop codon and are out of frame when compared to the main ORF. They are included in a DNA sequence comprising 289 bases which in pOp1 precedes the ATG of the main ORF.

The main ORF includes a total of 1380 bases (1362 in the genomic clone). The distance between the stop codon and the 3′ processing site is 79 bases. The predicted size

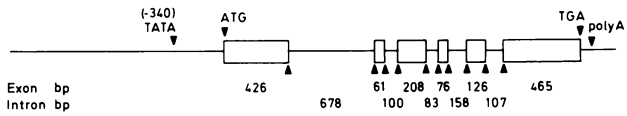

Fig. 3. Structure of the *O2* gene of maize derived from DNA sequences of both genomic and cDNA clones. The TATA box, the translation start, the stop codon and the polyadenylation signal are indicated. The ORF is split into six exons (blocks) interrupted by five introns [cDNA = 1751 bp; ORF = 1380 bp; protein = 460 amino acids (mol. wt 50 384); protein (genomic) = 454 amino acids; ▲ = consensus PuAGgtaPy...PyagGTAPu].

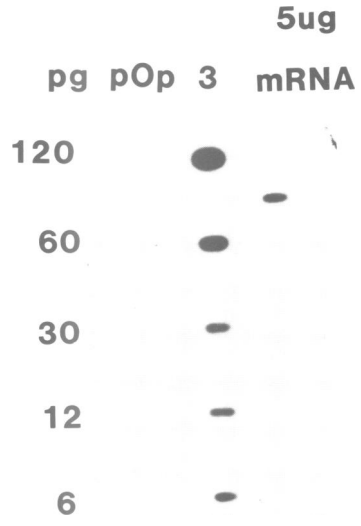

Fig. 4. Blot analysis of the abundance of *O2* mRNA. **Lane 1,** 5 μg of total mRNA. Lane 2, amounts of cDNA from pOp3 plasmid, 120 pg, 60 pg, 30 pg, 12 pg, 6 pg from top to bottom respectively.

of the *O2* mRNA as estimated from the genomic clones is 1732 nucleotides [poly(A) tail not included]. This corresponds to the 2.0 kb *O2* specific poly(A)$^+$ RNA that we have previously seen in Northern blots (Motto *et al.*, 1988).

The abundance of *O2* poly(A)$^+$ mRNA in extracts of wild-type endosperms was determined by blot analysis, using increasing amounts of *O2* cDNA insert from the plasmid pOp3 for comparison (Figure 4). The results show that there are ~2−3 copies of *O2* mRNA for every 10^5 copies of total mRNA.

Structural analysis of the O2-encoded protein

The sequence of the putative main protein encoded by the *O2* locus (Figure 5) was derived by theoretical translation of the cDNA into the corresponding amino acid sequence. The protein has 460 amino acid residues and a computed mol. wt of 50 384 d. No sequence with the characteristics of a signal peptide is observable after the start codon.

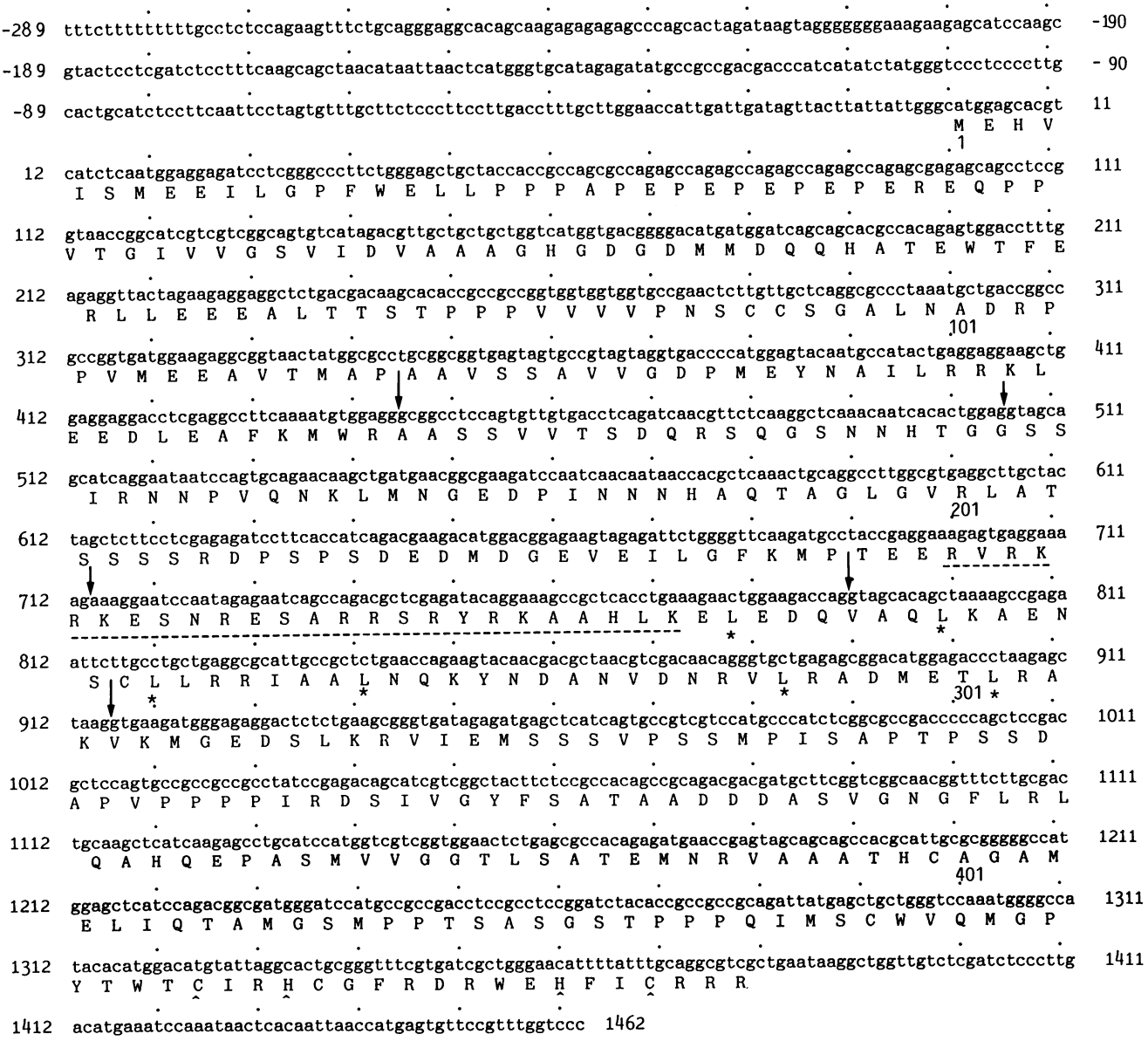

Fig. 5. cDNA sequence and predicted coding sequence of the opaque 2 polypeptide. A dotted line indicates the position of the basic domain preceding the putative leucine zipper(s). Leucine residues relevant to the putative zipper are marked by asterisks. The cysteine and histidine residues which may contribute to metal binding are indicated by chevrons. Arrows indicate exon−intron borders.

In the amino terminal part of the protein, amino acids frequently appear in pairs (i.e. Pro-Glu). The central part of the protein is rich in acid and basic amino acids. This makes the central domain of the protein highly hydrophilic. In addition, in the carboxy terminus region an interesting periodicity of leucine can be observed. Comparisons of the cDNA sequence in this region revealed homology with the DNA sequence of the yeast transcriptional activator gene *GCN4*. Further inspection of the homologous region at the protein sequence level (Figure 6) confirmed the periodic repetition of leucine residues which has homology with the leucine motif that occurs in the protein product of *myc*, *fos*, *jun* proto-oncogenes from human and mouse, the protein C/EBP (enhancer binding protein) from rat liver nuclei and the transcriptional activator *GCN4* of yeast (cf. Landschulz *et al.*, 1988a). In the *O2* sequence presented from amino acid 260 to 281 (Figure 6) the sequence Leu-X_6-Leu-X_6-Leu-X_6-Leu is present, followed by three other groups of seven amino acids, one of which has an alanine as the first residue and the other two a leucine (from amino acid 288 to 308).

It was also noted that the region of 30 amino acid residues next to the leucine repeats towards the amino terminus of the protein is rich in basic amino acids. This region is also homologous with similar DNA sequences found in *jun*, *GCN4* and *fos*. It is also notable that the leucine repeats and the highly positively charged region are encoded separately by exons 4 and 5, with an intron located between the two domains (the first leucine of the first heptamer belongs, however, to exon 4).

Close to the carboxy terminal region a sequence is present which may function as a metal binding site having two closely spaced cysteines followed by two histidines (Berg, 1986).

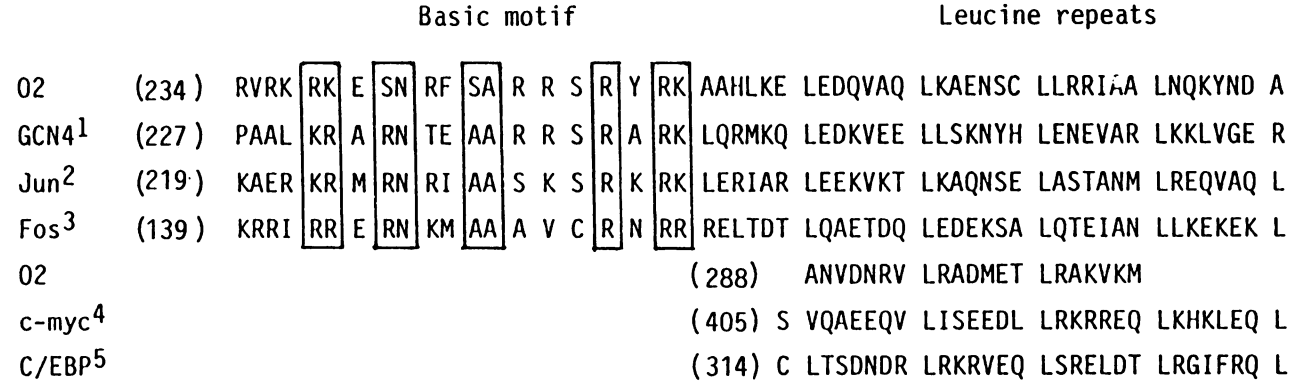

Fig. 6. Sequence comparison of putative 'basic' and 'leucine repeat' motifs of *O2* with corresponding regions of other proteins involved in transcriptional regulation. The second *O2* sequence starting at 288 shows a possible extension of the zipper region. (1) Hinnebusch (1984); (2) Maki *et al.* (1987); (3) Van Beveren *et al.* (1983); (4) Watson *et al.* (1983); (5) Landschulz *et al.* (1988b).

Discussion

Several plant genes highly regulated either in specific cell types or during the development of a particular tissue or organ, are currently under study in several laboratories (reviewed in Kuhlmeier *et al.*, 1987). The goal of these investigations is the discovery and description of specific proteins involved in transcriptional activation and of their sequences located in the 5′ region of genes under control which act as consensus motifs for protein binding. Accumulation of zein proteins in maize endosperm presents an ideal model system to study plant gene regulation. These proteins are synthesized only in the endosperm tissue, they are coded by a multigene family consisting of several subfamilies, the structural genes encoding zein polypeptides are developmentally regulated and mutants are known which control the expression of subfamilies of zein genes.

The sequence of the *Opaque 2* gene as presented here contributes significantly in this matter. All consensus sequences necessary for the expression of the gene have been found in the nucleotide sequence of the cloned DNA and they correspond to conventional signals present in other plant genes (cf. Joshi, 1987). A clear CAAT or AGGA box, however, is not evident, but the absence of such a consensus is not surprising because a similar situation is also found in other plant genes (Heidecker and Messing, 1986). The untranslated leader sequence preceding the major ORF coding for a putative polypeptide of 50 kd is quite long compared with other plant genes. An inspection of the possibility of additional coding units located in this region, revealed three ORFs with the coding capacity of 3, 21 and 20 amino acids respectively. We have no evidence for the presence of such peptides *in vivo*, but it is nevertheless tempting to speculate on the possibility that they can be involved in a regulatory mechanism controlling the expression of *O2*. In this connection, long 5′ untranslated regions have also been reported for *GCN4* and *jun*, *trans*-acting factors with leucine zipper motifs (Hinnebusch, 1984; Hattori *et al.*, 1988) homologous to the one existing in *O2*. The coincidence extends to the presence in *jun* of an unusual TATA box (Hattori *et al.*, 1988), a situation again encountered in *O2*.

The protein translation of the nucleotide sequence of the main ORF present in *O2* reveals further interesting features. As shown in Figure 6, a striking homology of this amino acid sequence was found with proteins known to be involved in transcriptional activation, such as the gene products of *GCN4*, *jun*, *fos*, *myc* and *C/EBP* (Landschulz *et al.*, 1988a). The region of homology is restricted to *O2* protein domains encoded by exons 4 and 5. In the first of these two domains, 11 residues out of 30 are positively charged. This amino acid sequence is particularly well conserved compared, for instance, with the ones present in *jun*, *fos* and *GCN4* (for instance, at the nucleotide level, from amino acid 253 to 260 the homology between *GNC4* and *O2* is 86%). That this region is required for DNA binding is confirmed by *in vitro* studies with *fos* where mutations in this region abolished binding (Neubert *et al.*, 1989). This DNA binding, basic domain is common to other DNA binding proteins possessing leucine repeats (Kouzarides and Ziff, 1988). In the O2 protein basic domain, the presence of lysine and arginine residues which favor α-helices, and the absence of proline which is not present in this type of secondary structure, reinforces at a functional level the cited homology with the transcription regulatory proteins of yeast and mammals.

The second protein domain of Opaque 2 that, with the exception of a first critical leucine, is encoded by exon 5, is characterized by a periodic repetition of leucine conforming to the Leu-X_6-Leu motif reported for the so-called leucine zipper present in proteins encoded by several transcriptional activators (Landschulz *et al.*, 1988a). A significant homology with GCN4 is moreover evident at intermediate positions between two leucines, particularly in the region coding the amino acids from 260 to 276. In the same protein domain of *O2* containing the described leucine zipper, a second stretch of amino acids conforms to the leucine periodicity Leu-X_6-Leu and extends for two further groups of heptamers.

The very C-terminal portion of the major *O2* protein may have a third domain with interesting binding properties. In exon 6 two cysteines are present which are followed by two other residues and then by histidine. It is known that such sequences may constitute a metal binding core (Miller *et al.*, 1985). The sequence reported, for instance, is a component of the binding site of plastocyanin, a Cu metalloprotein (Haehnel, 1986). A similar Cys-X_2-His motif participates in DNA fingers capable of nucleic acid binding (Berg, 1986), or flanks a Cys-X_5-Cys loop responsible for Zn binding (the case of aspartate transcarbamoylase, Berg, 1988). The

presence in the major *O2*-encoded product of the two Cys-X_2-His repeats spaced by eight amino acids suggests a folding or at least a binding capacity of the molecule which may be of interest for gene regulation. The association of finger structures with transcriptional activation is in fact so well accepted that the presence of the corresponding motifs may be considered diagnostic of new transcription factors (Evans and Hollenberg, 1988).

To conclude, we would like to point out that in maize gene tagging via transposon-mutagenesis is contributing relevant tools for understanding transcriptional activation in plants. The maize genes *C1* and *O2*, known from classical genetic studies to have the capacity to activate other genes *in trans* (Coe and Neuffer, 1977), after cloning and sequencing have all been found to encode DNA binding functions. The *C1* gene encoded protein has homology with a *myb* proto-oncogene product (Paz-Ares *et al.*, 1987) and the *O2* gene to *jun* related DNA binding motifs. These clones may provide a probe toward the isolation of other regulatory genes. The use of heterologous probes derived from critical DNA sequences of the maize *C1* gene has already made possible the isolation of a small gene family of putatively regulated genes in barley (Marocco *et al.*, 1989).

Materials and methods

Enzymes and chemicals

Restriction enzymes, T4 DNA ligase and nick translation kit were obtained from Bethesda Research Laboratories. Sequenase® enzyme was purchased from United States Biochemical. All enzymes were used as indicated by the manufacturer.

Deoxynucleotides, dideoxynucleotides and sequence primer were obtained from United States Biochemicals; reverse sequence primer, ^{32}P-labeled nucleoside triphosphate and [^{35}S]dATP were purchased from Amersham International.

The phagemids pGEM3Zf(+) and pGEM3Zf(−), helper phages R408 and M13K07 as well as bacterial strain JM109 were obtained from Promega Biotech.

Construction of an A69Y cDNA library

Total RNA was extracted from endosperms 20 days after pollination from the inbred line A69Y$^+$ and purified as described by Dean *et al.* (1985). Poly(A)$^+$ RNA was then obtained by two cycles of oligo(dT) cellulose chromatography (Aviv and Leder, 1972). Subsequently a cDNA library was prepared using the cDNA synthesis kit from Boehringer, Mannheim. The synthesized cDNA was size selected (>700 bp) by agarose gel electrophoresis and remaining *Eco*RI linkers were removed by selective precipitation. The *Eco*RI-linked cDNA was ligated to *Eco*RI-digested λ NM1149 arms, prepared as described by Maniatis *et al.* (1982) and packaged *in vitro*. Approximately 2 × 10^6 plaque forming units were plated on the selective strain pop13. This library was then screened with the 0.9 *Xho*I fragment of the *02-m5* clone number 6IP (Motto *et al.*, 1988) and the isolated clone was further used to rescreen the same library. Two clones, pOp1 and pOp2 which were selected from this library, when overlapped gave a full length cDNA sequence. They were subcloned in the pGEM3Zf(+) cloning vector and used for sequence analysis.

Screening of a genomic library

A genomic library (a kind gift of Alfons Gierl, Max-Planck Institut für Züchtungsforschung, Cologne), of the maize accession AC1503 GM1417 MPI, cloned, after partial digestion with *Sau*3A, in the λ vector EMBL4 was screened using a probe derived from a 0.9 kb *Xho*I fragment of the *o2-m5* clone number 6IP (Motto *et al.*, 1988). Two clones containing overlapping inserts were used in restriction map analysis of the *O2* gene and its flanking regions. Fragments containing the coding and flanking regions were subcloned in the pGEM3Zf(+) vector.

DNA sequence analysis

Sequence analysis was performed following the dideoxynucleotide chain terminator method (Sanger *et al.*, 1977) using the phagemids pGEM3Zf(+) and pGEM3Zf(−) in combination with Sequenase (modified T7 DNA polymerase) enzyme. The cDNA clones were sequenced on both strands, subcloning large overlapping fragments in order to avoid sequence ambiguities.

Hybrid-selected translation

The methods used were as previously described by Di Fonzo *et al.* (1988).

In vitro transcription and translation

Starting from pOp1 and pOp2, a third clone pOp3 was derived by cloning in pGEM3Zf(+) which contains the full length coding capacity of the *O2* gene.

500 ng of plasmid DNA pOp3 was cut with the appropriate restriction enzyme, precipitated, washed, dried and subsequently resuspended in 40 mM Tris−HCl (pH 7.4), 6 mM $MgCl_2$, 2.5 mM Spermidine, 10 mM NaCl, 10 mM DTT, 0.5 mM each of ATP, GTP, UTP, CTP, 15 units RNasin® (Ribonuclease inhibitor) in a total volume of 20 μl. Sixteen units of SP6 RNA polymerase (Promega) were added and incubation was carried out at 37°C for 1 h; after phenol extraction 20 μg of tRNA was added and the RNA precipitated overnight at −30°C, adding 1/4 vol of NH_4 acetate (7.5 M) and 2.5 vol of absolute ethanol. Total RNA was translated *in vitro* using the rabbit reticulocyte lysate system as previously described (Di Fonzo *et al.*, 1988).

DNA preparation and probes

The hybridization probes were prepared from gel-purified DNA inserts and labeled with ^{32}P by nick translation (Rigby *et al.*, 1977) to a specific activity of 1 × 10^8 c.p.m./μg DNA. Plasmid DNA was prepared in small amounts by alkaline lysis (Maniatis *et al.*, 1982) and in large quantity by the method of Clewell and Helinski (1969). Purification of bacteriophage λ and extraction of phage DNA were as described by Yamamoto *et al.* (1970).

Acknowledgements

This work was supported by EEC contract number BAP-0214-I(A) in the framework of Biotechnology Action Programme and by Ministero dell'Agricoltura e delle Foreste, Rome, Italy. Dr A.Gierl, U.Wienand and P.Starlinger are acknowledged for their contributions of gene libraries and valuable discussion of results. At the recent maize meeting held in Monticello, WI, USA, R.Schmidt (University of California) reported independent results on the leucine repeats similar to those discussed here.

References

Aviv,H. and Leder,P. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 1408−1412.
Berg,J.M. (1986) *Science*, **232**, 485−487.
Berg,J.M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 99−102.
Berget,S.M. (1984) *Nature*, **309**, 179−182.
Boston,R.S., Kodrzycki,R. and Larkins,B.A. (1986) In Shannon,L.M. and Chrispeels,M.J. (eds), *Molecular Biology of Seed Storage Proteins and Lectins*. University of California, Riverside, CA, pp. 117−126.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349−383.
Burr,B. and Burr,F.A. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 515−519.
Burr,F.A. and Burr,B. (1982) *J. Cell Biol.*, **94**, 201−206.
Clewell,D.B. and Helinski,D.R. (1969) *Proc. Natl. Acad. Sci. USA*, **62**, 1159−1166.
Coe,E.H.J. and Neuffer,M.G. (1977) In Sprague,G.F. (ed.), *Corn and Corn Improvement*. Madison, WI, USA, pp. 111−223.
Dean,C., Van Den Elzen,P., Tamaki,S., Dunsmuir,P. and Bedbrook,J. (1985) *EMBO J.*, **4**, 3055−3061.
Di Fonzo,N., Hartings,H., Brembilla,M., Motto,M., Soave,C., Navarro,E., Palau,J., Rohde,W. and Salamini,F. (1988) *Mol. Gen. Genet.*, **212**, 481−487.
Evans,R.M. and Hollenberg,S.M. (1988) *Cell*, **52**, 1−3.
Gianazza,E., Righetti,P.G., Pioli,F., Galante,E. and Soave,C. (1976) *Maydica*, **21**, 1−17.
Haehnel,W. (1986) In Pirson,A., Zimmermann,M.H. (eds), *Encyclopedia of Plant Physiology*. Springer, NY, pp. 547−559.
Hagen,G. and Rubenstein,I. (1981) *Gene*, **13**, 239−249.
Hattori,K., Angel,P., Le Beau,M.M. and Karin,M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 9148−9152.
Heidecker,G. and Messing,J. (1986) *Annu. Rev. Plant Physiol.*, **37**, 439−466.
Hinnebusch,A.G. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6442−6446.
Jones,R.A., Larkins,B.A. and Tsai,C.Y. (1977) *Plant Physiol.*, **59**, 525−529.

Joshi,C.P. (1987) *Nucleic Acids Res.*, **15**, 6643–6653.
Kodryzcki,R., Boston,R.S. and Larkins,B.A. (1989) *Plant Cell*, **1**, 105–114.
Kouzarides,T. and Ziff,E. (1988) *Nature*, **336**, 646–651.
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.
Kuhlmeier,C., Green,P.J. and Chua,N.-H. (1987) *Annu. Rev. Plant Physiol.*, **38**, 221–257.
Landschulz,W.H., Johnson,P.F. and McKnight,S.L. (1988a) *Science*, **240**, 1759–1764.
Landschulz,W.H., Johnson,P.F., Adashi,E.Y., Graves,B.J. and McKnight,S.L. (1988b) *Genes Dev.*, **2**, 786–800.
Larkins,B.A. and Hurkman,W.J. (1978) *Plant Physiol.*, **62**, 256–263.
Lee,K.H., Jones,R.A., Dalby,A. and Tsai,C. (1976) *Biochem. Genet.*, **14**, 641–650.
Maddaloni,M., Di Fonzo,N., Hartings,H., Lazzaroni,N., Salamini,F., Thompson,R. and Motto,M. (1989) *Nucleic Acids Res.*, **17**, in press.
Maki,Y., Bos,T.J., Davis,C., Starbuck,M. and Vogt,P.K. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 2848–2852.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Marks,M.D., Lindell,J.S. and Larkins,B.A. (1985) *J. Biol. Chem.*, **260**, 16445–16450.
Marocco,A., Wissenbach,M., Blecker,D., Paz-Ares,J., Saedler,H., Salamini,F. and Rohde,W. (1989) *Mol. Gen. Genet.*, **216**, 183–187.
McLauchlan,J., Gaffney,D., Whitton,J.L. and Clements,J.B. (1985) *Nucleic Acids Res.*, **13**, 1347–1368.
Mertz,E.T., Bates, L.S. and Nelson,O.E. (1964) *Science*, **145**, 279–280.
Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1614.
Motto,M., Maddaloni,M., Ponziani,G., Brembilla,M., Marotta,R., Di Fonzo,N., Soave,C., Thompson,R.D. and Salamini,F. (1988) *Mol. Gen. Genet.*, **2121**, 488–494.
Motto,M., Di Fonzo,N., Hartings,H., Maddaloni,M., Salamini,F., Soave,C. and Thompson,R.D. (1989) *Oxford Surveys of Plant Molecular and Cell Biology*. Vol. 6, in press.
Neuberg,M., Schuermann,M., Hunter,J.B. and Müller,R. (1989) *Nature*, **338**, 589–590.
Paz-Ares,J., Ghosal,D., Wienand,U., Peterson,P.A. and Saedler,H. (1987) *EMBO J.*, **6**, 3553–3558.
Pedersen,K., Bloom,K.S., Anderson,J.N., Glover,U.V. and Larkins,B.A. (1980) *Biochemistry*, **19**, 1644–1650.
Rigby,P.W., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237–251.
Righetti,P.G., Gianazza,E., Viotti,A. and Soave,C. (1977) *Planta*, **136**, 115–123.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schmidt,R.J., Burr,F.A. and Burr,B. (1987) *Science*, **228**, 960–963.
Soave,C. and Salamini,F. (1984) *Phil. Trans. R. Soc. Lond.*, **304**, 341–347.
Struhl,K. (1987) *Cell*, **50**, 841–846.
Van Beveren,C., Van Straaten,F., Curran,T., Müller,R. and Verma,I.M. (1983) *Cell*, **32**, 1241–1255.
Vogt,P.K., Bos,T.J. and Doolittle,R.F. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3316–3319.
Watson,D.K., Psallidopoulos,M.C., Samuel,K.P., Dalla-Favera,R. and Papas,T.S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3642–3645.
Yamamoto,K.R., Alberts,B.M., Benzinger,R., Lawhorne,L. and Treiber,G. (1970) *Virology*, **40**, 734–740.

Received on May 12, 1989; revised on June 19, 1989