The EMBO Journal vol.10 no.3 pp.617-624, 1991

# The maize regulatory locus *Opaque-2* encodes a DNA-binding protein which activates the transcription of the *b-32* gene

**Stefan Lohmer, Massimo Maddaloni, Mario Motto[1], Natale Di Fonzo[1], Hans Hartings[1], Francesco Salamini and Richard D.Thompson**

Max-Planck-Institut für Züchtungsforschung, Carl von Linne Weg 10, D-5000 Köln 30, FRG and [1]Istituto Sperimentale per la Cerealicoltura, Sezione di Bergamo, Via Stezzano 24, 24100 Bergamo, Italy

Communicated by J.Schell

**The maize locus, *Opaque-2*, controls the expression in developing endosperm of structural genes encoding a family of storage proteins, the 22 kd zeins, and an abundant albumin, termed b-32. It is shown that the promoter of the *b-32* gene is activated *in vivo* in the presence of the *O2* gene product and that the information necessary for this activation resides in a 440 bp DNA fragment containing five *O2* binding sites (GATGAPyPuTGPu). Two of these sites are embedded in copies of the 'endosperm box', a motif thought to be involved in endosperm-specific expression, which is also represented in 22 kd zein promoters. The O2 protein is also shown to be capable of binding *in vitro* and activating *in vivo*, its own promoter.**

*Key words: b-32* gene/DNA binding protein/*O2* gene/transactivation/*Zea mays*

## Introduction

The endosperm of maize (*Zea mays*), in common with that of other cereals, is a tissue primarily devoted to the accumulation of starch and proteins which later provide a nitrogen and energy source for the germinating seedling (for a review, see Motto *et al.*, 1989). In this tissue between days 15 and 40 after fertilization, the major storage proteins, the zeins, are synthesized by membrane-bound polyribosomes and transported into the lumen of the endoplasmic reticulum, where they are assembled into protein bodies (Khoo and Wolf, 1970; Larkins and Hurkman, 1978). At maturity ~55% of the total protein content of the tissue consists of zeins.

The zeins are a mixture of polypeptides, encoded by at least four major gene families. Two families are responsible for the synthesis of the 19-20 kd and 22 kd zeins to which 15-22 polypeptides belong (Burr and Burr, 1981; Hagen and Rubenstein, 1980). In wild-type endosperm, expression of zein genes is under the control of several regulatory genes (for a review see Soave and Salamini, 1982 and Motto *et al.*, 1989). One of these is the *Opaque-2* (*O2*) locus, located on chromosome 7, which exerts its major effect on the level of the 22 kd zein polypeptides, for which there are several structural genes located on chromosome 4 (Soave *et al.*, 1982). In homozygous *o2* mutants this class of zeins is greatly reduced (Soave *et al.*, 1976; Pedersen *et al.*, 1980; Burr and Burr, 1982), whereas the levels of other zein classes are only weakly affected. The effect of the mutation is a reduction of the overall zein content to a level of 50-70% of wild-type, depending on the genetic background in which the mutation is present (Soave *et al.*, 1976). The *o2* mutation results in a reduction in the level of 22 kd zein transcripts as shown by *in vitro* nuclear run-on studies (Kodrzycki *et al.*, 1989) and is therefore likely to be involved in transcriptional activation.

The *o2* mutant also lacks a number of non-zein polypeptides present in the wild-type endosperm, the most abundant of which is a 32 kd albumin termed b-32. The b-32 protein is linked to and may be regulated or encoded by the *Opaque-6* (*O6*) locus (Soave *et al.*, 1981; Manzocchi *et al.*, 1986). It is also known that the expression of b-32 during seed development is temporally and quantitatively coordinated with the deposition of storage proteins and that in the *o2* mutant, the level of b-32 mRNA is reduced to <5% of wild-type. The gene encoding the b-32 protein has been cloned and the complete amino acid sequence of this protein derived (Di Fonzo *et al.*, 1988; Hartings *et al.*, 1990). The function of b-32 is currently undefined. However, a recent search for homology to available protein sequences revealed a region of strong homology (33.5% in a 173 amino acid long region, our unpublished observations) with the protein synthesis inhibitor II of barley seeds. The barley protein has a mol. wt of 30 kd and inhibits protein synthesis in the cell-free reticulocyte lysate system (Asano *et al.*, 1986). While it remains to be unequivocally demonstrated that the b-32 protein is the product of the *O6* gene, it is nevertheless clear that the activation of the *b-32* gene requires an *O2*-encoded function.

The *O2* gene has been cloned by using a transposon tagging strategy with the help of the mobile elements *Spm* (Schmidt *et al.*, 1987) and *Ac* (Motto *et al.*, 1988). Translation of the corresponding cDNA (isolated by Hartings *et al.*, 1989) results in the production of a 48 kd protein, which possesses characteristic features in common with the 'leucine zipper' class of *trans*-acting factors. Members of this class include the proteins c-jun, c-fos, C/EBP, GCN-4 and CREB (Bohmann *et al.*, 1987; Angel *et al.*, 1988a; Curran *et al.*, 1982; Landschulz *et al.*, 1988; Montminy and Bilezikjian, 1987; Hoeffler *et al.*, 1988). *O2* shares with these transcription factors the presence of a region rich in basic amino acids immediately preceding the leucine repeat.

In this paper it is shown that the *O2* gene product is able to transactivate the *b-32* gene as mediated by a direct protein-DNA interaction with the b-32 promoter region.

## Results

### *The Opaque-2 protein transiently expressed in tobacco protoplasts transactivates the b-32 promoter*

To verify experimentally the *trans*-acting property of the O2 protein, several potential target genes are available. These

include the zein genes, which are grouped at multigene loci also containing pseudogene members (Pedersen *et al.*, 1980; Spena *et al.*, 1983). As a consequence, the specific zein polypeptide encoded by a particular cloned gene has not in general been identified. It is therefore difficult to assess whether a given zein gene is active and whether its promoter is under the control of O2. We have chosen instead to use the b-32 promoter as a potential target site for the O2 protein. It is known that only three copies of this gene are present in the maize genome (Hartings *et al.*, 1990) and one of these is active as shown by the isolation of a homologous cDNA clone (Di Fonzo *et al.*, 1988).

An assay for transient gene expression in tobacco protoplasts has been employed to investigate the possible activation of a *b-32* promoter by the *O2* product. The assay is based on cotransfection of mesophyll protoplasts with an expression and reporter plasmid (Figure 1). The expression plasmid, pCaMVO2, consisted of the full length *O2* cDNA placed as a transcriptional fusion under the control of the strong promoter of the 35S gene from cauliflower mosaic virus (CaMV), and inserted into the binary vector pPCVOO2 (Koncz and Schell, 1986; Figure 1A). The reporter plasmid pB32GUSI (Figure 1B) was based on the pBI201.2 vector (Jefferson, 1987) and was constructed by fusing the *b-32* promoter region (from −1283 to +4, see clone b-32.129 in Hartings *et al.*, 1990) to the coding region of the bacterial β-glucuronidase (*GUS*) gene (Jefferson, 1987).

The plasmids were used separately or as a mixture to transfect tobacco mesophyll protoplasts, and GUS activity was measured after 17 h. Transfection of the reporter plasmid or of the expression plasmid alone resulted in weak basal-level GUS activity, while the cotransfection of pB32GUSI together with the pCaMVO2 expression plasmid stimulated GUS activity ~80-fold over background. The results of the transient expression experiments are reported in Table I. They demonstrate the striking activation of the b-32 promoter in the presence of the *O2* gene product.

In order to define more precisely the region of the *b-32* promoter which is responsible for being transactivated by O2, a truncated version of pB32GUSI was constructed, containing 443 bp of the promoter sequence (pB32GUSII, Figure 1C). This was also able to be activated at high levels (Table I). The promoter region it contains was used in DNA binding studies (see below).

From this experiment it was also evident that tobacco mesophyll protoplasts lack the appropriate factors at concentrations which could substitute for the stimulatory effect of O2 on the transcription of endosperm genes. This finding may be helpful for the analysis by transient expression in tobacco protoplasts of other endosperm-specific genes.

### *Mapping O2 protein binding sites*

To determine whether the transactivation of *b-32* by O2 observed in the cotransfection assay might have been mediated through direct binding of the O2 product to the *b-32* promoter, DNA-binding studies were carried out. In order to obtain purified O2 protein, the *O2* gene was expressed as an IPTG-inducible glutathione-S-transferase (GST) fusion protein in *Escherichia coli* (Smith and Johnson, 1988). It was evident from preliminary experiments that the expression level of the complete O2 protein fused to GST was very low (data not shown). Using several truncated

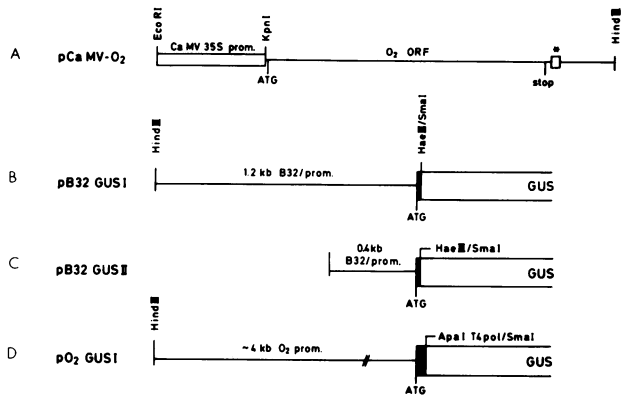


**Fig. 1.** Constructs used for the transient assays. **A.** Activator plasmid carrying the complete *O2* coding region. Asterisk indicates the *O2* polyadenylation sequence; **B.** Reporter plasmid carrying the 1.2 kb *b-32* promoter; **C.** Reporter plasmid carrying the shortened version of B; **D.** Reporter plasmid carrying the *O2* promoter region.

**Table I.** β-glucuronidase activity in transiently transformed mesophyll protoplasts

| Transfected plasmid(s) | $\frac{\text{nmol MU}}{\text{min} \times \text{mg prot}}$ | Fold activation |
|---|---|---|
| 35SGUS | 1470 | – |
| pCaMV-O2 | 8.9 | – |
| B32GUSI | 14 | 82 |
| B32GUSI + pCaMV-O2 | 1150 | |
| B32GUSII | 16.8 | 55.3 |
| B32GUSII + pCaMV-O2 | 930 | |

The β-glucuronidase activity is expressed in nanomoles of 4-methylumbelliferone (MU) produced/min/mg total protein.

versions of the *O2* cDNA fused to the *GST* gene, it appeared that the low expression level of the complete fusion protein was due to a sequence of ~60 amino acids present in the N-terminal region of O2. All constructs possessing the O2 N-terminus were expressed at low level whereas constructs of almost similar length but lacking the N-terminus region were expressed at high level. Therefore we have expressed a shortened version of O2 (GST-$O2_{55}$), where 55 indicates a fusion protein of a total mol. wt of 55 kd, which still contains 60% of the wild-type protein, including the basic domain and the leucine zipper, but lacks the N-terminal region (Figure 2B). The fusion protein was isolated via affinity chromatography on glutathione−agarose beads and the integrity and purity of the isolated product was checked by SDS−PAGE (Figure 2C). In this experiment a second construct, GST-$O2_{33}$, was also included. Figure 2C shows (in lane 3 and 4) that plasmids GST-$O2_{33}$ and GST-$O2_{55}$ direct in *E.coli* the synthesis of fusion products of the predicted size. The purified protein products from both constructs consisted of several bands, of which the two upper ones present in lane 7 are specifically derived from GST-$O2_{55}$, whereas the remaining bands are also present in purified protein products of GST-$O2_{33}$. The bands with lower mol. wts in lanes 6 and 7 must be the result of specific protein degradation or premature translation stop products of the fusion protein, because purified protein extracts from cells expressing GST alone (lane 5) lack these bands and an antibody raised against the GST component of the fusion protein recognizes all of these bands (data not shown).

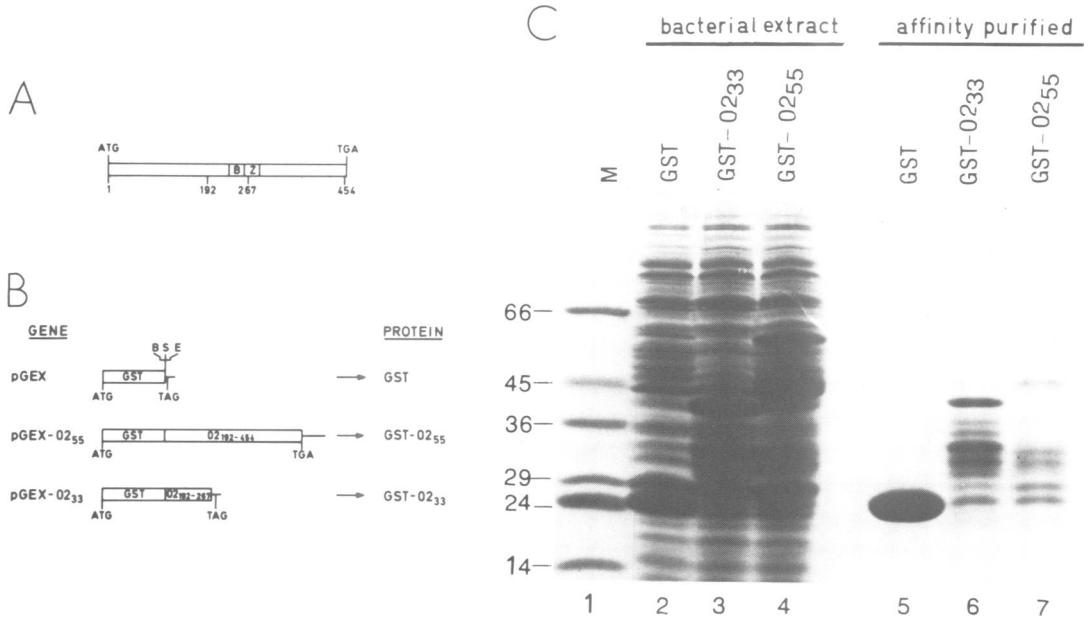


**Fig. 2.** Expression and purification of the GST-O2 fusion protein from *E.coli*. **A**. Diagram of the full-length Opaque-2 protein as deduced from the cDNA sequence (see Hartings *et al.*, 1989); B = indicates the basic domain, Z marks the leucine zipper region; numbers represent the amino acid position in the full-length sequence. **B**. Diagrams of the constructs utilized for the expression of the O2 protein in *E.coli*. pGEX-O2$_{55}$ includes amino acids 192–454 of the O2 protein. pGEX-O2$_{33}$ includes amino acids 192–267; B = *Bam*HI, E = *Eco*RI, S = *Sma*I. **C**. SDS–PAGE of protein extracts. Lane 1 = molecular weight markers; lane 2 = extract from B1-21 harbouring pGEX; lane 3 = extract from B1-21 (pGEX-O2$_{33}$); lane 4 = extract from B1-21 (pGEX-O2$_{55}$); lane 5 = purified GST; lane 6 = purified GST-O2$_{33}$; lane 7 = purified GST-O2$_{55}$.

In the following DNA-binding studies we have used the affinity-purified GST-O2$_{55}$ protein mixture rather than gel-purified 55 kd fusion polypeptide for several reasons: (i) In a south-western analysis only the highest molecular weight protein of lane 7 was capable of binding DNA containing specific consensus sequences (Figure 4C). (ii) The affinity-purified GST-O2$_{55}$ products showed the same behaviour in DNA-binding experiments as the isolated 55 kd fusion protein (data not shown). (iii) The affinity-purified GSTO2$_{55}$ protein mixture, representing the fusion protein in its native conformation, was readily available in detergent-free preparation.

As an initial test for the potential DNA binding capacity of O2 protein, gel retardation assays were performed using purified GST-O2$_{55}$ protein mixture and a fragment from the b-32 gene containing 249 bp of the promoter region between −265 and −16 upstream from the TATA box. We have chosen this particular promoter region because (i) when three different b-32 genomic clones were compared (see Hartings *et al.*, 1990), this region was highly conserved, and therefore may include sequences necessary for expression and (ii) this fragment is included in the promoter sequence of the construct pB32GUSII, which is sufficient for transactivation by O2 in the transient assay (see Table I). Incubation of purified GST-O2$_{55}$ proteins with the labelled *O2* promoter fragment resulted in two DNA–protein complexes of reduced electrophoretic mobility relative to free DNA (Figure 3A). The formation of the retarded DNA–protein complexes was specific: they were resistant against a 1000-fold molar excess of unspecific competitor DNA (lane 2), whereas a 100-fold molar excess of unlabelled *b-32* promoter fragment abolished the formation of the complexes (lane 3). To demonstrate that the O2 component of the GST-O2$_{55}$ fusion protein was responsible for the DNA-binding activity, the labelled b-32 promoter was incubated in the presence of purified GST (lane 4). Even at the high GST concentration, no DNA–protein complex formation was seen, supporting the conclusion that the O2 part of the GST-O2$_{55}$ fusion protein was responsible for DNA binding.

The site of interaction of GST-O2$_{55}$ on the *b-32* promoter fragment was mapped using DNase I footprinting (Galas and Schmitz, 1978). For this purpose the same protein mixture and end-labelled promoter fragment as utilized in the bandshift assays were used (Figure 3B). The promoter fragment tested contains five regions which are protected by O2 against DNase I digestion which we have termed B1, B2, B3, B4 and B5. Analysis of the protected regions reveals that the two protected regions B1 and B4 differ by only one nucleotide (see Table II). The sites B2 and B3 are also similar to each other (see Table II), whereas B5 is nearly homologous to the binding site of the mammalian transcription activator CREB (TGACGTCA, Montminy *et al.*, 1986; Yamamoto *et al.*, 1988). In all five sites the sequence GATGAPyPuTG is conserved. This consensus may reflect the general binding requirement of O2 protein at the nucleotide level. Inspection of the complete b-32 promoter sequence as reported by Hartings *et al.* (1990) revealed no other GATGAPyPuTG-like motifs. The *in vitro* experiments show conclusively that the *O2* gene product binds to the promoter region of *b-32* gene, a finding supported by the cotransfection experiments and consistent with the available genetic data.

### *The Opaque-2 protein binds to its own promoter and activates it in plant cells*

Inspection of the *O2* promoter by computer analysis for the presence of an O2 protein binding site, based on the footprinting data obtained for the *b-32* promoter, revealed

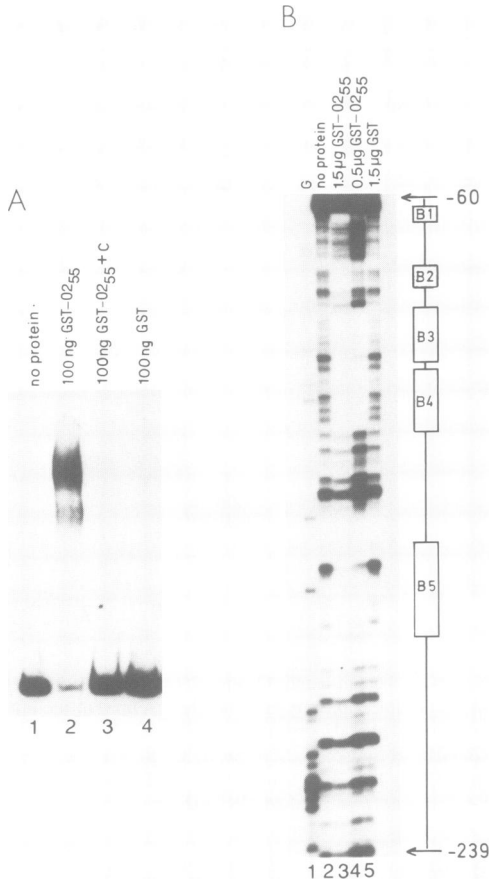


**Fig. 3.** Binding of the GST-O2$_{55}$ fusion protein to the *b-32* promoter. **A**. Gel retardation of the GST-O2$_{55}$ fusion protein on the *b-32* promoter. A lower strand end-labelled 249 bp genomic fragment, containing the promoter region between −265 and −16 upstream of the *b-32* TATA box, was used. Lane 1 = *b-32* promoter fragment alone; lane 2 = 100 ng GST-O2$_{55}$ added; lane 3 = the same as in lane 2 plus 100-fold excess of unlabelled *b-32* promoter fragment as specific competitor (C); lane 4 = the same as in lane 1 plus 100 ng of GST protein. **B**. Footprinting analysis of the DNA binding specificities of O2 fusion protein on the *b-32* promoter fragment. The same *b-32* promoter fragment was used as in the gel retardation assay. Lane 1: Maxam and Gilbert G-ladder; 2 ng end-labelled *b-32* promoter fragment was incubated before DNase I digestion with no protein (lane 2), with 1.5 µg purified GST-O2$_{55}$ mixture (lane 3), with 0.5 µg purified GST-O2$_{55}$ mixture (lane 4) or with 0.5 µg purified GST.

**Table II.** Alignment of the Opaque-2 binding sites

| Box | DNA sequence | Location relative to TATA box |
|---|---|---|
| B1 | GATGACATGG | −68 to −59 |
| B2 | GATGATATGG | −117 to −108 |
| B3 | GATGATGTGG | −143 to −134 |
| B4 | GATGACATGA | −167 to −158 |
| B5 | GTTGACGTGA | −210 to −201 |
| Consensus | GATGAPyPuTGPu (T) | |

DNA sequences on the *b-32* promoter protected by Opaque-2 in the DNase I footprinting experiment; Py = Pyrimidine, Pu = Purine.

a sequence (GTTGACGTTG) resembling the B5 binding site GTTGACGTGA. The sequence is positioned between −251 and −242 bp upstream of the TATA box. Because autoregulation is a feature of some eukaryotic transcription factors (for a review see Serfling, 1989), the existence of this DNA sequence prompted us to test whether the *O2* gene product interacts with its own promoter, thereby affecting its own expression.

It was first demonstrated in a gel retardation assay that the 377 bp *O2* promoter fragment containing the putative binding site was specifically retarded in the presence of the O2 protein (Figure 4A). Using the DNase I footprint technique, a single site was protected by the GST-O2$_{55}$ fusion protein on this fragment, which was located, as expected, in the B5-homologous region (Figure 4B). This protected site is not only nearly homologous to the B5 site of the *b-32* promoter, but is also positioned at a similar distance from the TATA box, supporting its possible role in transcriptional regulation. In addition, a south-western analysis (Figure 4C) carried out with this fragment and the affinity purified polypeptides of GST-O2$_{55}$ (Figure 2C, lane 7) showed that only the 55 kd protein is able to bind to this promoter fragment i.e., the DNA binding is not due to interaction with a degradation product of the O2 fusion.

*O2* gene auto-activation was tested in transient assay experiments similar to those already described. The reporter plasmid pO2GUSI consisted of a pUC19 vector to which a chimeric gene was added composed of the *O2* promoter region (from −4.1 kb to +34; see Maddaloni *et al.*, 1989), fused translationally to the coding region of GUS (Figure 1D). When the expression plasmid pCaMV-O2 was used in a mixture with the reporter plasmid pO2GUSI to transfect mesophyll protoplasts of tobacco, GUS activity was stimulated at the level of 2- to 4-fold over the basal reading (Table III). Although weak, the activation induced by the *O2* gene on its own promoter was consistently observed in several independent experiments.

## Discussion

The study of the regulation of gene expression in the developing maize endosperm is facilitated by the existence of a series of viable mutants, many of which affect storage protein synthesis. Genetic analysis suggests that some of these mutants may correspond to regulatory genes, which control the timing and level of zein gene expression. We have focused our attention on one putative regulatory locus, *Opaque-2*, and investigated its interaction with one of its genetically identified target genes, *b-32*.

### *Opaque-2 functions as a trans-activator in vivo*

The data presented here, showing that O2 can transactivate the *b-32* promoter *in vivo* and binds to it *in vitro*, indicate that the interaction between *O2* and *b-32* is indeed at the transcriptional level. This confirms a similar prediction made by genetic analysis (for a review, see Motto *et al.*, 1989) and on the basis of sequence homology of O2 to the leucine zipper family of transcriptional regulators (Hartings *et al.*, 1989; Schmidt *et al.*, 1990). The expression of O2 protein in tobacco mesophyll protoplasts brings about a striking activation of the *b-32* promoter coupled to a reporter gene, *GUS*. This system holds promise for investigating in more detail the structural requirements of the activator and *cis*-element components, particularly as the endogenous expression level of the reporter plasmid is close to zero.

Schernthaner *et al.* (1988) have shown previously that a

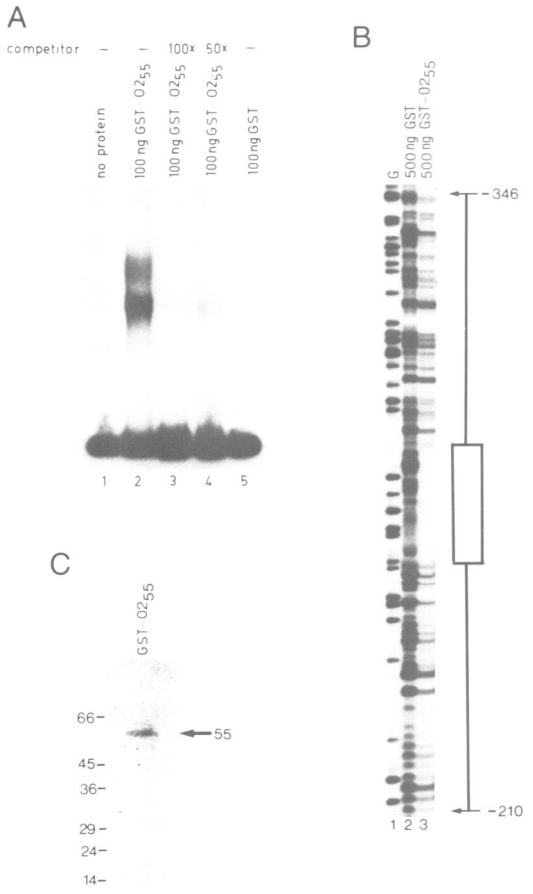


**Fig. 4.** GST-O2$_{55}$ fusion protein binds specifically to the *O2* promoter. **A**. Gel retardation assay. 1 ng of an upper strand end-labelled, 377 bp long *O2* promoter fragment, containing the region between −560 and −183 with respect to the TATA box, was incubated without protein (lane 1), with 100 ng GST-O2$_{55}$ mixture (lane 2), with 100 ng GST-O2$_{55}$ + 100-fold excess of unlabelled fragment (lane 3), with 100 ng GST-O2$_{55}$ + 50-fold excess of unlabelled fragment (lane 4) or with 100 ng of purified GST (lane 5). **B**. DNase I protection analysis of the *O2* promoter region. The same end-labelled promoter fragment of *O2* as in the bandshift experiment shown in Figure 4A was used. Lane 1: Maxam and Gilbert G-ladder. 2 ng of the end-labelled promoter fragment was incubated with 0.5 μg GST (lane 2), or with 0.5 μg purified GST-O2$_{55}$ mixture (lane 3), followed by digestion with DNase I. **C**. DNA binding capacity of the isolated GST-O2$_{55}$ fusion protein mixture. The separated GST-O2$_{55}$ fusion protein mixture (lane 7 in Figure 2C) was blotted onto nitrocellulose and incubated with the labelled *O2* promoter fragment used in A and B. Molecular weights are given in kd. The arrow indicates the position of the protein with the highest molecular weight (55 kd).

19 kd zein gene was expressed only in the endosperm of transgenic tobacco plants. Other authors (Colot *et al.*, 1987; Marris *et al.*, 1988; Robert *et al.*, 1989) have also demonstrated endosperm specific expression of prolamin genes in transgenic plants. This implies that factors necessary for zein gene expression are absent from tobacco mesophyll and other non-endosperm cells. It must be borne in mind, however, that the template concentration of the expression plasmid in the experiments reported here is much higher than that of the *O2* gene in transgenic plants. It is therefore still possible that expression of the reporter constructs in trangsenic plants are limited by factors other than Opaque-2. Another difference between the systems is that in the transient

**Table III.** β-glucuronidase activity in mesophyll protoplasts transformed with the expression plasmid and a reporter plasmid bearing the *O2* promoter coupled with the *GUS* gene

| Transfected plasmid(s) | $\frac{\text{nmol MU}}{\text{min} \times \text{mg prot}}$ | fold activation |
|---|---|---|
| pO2GUSI | 13.7 | 4.5 |
| pO2GUSI + pCaMVO2 | 62.7** | |
| pO2GUSI | 4.6 | 2.0 |
| pO2GUSI + pCaMVO2 | 9.2* | |

The β-glucuronidase activity is expressed as nanomoles of 4-methylumbelliferone (MU) produced/min/mg total protein. The values are means of nine internal repetitions either for both control and treatment. The two experiments presented are those with the strongest and the weakest activation obtained. To compare means, a standard analysis of variance (Moroney, 1951) was performed; *significant at 0.05; **significant at 0.01 levels of probability, respectively.

expression system the DNA is supplied naked and unmethylated. Any epigenetic imprinting is therefore bypassed.

***Opaque-2* binds to specific promoter elements *in vitro***

The O2 binding sites in the *b-32* promoter, established by DNase I footprinting, correspond to a highly conserved decamer, present at five sites, between −210 and −59 (relative to the TATA box), with the general consensus sequence GATGAPyPuTGPu. These binding sites contain as a core sequence the motif TGACGT, which is also present in the binding site of the leucine zipper activators CREB (Montminy *et al.*, 1987) and of the TGA1A (Katagiri *et al.*, 1989) from tobacco. It is also interesting to note that the homology between the O2-binding sites in the *b-32* promoter B1 and B4 extends to the neighbouring 5′ upstream region (see Table IV). The conserved B1/B4 region is also homologous to the so-called endosperm box (see Table IV). This is a 20 bp long DNA sequence (Forde *et al.*, 1985), present in nearly all storage protein promoters of the prolamin class in cereals, and previously suggested as a possible mediator of endosperm specific expression. The significance of the endosperm box is supported by deletion analysis of storage protein promoters in transgenic tobacco (Colot *et al.*, 1987; Marris *et al.*, 1988; Robert *et al.*, 1989) in which the sequence is present in the minimum region necessary for endosperm-specific expression, albeit in the case of *Glutenin-1* promoters only as a truncated version (TG$^{T}_{C}$AAAG). Moreover, a comparison of the endosperm box region of a 22 kd zein gene with that of a 19 kd zein gene shows (cf. Table IV; Spena *et al.*, 1983; Forde *et al.*, 1985) that the sequence in the 5′ half of the endosperm box is nearly identical in both genes, whereas the 3′ half is very divergent. The divergent 3′ part of the 22 kd zein endosperm box matches the B1 and B4 O2-binding sites in the *b-32* promoter, and fits the consensus O2-binding sequence obtained with the exception of a 2 bp insertion in the middle of the sequence. The presence (22 kd zein) or absence (19 kd zein) of a potential O2-binding site in the otherwise similar endosperm box is in agreement with the genetic data, which suggest that the 22 kd zein genes are under the control of the *O2* gene (Soave *et al.*, 1976), whereas the 19 kd class of zeins are regulated by the *Opaque-7* (Di Fonzo *et al.*, 1979) gene. The 5′ endosperm box region common to the 19 and 22 kd zein genes and the *b-32* gene probably represents the binding site of another *trans*-acting factor,

**Table IV.** Comparison of the *b-32* endosperm box with that of 19 kd zeins, 22 kd zeins and the consensus endosperm box of several other cereals

| Gene | DNA sequence | Homology | Reference |
|---|---|---|---|
| Z19 | TGTAAAGGTATTGC*ATCACACTAT | 16/25 | Spena *et al.*, 1983 |
| Z22 | TGTAAAGGTGAAGAGATGATGCATG | 23/25 | Kridl *et al.*, 1984 |
| *b-32* (B1) | TGGAAGGTTAGTATGATGA**CATG | 20/25 | Hartings *et al.*, 1990 |
| *b-32* (B4) | TTGGATGTTAGTGGGATGA**CATG | 18/25 | Hartings *et al.*, 1990 |
| Consensus | TGTAAAGwwAATnnGATGAnwCATG | 25/25 | Forde *et al.*, 1985 |

The DNA sequence characteristic of the endosperm box of three genes is represented with the degree of homology compared to the endosperm consensus sequence. The Opaque-2 binding sites B1 and B4 in the *b-32* endosperm box are underlined, and a putative O2 binding site in the endosperm box of the 22 kd zein (Z22) is indicated with a double line. * = deletion, w = G or T, n = A, C, T or G.

which is in agreement with data obtained by Maier *et al.* (1987), who demonstrated that a maize nuclear factor binds to that part of the endosperm box.

It has been suggested that multiple binding sites have a cooperative effect in facilitating DNA binding via DNA looping between two or more regulator molecules bound to different target sites (Ptashne, 1988; Oehler *et al.*, 1990). The different affinities of the O2 binding sites in the *b-32* promoter (data not shown), which may reflect a mechanism of gene regulation *per se*, are consistent with a cooperative effect.

From the *in vitro* binding studies, it is clear that O2 is able to bind to DNA without the involvement of another protein, resembling the situation for the mammalian factor c-jun (Angel *et al.*, 1988b). It is not currently known whether O2 forms a heterodimer with another partner. A plausible explanation for the presence of two retarded complexes in the band shift assays is that they result from the interaction of differing homomeric O2 complexes with the DNA fragment.

### *Regulation of the activity of the O2 gene product*

The binding of O2 to its own promoter *in vitro*, and the autoactivation observed in protoplasts, suggests that *O2* is weakly positively autoregulated by its own gene product. Autoregulation has also been observed for fos and jun. The jun-binding site on its own promoter differs from the c-jun target sequence in other promoters. This is also the case for O2, indicating that the mode of DNA–protein interaction, and its consequences, may differ in the auto- as opposed to the *trans*-regulation. This finding provokes the more general question of how the level of O2 protein activity is regulated. It has been shown that the mRNA is present at day 10 after pollination (Motto *et al.*, 1988) and is endosperm-specific. *Opaque-2* gene activity is therefore subject to strict transcriptional regulation. The presence of a potential casein kinase II target site in an acidic region of the O2 protein adjacent to the basic region, and the efficient phosphorylation of O2 in the presence of a nuclear extract from endosperm cells (our unpublished data), suggest that a possible site-specific phosphorylation event may also be involved.

The location of an activation domain within O2 is not yet determined. Activation domains identified in other transcription factors (Mitchell and Tjian, 1989) share overall amino acid composition rather than precise sequence homology. As none of the previously characterized activation motifs are clearly present in O2, the question is being addressed via dissection experiments.

The role of the b-32 protein itself remains enigmatic. A direct role in transcription activation either of or with O2 as has previously been suggested (Soave *et al.*, 1981) seems very unlikely. The development of *o2/o2* mutant endosperm rules out any vital metabolic function in this tissue. One possibility is that b-32 assists in the efficient translation/translocation of zeins. This would fit with its homology to ribosome-binding proteins.

The combination of classical genetics, transposon tagging and molecular analysis has resulted in the recent identification of a number of genes regulating plant development and encoding putative transcription factors: *C1* (Paz-Ares *et al.*, 1987), *R* and *B* (Ludwig *et al.*, 1989), which regulate anthocyanin biosynthesis, *Vp1* (McCarty *et al.*, 1989) mediating ABA-regulation, from maize and two genes regulating flower development, *Def-A* (Sommer *et al.*, 1990) and *agamous* (Yanofsky *et al.*, 1990). In the case of *C1*, *R*, *B* and *Vp1*, like O2, target genes are known and *trans*-activation experiments are feasible. To date results demonstrating *trans*-activation of a *Bronze-1* promoter/reporter construct via transient expression in wild-type maize endosperm represent the only other example of *trans*-activation involving homologous factors (Goff *et al.*, 1990).

## Materials and methods

### *Molecular cloning*

Manipulations of recombinant DNA were carried out by standard procedures (Maniatis *et al.*, 1982). Enzymes used in molecular cloning were obtained from BRL, Boehringer or New England Biolabs. DNA fragments were isolated from agarose gels using NA45 DEAE membrane (Schleicher & Schuell).

### *Construction of plasmids for transient expression*

pCaMVO2 was assembled via a tripartite ligation starting from (i) the vector pPCV002 (Koncz *et al.*, 1986) cut with *Eco*RI and *Hin*dIII, (ii) the plasmid PDG2 (Pietrzak *et al.*, 1986) cut with *Eco*RI and *Kpn*I to obtain the CaMV 35S promoter region and (iii) the *O2* coding region, modified by placing a *Kpn*I site just before the ATG and a *Hin*dIII site downstream from the polyadenylation signal. pB32GUSI was constructed by fusing the *Hin*dIII–*Hae*III promoter region of *b-32* clone 129 (Di Fonzo *et al.*, 1990) to pBI201.2 (Jefferson, 1987) cut with *Hin*dIII and *Sma*I. pB32GUSII was derived from pB32GUSI by cutting with *Hin*dIII and *Bgl*II, making flush-ends and self-ligating the larger product of the double digestion. To construct pO2GUSI, the plasmid pScO212, bearing sequences 5′ to the *O2* coding region as well as part of the coding region, was cut with *Apa*I, the ends made flush with T4 DNA polymerase, and then cut with *Hin*dIII; the resulting 4.2 kb fragment was separated on an agarose gel and ligated to pBI201.2 cut with *Hin*dIII and *Sma*I.

### *Transient expression in protoplasts*

Around 3 g of young leaves of *in vitro* cultured *Nicotiana tabacum*, cultivar Petit-Havana SR1, were digested overnight in K3 medium with cellulase and Macerozyme (Serva, highest quality). Protoplasts were floated, washed,

transfected and incubated for transient expression as described in Negrutiu *et al.* (1987). GUS assays were performed as previously described (Jefferson, 1987).

### *Production of GST-O2 fusion proteins*

A 1.6 kb long *Hin*dIII–*Pvu*II fragment of *O2* cDNA, extending from the first amino acid to 225 nucleotides downstream of the terminal amino acid, was cloned into the *Eco*RI site of the glutathione-S-transferase expression vector (pGEX-2; Smith and Johnson, 1988), generating the plasmid pGEX-O2, to get an in-frame translational fusion. The plasmid pGEX-O2 was cut with *Sma*I and *Stu*I, religated to yield a fusion pGEX-$O2_{55}$ open reading frame downstream from the *tac* promoter, in which the sequence coding for the first 192 amino acids of O2 has been deleted. For generating the expression plasmid pGEX-$O2_{33}$ a 600 bp long *Eco*RI fragment from construct pGEX-$O2_{55}$ was deleted, retaining the coding region for amino acids 267–454. The contiguity of the GST and *O2* open reading frames were in each case checked by DNA sequencing (Sanger *et al.*, 1977).

Wild-type glutathione-S-transferase and fusion proteins were purified as follows: *E.coli* strain B1-21 (Studier and Moffat, 1986), transformed with various expression constructs was grown overnight in 50 ml culture of L-broth plus 150 μg/ml of ampicillin at 37°C. 450 ml of fresh L-broth containing 150 μg/ml of ampicillin was added, and the culture was grown for 1 h before adding IPTG to 0.1 mM. After a further 4 h of growth, cells were harvested by centrifugation (4000 *g*, 10 min) and resuspended with 4.5 ml MTPBS (150 mM NaCl, 16 mM $Na_2HPO_4$, 4 mM $NaH_2PO_4$ pH 7.3). Addition of 12.5 mg lysozyme and 10 μl PMSF (10 mM) was followed by incubation for 15 min on ice. The cells were treated with 500 μl 10% Triton X-100 and after incubation for 10 min on ice the cells were frozen at −70°C for 30 min. To ensure complete bacterial lysis the thawed suspension was sonicated three times for 15 s. After centrifugation for 15 min (12 000 *g*, 4°C), the supernatant was incubated with 1 ml of a 1:1 slurry of reduced glutathione–agarose (Sigma Chemical Company, in MTPBS) and rocked for 30 min at 4°C. The agarose beads were spun out of solution (500 *g*, 1 min) and batch-washed six times with 20 volumes of MTPBS (containing 1 mM PMSF). Wild-type glutathione-S-transferase and fusion proteins were eluted from beads wth 5 mM reduced glutathione (Sigma Chemical Company, in 50 mM Tris pH 8.0).

### *DNA fragments used for studying protein–DNA interactions*

A 249 bp *Dra*I–*Sph*I fragment, containing the promoter region of the *b-32* gene between −265 and −16 upstream of the TATA box, was inserted in the *Sma*I site of pUC18, generating the plasmid pUCb-32. For end-labelling of the coding strand, pUCb-32 was cut with *Hin*dIII, filled in using [$\alpha$-$^{32}$P] dATP as a substrate for Klenow polymerase and recut with *Eco*RI. The labelled fragment was isolated via agarose gel electrophoresis. The same subcloning and isolation procedure was employed for end-labelling the non-coding strand of a 377 bp *Pvu*II–*Dra*I *O2* promoter fragment, containing the region between −560 and −183.

### *Gel-shift assays*

For analytical shifts with bacterially produced fusion proteins, ~5 × $10^4$ c.p.m. of end-labelled fragments (~1 ng) was incubated with 100 ng (or as indicated in the figures) GST or O2 fusion proteins for 10 min at room temperature in 20 mM HEPES pH 7.9, 50 mM KCl, 1 mM EDTA, 10% glycerol and 1 μg poly[d(I-C)] (Boehringer) in a final volume of 20 μl. The mixture was loaded onto a 3.5% native acrylamide gel (30:0.8) containing 50 mM Tris base, 50 mM boric acid and 1 mM EDTA (same solution in running buffer) and run at room temperature at 10 V/cm.

### *DNase I footprinting*

1 ng of end-labelled DNA was incubated for 15 min at 0°C plus 2 min at 25°C with different amounts (as indicated in the figures) of purified fusion protein or GST in 50 μl binding buffer containing 20 mM HEPES pH 7.9, 40 mM KCl, 0.1 mM EDTA, 0.5 mM DTT, 10% glycerol and 1 μg poly[d(I-C)]. An equal volume of 5 mM $CaCl_2$, 1.5 mM EDTA was added, followed by 3 μl of freshly diluted DNase I (Boehringer, 0.3 U/μl), and the mixture was incubated at room temperature for 90 s. Reactions were stopped by the addition of 200 μl STOP solution (200 mM NaCl, 20 mM EDTA, 1% SDS, 50 μg tRNA) and then 400 μl of 1:1 phenol–chloroform mixture. The DNA was ethanol precipitated from the aqueous phase and electrophoresed on an 8% sequencing gel. Parallel lanes containing similar DNA treated with the chemical sequencing reactions of Maxam and Gilbert (1980) with modifications (Green *et al.*, 1989) were also run on the same gel.

### *South-western analysis*

1 μg purified GST-$O2_{55}$ protein mixture was separated on a 10% SDS–PAA gel, blotted onto nitrocellulose, stained with Ponceau S to visualize the protein bands, hybridized with 1 ng of endlabelled *O2* promoter fragment, and washed as described in Miskimins (1985).

## Acknowledgements

The authors wish to thank Dirk Prüfer for instruction in transient expression methodology. The project is supported by the Deutsche Forschungsgemeinschaft (SFB 1333). One of us (M.M.) was in receipt of a training fellowship from the European Community.

The first two authors contributed equally to this work.

## References

Angel,P., Allegretto,E.A., Okino,S.T., Hattori,K., Boyle,W.J., Hunter,T. and Karin,M. (1988a) *Nature*, **332**, 166–171.
Angel,P., Hattori,K., Smeal,T. and Karin,M. (1988b) *Cell*, **55**, 875–885.
Asano,K., Svensson,B., Svendsen,I., Poulsen,F.M. and Roepstorff,P. (1986) *Carlsberg Res. Commun.*, **51**, 129–141.
Bohmann,D., Bos,T.J., Admon,A., Nishimura,T., Vogt,P.K. and Tjian,R. (1987) *Science*, **238**, 1386–1392.
Burr,B. and Burr,F.A. (1981) *Stadler Genet. Symp.*, **13**, 79–92.
Burr,F.A. and Burr,B. (1982) *J. Cell Biol.*, **94**, 201–206.
Colot,V., Robert,L.S., Kavanagh,T.A., Bevan,M.W. and Thompson,R.D. (1987) *EMBO J.*, **6**, 3559–3564.
Curran,T., Peters,G., van Beveren,C., Teich,N.M. and Verma,I.M. (1982) *J. Virol.*, **44**, 674–682.
Di Fonzo,N., Fornasari,E., Salamini,F., Reggiani,R. and Soave,C. (1979) *J. Hered.*, **71**, 397–402.
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.
Forde,B.G., Heyworth,A., Pywell,J. and Kreis,M. (1985) *Nucleic Acids Res.*, **13**, 7327–7339.
Galas,D.J. and Schmitz,A. (1978) *Nucleic Acids Res.*, **5**, 3157–3170.
Goff,S.A., Klein,T.M., Roth,B.A., Fromm,M.E., Cone,K.C., Radicella,J.B. and Chandler,V.L. (1990) *EMBO J.*, **9**, 2517–2522.
Green,P.J., Kay,S.A., Lam,E. and Chua,N.-H. (1989) In Gelvin,S.B., Schilperoort,R.A. and Verma,D.P.S. (eds), *Plant Molecular Biology Manual*. Kluwer Academic Publishers, Dordrecht, pp. B11/1–B11/22.
Hagen,G. and Rubenstein,I. (1980) *Plant Sci. Lett.*, **19**, 217–223.
Hartings,H., Maddaloni,M., Lazzaroni,N., Di Fonzo,N., Motto,M., Salamini,F. and Thompson,R. (1989) *EMBO J.*, **8**, 2795–2801.
Hartings,H., Lazzaroni,N., Marsan,P.A., Araguay,A., Thompson,R., Salamini,F., Di Fonzo,N., Palau,J. and Motto,M. (1990) *Plant Mol. Biol.*, **14**, 1031–1040.
Hoeffler,J.D., Meyer,T.E., Yun,Y., Jameson,J.L. and Habener,J.F. (1988) *Science*, **242**, 1430–1433.
Jefferson,R.A. (1987) *Plant Mol. Biol. Rep.*, **5**, 387–405.
Katagiri,F., Lam,E. and Chua,N.H. (1989) *Nature*, **340**, 727–730.
Khoo,V. and Wolf,M.J. (1970) *Am. J. Bot.*, **57**, 1042–1050.
Kodrzycki,R., Boston,R.S. and Larkins,B.A. (1989) *Plant Cell*, **1**, 105–114.
Koncz,C. and Schell,J. (1986) *Mol. Gen. Genet.*, **204**, 383–396.
Kridl,J.C., Vieira,J., Rubenstein,I. and Messing,J. (1984) *Gene*, **28**, 113–118.
Landschulz,W.H., Johnson,P.F., Adashi,E.Y., Graves,B.J. and McKnight,S.L. (1988) *Genes Dev.*, **2**, 786–800.
Larkins,B.A. and Hurkman,W.J. (1978) *Plant Physiol.*, **62**, 256–263.
Ludwig,S.R., Habera,L.F., Dellaporta,S.L. and Wessler,S.R. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 7092–7096.
Maddaloni,M., Di Fonzo,N., Hartings,H., Lazzaroni,N., Salamini,F., Thompson,R. and Motto,M. (1989) *Nucleic Acids Res.*, **17**, 7532.
Maier,U.G., Brown,J.W.S., Toloczyki,C. and Feix,G. (1987) *EMBO J.*, **6**, 17–22.
Maniatis,T., Frisch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Manzocchi,L., Tonelli,C., Gavazzi,G., Di Fonzo,N. and Soave,C. (1986) *Theor. Appl. Genet.*, **72**, 778–781.
Marris,C., Gallois,P. and Kreis,M. (1988) *Plant Mol. Biol.*, **10**, 359–366.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
McCarty,D.R., Carson,C.B., Stinard,P.S. and Robertson,D.S. (1989) *Plant Cell*, **1**, 523–532.
Miskimins,W.K., Roberts,M.P., McClelland,A. and Ruddle,F.H. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6741–6744.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.

Montminy,M.R. and Bilezikjian,L.M. (1987) *Nature*, **328**, 175–178.
Montminy,M.R., Sevarino,K.A., Wagner,J.A., Mandel,G. and Goodman,R.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6682–6686.
Moroney,M.J. (1951) *Facts from Figures*. Penguin Books, Harmondsworth.
Motto,M., Maddaloni,M., Ponziani,G., Brembilla,M., Marotta,R., Di Fonzo,N., Soave,C., Thompson,R. and Salamini,F. (1988) *Mol. Gen. Genet.*, **212**, 488–494.
Motto,M., Di Fonzo,N., Hartings,H., Maddaloni,M., Salamini,F., Soave,C. and Thompson,R.D. (1989) *Oxford Surveys Plant Mol. Cell Biol.*, **6**, 87–114.
Negrutiu,I., Shillito,R., Potrykus,I., Biasini,G. and Sala,F. (1987) *Plant Mol. Biol.*, **8**, 363–373.
Oehler,S., Eismann,E.R., Krämer,H. and Müller-Hill,B. (1990) *EMBO J.*, **9**, 973–979.
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,D.V., Glover,D.V. and Larkins,B.A. (1980) *Biochemistry*, **19**, 1644–1650.
Pietrzak,M., Shillito,R.D., Hohn,T. and Potrykus,I. (1986) *Nucleic Acids Res.*, **14**, 5857–5868.
Ptashne,M. (1988) *Nature*, **335**, 683–689.
Robert,L., Thompson,R.D. and Flavell,R.B. (1989) *Plant Cell*, **1**, 569–578.
Sanger,F., Nicklen,S. and Coulsen,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schernthaner,J.P., Matzke,M.A. and Matzke,A.J.M. (1988) *EMBO J.*, **7**, 1249–1255.
Schmidt,R.J., Burr,F.A. and Burr,B. (1987) *Science*, **238**, 960–969.
Schmidt,R.J., Burr,F.A., Aukerman,M.J. and Burr,B. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 46–50.
Serfling,E. (1989) *Trends Genet.*, **5**, 131–133.
Smith,D.B. and Johnson,K.S. (1988) *Gene*, **67**, 31–40.
Soave,C. and Salamini,F. (1982) *Phil. Trans. R. Soc. Lond.*, **304**, 341–347.
Soave,C., Righetti,P.G., Lorenzoni,C., Gentinetta,E. and Salamini,F. (1976) *Maydica*, **21**, 61–75.
Soave,C., Tardani,I., Di Fonzo,N. and Salamini,F. (1981) *Cell*, **27**, 403–410.
Soave,C., Reggiani,R., Di Fonzo,N. and Salamini,F. (1982) *Biochem. Genet.*, **11**, 1027–1037.
Sommer,H., Beltran,J.P., Huijser,P., Pape,H., Lönnig,W.E., Saedler,H. and Schwarz-Sommer,Z. (1990) *EMBO J.*, **9**, 605–613.
Spena,A., Viotti,A. and Pirotta,V. (1983) *J. Mol. Biol.*, **169**, 799–811.
Studier,F.W. and Moffatt,B.A. (1986) *J. Mol. Biol.*, **189**, 113–130.
Yamamoto,K.K., Gonzalez,G.A., Biggs,W.H.III and Montminy,M.R. (1988) *Nature*, **334**, 494–498.
Yanofsky,M.F., Ma,H., Bowman,J.L., Drews,G.N., Feldmann,K.A. and Meyerowitz,E.M. (1990) *Nature*, **346**, 35–39.

*Received on November 28, 1990*