The EMBO Journal vol.11 no.9 pp.3367-3374, 1992

Zmhox1a, the product of a novel maize homeobox gene, interacts with the *Shrunken* 26 bp *feedback* control element

Regina Bellmann and Wolfgang Werr

Institut für Genetik der Universität zu Köln, Weyertal 121, 5000 Köln 41, Germany

Communicated by P.Starlinger

A new maize homeobox gene was isolated by screening a λgt11 expression library with the 26 bp *Shrunken feedback* control element. *Zmhox1a* (Zea mays homeobox) is an unidentified maize gene mapping to the long arm of chromosome 8. It is a member of a new class of maize homeobox genes only distantly related to the *Knotted* class. The 3.1 kb *Zmhox1a* transcript can be detected in different maize tissues and encodes a polypeptide of 719 amino acids. Western blotting experiments detect the native 112 or 115 kDa protein in nuclear protein extracts, the nuclear localization being compatible with a function in transcriptional control. No Zmhox1a protein is detected in maize roots despite the presence of the *Zmhox1a* transcript; this may indicate a post-transcriptional control mechanism. A highly acidic central region of the Zmhox1a polypeptide implies a transcriptional activator function. The carboxy-terminal part of the maize homeodomain protein is related to the human Oct2 transcription factor, but homology to the POU specific domain is restricted to the POU-B subdomain. It was confirmed by DNase I footprinting experiments that DNA binding of the Zmhox1a homeodomain was at three sites flanking the TATA-box of the *Shrunken* promoter.

Key words: DNA binding/leucine zipper/maize homeobox gene/post-transcriptional control/POU domain

Introduction

Protein-DNA interactions are of fundamental importance for transcriptional regulation (reviewed in Mitchell and Tjian, 1989; Lamb and McKnight, 1991). Our knowledge about transcriptional control in eukaryotes and the factors involved is mainly derived from yeast or animal systems, but during the last few years an increasing number of plant transcription factors have been identified or cloned. Amino acid sequence comparison of plant and animal transcription factors has revealed, interestingly, that functional protein domains are quite conserved between the two kingdoms. These motifs include the helix-loop-helix or leucine zipper (for dimerization) and the basic regions, zinc finger or myb-related domains for DNA binding (Sommer *et al.*, 1990; Lohmer *et al.*, 1991; Paz Arez *et al.*, 1987; reviewed by Katagiri and Chua, 1992). Recently, sequencing of the maize *Knotted* (*Kn*) gene showed that the homeodomain is also conserved in plants (Vollbrecht *et al.*, 1991). The homeodomain is responsible for sequence specific DNA recognition and exhibits some structural similarity to the prokaryotic helix-turn-helix DNA binding motif (Otting *et al.*, 1990). The 61 amino acid sequence of the homeodomain is encoded by the homeobox, a 183 bp DNA element shared by many genes involved in the control of insect or vertebrate development. The phenotype of the *knotted* mutation is an alteration of leaf morphology, therefore it is tempting to speculate that homeodomain proteins in plants are involved in differentiation and/or developmental control as they are in the animal kingdom. The isolation of additional homeobox genes from maize in hybridization experiments with the *Kn* homeobox (Vollbrecht *et al.*, 1991) may also support this hypothesis.

We are interested in the transcriptional regulation of the maize *Shrunken* (*Sh*) gene located on chromosome 9 of the maize genome and encoding the enzyme sucrose synthase (EC 2.4.1.13, which catalyses the reaction sucrose + UDP ⟷ UDP-glucose + fructose, Chourey and Nelson, 1976). In maize the *Sh* gene is preferentially transcribed in sink organs depending on the import of sucrose (Springer *et al.*, 1986). Transient gene expression experiments performed in maize protoplasts with the *Sh* promoter fused to the neomycin phosphotransferase II marker gene revealed that transcription is repressed by high sucrose concentrations and activated at high rates of UDP-glucose turnover. Sucrose and UDP-glucose are either substrate or product of the reaction catalysed by the sucrose synthase enzyme. Transcription of the *Sh* promoter consequently seems to be linked directly to the consumption of carbohydrates in cellular metabolism. Deletional analysis of the *Sh* promoter identified a *feedback* control element in a 26 bp region spanning the transcription start site which is sufficient to initiate transcription in the absence of the TATA-box in the protoplast transient expression system (Maas *et al.*, 1990).

At present it is not clear how the physiological signal is mediated to the *feedback* control element at the *Sh* transcription start site. Most probably, however, the internal sequences of the 26 bp *Sh* initiator element are recognized by a transcription factor which is at the end of the signal transduction pathway. One approach that has often been used successfully to isolate genes for sequence specific DNA-binding proteins is screening of a λgt11 expression library with oligonucleotides containing potential target sequences. By this method we have identified several clones recognizing the 26 bp feedback control element (SP20+6) in a maize shoot λgt11 cDNA library. Here we describe the analysis of one isolated cDNA clone encoding a maize homeobox protein. We name the gene *Zmhox1a* for *Zea mays homeobox* gene and the encoded protein Zmhox1a.

Results

Isolation and sequence analysis of the Zmhox1a gene

Screening of the maize shoot λgt11 cDNA expression library (9×10^5 recombinant p.f.u./ml) was performed with a pentamerized SP20+6 oligonucleotide according to Singh

et al. (1988). Phage RB11 contained a cDNA insert of 1061 bp terminating in an 83 bp poly(A) tail. An open reading frame of 185 amino acids was expressed in-frame with the β-galactosidase protein. Computer assisted comparison of the deduced amino acid sequence with known protein sequences showed homology to a number of homeodomain-containing proteins. The similarity was generally restricted to this DNA binding region. Highest homology was observed in the recognition helix of the helix−turn−helix motif where the WFXNRR hexapeptide in helix 3 is most conserved among all known homeodomains. The identity of the carboxy-terminal 44 amino acids of the homeodomain to the *Caenorhabditis elegans ceh-5* gene (Buerglin *et al.*, 1989) was 45%; that to the *Drosophila melanogaster invected*, *even-skipped* or *paired* genes varied between 31% and 27% (Coleman *et al.*, 1987; Frasch *et al.*, 1987; Frigerio *et al.*, 1986).

Several cDNA libraries were screened for clones extending further 5′ to the mRNA cap site. The number of hybridizing clones in the different cDNA libraries was always in the range 5×10^{-4} indicating a rare mRNA transcript. The longest clone obtained from cDNA libraries started 2033 bp upstream of the poly(A) addition site. Additional upstream sequences were obtained by rapid amplification of cDNA ends (RACE, Frohman and Martin, 1989). As several RACE clone termini are clustered at the 5′ end of the *Zmhox1a* cDNA sequence, we suspect that the mRNA cap site is in close proximity. However, the precise transcription start will have to be determined in subsequent experiments in comparison with the genomic sequence. The Zmhox1a translation start codon at position 264 is in good agreement with a consensus sequence established for monocot genes (Cavener and Ray, 1991) but it is preceded by three AUG codons opening small ORFs upstream of the coding frame (see Figure 1). The *Zmhox1a* ORF is 2157 nucleotides long and is followed by an untranslated 3′ sequence of 431 nucleotides before the poly(A) tail.

DNA binding and sequence specificity of the Zmhox1a protein

To confirm that the maize homeodomain binds DNA, the 185 carboxy-terminal amino acids of Zmhox1a encoded by phage RB11 were cloned into the expression vector pET3a (Studier *et al.*, 1990). After transformation into *Escherichia coli* strain BL21 and induction of protein biosynthesis, crude protein extracts were analysed for the expression of a protein interacting with the SP20+6 *feedback* control element. The gel retardation experiment (Fried and Crothers, 1981) in Figure 2a shows that bacteria expressing the cDNA construct in the sense orientation synthesize a protein recognizing the monomeric SP20+6 oligonucleotide. No nucleoprotein complexes are detectable with protein extracts obtained from bacteria transformed with the antisense construct or from untransformed *E.coli* strain BL21. Analysis of the *E.coli* protein extracts on SDS−polyacrylamide gels showed an abundant polypeptide expressed only with the sense construct (Figure 2b). The apparent molecular weight of 31 kDa is larger than that deduced from the amino acid sequence (20 kDa). The 31 kDa polypeptide was purified to homogeneity for the production of antisera and for DNA binding studies (Figure 2b).

Sequence specificity of the homeodomain polypeptide was determined by DNase I footprinting (Galas and Schmitz,

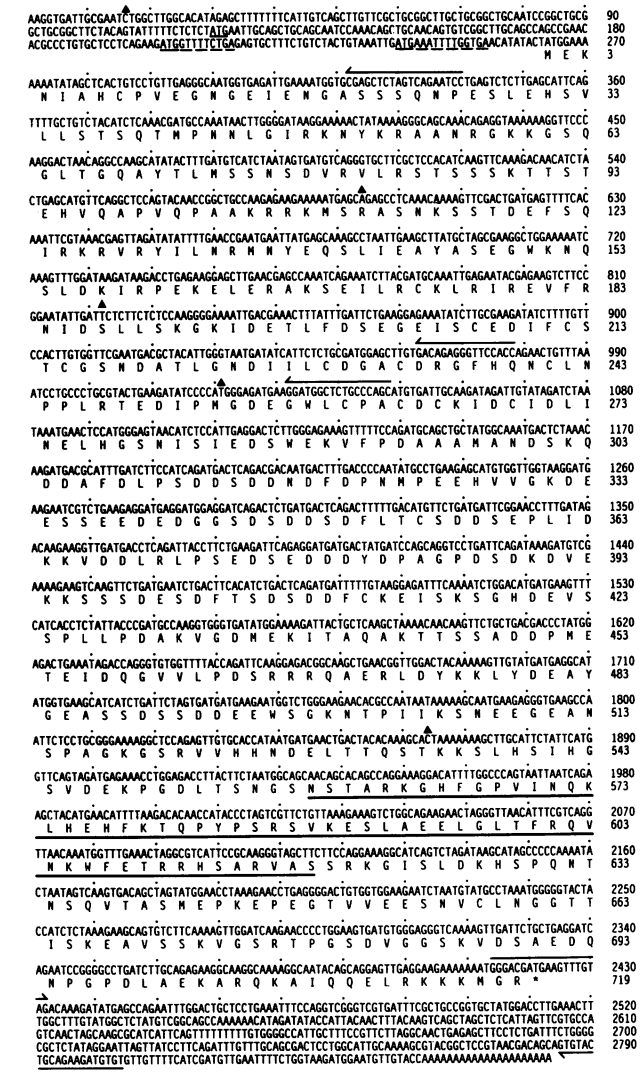


Fig. 1. Sequence of the *Zmhox1a* cDNA and translation of the 719 amino acid ORF. The homeobox region is underlined and the 5′ termini of different cDNA clones are indicated by triangles above the DNA sequence. The primary clone RB 11 isolated during sreening of the λgt11 expression library is the shortest cDNA clone. Three ATG codons upstream of the Zmhox1a translation start are underlined. The short ORFs are marked by interrupted lines, ATG 1 is in-frame with ATG 3. The position and direction of different primers used for RACE cloning or RT-PCR are indicated by arrows.

1978) experiments with the radioactively labelled SP47+42 *Sh* promoter fragment which contains the 26 bp SP20+6 *feedback* control element within its natural sequence surrounding. The homeodomain polypeptide protects three regions within SP47+42 (Figure 3a) against nuclease digestion. Protection between positions −17 to −7 within the SP20+6 *feedback* element was observed only in the upper strand. The purified polypeptide also recognized sequences immediately upstream of the TATA-box in the *Sh* promoter. The sequences protected in the upper or lower strand only partially overlap, but the entire region exhibits a palindromic structure (see schematic drawing in Figure 3a), suggesting that two polypeptide molecules bind. Three different nucleoprotein complexes with the SP47+42 *Sh* promoter fragment also form in gel retardation experiments if the concentration of the purified maize homeodomain polypeptide is increased (Figure 3b).

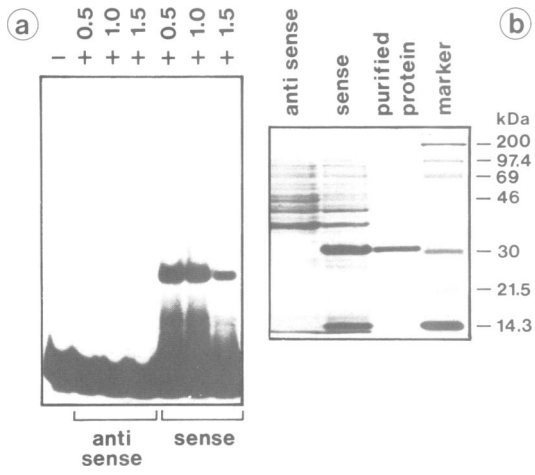


Fig. 2. (**a**) Gel retardation experiment performed with the radioactively labelled monomeric SP20+6 promoter fragment. The numbers indicate that the oligonucleotide contains the sequences between −20 and +6 in relation to the *Sh* transcription start site. Protein extracts were obtained from *E.coli* strains transformed with pET3a constructions expressing the RB11 cDNA in either the sense or antisense orientation. The kind of protein extract incubated with labelled SP20+6 is indicated below the autoradiograph. The absence (−) or presence (+) of protein in the binding reaction and the excess of unlabelled non-specifc CT-DNA is indicated above each lane. There is 1 μg of *E.coli* protein extract in each protein-containing lane. (**b**) Protein extracts from *E.coli* BL21 strains transformed with either sense or antisense constructs of RB11 in pET3a are compared. The 31 kDa homeodomain polypeptide was only detected with the sense construct. The additional 14 kDa band in the sense lane is lysozyme which was added during large scale preparation used for purifying the 31 kDa polypeptide. The purified homeodomain polypeptide is free of contamination as judged by Coomassie staining.

Southern analysis and position within the genome

In Southern experiments performed with the RB11 cDNA insert at high stringency, *Zmhox1a* behaved as a single copy gene. As shown in the autoradiograph (Figure 4), unique bands were detected after digestion of genomic maize DNA with different restriction enzymes. At lower stringency at least one additional band could be detected in each lane (data not shown). Corresponding cDNA clones for this gene (*Zmhox1b*) were isolated and are being analysed presently. Preliminary sequence information indicates that the homology between the genes *Zmhox1a* and *Zmhox1b* is not restricted to the homeobox but may extend over a large part of the protein-coding region. The *Zmhox1b* homeodomain sequence is included in Table I.

To map the *Zmhox1a* gene within the maize genome we have analysed a segregating population of maize recombinant inbreds resulting of the cross between inbreds CO159 and Tx505 (Burr *et al.*, 1988) for restriction fragment length polymorphisms (RFLPs). A *Bam*HI polymorphism (CO159: 15 kb, Tx303: 10 kb) has allowed the position of the *Zmhox1a* gene to be mapped to position 8L017 on the long arm of chromosome 8. No known phenotypic mutant is located at this position in the genetic map, therefore *Zmhox1a* is an unidentified maize gene.

The Zmhox1a gene encodes a nuclear protein and is transcribed in different maize tissues

Western blotting experiments confirmed that the native Zmhox1a protein is in the nuclear compartment, implying that it has a function in transcriptional regulation. In the

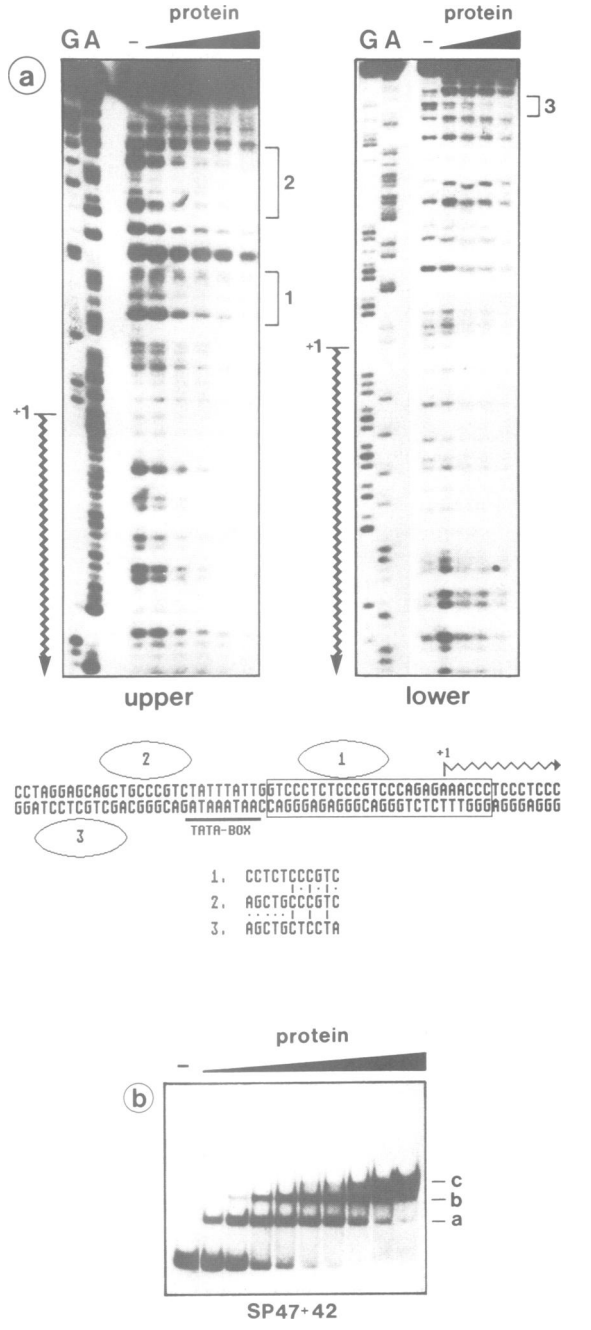


Fig. 3. (**a**) DNase I footprint analysis performed with the purified 31 kDa homeodomain polypeptide. The SP47+42 *Sh* promoter fragment was radioactively labelled at the +42 terminus with either Klenow or kinase. The protein concentration increases from 300 to 2400 ng as indicated above the lanes. Protected regions are indicated in the margin; numbering corresponds to the schematic drawing below the autoradiograph. Binding sites are compared below the scheme. Dots indicate nucleotide conserved between two binding sites, identical nucleotides are connected by bars. The *AGCTGC* sequence in binding sites 2 and 3 shows similarity to the octamer motif *ATTTGCAT* recognized by the Oct family of transcription factors. (**b**) Gel retardation performed with SP47+42 and an increasing concentration (70−630 ng) of purified 31 kDa homeodomain polypeptide. The three different protein−DNA complexes are indicated by a, b and c beside the autoradiograph.

experiment shown in Figure 5a, nuclear protein extracts from immature maize kernels (20 days after pollination) or our maize suspension cell line were separated on

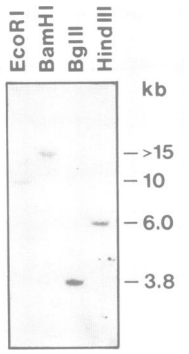
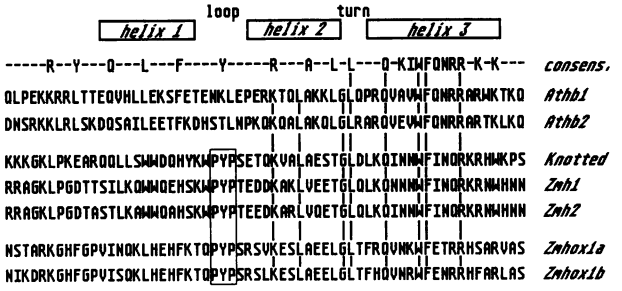


Fig. 4. Southern analysis performed with the RB11 cDNA insert. Restriction enzymes are indicated above the corresponding lane; sizes were estimated according to a size marker run in parallel.

Table I. Comparison of published plant homeodomain sequences.



The structure of the homeodomain and the consensus sequence determined for animal homeodomains is indicated above the plant sequences. Amino acids identical in all plant homeodomains are marked by bars, the PYP tripeptide present in the loop of all five maize homeodomains is boxed.

SDS−polyacrylamide gels and analysed with antisera raised against the 185 carboxy-terminal amino acids of Zmhox1a. In both protein extracts two polypeptides (see also Figure 5b) of 112 and 115 kDa were recognized by two different antisera. The size estimated from SDS−PAGE is ~33 kDa larger than the calculated molecular weight of the 719 amino acid Zmhox1a protein (79.1 kDa). The size determined from SDS−PAGE may be an overestimation because several Zmhox1a polypeptide parts expressed in *E.coli* migrate more slowly than predicted from their amino acid sequence. The predicted molecular weight of the 185 amino acid polypeptide in Figure 2b is 20 kDa but it migrates to a position corresponding to 31 kDa. As reported for various transcription factors (Jackson and Tjian, 1988), protein modifications such as glycosylation may in addition reduce electrophoretic mobility.

Northern experiments with double stranded DNA probes radioactively labelled by either nick-translation or random priming failed to detect a *Zmhox1a* transcript. Therefore a RNA probe covering the 3′ untranslated region of the *Zmhox1a* cDNA excluding the poly(A) tail was prepared; it detected a 3.1 kb *Zmhox1a* transcript (Figure 5b). The mRNA steady state level is high in maize kernels (20 days after pollination) and slightly reduced in roots or etiolated shoots of young seedlings. Lower transcript levels are observed in green young leaves and the maize suspension cell line used to identify the *feedback* control element (Maas *et al.*, 1990). To confirm the specificity of our riboprobe

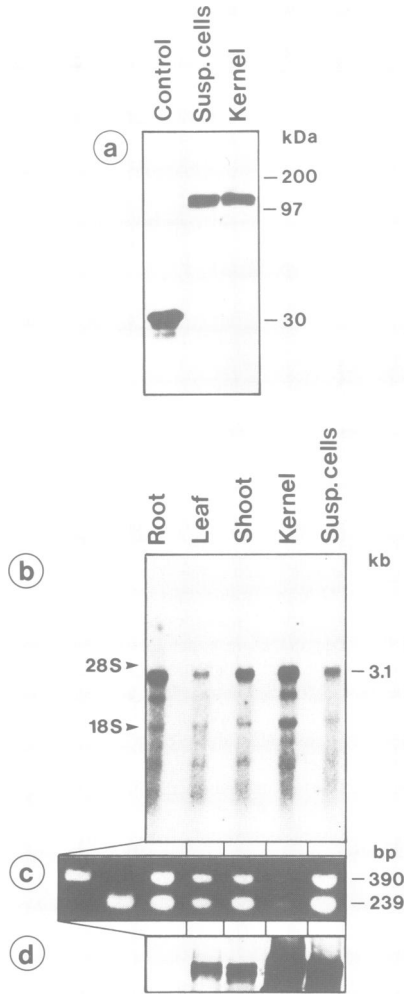


Fig. 5. (**a**) Western blot experiment performed with nuclear extracts (20 μg protein) of immature maize kernels (20 days after pollination) and our maize suspension cell line. The sizes of the two polypeptides are 112 and 115 kDa. The control lane contains 20 ng of the purified 31 kDa homeodomain polypeptide. The positions of some size markers are indicated to the right in the margin. (**b**) *Zmhox1a* transcript levels in different plant organs. Each lane contains 5 μg poly(A) RNA isolated from the tissues indicated above. The filter was hybridized with a riboprobe spanning the 3′ untranslated sequence of the *Zmhox1a* cDNA. The position of the ribosomal bands is indicated in the margin. We cannot distinguish whether the 2.6 kb transcript hybridizing in kernel and root mRNA is a degradation product or whether it is derived from a related gene. (**c**) RT-PCR performed with *Zmhox1a* and *Zmhox1b* specific primers. The *Zmhox1a* primer pair is marked in Figure 1, the upstream primer is common for *Zmhox1a* and *1b*. The sequence of the *Zmhox1b* specific primer is given in Materials and methods. The three leftmost lanes are RT-PCRs performed with (from left to right) root poly(A) RNA, the *Zmhox1a* specifc primer pair and the Zmhox1b specific primer pair; the other lanes contain a combination of all three primers. The reduced amount of amplification products in kernels is due to reduced first strand synthesis in the reverse transcriptase reaction as estimated by the incorporation of radioactively labelled [^{32}P]CTP. The size of the gene specific amplification products is indicated in the margin. (**d**) Amount of 112 and 115 kDa polypeptides in 30 μg nuclear protein extracts detected in the Western blot experiment with the Zmhox1a antiserum. Tissue lanes correspond to the Northern experiment.

and to distinguish the *Zmhox1a* transcript from transcripts of the *Zmhox1b* gene (unpublished data) we performed reverse transcriptase PCR (RT-PCR) with gene specific primers. The result in Figure 5c shows that the 390 bp

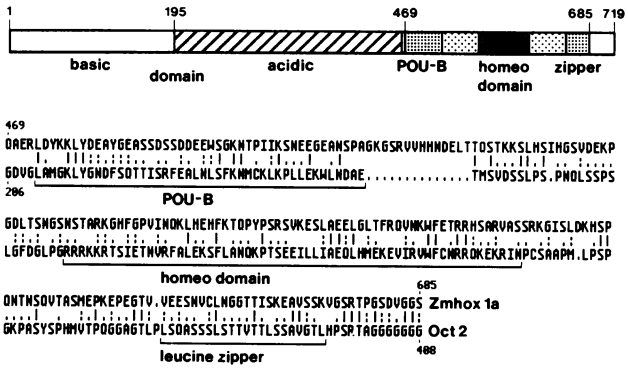


Fig. 6. Schematic drawing of the Zmhox1a protein. In the amino-terminal hydrophilic domain, basic residues predominate. The acidic domain is hatched, the region compared with Oct2 at the amino acid sequence level below is dotted. Different dot intensities indicate distinct protein domains. In the sequence comparison bars indicate amino acid identity, two dots isomorphic replacements and one dot side chains of similar character.

Zmhox1a specific PCR amplification product is clearly distinct from the 239 bp *Zmhox1b* specific product. Moreover, the relative amounts of the 390 bp *Zmhox1a* and 239 bp *Zmhox1b* specific amplification products indicate that both homeobox genes are transcribed coordinately at similar levels in the different tissues analysed. In respect to protein translation it is important that the 390 bp *Zmhox1a* specific amplification product is present in maize roots.

Translation of the *Zmhox1a* mRNA was analysed in different maize organs in Western blotting experiments (Figure 5d). The Zmhox1a protein was detected in nuclear protein extracts of leaves, shoots, kernel and suspension cells. In each tissue the 112 and 115 kDa polypeptide species were observed simultaneously; protein levels were lower in leaves and shoots than in suspension cells or maize kernels. However, no protein was detectable in root nuclei even with 10-fold more nuclear protein. Non-specific degradation of proteins in the root nuclear extracts was excluded by silver staining of the SDS−polyacrylamide gel. The *Zmhox1a* transcript level in maize roots as judged by two independent methods clearly contradicts the absence of the Zmhox1a protein in the root nuclei which therefore must be due to regulation at the post-transcriptional level.

Discussion

Zmhox1a belongs to a novel class of maize homeobox genes

Homeoboxes were first identified as a conserved DNA sequence element in many genes that play central roles in *Drosophila* development. Subsequently this motif has been found to be widespread in the animal kingdom. It encodes the homeodomain, the DNA binding motif in a still increasing number of developmental genes or transcription factors. Recently it was found that the *Zea mays Kn* gene encoded the first known plant homeodomain protein (Vollbrecht *et al.*, 1991). We describe here the isolation and characterization of the *Zmhox1a* gene encoding another maize homeobox protein. *Zmhox1a* is a so far unidentified single copy gene mapping to the long arm of chromosome 8. The gene encodes a homeodomain protein that recognizes sequences close to the transcription start site of the *Shrunken* promoter in *in vitro* DNase 1 footprinting experiments.

The amino acid sequences of homeodomains in the Zmhox1a and Kn proteins show only 26% identity, which is much lower than the 45% observed between Zmhox1a and the *C.elegans* Ceh-5 protein (Buerglin *et al.*, 1989). Other maize homeodomain sequences included in Table I are deduced from related genes which were isolated at reduced stringency in hybridization experiments with either the *Zmhox1a* or the *Kn* homeobox. The existence of these closely related genes implies that plant homeobox genes will fall into distinct homeodomain classes as in the animal kingdom (reviewed by Scott *et al.*, 1989). The two *Arabidopsis thaliana* homeobox genes listed in Table I were isolated using highly degenerate oligonucleotides (Ruberti *et al.*, 1991) and indicate that homeobox genes presumably exist in many plant species. Whether their roles in pattern formation in plants are comparable to those in animals will be one of the most interesting questions for future investigations.

The homeodomain is composed of three α-helices with helices 2 and 3 forming a helix−turn−helix motif showing similarity to prokaryotic DNA binding domains. Of these helices, helix 3 is the best conserved between plant and animal homeodomains (see Table I); in NMR studies performed with the *Antennapedia* homeodomain it has been shown to contact DNA (Otting *et al.*, 1990). At least three amino acid residues in the DNA recognition helix (WF---R) have up to now been found to be invariant in plants as they are in animal homeodomains. It will be interesting to see whether the amino acid residues exchanged in the *Zmhox1a* and *Zmhox1b* homeodomains will alter DNA sequence specificity of binding. This may be indicated by the asparagine to lysine exchange of residue 6 in helix 3 because the residue at this position in the *Antennapedia* homeodomain contacts DNA (Otting *et al.*, 1990). For evolutionary considerations it may be noteworthy that in all five maize homeobox genes the motif PYP is conserved in the loop between helices 1 and 2. This tripeptide is only rarely found in other known homeodomains, for example in the *Drosophila cut* gene (Blochinger *et al.*, 1988) and the yeast *mat2-P* gene (Kelly *et al.*, 1988).

Structure of the Zmhox1a protein

The different domains of the Zmhox1a protein are schematically drawn in Figure 6. The first 195 protein residues are rather hydrophilic with basic amino acids in excess. Two clusters of positively charged amino acids (residues 48−61 and 105−111) show some homology to nuclear localization signals which is consistent with the detection of the Zmhox1a protein in nuclear protein extracts. The amino-terminal domain is followed by an acidic region between amino acids 195 and 465 containing 30% acidic residues (82 in total). The highest density of 47 negatively charged side chains in 110 residues (300−409) coincides with a high α-helical prediction according to the Chou−Fasman algorithm which is characteristic of acidic transcription activation domains (Mitchell and Tjian, 1989). In slight contrast to other activation domains where acidic amino acids are clustered in relatively short stretches (for review see Mitchell and Tjian, 1989) the 82 negatively charged residues are distributed over 270 amino acids in the Zmhox1a protein. Therefore it may be interesting to test whether this indicates a multiplicity of activation domains as described in the transcription factor SP1 (Pascal and Tjian,

1991). Computer comparison of the Zmhox1a amino acid sequence with known protein sequences revealed extended similarity to the human Oct2 protein (Scheidereit *et al.*, 1988) at the carboxy-terminus. A stretch of 127 amino acids in the *Zmhox1a* gene product (see Figure 6) shows 39.7% similarity (20.6% identity) to the Oct2 carboxy-terminus; the similarity increases to 44.5% (25% identity) if the homeodomain is omitted. Sequence conservation extends into the leucine zipper located near the carboxy-terminus in Oct2. In contrast to Oct2, the presumptive Zmhox1a maize zipper, however, contains only one leucine. Additional support of functional significance may be the conserved distance between the presumptive zippers and the DNA binding homeodomain in both proteins. If an α-helical conformation is assumed, both zipper regions exhibit a weak but similar amphipathic character. Evidence for homodimer formation may be provided by the footprinting experiment in which the Zmhox1a 31 kDa polypetide protected a palindromic region on either the upper or lower strand upstream of the *Sh* TATA-box (Figure 3a, sites 2 and 3). Sites 2 and 3 also show some homology to the octamer motif recognized by the Oct2 protein (ATTTGCTA, see legend of Figure 3a).

Oct2 belongs to the POU family of transcription factors which contain a conserved POU-specific protein domain amino-terminal to the homeodomain. Only little sequence similarity is found outside of these two conserved regions between individual POU family members (for review see Rosenfeld, 1991; Schöler, 1991). Conservation between the Zmhox1a amino acid sequence and the POU specific domain in the POU family of transcription factors is only weak. The best homology was found in a 41 amino acid region (see Figure 6) where similarity reaches 46.4% (24% identity) to the POU-B subdomain of Oct2. Five out of a total of 10 amino acids are conserved at positions that are rather invariant in the known animal POU-B subdomains. In addition, the essential basic centre and the high α-helical density predicted at the carboxy-terminal boundary of animal POU-B subdomains (reviewed by Rosenfeld, 1991) are structurally conserved in the 41 amino acid Zmhox1a region. An incomplete POU-specific domain was first described in the rat transcription factor HNF-1, but similarity is restricted to POU subdomain A (Baumhueter *et al.*, 1990). The function of the POU domain is at present quite controversial. It may contribute to sequence specificity or DNA binding affinity as in the Oct1 transcription factor (Verrijzer *et al.*, 1990; Kristie and Sharp, 1990) but is also implicated in protein–protein interactions (Ingraham *et al.*, 1990). A contribution of this protein region in Zmhox1a to DNA binding specificity would have escaped detection because this region is absent in the 31 kDa homeodomain polypeptide used in the footprinting experiments. We take the carboxy-terminal structural similarity of Zmhox1a and human Oct2 as an indication that these plant and human transcription factors have a common ancestor, implying that this structural organization must be very old evolutionarily.

Expression of the Zmhox1a gene

The 3.1 kb *Zmhox1a* transcript was detectable in all tissues tested but mRNA levels differed slightly. We have so far been unable to detect any tissue lacking the *Zmhox1a* mRNA (RT-PCR, Figure 5c and data not shown). At the protein level, our antisera raised against the carboxy-terminal 185 amino acids of the Zmhox1a protein simultaneously detect two polypeptides of 112 and 115 kDa located in the nuclear compartment. These two protein species may arise by post-translational modification of the Zmhox1a protein, but other alternatives that presently cannot be excluded are that they arise from alternative usage of translation initiation sites or that they are products of differential splicing. As the antiserum raised against the Zmhox1a carboxy-terminus crossreacts with the Zmhox1b protein overexpressed in *E.coli* (data not shown), one polypeptide species may correspond to the closely related *Zmhox1b* gene product. This interpretation is at least compatible with the simultaneous transcription of both genes as determined by RT-PCR (Figure 5b); however, it does not explain the absence of both polypeptides in root protein extracts.

The finding that maize roots have high levels of *Zmhox1a* mRNA but no Zmhox1a protein will have to be analysed in more detail in future experiments. One explanation might be that translation is blocked, as observed in the *Drosophila* homeotic gene *achaete-scute* (Cabrera, 1990); however, the integrity of the entire *Zmhox1a* transcript has to be confirmed in roots. The Northern blotting data show a similar transcript size and the RT-PCR proves that the *Zmhox1a* mRNA is present. However, the root *Zmhox1a* mRNA may be untranslatable due to incomplete processing or differential splicing. The persistence of a small intron within the large *Zmhox1a* transcript would be almost undetectable. It was observed that there are additional small ORFs in front of the Zmhox1a AUG start codon (AUG 4, see Figure 1); since similar ORFs in the yeast GCN4 transcriptional regulator are involved in translational control (reviewed by Hinnebusch, 1990), this supports the notion that Zmhox1a protein levels are controlled translationally. At least the second AUG provides a good translation start signal in comparison to the monocot consensus sequence (Cavener and Ray, 1991). Independently of which mechanism will explain the absence of the *Zmhox1a* gene product in roots, our data suggest that homeodomain proteins in plants may be not solely under transcriptional control.

In summary, the *Zmhox1a* gene belongs to a new class of maize homeobox genes and should allow isolation of new members of the plant homeobox gene superfamily. The central acidic domain of Zmhox1a indicates that it may have transcriptional activator function, which may be exerted close to the transcription start site of the *Shrunken* promoter. As the Zmhox1a protein is present in the nuclei of our maize suspension cell line, it remains a potential candidate for mediating the transcriptional feedback control observed at the *Shrunken* promoter in the transient gene expression experiments (Maas *et al.*, 1990). The sequence similarity exhibited at the carboxy-termini of maize Zmhox1a and the human Oct2 transcription factor may indicate an evolutionarily ancient structure.

Materials and methods

Screening of the λgt11 cDNA expression library

The maize shoot cDNA library constructed in λgt11 was purchased from Clontech Laboratories Inc. and contained 9×10^5 independent clones. Phages were grown on *E.coli* host Y1090 at a density of 5×10^4 p.f.u./150 mm dish. Plates were incubated at 42°C for 3–4 h before applying a nitrocellulose filter saturated with IPTG. The first filter was removed after 6 h; replica filters were incubated overnight. Pretreatment of nitrocellulose filters and DNA binding followed the protocol outlined by Singh *et al.* (1988) with the following modifications. Non-specific competitor DNA was omitted from the solutions and DNA binding to the filter bound proteins was

performed simultaneously in one dish with up to 50 filters. Binding was performed for 2 h at 4°C with constant agitation; filters were washed at 4°C three times for 3 min. The SP20+6 pentamer was generated by oligomerization of a synthetic oligonucleotide and subcloning into the *Sma*I site of pUC19. The double stranded pentamer was isolated by *Bam*HI/*Asp*718 double digestion and radioactively labelled by filling in with the large fragment of DNA polymerase I. Supplying all four radioactively labelled nucleotides [α-^{32}P]dNTP (3000 Ci/mMol), a specific activity $>2\times10^8/\mu g$ double stranded SP20+6 pentamer was routinely achieved. The *Zmhox1a* cDNA clone (RB11) was unique among 27 positive clones isolated after three cycles of purification.

Isolation of longer cDNA clones, rapid amplification of cDNA ends (RACE), subcloning and DNA sequence analysis

Homologous *Zmhox1a* clones overlapping RB11 were isolated by hybridizing the λgt11 library with the most 5′ terminal 370 bp *Eco*RI−*Xba*I fragment in 5×SSPE, 5×Denhardt's, 250 μg/ml CT-DNA and washing with 0.2×SSC, 0.1% SDS at 68°C. The termini of individual isolated *Zmhox1a* cDNA clones are indicated in Figure 1. Closely related *Zmhox1b* clones were isolated after hybridization at 55°C and washing in 2×SSC, 0.1% SDS at 55°C. Clones extending to the mRNA 5′ end were obtained by the RACE method with the nested primers described by Frohman and Martin (1989). First strand cDNA synthesis was performed with maize kernel poly(A) RNA primed by random hexanucleotides using the Superscript preamplification kit (BRL). The positions of the three primers used (GCTGGGCAGAGCCATCC, GGTGGAACCCTCTGTC and GGATTCTGACTAGAGCTC) within the *Zmhox1a* sequence are indicated in Figure 1. Amplification was performed for 30 cycles [1 μM each primer, annealing at 65°C (1 min), elongation at 72°C (2 min) and denaturation at 93°C (1 min)] in the buffer system supplied with *Taq* DNA polymerase (BRL). PCR products were treated by filling in with Klenow enzyme and were 5′-phosphorylated with T4 polynucleotide kinase prior to subcloning into the *Sma*I site of pUC19. *Zmhox1a* clones were identified by colony filter hybridization. Synthetic oligonucleotides were prepared on an Applied Biosystems 394 DNA synthesizer. Double stranded DNA was sequenced on both strands by the dideoxy chain termination method using T7 polymerase and analysed on a VAX/VMS V5.4 system with the GCG sequence analysis software package (Devereux *et al.*, 1984).

Expression and purification of the maize homeodomain polypeptide and raising of antiserum

For the expression of the 185 amino acid ORF of RB11 the cDNA insert except for the poly(A) tail was cloned in both orientations into the expression vector pET3a. Transformation of *E.coli* host strain BL21 and induction of protein synthesis followed exactly the protocol outlined by Studier *et al.* (1990). Analysis of protein extracts on 10% SDS−polyacrylamide gels showed that the 31 kDa homeodomain polypeptide is soluble and not incorporated into inclusion bodies. Purification of the maize polypeptide started from 1 l *E.coli* culture. Cells were pelleted and resuspended in 50 ml 50 mM Tris pH 8, 50 mM NaCl, 10 mM EDTA supplemented with 20 μg/ml lysozyme. The suspension was frozen in liquid nitrogen, thawed and stirred for 30 min at room temperature after the addition of benzon nuclease (10 U/ml). Cell debris was pelleted by centrifugation and the protein supernatant was concentrated by ammonium sulfate precipitation (350 mg/ml). Protein pellets were dissolved in a minimal volume of 10 mM Tris pH 7.5, 10 mM NaCl, 10 mM $MgCl_2$ (RB), dialysed against 100 vol RB and loaded on a Mono S HR5/5 column equilibrated with RB. The 31 kDa homeodomain polypeptide eluted with 0.4 M NaCl together with a 14 kDa polypeptide (presumably lysozyme). The 14 kDa polypeptide was removed by gel filtration on a Pharmacia HiLoad Superdex 200 (16/50) column. The 31 kDa homeodomain polypetide eluted in one peak with the molecular weight of the monomeric polypeptide. The yield of 1 l *E.coli* culture was ~2 mg homogenous 31 kDa polypeptide.

For preparation of antisera rabbits were immunized with 500 μg of the purified polypeptide; boosting was performed with 200 μg after 4 weeks. Preimmuneserum did not show any affinity for maize nuclear protein extracts.

Gel retardation and DNase I footprinting analysis

Gel retardation analysis was performed as described by Werr *et al.* (1988) except that DNA−protein complexes were electrophoretically analysed on 6% polyacrylamide gels (29:1). For DNase I footprinting experiments DNA fragments were radioactively labelled at one end either by filling in with Klenow or with T4 polynucleotide kinase. After 10 min of protein binding, samples were treated with DNase I for 30 s at room temperature. The reaction was stopped by adding an equal volume of 20 mM EDTA. After extraction with phenol, nucleic acids were precipitated and analysed on 7% sequencing gels.

Southern analysis and RFLP mapping

Genomic DNA for Southern analysis was extracted from the variety Solo (Kleinwanzlebener Saatzucht AG) according to the method described by Saghai-Maroof *et al.* (1984) and purified once over a caesium chloride gradient. Hybridization was performed overnight under the conditions described for screening the expression library. The segregating population of maize recombinant inbreds was obtained from Dr B.Burr (Brookhaven Laboratory, Stonybrook). DNA from individual plants was prepared according to Saghai-Maroof *et al.* (1984) and digested with *Alu*I. Electrophoresis, blotting onto Hybond N nylon membrane, hyridization in 6×SSPE, 5×Denhardt's, 250 μg/ml CT-DNA and washing were performed according to Gebhardt *et al.* (1989). Filters were placed against Kodak XAR5 films for 2−3 days.

Northern, reverse transcriptase PCR and Western blot analyses

Preparation of RNA from maize tissues was performed as described in Springer *et al.* (1986). For Northern analysis 5 μg poly(A)$^+$ RNA from each tissue was electrophoretically separated on 1.2% agarose−formaldehyde gels, 2 μg of ethidium bromide were added to each sample prior to gel loading and the gel was photographed before blotting onto nitrocellulose. The radioactively labelled antisense RNA probe of the *Zmhox1a* 3′-untranslated trailer sequence was generated by transcription of an *Xba*I−*Pst*I fragment cloned into the vector pGEM4z with T7 polymerase.

First strand cDNA synthesis for the reverse transcriptase PCR was performed with the Superscript preamplification kit (BRL) and primed with oligo(dT). The positions of the primers in the *Zmhox1a* gene are indicated in Figure 1, the upstream primer within the protein coding region (GGGACGATGAAGTTTGTAG) is common to both the *Zmhox1a* and *Zmhox1b* genes. The specific downstream primers (ACACATCTTCTG-CAGTACAC for *Zmhox1a*, GAATGAATAGCCCCCACAG for *Zmhox1b*) overlap small gaps in the untranslated 3′ regions between both genes. The primer concentration was 1 μM and amplification was performed as described for the RACE experiment.

Nuclear protein extracts were performed according to Werr *et al.* (1988) except that in root and leaf tissue floating of the nuclei on 2.4 M sucrose was omitted. Nuclear proteins were separated on 10% SDS−polyacrylamide gels (37.5:1) and transferred to nitrocellulose by electroblotting. Filters were incubated with the Zmhox1a antisera (1:5000 dilution) for 60 min and specific complexes were visualized with the ECL system (Amersham) according to the supplier's manual.

Acknowledgements

We thank Dr Benjamin Burr (Brookhaven National Laboratory, Upton, New York) for providing the segregating population of maize recombinant inbreds and for computer analysis of RFLPs. The RACE primer set was kindly given by Thomas Christoph (Institut für Genetik, Universität zu Köln). We also thank Dr Ellen Barzen (Max-Planck-Institut für Züchtungsforschung, Abt. Prof. Salamini, Köln) for theoretical and practical help during RFLP mapping. This work was supported by the Deutsche Forschungsgemeinschaft through SFB 274.

References

Baumhueter,S., Mendel,D.B., Conley,P.B., Kuo,C.J., Turk,C., Graves,M.K., Edwards,C.A., Courtois,G. and Crabtree,G.R. (1990) *Genes Dev.*, **4**, 372−379.

Blochinger,K., Bodmer,R., Jack,J., Jan,L.Y. and Jan,Y.N. (1988) *Nature*, **333**, 629−635.

Buerglin,T.R., Finney,M., Coulson,A. and Ruvkin,G. (1989) *Nature*, **341**, 239−243.

Burr,B., Burr,F.A., Thompson,K.H., Albertson,M.C. and Stuber,C.W. (1988) *Genetics*, **118**, 519−526.

Cabrera,C.V. (1990) *Development*, **109**, 733−742.

Cavener,D.R. and Ray,S.C. (1991) *Nucleic Acids Res.*, **19**, 3185−3192.

Chourey,P. and Nelson,O. (1976) *Biochem. Genet.*, **14**, 1041−1055.

Coleman,K.G., Poole,S.J., Weir,M.P., Soeller,W.C. and Kornberg,T. (1987) *Genes Dev.*, **1**, 19−28.

Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acid Res.*, **12**, 387−395.

Frasch,M., Hoey,T., Rushlow,C., Doyle,H. and Levine,M. (1987) *EMBO J.*, **6**, 749−759.

Frigerio,G., Burri,M., Bopp,D., Baumgartner,S. and Noll,M. (1986) *Cell*, **47**, 735−746.

Fried,M. and Crothers,D.-M (1981) *Nucleic Acid Res.*, **9**, 6505−6525.

Frohman,M.A. and Martin,G.R. (1989) *Techniques*, **1**, 165−170.

Galas,D. and Schmitz,A. (1978) *Nucleic Acid Res.*, **5**, 3157–3169.
Gebhardt,C., Ritter,E., Debener,T., Schachtschabel,U., Walkemeier,B., Uhrig,H. and Salamini,F. (1989) *Theor. Appl. Genet.*, **78**, 65–75.
Hinnebusch,A.G. (1990) *Trends Biochem. Sci.*, **15**, 148–152.
Ingraham,H.A. Flynn,S.E., Voss,I.W., Albert,V.R., Kapiloff,M.S., Wilson,L. and Rosenfeld,M.G. (1990) *Cell*, **61**, 1021–1033.
Jackson,S.P. and Tjian,R. (1988) *Cell*, **55**, 125–133.
Katagiri,F. and Chua,N.H. (1992) *Trends Genet.*, **8**, 22–27.
Kelly,M., Burke,J., Smith,M., Klar,A. and Beach,D. (1988) *EMBO J.*, **7**, 1537–1547.
Kristie,T.M. and Sharp,P.A. (1990) *Genes Dev.*, **4**, 2383–2396.
Lamb,P. and McKnight,S.L. (1991) *Trends Biochem. Sci.*, **16**, 417–422.
Lohmer,S., Madaloni,M., Motto,M., DiFonzo,N., Hartings,H., Salamini,F. and Thompson,R.D. (1991) *EMBO J.*, **10**, 617–624.
Maas,C., Schaal,S. and Werr,W. (1990) *EMBO J.*, **9**, 3447–3452.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.
Otting,G., Qian,Y.Q., Billter,M., Müller,M., Affolter,M., Gehring,W. and Wüthrich,K. (1990) *EMBO J.*, **9**, 3085–3092.
Pascal,E. and Tjian,R. (1991) *Genes Dev.*, **5**, 1646–1656.
Paz Arez,J., Ghosal,D., Wienand,U., Peterson,P.A. and Saedler,H. (1987) *EMBO J.*, **6**, 3553–3558 .
Rosenfeld,M.G. (1991) *Genes Dev.*, **5**, 897–907.
Ruberti,I., Sessa,G., Lucchetti,S. and Morelli,G. (1991) *EMBO J.*, **10**, 1787–1791.
Saghai-Maroof,M.A., Soliman,K.M., Jorgensen,R.A. and Allard,R.W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 8014–8018.
Scheidereit,C., Cromlish,J.A., Gerster,T., Kawakami,K., Balmaceda,C.G., Currie,R.A. and Roeder,R.G. (1988) *Nature*, **336**, 551–557.
Schöler,H.R. (1991) *Trends Genet.*, **7**, 323–328.
Scott,M.P., Tamkun,J.W. and Hartzell,G.W. (1989) *Biochem. Biophys. Acta*, **989**, 25–48.
Singh,H., LeBowitz,J.H., Baldwin,A.S. and Sharp,P.A. (1988) *Cell*, **52**, 415–423.
Sommer,H., Beltran,J.P., Huiser,P., Pape,H., Lönnig,W.E., Saedler,H. and Schwarz-Sommer,Z. (1990) *EMBO J.*, **9**, 605–613.
Springer,B., Werr,W., Starlinger,P., Bennett,D.C., Zokolica,M. and Freeling,M. (1986) *Mol. Gen. Genet.*, **205**, 471–468.
Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
Verrijzer,C.P., Kal,A.J. and van der Vliet,P.C. (1990) *Genes Dev.*, **4**, 1964–1974.
Vollbrecht,E., Veit,B., Sinha,N. and Hake,S. (1991) *Nature*, **350**, 241–243.
Werr,W., Springer,B., Schürmann,J. and Bellmann,R. (1988) *Mol. Gen. Genet.*, **212**, 342–350.

Received on February 13, 1992; revised on June 4, 1992