The transcriptional regulator CPRF1: expression analysis and gene structure

Abstract Many eukaryotic DNA-binding proteins share a conserved amino acid sequence known as the basic region leucine zipper (bZIP) domain. bZIP proteins recognise DNA, upon dimerization, in a sequence-specific manner. The Common Plant Regulatory Factor 1 (CPRF1) is a bZIP transcription factor from parsley (Petroselinum crispum), which recognises defined elements containing ACGT cores. CPRF1 genomic DNA was cloned and the gene was sequenced. Analysis of the sequence data revealed the existence of 12 exons and 11 introns within a stretch of about 9 kb. A second RNA species hybridising to CPRF1 probes was identified as an alternatively spliced, additional CPRF1 transcript containing intron 8. This polyadenylated RNA species showed accumulation characteristics very similar to those of the CPRF1 mRNA. CPRF1 specifically binds an ACGT-containing element which is located within the composite regulatory unit that is necessary and sufficient for light activation of the parsley chalcone synthase (CHS) minimal promoter. Expression studies at the mRNA level demonstrated that CPRF1 mRNA is present in all organs of light-grown plants in which CHS mRNA expression is detectable, and light-dependent CHS mRNA accumulation was shown to be blocked by cycloheximide. Therefore, translation of a protein factor, possibly CPRF1, may be a prerequisite for CHS promoter activation.

Key words Trans-acting factor • Light induction • Cycloheximide • Chalcone synthase • ACGT element

Introduction

The regulation of gene expression requires the precise targeting of transcription factors to specific binding sites present in promoter DNA. Transcription factors that specifically recognise DNA sequences often bind as dimers and show a modular domain structure (Frankel and Kim 1991). Depending on the nature of their DNA-binding domains, these factors have been classified into several families (Johnson and McKnight 1989; Mitchell and Tjian 1989; Pabo and Sauer 1992). One important family is characterised by the basic region-leucine zipper (bZIP) domain (Landschulz et al. 1988; Vinson et al. 1989). The bZIP domain is a compact structure in which the leucine zipper (ZIP) acts as a dimerization interface and the basic region (b) contacts DNA. Dimerization is a prerequisite for DNA binding of bZIP proteins (Ellenberger 1994).

More than 30 plant bZIP proteins have been identified, most of them either by screening expression libraries with DNA ligand probes or by low stringency hybridisation with probes based on the conserved bZIP region (Foster et al. 1994). The first bZIP proteins isolated from plants were TGAla, TGAlb (Katagiri et al. 1989) and HBP1 (Tabata et al. 1989); Opaque2 was identified genetically (Hartings et al. 1989). All plant bZIP proteins characterised so far recognise ACGT-Containing Elements (ACE) (Weisshaar et al. 1991; Armstrong et al. 1992; Feldbrügge et al. 1994; Foster et al. 1994). The affinity with which a given plant bZIP protein binds to an ACE is mainly determined by the nucleotide triplets 5' and 3' to the ACGT core (Izawa et al. 1993; Suckow et al. 1993; Foster et al. 1994). In an analysis of the regulation of chalcone synthase (CHS) gene expression in parsley (Petroselinum crispum), in response to UV-containing white light, we identified several parsley bZIP proteins which recognise a distinct regulatory element in the CHS gene promoter. These parsley bZIP factors were designated Common...
Plant Regulatory Factors (Weisshaar et al. 1991). CHS catalyses the first specific step in the biosynthesis of flavonoids (Hahlbrock and Scheel 1989). In parsley, CHS is encoded by a single gene (Herrmann et al. 1988). Analyses of the CHS promoter defined a 52-bp light regulatory unit (LRU1) consisting of at least two cis-acting elements which were designated BoxI and ACE<sub>CIS</sub> (Schulze-Leffert et al. 1989; Feldbrügge et al. 1994). CPRF1 recognises ACE<sub>CIS</sub> both in vitro (Armstrong et al. 1992) and in vivo (Feldbrügge et al. 1994).

Here, we have analysed the structure of the CPRF1 gene and its expression pattern in parsley plants. An additional CPRF1 transcript was identified and shown to be the result of the simplest form of alternative splicing. Additionally, data are presented demonstrating the existence of a CPRF1-related pseudogene within the parsley genome. Finally, experiments using the protein synthesis inhibitor cycloheximide (Chx) are presented and discussed in the context of the possible requirement of CPRF1 expression for CHS mRNA accumulation.

**Materials and methods**

Standard techniques and materials

Basic molecular biology techniques were carried out as described by Sambrook et al. (1989). Hybridisation experiments were performed as outlined before (Kawalleck et al. 1993). Conditions for stringent washes were 0.2× SSC/0.1% SDS at 68°C, washes at reduced stringency were with 2× SSC/0.5% SDS at 50°C. X-ray films were exposed at −80°C with intensifying screens. Membranes were stripped between the hybridizations by incubation in 0.1% SDS solution for 20 min at 80°C. Exposure times of several days were required to detect the CPRF1 transcripts in contrast to a few hours for CHS or UB4. Oligonucleotides used for sequencing were chemically synthesised on an Applied Biosystems 392 DNA Synthesiser (Foster City, Calif.) and separated from low molecular weight contaminants on Sephadex (Pharmacia, Freiburg, Germany) G-25 columns. Restriction endonucleases and other DNA-modifying enzymes were purchased from either Boehringer Mannheim, New England Biolabs (Schwalbach-Taunus, Germany), or Stratagene (Heidelberg, Germany). Plant material and RNA isolation

For isolation of RNA, from plant organs, parsley (<i>P. crispum</i>) plants were grown for 6 months in a greenhouse. Three-week-old seedlings were grown in a phytochamber (12 h white light per 24 h). Samples from both sources were taken about 4 h after onset of illumination. Plant material was frozen in liquid nitrogen immediately after harvesting and stored until use at −80°C. Total RNA was prepared as described previously (Feldbrügge et al. 1994). Poly(A)<sup>+</sup> RNA was isolated using oligo(dT)-Sephadex columns (Pharmacia). The RNA from cultured parsley cells used for the experiment shown in Fig. 4 is identical to that used previously (Weisshaar et al. 1991).

Probes used in hybridisation experiments

All labelled probes were prepared from gel-purified DNA fragments. The DNA used for the CPRF1 probe employed in the screening experiments as well as in the RNA gel blot experiments was a 1.7-kb EcoRI fragment isolated from the original CPRF1 cDNA (Weisshaar et al. 1991). The DNAs used as intron 8, CHS, and UB4 probes were a 299-bp BstBI fragment (representing only intron 8 sequences) isolated from a plasmid containing the respective genomic DNA sequences, two Pair fragments representing the complete parsley CHS cDNA (Reimold et al. 1983), and a 1-kb EcoRI fragment isolated from a parsley polyubiquitin cDNA (Kawalleck et al. 1993), respectively.

Cloning procedures and sequence analysis

To obtain recombinant λ clones representing the complete CPRF1 gene, three phage libraries were screened: an amplified λEMBL4 library (Herrmann et al. 1988), an newly prepared, amplified parsley genomic DNA library constructed in λGEM12 (Promega, Heidelberg, Germany) using partially Mbol-restricted parsley DNA cloned via the "half site approach" (Zabarovsky and Allikmets 1986), and an unamplified λGEM12 library (prepared as mentioned above) of 10<sup>6</sup> plaques plated on E.coli strain K2571 (Promega) and packaged using Stratagene Gigapack Gold extracts (Graham et al. 1990). Restriction fragments of the phage DNA were subcloned into pBlue-scriptSK<sup>11</sup> (Stratagene). From one of the members of a family of five additional λ clones that hybridised at reduced stringency to the CPRF1 probe, a 2-kb SpeI fragment was subcloned. This SpeI fragment contains the complete 1.1-kb Xbal fragment, as well as a part of the adjacent 1.7-kb Xbal fragment, identified in DNA gel blot experiments with genomic DNA (see Fig. 2). The sequenced CPRF1 cDNA containing intron 8 was identified in the same way as the nine other CPRF1 cDNAs (Feldbrügge et al. 1994) which were analysed to obtain a full-length cDNA.

Sequencing was carried out by the dideoxy chain-termination method (Sanger et al. 1977) using double-stranded plasmid DNA and the T7 DNA sequencing kit (Pharmacia). Most of the sequencing was done by creating the appropriate subclones using oligonucleotide primers which anneal to vector sequences. Gaps were filled in by the use of synthetic oligonucleotides annealing to CPRF1 sequences. The sequencing reaction products were resolved on gels with the addition of 1 M sodium acetate in the lower (anode) buffer chamber (Sheen and Seed 1988). All new sequence information in this work has been verified on both strands. Nucleotide and amino acid sequence analysis was performed on a DEC 3000 computer using the University of Wisconsin Genetics Computer Group (UWCG) program package (Devereux et al. 1984).

**Results and discussion**

The CPRF1 gene spans about 9 Kb

To determine the structure of a plant gene encoding a bZIP regulatory protein, we cloned CPRF1 genomic DNA. Two amplified parsley genomic libraries and one unamplified library were screened for λ clones representing the CPRF1 gene. In the first two attempts, only the 3′ region of the CPRF1 gene was obtained from an amplified λEMBL4 library as well as from a newly prepared, amplified parsley genomic DNA library constructed in λGEM12. A λ clone containing the 5′ region of the CPRF1 gene was ultimately obtained by screening an unamplified λGEM12 library. Analysis of this clone indicated the presence of repetitive DNA sequences in the region 5′ to the CPRF1 gene. Inserts from phages which displayed positive signals when