A G-box element from the \textit{Catharanthus roseus} strictosidine synthase (\textit{Str}) gene promoter confers seed-specific expression in transgenic tobacco plants

Received: 6 August 1998 / Accepted: 22 January 1999

Abstract The enzyme encoded by the strictosidine synthase (\textit{Str}) gene from \textit{Catharanthus roseus} catalyses a key step in the biosynthesis of the pharmaceutically important terpenoid indole alkaloids. \textit{Str} cDNA and genomic clones have already been isolated, allowing us to study the regulation of \textit{Str} gene expression. Here we focus on the role of a putative \textit{cis}-acting element, CA CGTG, in the \textit{Str} promoter. This sequence is known as a G-box, and functions as a transcription-regulating sequence in a number of other promoters. By means of electrophoretic mobility shift assays it was demonstrated that the \textit{Str} G-box is capable of interacting with nuclear factors in tobacco and with the cloned tobacco G-box-binding factor TAF-1. Disruption of the \textit{Str} G-box sequence by two single-nucleotide mutations prevented binding of factors, thereby demonstrating the specificity of the observed interactions. Functional analysis in transgenic tobacco plants demonstrated that these mutations also reduced the transcriptional activity of constructs containing tetramers of the \textit{Str} G-box sequence. Expression directed by a tetramer of the \textit{Str} G-box fused to a truncated promoter containing only a TATA box was confined to seeds and was found to increase during seed maturation. Thus, the \textit{Str} G-box tetramer is able to direct seed-specific expression independently of other regulatory sequences. G-box-directed expression in leaves required the presence of an enhancer region from the cauliflower mosaic virus (CaMV) 35S promoter. The results indicate that the G-box needs to interact with other elements to drive expression in leaf, and that it can by itself confer seed-specific expression as a multimer. The fact that only some of the G-boxes found in different promoters serve as seed-specific elements indicates that sequences flanking the G-box determine the transcriptional activity in different tissues. Based on sequence comparisons we propose that the nucleotides at positions −4, −3, −2 and/or +4 are important in determining seed-specific expression.

Key words \textit{Catharanthus roseus} \cdot G-box \cdot Seed-specific expression \cdot \textit{Str} gene \cdot TAF-1

Introduction

An important aspect of the regulation of gene expression is the specific interaction of transcription factors with \textit{cis}-acting elements present in gene promoter regions. These \textit{cis}-acting elements determine gene activity during plant development, and in response to external stimuli such as hormones, pathogens or light. Functional studies on plant promoters have resulted in the identification of a number of \textit{cis}-acting elements and their cognate \textit{trans}-acting DNA-binding factors (reviewed by Foster et al. 1994; Menkens et al. 1995; Meshi and Iwabuchi 1995). We are interested in the transcriptional regulation of genes involved in the biosynthesis of the pharmaceutically important terpenoid indole alkaloids in the tropical plant \textit{Catharanthus roseus} (Madagascar periwinkle, family Apocynaceae). This group of secondary metabolites is thought to have anti-herbivore activity and to function in protection against pathogens and/or UV radiation (Aerts et al. 1991; Hartmann 1991; Luijendijk 1995; Thomas et al. 1995). One of the key steps in the biosynthesis of terpenoid indole alkaloids is the condensation of the amino acid derivative tryptamine and the terpenoid secologanin, catalysed by the enzyme strictosidine synthase (EC 4.3.3.2.) (reviewed in Meijer et al. 1993; Hashimoto and Yamada 1994). The product of this reaction is strictosidine, which is the precursor for all other terpenoid indole alkaloids. cDNA and genomic clones of \textit{Str} have been obtained from \textit{C. roseus}, al-
allowing studies of the molecular mechanisms that control Str gene expression (McKnight et al. 1990; Pasquali et al. 1992, 1999). In C. roseus the Str gene was shown to be expressed in various tissues, with highest levels in roots. Str was also found to be induced by fungal elicitors, consistent with its presumed role in plant defence.

A search of the Str promoter sequence for putative cis-regulatory DNA-sequences that might be involved in determining the Str expression pattern revealed the presence of the palindromic sequence CACGTG, which is known as the G-box. The G-box was originally found in RbcS promoters from various plants species, as a conserved region which is protected by plant nuclear extracts from DNase digestion in vitro (Giuliano et al. 1988). Since then, G-box and G-box-like sequences have been studied for their functional significance in vivo; they have been shown to be involved in the regulation of a variety of unrelated genes in various responses, following exposure to visible and UV light (Schulze-Lefert et al. 1989a, 1989b; Block et al. 1990; Chattopadhyay et al. 1998), dehydration-stress (Lam and Chua 1991; Dolf erus et al. 1994), cold-stress (Dolf erus et al. 1994), abscisic acid (Marcotte et al. 1989; Lam and Chua 1991) and sucrose (Urwin and Jenkins 1997). Moreover, G-box-(like) sequences were also found to determine tissue-specific expression patterns, for example in seeds and roots (Salinas et al. 1992; Thomas 1993).

A number of cDNAs encoding nuclear proteins which specifically interact with G-box-(like) sequences have been identified. Nearly all of these proteins are members of the basic leucine zipper (bZIP) family (Guillotin et al. 1990; Oeda et al. 1991; Weisshaar et al. 1991; Schindler et al. 1992; Memelink and Gruissem 1994; Meier and Gruissem 1994; Hong et al. 1995; Izawa et al. 1994; for a review see Menkens et al. 1995). In addition, a trans-acting factor of the MYC type has been shown to interact with a G-box sequence (De Pater et al. 1997).

It is still not clear how identical G-box sequences can be involved in the specific responses of a wide variety of unrelated genes to a number of completely different stimuli. One important factor to consider is that the affinities of G-box binding trans-acting factors (GBFs) are also dependent on the sequence context flanking the G-box (Williams et al. 1992; Izawa et al. 1993). This is demonstrated by the fact, among others, that some perfect G-boxes have been found not to bind GBFs at all, whereas other G-boxes form either characteristic low- (type A) or high- (type B)-mobility complexes in electrophoretic mobility shift assays (Williams et al. 1992). Several promoter studies that have focused on the function of G-box sequences in vivo have revealed that additional cis-acting elements are necessary for trans-activation of gene expression, thereby demonstrating the existence of synergism between certain G-boxes and other cis-acting elements (Donald and Cashmore 1990; Loake et al. 1992; Kawagoe et al. 1994; Shen and Ho 1995). Based on the data currently available, the sequence context of a particular G-box is insufficient to allow one to predict whether that G-box is recognized by GBFs, whether it is a functional cis-acting element, and which in vivo activity it may confer.

The aim of the experiments described here was to determine whether the G-box in the C. roseus Str promoter binds GBFs and whether this element can drive gene expression. Previous studies showed that 0.6 kb of Str upstream sequences could confer expression on the gusA reporter gene in transgenic tobacco plants (Pasquali et al. 1994, 1999). Here we show that the Str G-box is recognized in vitro by nuclear factors from tobacco. Furthermore, we demonstrate that a tetramer of the Str G-box directs seed-specific expression when combined with a truncated promoter containing only a TATA box, and elevates the expression level in leaves in a synergistic interaction with the as-1 enhancer-containing region of the CaMV 35S promoter.

Materials and methods

Vector construction and generation of transgenic plants

Head-to-tail tetramers of the wild-type (4SW) and mutant (4SM) Str G-box were made using 18-bp oligonucleotides flanked by SfiI and XhoI sites (underlined in Fig. 1A), as described by Ouwerkerk and Memelink (1997), and were subsequently cloned upstream of the truncated CaMV 35S (−47 to −90) promoters using the SfiI sites in the plasmids GusSH−47 and GusSH−90 (Pasquali et al. 1994). These vectors contain the −47 to +27 and −90 to +27 sequences, respectively, from the CaMV 35S promoter, in a

Fig. 1 A Schematic representation of the Str −531 promoter from C. roseus. Indicated are the TATA box (−29 to −26), the mRNA transcription start site at +1 and the translation start codon at +57. The SfiI and RsaI restriction sites generate the G-box-containing fragment SR (−115 to −101) which was used in EMSAs. The sequence of SR is presented and the G-box is underlined. The nucleotide sequence of the Str gene is available from the EMBL/GenBank/ DDBJ database under accession number Y10182. The synthetic oligonucleotides SW and SM represent wild-type and mutant Str G-box sequences, respectively. The sequences in lower case at the 5’ and 3’ ends represent SfiI and XhoI sites used for tetramerization. The GT to CA mutation in the G-box is also shown in lower case. B Structure of the chimeric Str G-box-gusA fusion constructs. Tetramers of wild type (SW) or mutant (SM) Str G-box sequences were cloned upstream of the truncated CaMV 35S −90 or −47 promoters