Invited Trends Article

Towards an Understanding of Plant Gene Regulation: The Action of Nuclear Factors

Kurt Weising and Günter Kahl

Pflanzliche Molekularbiologie, Botanisches Institut der Universität Frankfurt, Siesmayerstraße 70, D-6000 Frankfurt, Bundesrepublik Deutschland

Z. Naturforsch. **46c**, 1–11 (1991); received December 7, 1990

Over the last decade an intensive research on the regulation of gene expression in viral and animal systems has led to the discovery of *cis*-acting regulatory sequences, the identification of sequence-specific DNA-binding proteins (*trans*-acting factors), the characterization of protein domains involved in DNA-protein recognition and binding as well as in protein-protein interactions, and the cloning and sequencing of genes encoding regulatory proteins. The tremendous progress in this field is now being complemented by advances in our understanding of how plant genes are regulated. A wealth of data has accumulated in the past few years witnessing basic similarities in the transcriptional regulation of various eukaryotic genes, but also specific features of plant genes. This article collects presently available data, focusses on DNA-protein interactions in plant genes, particularly in light-regulated and "constitutively expressed" genes, reports on the isolation of plant genes encoding regulatory proteins, and is meant to induce further activities in plant gene research.

Introduction

Eukaryotic transcriptional initiation is regulated by complex interactions between *cis*-acting DNA motifs and *trans*-acting protein factors [1–3]. Among the *cis*-regulatory regions, *promoters* are located close to the transcription initiation site and usually consist of proximal (*e.g.* the TATA box) and more distal elements (*e.g.* that CCAAT box). *Enhancers,* on the other hand, can be located far up- and down-stream from the initiation site (or even *in trans* [4]) and may act independently of their position and orientation. Promoters and enhancers are usually composed of several discrete, often redundant elements [5], each of which may be specifically recognized by one or more *trans*-acting proteins. At least three separate domains have been identified within such regulatory proteins. One is necessary for sequence-specific DNA recognition, one for activation of transcriptional initiation, and one for the formation of protein-protein interactions (*e.g.* dimerization). Four "motifs" involved in DNA sequence recognition and/or factor dimerization have been characterized: zinc fingers [6], helix-turn-helix [7], leucine-zipper [8, 9] and helix-loop-helix motifs [10]. Since leucine-zipper and helix-loop-helix proteins form homo- or heterodimers as a prerequisite for DNA binding, the combination of factors that differ in their activating capacity provides a flexible tool for the fine-tuning of transcriptional regulation [11, 12]. "Activating" domains are rich in either proline, glutamine, or acidic amino acids [1–3]. They are thought to interact with the TATA box-binding transcription factor TFIID [13], RNA polymerase [14], or another protein of the general transcription apparatus, either directly or *via* an "adaptor" protein not directly attaching to DNA [15, 16]. Development- and tissue-specific gene expression is thought to be regulated by the interaction of enhancer- and promoter-bound, general and tissue-specific factors with DNA, with one another, and possibly with intermediary factors, thereby forming a preinitiation complex with properties comparable to a jigsaw puzzle [3].

While most of our current knowledge on transcriptional regulation derives from work on yeast, *Drosophila,* and mammalian cells, it is now complemented by considerable progress in plant systems. *Trans*-acting factors binding to *cis*-regulatory DNA sequences of a variety of plant genes have been characterized *in vivo* and *in vitro*. Genes coding for regulatory proteins have already been isolated and are currently being characterized. In some instances, mammalian and yeast transcription factors were shown to function in plants and *vice versa*.

Reprint requests to Günter Kahl.

Verlag der Zeitschrift für Naturforschung, D-7400 Tübingen
0939–5075/91/0100–0001 $ 01.30/0

DNA-protein interactions in plant genes

A variety of *in vitro* methods has been developed for the analysis of protein-DNA interactions. After performing the initial binding reaction (nuclear extracts are mixed with DNA fragments carrying the native or mutagenized sequence of interest in the presence of specific or nonspecific competitors), resulting complexes can be characterized by gel retardation [17], nitrocellulose filter binding [18], or "footprinting", *i.e.* limited digestion with DNase I, exonuclease III, or treatment with DNA-reactive chemicals followed by the analysis of the protected DNA "footprints" on a sequencing gel. A factor that binds DNA *in vitro* will not necessarily do so in the intact cell. To test for *in vivo* interactions, *in vivo* footprinting procedures based on the genomic sequencing technique have been developed [20, 21].

Non-sequence-specific interactions

DNA interacts more or less unspecifically with a variety of nuclear proteins. For example, the core *histones* build up the basic unit of eukaryotic chromatin, the nucleosome, and histone H 1 is in part responsible for higher order chromatin structure. Histone-DNA interactions and their role in active plant chromatin formation have been reviewed extensively [22, 23] and will not be discussed here.

High-mobility-group-(HMG-)proteins also bind DNA with little sequence preference. These proteins are major constituents of eukaryotic chromatin and have also been characterized in some plant species (reviewed in [22, 23]). In mammals, several classes of these proteins are known. Among these, HMG 14 and 17 are thought to be involved in the formation of active chromatin, and HMG I, a protein that binds preferentially to AT-rich DNA sequences, was suggested to function in nuclear matrix attachment of DNA [24]. Interestingly, two nuclear factors (LAT 1 and NAT 1) that have recently been characterized in different organs of soybean fulfill the operational criteria for an HMG I-like protein: they are released from chromatin with a low-salt buffer, are soluble in 2% trichloroacetic acid, and bind to AT-rich regions in soybean nodulin gene promoters [25]. Although no clear function could be assigned to LAT 1 and NAT 1, a role in modulating the chromatin conformation of nodulin promoters as a prerequisite for the organ-specific interaction with other factors was suggested [25]. HMG-like proteins binding to AT-rich promoter regions of a member of the zein gene family have also been found in maize nuclear extracts [26].

Another category of nuclear factors interacting with DNA non-sequence-specifically comprises proteins recognizing methylated DNA. Cytosine methylation, especially if it occurs in the sequence context of CpG or CpXpG, is thought to be responsible for gene inactivation in most eukaryotes including plants (reviewed in [27]). In mammals, a protein was identified that binds unspecifically to any cluster of methylated CpG [28]. Binding of this protein supposedly stabilizes the 30 nm solenoid fiber, so that the DNA is inaccessible for transcription factors. This would result in general repression of genes that possess methylated CpG islands in their promoters [29]. A recent report provides evidence that similar proteins may also exist in plants [30]. Gel retardation experiments with nuclear extracts from pea seedlings revealed that a factor (DBP-m) recognizes 5-methylcytosine residues in DNA without appreciable DNA sequence specificity. Future research on plant factors binding to methylated DNA might shed light on the mechanism of transcriptional regulation of plant genes by DNA methylation and help to answer the question as to why plant DNA is so much richer in 5-methylcytosine than mammalian DNA [31].

Non-sequence-specific interactions with DNA could also occur with conformation-specific regulatory proteins recognizing DNA sequence elements in a non-B-conformation (*e.g.* Z-DNA and H-DNA). The detection of nuclear factors interacting with S 1 nuclease-sensitive polypurine/polypyrimidine stretches was reported for several mammalian promoters (*e.g.* [32]). However, no such data do yet exist for plants.

Sequence-specific interactions

In recent years, much work has been devoted to the sequence-specific *in vivo* and *in vitro* interactions between nuclear plant factors and *cis*-regulatory sequences from plant, viral and T-DNA promotors and enhancers (summarized in Table I). Several conserved motifs or "boxes" were recognized, and complex patterns of interactions were observed in many cases. We confine the pres-

Table I. *In vitro* and *in vivo* interactions between nuclear proteins and plant-, plant virus- and T-DNA-derived upstream DNA sequence elements.
Abbreviations of methods: GR (gel retardation), DF (*in vitro* DNase I footprinting); CF (*in vitro* chemical footprinting); DMS (*in vitro* dimethyl sulfate binding interference); EP (*in vitro* exonuclease protection); GF (*in vivo* genomic footprinting). The sources of nuclear extracts are mentioned only if isolated from a species different from the origin of the investigated genes.

Origin of gene(s) and nuclear extract	Method	Recognized motif(s)	Binding factor	Reference
Cauliflower mosaic virus (CaMV) 35S promoter, tobacco extracts	GR, DF, DMS	2× TCACG		85
CaMV 35S; pea extracts	GR, DF	2× TCACG	ASF-1	86
CaMV 35S; wheat and sunflower extracts	GR, DF	ACGTCA	HBP-1	87
CaMV 35S; tobacco extracts	GR, DF, DMS	GATGTGATA	ASF-2	79
MSV upstream activating sequence; maize extracts	GR			77
T-DNA octopine synthase (ocs) gene; tobacco and maize extracts	GR, CF	ACGTAAGCGCTTACGT (ocs enhancer)	OCSTF OCSBF-1	88 89 90 92
T-DNA ocs gene; tobacco extract	GR	ocs enhancer	ASF-1	91
T-DNA nopaline synthase (nos) gene; tobacco	GR, DF	TGAGCTAAGCACATACGTCAG (nos enhancer)	ASF-1	93
T-DNA nos gene; wheat and sunflower extract	GR, DF	ACGTCA	HBP-1	87
Wheat histone H3 and H4 genes	GR, DF, DMS	ACGTCA	HBP-1	94 97
Wheat histone H3 genes	GR		HSBF	93
Maize alcohol dehydrogenase 1 (Adh1) gene (anaerobically induced)	GF	A: GGTGTCCGCC B1: GTGG B2: CCCCGG C: GGTGC		119
Maize Adh1 gene	GR	B2: CCCCGG	ARF-B2	120
Arabidopsis Adh gene	GF, GR, DF	CCCC-motif CCACGTGG (G-box)	GBF	47 48 49
Rice rab-16A gene (ABA-responsive)	GR, DF	I: TACGTGG (G-box) II: CCGCCGCGCCT		53
Wheat Em gene (ABA-responsive)	GR, DMS	CACGTGGC (G-box)	EmBp-1	54
Tomato E4 and E8 gene (ethylene-responsive)	GR, DMS	AT-rich motifs		67 68
Rice alpha-amylase gene (gibberellin-responsive)	GR, EP			75
Carrot extensin gene (wound-responsive)	GR, EP	AT-rich motif TTTTTTTT TGACGT	EGBF-1	69 70

Origin of gene(s) and nuclear extract	Method	Recognized motif(s)	Binding factor	Reference
Potato proteinase inhibitor II gene (wound-responsive), tomato extract	GR	AAGCTAAGT		121
Parsley chalcone synthase (chs) gene (UV-induced)	GF	AACCTAACCT TCCACGTGGC (G-box) ACGTGGA (G-box) CTTCACTTGATGTATC		45 51 52
Snapdragon chs gene; snapdragon, petunia, tobacco and Arabidopsis extracts	GR	CACGTG (G-box)	CG-1	50
Snapdragon chs gene	GR	47 bp repeat		122
French bean chs gene	GR, DF	2× CCTACC (N 7) CT		123
Parsley phenylalanine ammonia-lyase and 4-coumarate: CoA ligase genes (UV- or elicitor-induced)	GF, DMS	2 constitutive: TCTCCAC; TGTCCACGT 3 inducible: CTCCAACAAACCCCTTC CCTACC, CCGTCC		124
Pea ribulose-bisphosphate carboxylase small subunit (rbcS) 3 A gene (light-induced)	DR, DF, DMS	GTGTGGTTAATATG (box II) ATCATTTTCACT (box III)	GT-1	35 36 37 38 39
Pea rbcS 3 A gene; tobacco extracts	GR, DF	AT-rich motif	3 AF 1	59
Pea rbcS 3 A gene; tobacco extracts	GR	ATGATAAGG (I-box)	GAF-1	33
Pea, tomato and Arabidopsis rbcS genes	GR, DF	TCTTACACGTGGCA (G-box)	GBF	40
Pea rbcS-3.6 gene	GR	AATATTTTTATT	AT-1	41
Lemna rbcS gene	GR, DF	GATAAG (I-box)	LRF-1	78
Arabidopsis rbcS gene; yeast extract	GR, DMS	CACGTGGC (G-box) GATAAG (I-box)	GBF GA-1	57
Rice phytochrome gene (light-repressed)	GR	2 GGTTAA-motifs	GT-1	44
Petunia chlorophyll a/b-binding protein (cab) gene (light-induced)	GR, DF DMS	GATATAGATA (I-box?)	ASF-2	79
Tobacco cabE gene	GR	ATAAAAATAATT	AT-1	34
Tobacco cabE gene	GR	GATATAGATA GATAAG (I-box)	GA-1	34
Tobacco cabE gene	GR	GGGCCGG	GC-1	34
Tobacco cabE gene	GR, DMS	AGACGTGG (G-box)	GBF	34
Tobacco cabE gene	GR, DMS	7 GGTTAA-like motifs	GT-1	34
Soybean nodulin N 23 gene (nodule-specific)	GR, DF	AT-rich motifs	LAT 1 NAT 1 NAT 2	25

Origin of gene(s) and nuclear extract	Method	Recognized motif(s)	Binding factor	Reference
Soybean leghemoglobin (lbc 3) gene (nodule-specific)	GR	AT-rich motifs		71 72
Sesbania rostrata leghemoglobin gene; soybean, alfalfa and Sesbania extracts	GR	AT-rich motifs		72 73
French bean glutamine synthetase gene (nodule-specific subunit)	GR	AT-rich motifs	PRF-1 PNF-1	74
Soybean β-conglycinin gene, alpha-subunit (embryo-specific)	GR, EF	2 AACCCA-motifs AT-rich motifs	SEF-3 SEF-1	62
Sunflower helianthinin gene (embryo-specific)	GR, CF	AT-rich motifs		61
Soybean lectin gene (embryo-specific)	GR, DF	2 AT-rich motifs		60
French bean β-phaseolin gene (embryo-specific)	DR, DF, CF	AT-rich motifs		63
French bean β-phaseolin gene; carrot extract	GR	AT-rich motifs		64
French bean phytohemagglutinin genes (embryo-specific)	GR	AT-rich motifs		65
Carrot DC 59 gene (embryo-specific)	GR, DF	AT-rich motifs		66
Maize sucrose synthase gene (transcribed in various tissues)	GR, DF, DMS	AT-rich motifs CAGCATATGCTA TTGCCGATCA	MNP-1 MNP-2	76 125
Maize zein gene pMS 1 (endosperm-specific)	GR, NC, DF	AT-rich motifs CACATGTGTAAAGGT AAAGGTAAAGGTGTGT		126 127 26
Maize ribosomal DNA	GR, DF, EF, NC	possible hairpin		128
Maize En-1 transposon	DMS	2× CCGACACTCTTA	tnpA	129 130
Pea ferredoxin gene (Fed-1)	GR		BC 1 BC 2	131

ent discussion to some well-characterized elements originally detected 5′ to the transcriptional start site of some light-regulated genes, and the cauliflower mosaic virus 35S gene. Some of these motifs, however, occur in a variety of other, and sometimes seemingly unrelated, promoter and enhancer elements.

Light-regulated genes: a variety of cis- and trans-acting elements

At least five different factor-binding motifs have been detected 5′ to the transcription start site of light-regulated genes, *e.g.* within the promoters of rbcS (ribulose-bisphosphate carboxylase) and cab (chlorophyll-a/b-binding protein) genes from different species (reviewed in [33, 34]):

- "box II" and "box III" bind a factor called GT-1 [35–39];
- the "G-box" binds a factor called GBF [34, 40];
- AT-rich elements bind a factor called AT-1 [34, 41];
- a GC-rich element binds a factor called GC-1 [34];
- the "I-box" binds a factor called GA-1 [34].

GT-1 binds to redundant box II- and box III-elements of the pea rbcS 3 A gene promoter and may be involved in both positive and negative photoregulation of the gene [42, 43]. Critical spacing requirements for GT-1 binding sites have been observed for GT-1-mediated reporter gene transcription in transgenic plants [37]. Changing the distance between box II and III had a profound effect on transcription, but not on GT-1 binding. Obviously, GT-1 binding may be productive or nonproductive. The factor is thus necessary, but not sufficient for light-regulated transcription. Although itself being probably involved in light regulation [38], GT-1 binding requires other factors for the formation of a stable transcription initiation complex. Interestingly, GT-1 binds also to promoter elements of the rice phytochrome gene, which is negatively regulated by light [44]. Binding of one and the same factor may thus exert opposite effects in different genes.

Although detected first in the promoter of a tomato rbcS gene [40] and present in the 5′ upstream region of at least 13 rbcS genes from different species [45] and a cab gene from *Nicotiana plumbaginifolia* [34, 46], "G-box" motifs (CCACGTGG) are not restricted to photosynthetic genes, and not even to plants. G-box-like sequence elements occur in the promoters of genes responsive to various kinds of stimuli, such as the *Arabidopsis* alcohol dehydrogenase gene [47–49], *Antirrhinum* [50] and parsley chalcone synthase genes [45, 51, 52], abscisic acid-responsive genes from rice [53] and maize [54], tissue-specific genes such as the potato patatin gene [45], the yeast pho4 gene [55] and even mammalian genes [56]. G-box motifs commonly interact with nuclear factors. *In vitro* studies with the tomato rbcS promoter identified a G-box-binding protein called GBF [40]. A probably related factor called CG-1 was found to bind *in vitro* to a G-box within a UV-responsive element of a snapdragon *(Antirrhinum majus)* chalcone synthase promoter [50]. A yeast factor bound *in vitro* to a G-box sequence derived from an rbcS gene of *Arabidopsis* [57]. Moreover, the presence of G-boxes upstream of a truncated promoter activated transcription of an adjacent gene in transformed yeast cells [57]. Obviously, both the G-box and its corresponding binding factor(s) are ubiquitous elements for transcriptional regulation in a variety of evolutionary distant organisms.

The presence of a G-box-binding factor in a nuclear extract, as determined by *in vitro* binding assays, does not necessarily implicate its binding to the DNA *in vivo*. In the *in vitro* study with the snapdragon chalcone synthase gene [50], CG-1 was present irrespective of UV induction, and binding was disturbed by cytosine methylation. In contrast, *in vivo* footprinting of the parsley chalcone synthase promoter region revealed UV-inducible factor binding to a G-box sequence [45, 51]. In comparative studies on the *Arabidopsis* Adh promoter G-box it was shown that *in vivo* binding occurred in suspension-cultured cells, but not in leaves [48]. However, a factor binding to the G-box *in vitro* was present in nuclear extracts from both cell types [48, 49]. The observed differences between *in vivo* and *in vitro* experiments may be explained by more complex *in vivo* interactions with accessory factors, and also by *in vivo* modification(s) of GBF.

AT-rich elements in pea rbcS and tobacco cab genes bind a factor *in vitro*, called AT-1 [34, 41]. Interestingly, a reversible modulation of binding capacity by phosphorylation was observed in this case: AT-1 binds in its non-phosphorylated state and loses all DNA-binding capacity upon phosphorylation. Deletion of the AT-1 binding site from the tomato rbcS-3 A promoter abolishes transcription in transgenic plants [58]. AT-1 may therefore be involved in positive control of gene activity. On the other hand, three AT-1 boxes reside within the negative regulatory element of the *Nicotiana plumbaginifolia* cabE promoter [34, 46], so that the influence of AT-boxes on transcription *in vivo* remains to be determined. Another AT-binding factor called 3 AF 1 was recently identified in nuclear extracts from tobacco [59]. This factor attached to an AT-rich motif in the vicinity of the pea rbcS 3 A TATA box. Since a tetramer of the binding motif does not confer light-regulated expression in transgenic plants, and extracts from roots as well as leaves do show binding capacity, factor 3 AF 1 is probably an accessory factor not directly involved in light regulation [59].

AT-rich elements that bind nuclear proteins *in vitro* are quite common among non-photosynthetic genes (see Table I). They have been observed 5′ to the transcription start site of a variety of embryo-specific genes [60–66], ethylene- and wound-responsive genes [67–70], nodule-specific genes

[25, 71–74], as well as a rice amylase [75] and a maize sucrose synthase gene [76]. Although cross-competition was sometimes observed, *e.g.* between factors binding to AT-rich regions from sunflower helianthinin and French bean phaseolin genes [61, 63], or from different nodule-specific genes [25, 74], the relationship of all these factors among each other and to AT-binding factors from light-regulated promoters is still unclear. While stringent sequence requirements for binding were sometimes observed (*e.g.* for PNF-1 binding to the TATTT(T/A)AT-motif in the French bean glutamine synthetase promoter region; [74]), abundance of AT in a target sequence *per se* seems to be sufficient for binding in other cases [63]. Since proteins binding to AT-rich promoter regions of a soybean nodulin gene [25] and a maize zein gene [26] were identified as HMG I-like, this class of nuclear proteins may also be involved in reported AT-binding with limited sequence specificity.

A GC-rich motif bound by a nuclear factor called GC-1 has been detected upstream of a *Nicotiana plumbaginifolia* cab gene [34]. No GC-rich motifs have yet been reported from other photosynthetic genes. A similar motif, however, occurs twice in a maize streak virus promoter and is bound by a maize nuclear protein [77].

GATA motifs upstream of the *Nicotiana plumbaginifolia* cab gene bind GA-1 [34]. A related sequence, referred to as "I-box" [40], occurs in other promoters and also binds nuclear factors. A GATA-motif within an rbcS gene promoter from *Lemna gibba* binds a factor called LRF-1 [78]. In the *Lemna* rbcS promoter, the concentration of the factor was dependent on light, and binding could be enhanced by a 2-min exposure of the plant to red light. Protein binding to GATA-related motifs may thus be involved in gene regulation by phytochrome. Another factor was shown to bind to a conserved GATA motif in a petunia cab promoter and, surprisingly, also to a similar motif between position −90 and −98 of the cauliflower mosaic virus (CaMV) 35S promoter ([79], see below). The factor, present in tobacco leaf, but not in root extracts, was called ASF-2 [79]. Finally, a factor called GAF-1 has been prepared from tobacco leaves that binds to the I-box (ATGATAAGG) of the pea rbcS3A gene [33]. This factor is present in greater abundance in extracts from light-grown as opposed to dark-grown plants [33]. The structural and functional relationships between GA-1, GAF-1, ASF-2 and LRF-1 are as yet unknown.

Which implications do all these motifs and their binding factors bear for light regulation? With the exception of LRF-1 in *Lemna* [78], and GAF-1 in tobacco [33], all the factors were similarly abundant in nuclear extracts from light- as well as dark-grown plants. Light-dependent modification mechanisms (*e.g.* repressor binding or phosphorylation) may thus convert a putative light-responsive factor from an active to an inactive configuration and vice versa. Some factors may not be light-responsive at all, but rather influence tissue specificity or rate of transcription. The ubiquitous G-box may have a general function, and may only work in concert with gene-specific transcription factors. It seems plausible that complex interactions between light-responsive (GT-1? LRF-1?), general (GBF?), and accessory factors (3AF1?) occur prior to the formation of a functional initiation complex upon light induction. The availability of an increasing number of cloned genes coding for plant *trans*-acting factors (see below), as well as of a plant *in vitro* transcription system [33] will further accelerate research in that area.

"Constitutive" promoters: TGACG-binding factors

One of the best characterized promoters functioning in plants is the 35S promoter from cauliflower mosaic virus that directs the constitutive expression of genes in transgenic plants and protoplasts [80, 81]. Extended deletion and dissection analyses provided evidence that this promoter is composed of several functional elements. Each element directs a distinct pattern of tissue- and development-specific transcription in transgenic plants [82–84]. Constitutive expression thus results from the combined action of all the regulatory units involved [84]. At least two of these elements were shown to bind transacting factors:

- one element (as-1) contains two TGACG-motifs between −82 and −66, binds a factor called ASF-1, and confers root-specific expression [85–87];
- another element (as-2) contains a GATGTGA-TA-motif between −98 and −90, binds a factor called ASF-2 and confers tissue-specific expression in trichomes, vascular elements, epidermal and mesophyll cells [79].

Interestingly, both these motifs and the corresponding factors do not only appear in the CaMV 35S promoter. GATA motifs are present in the 5′ region of various light-regulated genes (see above), and TGACG-related motifs that bind nuclear proteins have also been found upstream of the T-DNA-derived octopine [88–92] and nopaline synthase genes [93], and a wheat histone 3 gene [87, 94–96]. All these factors compete specifically with each other for binding in gel retardation assays [89, 96]. However, their binding domains are probably not identical, since TGACG-related motifs differ considerably in their number (monomer *vs.* dimer binding), distance, and orientation to one another [93]. Cloning and sequencing of the putative cDNAs coding for ASF-1 from tobacco [97], HBP-1 from wheat [98], and OCSBF-1 from maize [92] showed that these factors are members of the group of leucine-zipper proteins, together with a variety of transcription factors derived from yeast (*e.g.* GCN4) and mammals (*e.g.* the cAMP-responsive element binding protein; CREB). It seems likely that all these TGACG-binding factors belong to a family of regulatory proteins widely distributed in the eukaryotic kingdoms. A recent study on a specific factor (EmBP-1), binding to an abscisic acid-responsive element from the wheat Em gene [54] revealed that TGACG-binding factors and G-box binding factors may be related to one another. A DNA fragment that contained the recognition sequence for wheat HBP-1 competed with the abscisic acid-responsive element for EmBP-1 binding. The recognition sequence of EmBP-1 (CACGTGG), however, perfectly matches the critical G-box core of seven bases [52]. By aligning G-box- and TGACG-related motifs (see Table I), it is obvious that an ACGT-sequence is common to both motifs in all cases. The possibility of a close relationship is further supported by the fact that both motifs are recognized by leucine-zipper proteins [54, 92, 97, 98]. The binding of dimeric leucine-zipper factors requires binding sites of dyad symmetry [8, 15, 99], which are represented by the perfect G-box as well as, *e.g.*, the ocs enhancer. Substantial evidence for dimer binding was indeed obtained in the latter case [91]. It is tempting to speculate that, in analogy to mammalian and yeast leucine-zipper factors [8, 9, 11], families of G-box- and TGACG-binding factors exist in plants that are able to form homo- and heterodimers. If these protein complexes differ in their capacity to stimulate transcription, different combinations of a limited number of family members may provide the plant cell with a wide spectrum of regulatory potentials. Such potentials may be further expanded by post-transcriptional modification of factors in response to external stimuli, *e.g.*, by phosphorylation [41, 100, 101] or by cytosin methylation of target sites which inhibits factor binding [50, 102].

The isolation of genes encoding plant trans-acting factors

Several strategies were applied to identify genes coding for plant regulatory proteins. One approach was to analyze regulatory mutants of maize [103–110]: several maize genes involved in the regulation of the anthocyanin biosynthesis pathway have been cloned by transposon tagging, and their cDNA or genomic sequences have been determined. Their deduced protein sequences exhibited considerable similarities to mammalian transcription factors. For example, the c1 regulatory locus of maize encodes a protein with similarities to the DNA binding domain of human proto-oncoproteins such as c-myb [104]. A variety of cDNAs from maize and barley share this similarity [105]. B-I and B-peru, two members of the regulatory B gene family, as well as Lc, belonging to the R gene family, encode proteins that exhibit similarities to the helix-loop-helix DNA binding/dimerization motif of human L-myc gene products [106, 107]. B-I, B-peru, c1 and Lc-encoded proteins possess regions rich in acidic amino acids reminiscent of activation domains [1–3]. The observed similarities to known DNA-binding and activation domains support the proposed role of c1, B and R gene products as *trans*-acting factors regulating the activity of anthocyanin biosynthesis genes. In fact, transactivation of these genes leading to pigmentation of transgenic maize tissues was observed after transfer of B or Lc regulatory genes to maize cells by particle bombardment [107, 108]. Another regulatory gene from maize, the opaque-2 gene, that is involved in zein deposition in maize endosperm, encodes a protein with structural similarities to the leucine-zipper class of mammalian transcription factors [109, 110]. A β-gal-opaque-2 fusion protein synthesized in

E. coli was able to bind to the 5′ region of a genomic zein gene clone [110].

Homeotic genes involved in development-specific gene expression have also been isolated from other plant species. The *Arabidopsis* homeotic gene *agamous* was isolated by T-DNA tagging [111]. This gene encodes a flower-specific protein that shares similarities with the human serum-responsive factor, a yeast factor involved in regulating mating-type-specific genes, and the gene product of *deficiens* (an *Antirrhinum majus* gene involved in the control of flower development; [112]). Obviously, the cloning of regulatory loci by transposon or T-DNA tagging is a promising strategy for identifying and characterizing genes coding for plant *trans*-acting factors.

Another approach for the isolation of genes encoding DNA-binding regulatory proteins is just as promising. It is based on the screening of an expression library with a recognition site DNA, *i.e.*, a labeled oligonucleotide that carries a known *cis*-regulatory DNA sequence motif [113, 114]. Using this strategy, a variety of genes encoding plant nuclear factors could be isolated and characterized: genes for putative leucine-zipper proteins from tobacco [97], wheat [54, 98] and maize [92], as well as a putative zinc-finger protein from tobacco [59]. The feasibility of this technique will soon allow the isolation of many more genes encoding plant *trans*-acting factors with known recognition sequences.

Perspectives

The cloning of plant regulatory genes has thus far confirmed the evolutionary conservation of specific DNA-binding and dimerization motifs. Helix-loop-helix, leucine-zipper, and zinc-finger motifs previously characterized in mammalian, yeast and *Drosophila*-derived nuclear factors, have now also been identified in the deduced amino acid sequences of plant regulatory proteins [59, 92, 97, 98, 106, 109, 110]. In some cases, *trans*-kingdom factor binding or competition has been observed, *e.g.*, between a yeast factor and a plant-derived G-box [57], or between a plant factor and animal AP-1 and CREB sites [92]. The similarity of basic processes in the recognition of plant, fungal, and animal upstream elements and their interactions with the transcriptional machinery is further corroborated by

- the apparent conservation of domains within the TATA-box binding factor TFIID and the large subunit of RNA polymerase II recently cloned from *Arabidopsis* [115, 116],
- the transactivation of truncated plant promoters by the yeast factor GAL4 [117] or mammalian Fos and Jun oncoproteins [118].

Taken altogether, the present state of research demonstrates that not only dimerization and DNA-binding domains, but also activation domains of *trans*-acting factors, and the basic architecture of the transcriptional initiation complex are conserved among eukaryotes. With the availability of an *in vitro* transcription system [33] and cloned components of the transcriptional machinery [114, 115], it will soon be possible to detect specific features of plant gene regulation beyond the conserved mechanisms characterized so far.

[1] M. Ptashne, Nature **335,** 683–689 (1988).
[2] P. J. Mitchell and R. Tjian, Science **245,** 371–378 (1989).
[3] P. F. Johnson and S. L. McKnight, Annu. Rev. Biochem. **58,** 799–839 (1989).
[4] H. P. Müller and W. Schaffner, Trends Genet. **6,** 300–304 (1990).
[5] W. S. Dynan, Cell **58,** 1–4 (1989).
[6] A. Klug and D. Rhodes, Trends Biochem. Sci. **12,** 464–469 (1987).
[7] K. Struhl, Trends Biochem. Sci. **14,** 137–142 (1989).
[8] T. Abel and T. Maniatis, Nature **341,** 24–25 (1989).
[9] S. J. Busch and P. Sassone-Corsi, Trends Genet. **6,** 36–40 (1990).
[10] C. Murre, P. S. McCaw, and D. Baltimore, Cell **56,** 777–783 (1989).
[11] N. Jones, Cell **61,** 9–11 (1990).
[12] R. Benezra, R. L. Davis, D. Lockshon, D. L. Turner, and H. Weintraub, Cell **61,** 49–59 (1990).
[13] K. F. Stringer, C. J. Ingles, and J. Greenblatt, Nature **345,** 783–786 (1990).
[14] P. B. Sigler, Nature **333,** 210–212 (1988).
[15] M. Ptashne and A. A. F. Gann, Nature **346,** 329–331 (1990).
[16] B. Lewin, Cell **61,** 1161–1164 (1990).
[17] M. M. Garner and A. Revzin, Nucleic Acids Res. **9,** 3047–3060 (1981).
[18] J. F. X. Diffley and B. Stillman, Mol. Cell. Biol. **6,** 1363–1373 (1986).
[19] D. J. Galas and A. Schmitz, Nucleic Acids Res. **5,** 3157–3170 (1978).
[20] H. P. Saluz and J. P. Jost, Anal. Biochem. **176,** 201–208 (1989).
[21] G. M. Church and W. Gilbert, Proc. Natl. Acad. Sci. U.S.A. **81,** 1991–1995 (1984).

[22] G. Kahl (ed.), Architecture of Eukaryotic Genes, VCh Weinheim (1988).
[23] S. Spiker, Physiol. Plant. **75,** 200–213 (1988).
[24] M. J. Solomon, F. Strauss, A. Varshavsky, Proc. Natl. Acad. Sci. U.S.A. **83,** 1276–1280 (1986).
[25] K. Jacobsen, N. B. Laursen, E. O. Jensen, A. Marcker, C. Poulsen, K. A. Marcker, The Plant Cell **2,** 85–94 (1990).
[26] U.-G. Maier, K. D. Grasser, M. M. Haaß, and G. Feix, Mol. Gen. Genet. **221,** 164–170 (1990).
[27] H. Cedar, Cell **53,** 3–4 (1988).
[28] R. R. Meehan, J. D. Lewis, S. McKay, E. L. Kleiner, and A. P. Bird, Cell **58,** 499–507 (1989).
[29] F. Antequera, D. MacLeod, and A. P. Bird, Cell. **58,** 509–517 (1989).
[30] D. Zhang, K. C. Ehrlich, P. C. Supakar, and M. Ehrlich, Mol. Cell. Biol. **9,** 1351–1356 (1989).
[31] Y. Gruenbaum, T. Naveh-Many, H. Cedar, and A. Razin, Nature **292,** 860–862 (1981).
[32] E. K. Hoffman, S. P. Trusko, M. Murphy, and D. L. George, Proc. Natl. Acad. Sci. U.S.A. **87,** 2705–2709 (1990).
[33] P. M. Gilmartin, L. Sarokin, J. Mememlink, and N.-H. Chua, The Plant Cell **2,** 369–378 (1990).
[34] U. Schindler and A. R. Cashmore, EMBO J. **9,** 3415–3427 (1990).
[35] P. J. Green, S. A. Kay, and N.-H. Chua, EMBO J. **6,** 2543–2549 (1987).
[36] P. J. Green, M.-H. Yong, M. Cuozzo, Y. Kano-Murakami, P. Silverstein, and N.-H. Chua, EMBO J. **7,** 4035–4044 (1988).
[37] P. M. Gilmartin and N.-H. Chua, The Plant Cell **2,** 447–455 (1990).
[38] E. Lam and N.-H. Chua, Science **248,** 471–474 (1990).
[39] M. Cuozzo-Davis, M.-H. Yong, P. M. Gilmartin, E. Goyvaerts, C. Kuhlemeier, L. Sarokin, and N.-H. Chua, Photochem. Photobiol. **52,** 43–50 (1990).
[40] G. Giuliano, E. Pichersky, V. S. Malik, M. P. Timko, P. A. Scolnik, and A. R. Cashmore, Proc. Natl. Acad. Sci. U.S.A. **85,** 7089–7093 (1988).
[41] N. Datta and A. R. Cashmore, The Plant Cell **1,** 1069–1077 (1989).
[42] C. Kuhlemeier, R. Fluhr, P. J. Green, and N.-H. Chua, Genes Dev. **1,** 247–255 (1987).
[43] C. Kuhlemeier, M. Cuozzo, P. J. Green, E. Goyvaerts, K. Ward, and N.-H. Chua, Proc. Natl. Acad. Sci. U.S.A. **85,** 4662–4666 (1988).
[44] S. A. Kay, B. Keith, K. Shinozaki, M.-L. Chye, and N.-H. Chua, The Plant Cell **1,** 351–360 (1989).
[45] P. Schulze-Lefert, J. L. Dangl, M. Becker-André, K. Hahlbrock, and W. Schulz, EMBO J. **8,** 651–656 (1989).
[46] C. Castresana, I. Garcia-Luque, E. Alonso, V. S. Malik, and A. R. Cashmore, EMBO J. **7,** 1929–1936 (1988).
[47] R. J. Ferl and B. H. Laughner, Plant Mol. Biol. **12,** 357–366 (1989).
[48] W. L. McKendree, A.-L. Paul, A. J. DeLisle, and R. J. Ferl, The Plant Cell **2,** 207–214 (1990).
[49] A. J. DeLisle and R. J. Ferl, The Plant Cell **2,** 547–557 (1990).
[50] D. Staiger, H. Kaulen, and J. Schell, Proc. Natl. Acad. Sci. U.S.A. **86,** 6930–6934 (1989).
[51] P. Schulze-Lefert, M. Becker-André, W. Schulz, K. Hahlbrock, and J. L. Dangl, The Plant Cell **1,** 707–714 (1989).
[52] A. Block, J. L. Dangl, K. Hahlbrock, and P. Schulze-Lefert, Proc. Natl. Acad. Sci. U.S.A. **87,** 5387–5391 (1990).
[53] J. Mundy, K. Shinozaki-Yamaguchi, and N.-H. Chua, Proc. Natl. Acad. Sci. U.S.A. **87,** 1406–1410 (1990).
[54] M. J. Guiltinan, W. R. Marcotte, and R. S. Quatrano, Science **250,** 267–271 (1990).
[55] K. Vogel, W. Hörz, and A. Hinnen, Mol. Cell Biol. **9,** 2050–2057 (1989).
[56] H. Beckmann, L. Su, and T. Kadesch, Genes Dev. **4,** 167–179 (1990).
[57] R. G. K. Donald, U. Schindler, A. Batschauer, and A. R. Cashmore, EMBO J. **9,** 1727–1735 (1990).
[58] T. Ueda, E. Pichersky, V. S. Malik, and A. R. Cashmore, The Plant Cell **1,** 217–227 (1989).
[59] E. Lam, Y. Kano-Murakami, P. Gilmartin, B. Niner, and N.-H. Chua, The Plant Cell **2,** 857–866 (1990).
[60] K. D. Jofuku, J. K. Okamuro, and R. B. Goldberg, Nature **328,** 734–737 (1987).
[61] J. Jordano, C. Almoguera, and T. L. Thomas, The Plant Cell **1,** 855–866 (1989).
[62] R. D. Allen, F. Bernier, P. A. Lessard, and R. N. Beachy, The Plant Cell **1,** 623–631 (1989).
[63] M. M. Bustos, M. J. Guiltinan, J. Jordano, D. Begum, F. A. Kalkan, and T. C. Hall, The Plant Cell **1,** 839–853 (1989).
[64] M. J. Guiltinan, J. C. Thomas, C. L. Nessler, and T. L. Thomas, Plant Mol. Biol. **13,** 605–610 (1989).
[65] C. D. Riggs, T. A. Voelker, and M. J. Chrispeels, The Plant Cell **1,** 609–621 (1989).
[66] P. Hatzopoulos, G. Franz, L. Choy, and R. Z. Sung, The Plant Cell **2,** 457–467 (1990).
[67] J. Deikman and R. L. Fischer, EMBO J. **7,** 3315–3320 (1988).
[68] S. Cordes, J. Deikman, L. J. Margossian, and R. L. Fischer, The Plant Cell **1,** 1025–1034 (1989).
[69] M. J. Holdsworth and G. G. Laties, Planta **179,** 17–23 (1989).
[70] M. J. Holdsworth and G. G. Laties, Planta **180,** 74–81 (1989).
[71] E. O. Jensen, K. A. Marcker, J. Schell, and F. De Bruijn, EMBO J. **7,** 1265–1271 (1988).
[72] F. J. De Bruijn, G. Felix, B. Grunenberg, H. J. Hoffmann, B. Metz, P. Ratet, A. Simons-Schreier, L. Szabados, P. Welters, and J. Schell, Plant Mol. Biol. **13,** 319–325 (1989).
[73] B. A. Metz, P. Welters, H. J. Hoffmann, E. O. Jensen, and J. Schell, Mol. Gen. Genet. **214,** 181–191 (1988).
[74] B. G. Forde, J. Freeman, J. E. Oliver, and M. Pineda, The Plant Cell **2,** 925–939 (1990).
[75] T.-M. Ou-Lee, R. Turgeon, and R. Wu, Proc. Natl. Acad. Sci. U.S.A. **85,** 6366–6369 (1988).
[76] W. Werr, B. Springer, J. Schürmann, and R. Bellmann, Mol. Gen. Genet. **212,** 342–350 (1988).
[77] C. Fenoll, D. M. Black, and S. H. Howell, EMBO J. **7,** 1589–1596 (1988).
[78] J. S. Buzby, T. Yamada, and E. M. Tobin, The Plant Cell **2,** 805–814 (1990).

[79] E. Lam and N.-H. Chua, The Plant Cell **1,** 1147–1156 (1989).
[80] J. T. Odell, F. Nagy, and N.-H. Chua, Nature **313,** 810–812 (1985).
[81] D. W. Ow, J. D. Jacobs, and S. H. Howell, Proc. Natl. Acad. Sci. U.S.A. **84,** 4870–4874 (1987).
[82] P. N. Benfey, L. Ren, and N.-H. Chua, EMBO J. **8,** 2195–2202 (1989).
[83] P. N. Benfey, L. Ren, and N.-H. Chua, EMBO J. **9,** 1677–1684 (1990).
[84] P. N. Benfey, L. Ren, and N.-H. Chua, EMBO J, **9,** 1685–1696 (1990).
[85] S. Prat, L. Willmitzer, and J. Sanchez-Serrano, Mol. Gen. Genet. **217,** 209–214 (1989).
[86] E. Lam, P. N. Benfey, P. M. Gilmartin, R.-X. Fang, and N.-H. Chua, Proc. Natl. Acad. Sci. U.S.A. **86,** 7890–7894 (1989).
[87] K. Mikami, A. Sakamoto, H. Takase, T. Tabata, and T. Iwabuchi, Nucleic Acids Res. **17,** 9707–9717 (1989).
[88] K. Singh, J. G. Tokuhisa, E. S. Dennis, and W. J. Peacock, Proc. Natl. Acad. Sci. U.S.A. **86,** 3733–3737 (1989).
[89] D. Bouchez, J. G. Tokuhisa, D. J. Llewellyn, E. S. Dennis, and J. G. Ellis, EMBO J. **8,** 4197–4204 (1989).
[90] J. G. Tokuhisa, K. Singh, E. S. Dennis, and W. J. Peacock, The Plant Cell **2,** 215–224 (1990).
[91] H. Fromm, F. Katagiri, and N.-H. Chua, The Plant Cell **1,** 977–984 (1989).
[92] K. Singh, E. S. Dennis, J. G. Ellis, D. J. Llewellyn, J. G. Tokuhisa, J. A. Wahleithner, and W. J. Peacock, The Plant Cell **2,** 891–903 (1990).
[93] E. Lam, F. Katagiri, and N.-H. Chua, J. Biol. Chem. **265,** 9909–9913 (1990).
[94] K. Mikami, T. Tabata, T. Kawata, T. Nakayama, and M. Iwabuchi, FEBS L. **233,** 273–278 (1987).
[95] K. Mikami, T. Nakayama, T. Kawata, T. Tabata, and M. Iwabuchi, Plant Cell Physiol. **30,** 107–119 (1989).
[96] K. Mikami, H. Takase, T. Tabata, and M. Iwabuchi, FEBS L. **256,** 67–70 (1989).
[97] F. Katagiri, E. Lam, and N.-H. Chua, Nature **340,** 727–730 (1989).
[98] T. Tabata, H. Takase, S. Takayama, K. Mikami, T. Nakayama, and M. Iwabuchi, Science **245,** 965–967 (1989).
[99] C. R. Vinson, P. B. Sigler, and S. L. McKnight, Science **246,** 911–916 (1989).
[100] K. K. Yamamoto, G. A. Gonzalez, W. H. Biggs III, and M. R. Montminy, Nature **334,** 494–498 (1988).
[101] A. J. Berk, Biochim. Biophys. Acta **1009,** 103–109 (1989).
[102] S. M. M. Iguchi-Ariaga and W. Schaffner, Genes Dev. **3,** 612–619 (1989).
[103] J. Birchler, Trends. Genet. **6,** 231–233 (1990).
[104] J. Paz-Ares, D. Ghosal, U. Wienand, P. A. Peterson, and H. Saedler, EMBO J. **6,** 3553–3558 (1987).
[105] A. Marocco, M. Wissenbach, D. Becker, J. Paz-Ares, H. Saedler, F. Salamini, and W. Rohde, Mol. Gen. Genet. **216,** 183–187 (1989).
[106] S. R. Ludwig, L. F. Habera, S. L. Dellaporta, and S. R. Wessler, Proc. Natl. Acad. Sci. U.S.A. **86,** 7092–7096 (1989).
[107] S. R. Ludwig, B. Bowen, L. Beach, and S. R. Wessler, Science **247,** 449–450 (1990).
[108] S. A. Goff, T. M. Klein, B. A. Roth, M. E. Fromm, K. C. Cone, J. P. Radicella, and V. C. Chandler, EMBO J. **9,** 2517–2522 (1990).
[109] H. Hartings, M. Maddaloni, N. Lazzaroni, N. Di Fonzo, M. Motto, F. Salamini, and R. Thompson, EMBO J. **8,** 2795–2801 (1989).
[110] R. J. Schmidt, F. A. Burr, M. J. Aukerman, and B. Burr, Proc. Natl. Acad. Sci. U.S.A. **87,** 46–50 (1990).
[111] M. F. Yanofsky, H. Ma, J. L. Bowman, G. N. Drews, K. A. Feldmann, and E. M. Meyerowitz, Nature **346,** 35–39 (1990).
[112] H. Sommer, J. P. Beltrán, P. Huijser, H. Pape, W.-E. Lönning, H. Saedler, and Z. Schwarz-Sommer, EMBO J. **9,** 605–613 (1990).
[113] H. Singh, J. H. LeBowitz, A. S. Baldwin jr., and P. A. Sharp, Cell **52,** 415–423 (1988).
[114] C. R. Vinson, K. L. LaMarco, P. F. Johnson, W. H. Landschulz, and S. L. McKnight, Genes Dev. **2,** 801–806 (1988).
[115] A. Gasch, A. Hoffmann, M. Horikoshi, R. G. Roeder, and N.-H. Chua, Nature **346,** 390–394 (1990).
[116] C. Nawrath, J. Schell, and C. Koncz, Mol. Gen. Genet. **223,** 65–75 (1990).
[117] J. Ma, E. Prizibilla, J. Hu, L. Bogorad, and M. Ptashne, Nature **334,** 631–633 (1988).
[118] P. Hilson, D. de Froidmont, C. Lejour, S.-I. Hirai, J.-M. Jacquemin, and M. Yaniv, The Plant Cell **2,** 651–658 (1990).
[119] R. J. Ferl and H. S. Nick, J. Biol. Chem. **262,** 7947–7950 (1987).
[120] R. J. Ferl, Plant Physiol. **93,** 1094–1101 (1990).
[121] C. J. Palm, M. A. Costa, G. An, and C. A. Ryan, Proc. Natl. Acad. Sci. U.S.A. **87,** 603–607 (1990).
[122] D. Staiger, H. Kaulen, and J. Schell, Plant Physiol. **93,** 1347–1353 (1990).
[123] R. A. Dixon, M. J. Harrison, S. M. Jenkins, C. J. Lamb, M. A. Lawton, and L. Yu, UCLA Symp. Mol. Cell. Biol. **129,** 101–109 (1990).
[124] R. Lois, A. Dietrich, K. Hahlbrock, and W. Schulz, EMBO J. **8,** 1641–1648 (1989).
[125] B. Springer, R. Bellmann, and W. Werr, Mol. Gen. Genet. **221,** 94–101 (1990).
[126] U.-G. Maier, J. W. S. Brown, C. Toloczyki, and G. Feix, EMBO J. **6,** 17–22 (1987).
[127] U.-G. Maier, J. W. S. Brown, L. M. Schmitz, M. Schwall, G. Dietrich, and G. Feix, Mol. Gen. Genet. **212,** 241–245 (1988).
[128] M. L. Schmitz, U.-G. Maier, J. W. S. Brown, and G. Feix, J. Biol. Chem. **264,** 1467–1472 (1989).
[129] A. Gierl, S. Lütticke, and H. Saedler, EMBO J. **7,** 4045–4053 (1988).
[130] S. R. Grant, A. Gierl, and H. Saedler, EMBO J. **9,** 2029–2035 (1990).
[131] R. C. Elliott, T. J. Pedersen, B. Fristensky, M. J. White, L. F. Dickey, and W. F. Thompson, The Plant Cell **1,** 681–690 (1989).