

# Plant genes and their expression

Akhilesh K. Tyagi

Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi 110 021, India

---

**The transcription of protein encoding genes and their regulation in plant cells shows the signature of a typical eukaryotic system. The major outcome of research in this area is recognition of variability in components assembled to design constitutive or inducible promoters, including those constituting enhancers and boundary elements. Such *cis*-acting elements respond to diverse intrinsic and extrinsic signals to take care of the needs of sessile plants. While much remains to be defined about the basic transcription machinery in plants, it is expected to follow the pattern as deciphered in animals. At the same time, a large number of genes for regulatory proteins have been cloned leading to information about DNA binding domains, activation domains, nuclear localization signals and oligomerization sites. However, most of this information needs to be tested by experimentation. Availability of transgenic systems in plants and biotechnological needs have given impetus to research for designing tailor-made inducible/repressible promoters and to manipulate plant processes by deploying transcription factors. While considerable progress has been made, much remains to be learned about the biochemical and molecular basis of regulation of plant gene expression.**

---

THE drama of life is enacted by 'genes' which collectively act as a vast repository of information and 'elements' which regulate their activity. It is in the early 1960s, that the basic principles of gene regulation in prokaryotes were established<sup>1</sup>, leading some years later, to a theory of regulation of eukaryotic genes that evolved in the late sixties<sup>2</sup>. While in prokaryotes, the units of gene regulation were defined as promoters, operators and positive control elements, the analysis of eukaryotic promoters has to wait until the advent of recombinant DNA technology in 1970s which then led to the recognition of both positive and negative regulatory elements (enhancers and silencers) and in addition boundary elements. All these elements interact with proteins (repressors or activators) and RNA polymerase(s) to establish the dynamic state of regulation of gene expression. Notwithstanding the universal nature of components and basic principles, some fundamental differences have emerged in prokaryotic and eukaryotic gene regulation<sup>3</sup>. While in prokaryotes the ground state is nonrestrictive and regulation is mostly approachable

by repression, in eukaryotes the ground state is restrictive due to packaging of DNA in chromatin and, therefore, regulation has to be accessed mainly by activation. Despite the gross level differences, the mechanism involved in gene expression reflects the same pattern, of course, with increasing diversity and complexity.

Plants, by and large, also seem to follow typical eukaryotic norms for gene expression and regulation. Beginning in the late seventies, plant genes have been characterized in great detail. The sequencing of the complete genome of *Arabidopsis* is nearing completion and that of the rice genome is also progressing rapidly<sup>4</sup>. Based on information from these studies and also data on non-redundant ESTs, it is estimated that plants may contain approximately 25,000 to 50,000 genes. It has also been estimated that about one-third of these genes express in all the organs, albeit at variable levels. Another one-third may have their products present in a few, but not all, organs and the rest may belong to one or other unique organ as far as expression is concerned. This variability is reflected in the organization of promoters and regulatory elements as well as in genes characterized for the regulating factors. In this article, an effort has been made to give a bird's eye view of gene expression in plants. The information is restricted to the genes producing only mRNAs. Finally, emerging global trends in plant gene expression and the utility of regulated gene expression for biotechnological purposes has been taken up briefly. For obvious reasons, it is not possible to discuss all aspects and readers are advised to consult other references for details<sup>5-9</sup>.

## Basic mechanism of regulation of gene expression

Core promoters of eukaryotic protein-coding genes are defined as possessing TATA box, an initiator element as well as a transcription initiation site<sup>8</sup>. RNA polymerase II and general transcription factors, GTFs (TBP, TFIIB, TFIIE, TFIIIF, TFIIH) assemble at the core promoter and carry out accurate transcription *in vitro*. Other than the core elements, promoters also contain enhancer/regulatory elements which interact with activators in a sequence-specific manner. Since enhancers can act at a distance in both an orientation and position independent manner and chromatin organization can sequester distant regions together, it is necessary to de-

---

e-mail: aktpmb@hotmail.com

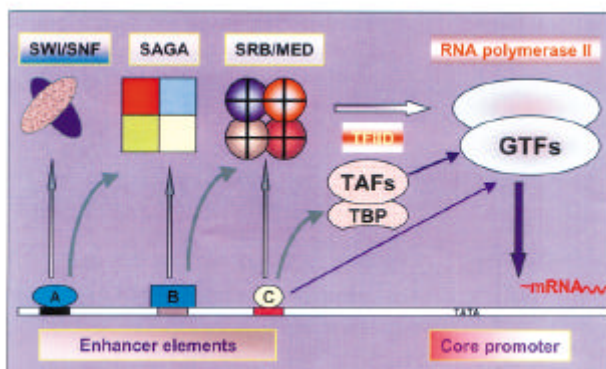
fine the boundary of action of enhancers. It has been found that certain sequences like the scaffold attachment regions (SARs), boundary elements (BEs) and locus control regions (LCRs) and their interactions with proteins may help delimit functional domains of chromatin<sup>10</sup>.

Another class of proteins, called coactivators or mediators, helps establish communication between activators and the RNA polymerase II machinery<sup>9</sup>. Certain coactivators like SWI/SNF act as chromatin remodelling factors. Others may have histone acetyltransferase activity (e.g. SAGA). It has been shown in yeast that a sequence-specific transcription factor SWI5 recruits SWI/SNF, which in turn recruits SAGA. Both are required to recruit the sequence-specific factor, SWI4-SWI6, which in turn may help assemble the pre-initiation complex to activate the *HO* gene (coding for an endonuclease). Several other coactivator complexes in yeast (e.g. SRB/MED) and their homologues in mammalian cells have been shown to mediate transcription through more than one activator proteins. TAFs (TBP-associated factors in TFIID) were the first to be identified as coactivators. But, they seem to perform such a role mostly in metazoan systems and not in yeast. In mammalian systems, however, TAFs seem to act in concert with other coactivators. It is possible that such observations reflect redundancy or division of labour among coactivators. This is also supported by shared subunits among different coactivator complexes<sup>11</sup>.

Turning to the mechanism of action, the coactivators mediate between activators and transcription machinery. The mechanistic steps in the process may include remodelling of chromatin, recruitment of additional activators, and/or direct interaction with RNA polymerase II or GTFs (Figure 1). Several coactivators (SRB/MED, SWI/SNF) have been found to be components of RNA polymerase II holoenzyme, raising the possibility of transcriptional activation by recruitment<sup>12</sup>. This is also supported by evidence for the presence of a giant complex 'transcriptosome' which may include RNA polymerase II, some GTFs and coactivators<sup>13</sup>. One may, therefore, visualize an immobile 'transcriptosome' and DNA moving through it for transcription and beyond<sup>14</sup>!

### Plant promoters and regulatory elements

In plants, the modern era of work on genes and their regulation began after the discovery of restriction endonucleases, advent of recombinant DNA revolution and arrival of DNA sequencing technology. These enabled the isolation and sequencing of first genes such as those coding for large and small subunit of Rubisco (by groups of L. Bogorad, R. B. Meagher and N.-H. Chua), of leghaemoglobin (by D. P. S. Verma and K. A. Marcker)



**Figure 1.** Transcriptional activation involves core promoter and enhancer elements. The core promoter interacts with TBP and the enhancer elements interact with activator proteins A, B and C. Activator proteins interact with coactivators (SWI/SNF, SAGA, SRB/MED, TAFs), general transcription factors (TFIID, GTFs) and RNA polymerase II (after Wu and Hampsey<sup>9</sup>).

or the seed storage proteins (by R. Goldberg and T. Hall). Because of the uniqueness of the process of photosynthesis in plants and early lead with genes encoding components of thylakoids and stroma, genes for such proteins have continued to attract attention of leading groups (like R. G. Herrmann, J. C. Gray, M. Sugiura, N.-H. Chua and T. Cashmore). Recently, emphasis has shifted to characterize components of regulatory system and transcription machinery.

The understanding of regulation of plant gene expression at *cis*-acting elements started with analysis of promoters from native plant genes and genes from viruses and *Agrobacterium* which express in plant cells<sup>15</sup>. It was found that, like their counterparts in animals, plant gene promoters contain TATA box, initiator element as well as well-defined transcription initiation site about 20 to 30 bp downstream to TATA element. In the upstream region, several gene promoters were found to contain positive or negative regulatory elements, some of which were characterized as enhancers or silencers<sup>16</sup>. The gain-of-function assay mostly served as evidence for the activity of an element. Thus, a minimal promoter (mostly -46 bp derivative of 35S CaMV promoter or its -90 bp derivative) is fused to a heterologous promoter sequence at its 5' end and to a reporter gene (*gus*, *gfp*, *luc*, *cat*) at its 3' end. Such constructs are introduced in plant cells by *Agrobacterium*, electroporation or biolistics and evaluated either for transient expression of the reporter gene or for its expression in stable transgenic system. The pattern of expression conferred by heterologous promoter sequences has been helpful in defining characteristics of developmental (spatial and temporal) as well as inducible regulatory promoters<sup>5,17</sup>. A few major findings from such experiments are discussed below.

*Constitutive promoters*

A few promoters, found to express in most of the cells of plants albeit at variable levels, have been termed constitutive promoters. These include promoters like 35S from CaMV, the octopine synthase, nopaline synthase and, 1'2' mannopine synthase gene promoters from *Agrobacterium*, as well as promoters from actin 1 and ubiquitin genes<sup>18-22</sup>. Investigations on the 35S promoter have served as paradigm for the analysis of activity of constitutive promoters<sup>23,24</sup>. The 35S promoter has been divided into two major domains, A and B. Domain A contains the minimal promoter (+8 to -46 bp) and A1 domain (-46 to -90 bp) characterized by the presence of *as-1* sequence. The B domain (-90 to -343 bp) has been defined as an enhancer and further divided into five subdomains, B1-B5. A combination of different B subdomains with minimal promoter or domain A revealed the modular organization of the enhancer element. Thus, different combinations gave distinct expression patterns of reporter gene activity in early and mature stages of plant development, unravelling synergistic interactions among various elements within the 35S enhancer, which ultimately result in constitutive expression under control of the full promoter.

Due to variable and weak expression of different constitutive promoters in certain tissues in transgenic plants, one can combine various modules to obtain appropriate constitutive expression. Thus, use of double 35S enhancer resulted in about 6-fold higher expression compared to single enhancer. Another 6-fold increase in expression became possible by combining +65 to -301 bp of mannopine synthase gene promoter and -90 to -941 bp of 35S promoter<sup>25</sup>. A combination of a trimer of upstream activating sequences of octopine synthase promoter and mannopine synthase promoter led to 156-fold improvement over 35S promoter<sup>26</sup> and expression was found to be more uniform.

*State-specific promoters*

The life cycle of a plant includes both gametophytic and sporophytic phases of development, each characterized by a unique spatial and temporal programme. In addition, due to the sessile nature, plants have to respond rapidly to environmental signals. Sometimes, plants adopt signals like light to control their developmental programme. All this requires a complex array of regulatory elements responding to signals like light, heat, cold, hypoxia, dehydration, hormones, wounding or to cues emanating from the pattern of plant development and differentiation. Characterization of several state-specific genes has been followed by analysis of their promoters to determine the *cis*-acting elements involved in regulated expression of genes<sup>5</sup>. Some of these are

being discussed below to illustrate the nature of elements and interactions involved.

*Organ-specific promoters:* Such promoters confer expression specific to an organ and show a well-defined temporal expression pattern. Organ-specific promoters have been isolated from genes for seed storage proteins, lectins, trypsin inhibitors, etc. It has been found that in most of the cases the first 200 bp are able to confer specificity of expression. However, distal regions may also have redundant quality control motifs or quantitative elements. In dicots, some elements, like the three tandem repeats of the 22 bp element containing TGACG motif, can confer on the 35S minimal promoter (-46 bp) seed-specific expression in a heterologous transgenic system<sup>27</sup>. Some elements present in seed-specific promoters like the G box (CACGTG), E box (CACCTG), legumin or RY motif work in combination with other elements to confer quantitatively high seed-specific expression. Other elements from the distal region have also been identified to work as quantitative elements. These include the AT-rich motifs acting as positive elements or the AGAA (A/C)A and CA-repeat motifs acting as negative elements.

In monocots also, seed-specific promoters and regulatory elements have been identified first in heterologous transgenic dicot system and more recently in homologous monocot systems. Promoters for glutelin genes from rice have been found to contain the prolamin box [TG(t/a/c)AAA(g/t)], GCN4 motif (5'-TGAGTCA-3'), ACGT motif and AACA motif (5'-AACAACTCAATC-3'). A 45 bp *GluA-3* promoter fragment contains a GCN4 motif acting as a positive regulatory element while the AACA motif acts as negative regulatory element in tissues other than endosperm<sup>28</sup>. Deletion of these motifs in *GluB-1* promoter results in loss of seed-specificity and a multimer of 21 bp containing GCN4 can confer endosperm-specific expression in transgenic rice on -46 bp 35S promoter<sup>29</sup>. Further, it has been found that several of these elements are present in multiples and show functional redundancy. It may, however, be noted that the other three elements are not able to direct expression even in multiple copies and a combinatorial interplay may be responsible for optimal expression<sup>30</sup>. Another class of promoters specific to seeds includes those controlling expression of zein in maize. The most important motif identified is called the opaque-2 binding site (5'-TTGACGTGG-3') which controls the expression of genes in developing endosperm. This site acts in conjunction with the prolamin box in establishing expression of 22 kDa zein protein<sup>6</sup>.

*Environment-responsive promoters:* Extrinsic cues like light play an important role in the life of a plant. Light-regulated promoters have been found to be present in a large number of genes encoding proteins related to pho-

tosynthesis, metabolic enzymes and developmental processes from seed germination to flowering<sup>5,31</sup>. In the life of a plant, light is a very important signal and it manifests through multiple forms of photoreceptors and a complex signal transduction cascade<sup>32</sup>. Among the light-regulated promoters, light-regulated elements of varying length have been found to act as positive regulatory elements or negative regulatory elements and show a high degree of variability and redundancy. In *RbcS* genes, the first few hundred base pairs of the promoters have been found to be sufficient for light regulation<sup>33,34</sup> and, in some cases, such elements worked as 'enhancer-like elements'<sup>16</sup>. Further work has delineated elements responding to specific photoreceptors. Phytochrome-responsive elements range from 10 to 40 bp in case of *Cab* genes. A fragment of *Cab* gene promoter (-176 to -36 bp) from tobacco could confer VLFR, LFR and HIR of phytochrome on the minimal CaMV 35S promoter<sup>35</sup>. Similar examples have been reported for blue-light-responsive elements.

An analysis of light-regulated promoter sequences has predicted 30 distinct conserved DNA module arrays<sup>36</sup>. These include GT-boxes, I-boxes, G-boxes and AT-rich sequences<sup>31</sup>. Despite the fact that each of these elements has been shown to influence the response of a promoter to light, no single element has been shown to confer light-responsiveness on a minimal promoter. While a tetramer of GT-1 site could confer light regulation on CaMV 35S (-90 bp) promoter containing *as-1* element, it failed to do so with CaMV 35S (-46 bp) promoter<sup>37</sup>. Analysis of several other promoters shows randomness about the presence of identified elements<sup>38,39</sup>. Some of these elements like G-box have been found to be present also in non-light-responsive promoters and others like the H-box from chalcone synthase gene may respond to light as well as an elicitor. Such results point towards the combinatorial nature of light-regulated promoters involving multiple elements and proteins binding to them. In an effort to decipher the nature of combinatorial action, activity of synthetic promoters containing paired or single LREs in wild-type and light-response mutants of *Arabidopsis* has been investigated<sup>40</sup>. It has been observed that only paired LREs are able to respond to a wide spectrum of light, while single LREs primarily respond to a specific wavelength of light.

### Chemically inducible promoters

Such promoters may be either native plant promoters which are induced by chemicals like nitrate, hormones or elicitors of wounding/pathogen attack responses or designer promoters involving inducible gene systems from heterologous organisms<sup>41,42</sup>. Inducible promoters are highly desirable for both basic and applied research

and should be able to show substantial difference between active and inactive states. In basic research, analysis of gene activity without interference in plant processes is possible, in addition to direct correlation between gene activity and phenotype, without the influence of secondary effects and homeostasis. Such promoters can also be deployed to engineer target-specific traits like conditional male-sterility and resistance to pests or pathogens.

The plant promoter from the *PR-1a* gene has been shown to be inducible by application of benzothiadiazole or isonicotinic acid. Certain agrochemicals are used to increase tolerance of plants to herbicides. Such a plant safener-inducible promoter, *In 2-2*, has been tested in *Arabidopsis*<sup>43</sup>. In addition, several other regulatory systems of plants can be tailored to provide chemically inducible systems for plants<sup>42</sup>.

Regulatory elements from heterologous systems may be unique and need not respond to chemicals commonly capable of eliciting a response in plants. Such promoters can be made to act in plants if *trans*-acting factors interacting with promoters in a chemically inducible manner are also transferred to plant systems.

In an effort to tailor a tetracycline-inducible promoter for plants, the bacterial repressor protein (TetR) gene was introduced into plants under the control of CaMV 35S promoter. This binds to *tet* operator in the absence to tetracyclines. The target promoter was engineered to contain *tet* operator sites flanking TATA box in CaMV 35S promoter. Application of tetracycline abolished interaction of *tet* operator-repressor, thereby resulting in 500- to 800-fold induction<sup>44</sup>. Alternatively, a tetracycline-inactivatable system made use of TetR and the acidic domain of *Herpes simplex* protein 16 (VP16) fusion (tTA) expression under the control of CaMV 35S promoter. The tTA interacted with the synthetic promoter containing several *tet* operators upstream of TATA box and resulted in high expression of the reporter gene. Application of tetracycline abolished tTA-promoter interaction and expression effectively<sup>44,45</sup>. Similarly, steroid-inducible promoters made use of the glucocorticoid receptor binding domain fused to plant transcription factors and its binding sites to yield a system which could be activated by application of the glucocorticoid dexamethasone<sup>46</sup>. A further advance was made by fusing the glucocorticoid receptor ligand-binding domain to DNA-binding domain of yeast GAL4 and the VP16 activation domain and targeting it to a promoter containing six GAL4 binding sites<sup>47</sup>. In another development, a certified agrochemical RH5992 inducible system makes use of ligand-binding domain of the ecdysone receptor from insects<sup>48</sup>. A novel inducible system makes use of control elements from copper-inducible genes in yeast and transcription factor ACE1. The chimeric promoter containing CaMV 35S minimal promoter could be induced 50-fold on application of

copper<sup>49</sup>. More recently, an alcohol-inducible system has made use of a regulator element from *alcA* promoter of *Aspergillus nidulans* in combination with minimal CaMV 35S promoter and the AlcR transcription factor<sup>50</sup>. The system has been reported to work on application of 0.1% alcohol in hydroponic culture and by soil drenching with 1% alcohol. It is hoped that such efforts would lead to efficient inducible and inactivatable systems which could be further made target-specific in combination with organ-specific or temporal elements from plant systems.

### Plant RNA polymerases and basal transcription factors

In plants, early efforts related to biochemical and immunological characterization of RNA polymerases, which like their counterparts in animals, turned out to be multimeric enzyme complexes<sup>51-54</sup>. Interestingly, in monocots snRNAs are transcribed by both RNA polymerase II and III, rather than the latter alone, and their genes contain TATA box at -30 bp and an upstream element. The distance of the upstream element to TATA box determines the specificity, i.e. 3 or 4 helical turns for RNA polymerase II or III, respectively<sup>55</sup>. In 1990, genes for the largest subunit of RNA polymerase II were cloned from *Arabidopsis* and soybean<sup>56,57</sup>. The polypeptide is encoded by a single copy gene in *Arabidopsis* and from a multigene family in soybean. The deduced amino acid sequence shows eight regions of similarity with its counterparts and the order of Zn<sup>2+</sup>-finger, the  $\alpha$ -amanitin binding motif and CTD (with 41 repeats) is conserved. Recently, a detailed analysis of the complex of DNA and a RNA polymerase I holoenzyme activity from *Brassica* has revealed that rRNA gene promoters can interact with the holoenzyme activity in a single DNA-binding event<sup>58</sup>. Such an evidence suggests that RNA polymerase I holoenzyme exists in a form that is self-sufficient for its activity on rRNA gene promoters. It correlates with the view that RNA polymerase holoenzymes are recruited rather than assembled<sup>12</sup>.

To function effectively, plant RNA polymerases also require basal transcription factors. TFIID is considered important for the recruitment of RNA polymerase II. Interestingly, its crucial component, TBP, is important not only for RNA polymerase II but also for two other RNA polymerases. The gene for TBP from *Arabidopsis* was amongst the first to be cloned from a two-gene family<sup>59</sup>, followed by that from potato and maize. The gene codes for a polypeptide of about 180 amino acid residues, showing conservation with the C-terminal regions of polypeptides from other sources. Further, co-crystallization of *Arabidopsis* TBP and a TATA box sequence from the adenovirus major late promoter shows that TBP binds to the minor groove of DNA,

causing major distortions in the contacted region<sup>60</sup>. The polypeptide takes the form of a saddle-shaped structure from the organization of the ten-stranded anti-parallel **b**-sheets and four **a**-helices lying on its upper surface. On either side of the saddle are present anti-parallel pairs of **b**-strands giving a stirrup-like structure<sup>60</sup>. TBP interacts with several other basal transcription factors, coactivators and activators to regulate gene expression. Evidence for the presence of such coactivators and TBP-associated factors (TAFs) in wheat has been provided recently<sup>61</sup>. In view of the availability of cDNAs for TFIIB from plants<sup>62</sup>, such an interaction for TBP-TFIIB was evaluated in maize cells and amino acids from the 'stirrups' (E-144, E-146) were found to be essential for TFIIB binding *in vitro*. Such interactions were dispensable for basal transcription and also in case of constitutive promoters (*35S*, *Ubiquitin* from maize), but they help significantly in regulated expression<sup>63</sup>.

TFIIA from wheat has been characterized and found to be similar to the human factor in action with both plant and animal viral promoters<sup>64</sup>. Recently, *Arabidopsis* has been shown to contain two genes for the large subunit and one for the small subunit of TFIIA<sup>65</sup>. The large subunit interacts with the small subunit via its N-terminal part and TFIIA also interacts with TBP-DNA complex. The TFIIA is functionally capable of stimulating transcription in yeast and plant cells and the activity is localized in the evolutionarily unconserved central region of the polypeptide which could be delimited to a 35-amino acid segment.

One major limiting factor in the progress of work on RNA polymerases in plants is the lack of an appropriate and efficient *in vitro* transcription system, despite several efforts<sup>66</sup>. Although a system for transcription of RNA polymerase I-dependent genes shows species specificity, its efficiency is very poor, ~0.001 compared to 1.4-4 per template for HeLa cell extracts. For RNA polymerase II, an efficient and faithful system is highly desirable.

### Regulatory proteins

Besides RNA polymerases and basal transcription factors, regulated expression of genes also involves several regulatory proteins. An estimate based on the genome sequence of *Arabidopsis* has revealed that 15% proteins in plants may be involved in transcription<sup>67</sup>. Some regulatory proteins have been identified that are capable of remodelling chromatin architecture which in turn facilitates subsequent steps of transcription. Histone acetylation is a common event in this direction and can lead to unfolding of the nucleosome and transcription also in plants. Similarly, the activity of histone deacetylases, also found to be present in plants, may have the reverse effect. Recently, genes for histone deacetylases have

been isolated from *Arabidopsis* and shown to be important in plant gene expression and development<sup>68</sup>. Other regulatory proteins bind to *cis*-acting elements present in the promoter and act directly on the transcription machinery or through protein–protein interaction via coactivators. A large number of such proteins and their genes have been identified by genetic/mutant studies, affinity of proteins to *cis*-acting elements or simply by homology to already known proteins<sup>7</sup>. Their functional activity in terms of DNA binding or transcription activation has been analysed by *in vitro/in vivo* assays, including transgenic systems. In addition, transcription factors are known to have nuclear localization signals, oligomerization domains as also other domains of still unspecified functions (Figure 2).

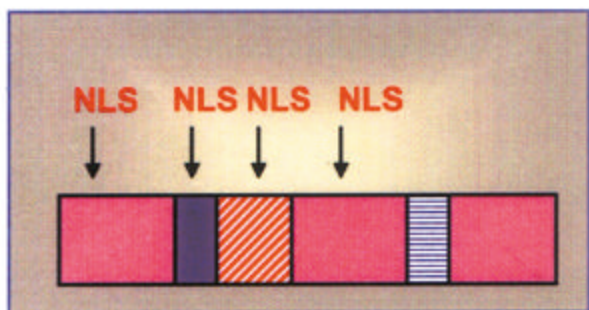
Most of the plant regulatory factors have been classified on the basis of conserved domains which also show similarity to conserved domains of other organisms. The basic-region leucine-zipper (bZIP) proteins contain a basic amino acid-rich region and a leucine-rich region capable of forming a zipper during oligomerization. In plants, genes for TGA1 factor and O2 factor were isolated by screening cDNA libraries with binding oligonucleotides<sup>69</sup> and by mutant analysis<sup>70</sup>, respectively. Another group of proteins, bHLH, containing a cluster of basic amino acids along with a helix–loop–helix is represented by the *R/B* family of maize and also acts by oligomerization<sup>71</sup>. The zinc finger motif is organized around a zinc ion with the help of cysteine and/or histidine residues and it may have different configurations identified as C<sub>2</sub>H<sub>2</sub>, C<sub>3</sub>H, C<sub>2</sub>C<sub>2</sub>, RING finger and LIM finger<sup>72</sup>. The homeodomain contains about 60 amino acids organized into three or four *a*-helices of which the terminal two helices form a helix–turn–helix (HTH) DNA-binding motif<sup>73</sup>. Analysis of *C1* gene from maize revealed presence of MYB-like proteins in plants<sup>74</sup>. Subsequently, more than 150 plant MYB-like proteins have been isolated which contain only one or two HTH motifs, instead of three such motifs found in animals. It is assumed that animal and plant MYB-like proteins

containing three motifs existed before divergence of the two groups and subsequently MYB-like proteins with one or two motifs evolved to predominantly regulate plant-specific processes<sup>75</sup>. A unique trihelix DNA-binding motif containing basic, acidic and proline/glutamine-rich regions has been identified in GT box-binding proteins<sup>76</sup>.

Two proteins involved in flower development, AGAMOUS and DEFICIENS were found to have a conserved domain of about 57 amino acids forming a long *a*-helix and two *b*-strands similar to MCM1 from yeast and human SRF<sup>77,78</sup>. This domain is known as the MADS box and several MADS box-containing regulatory proteins have been identified from plants. In addition, transcriptional proteins from plants are known to contain conserved motifs like AT-hook motif, HMG-box, AP2/EREBP, B3, ARF<sup>7</sup>. It may be mentioned that instead of an overall amino acid sequence conservation, various motifs show more conservation in secondary structure and location of specific amino acids.

The regulatory proteins have been shown to contain activation and repression domains which are effective across taxonomic borders. Variability in the regulatory domains may allow proteins with similar DNA-binding domains to act in a unique manner. Yeast GAL4 and *Herpes simplex* virus VP16 activation domains have been shown to be functional in plants<sup>79,80</sup>. Further, acidic domain of maize C1 protein<sup>80</sup> and proline-rich region of GBF1 (ref. 81) also represent typical motifs known to be involved in regulatory domains of eukaryotic transcription factors. While such activation domains are variable in primary sequence, the GCB motif of HBP-1a/GBF shows a high degree of conservation in consensus sequence<sup>82</sup>. On the other hand, the role of unique amino acids at selective positions in the activation domain<sup>83</sup> or its modular structure<sup>84</sup> has also been revealed. Evidence also exists for the presence of repressor domains since N-terminal domain of ROM2 (regulator of maturation-specific protein) has been shown to inhibit PvAF-activated transcription of phytohemagglutinin gene *DLEC2* from French bean<sup>85</sup>. Precious little is known about the target of regulatory proteins. Recently, GT-1 has been shown to interact with TFIIA–TBP–TATA complex<sup>86</sup>. It is obvious that much remains to be learned about the nature of regulatory domains, particularly in plants, and their sites of action.

In several instances, the quantity and availability of regulatory proteins may depend on their own expression patterns. Such controls may be exerted at transcriptional, post-transcriptional or translational levels<sup>7</sup>. The translated products should be able to reach the nucleus in order to be effective in transcription. Translocation of several regulatory factors is known to be activated in a stimuli-dependent manner<sup>87</sup>. Plant regulatory factors are also known to move from cell to cell via plasmodesmata to exert their effects<sup>88</sup>.

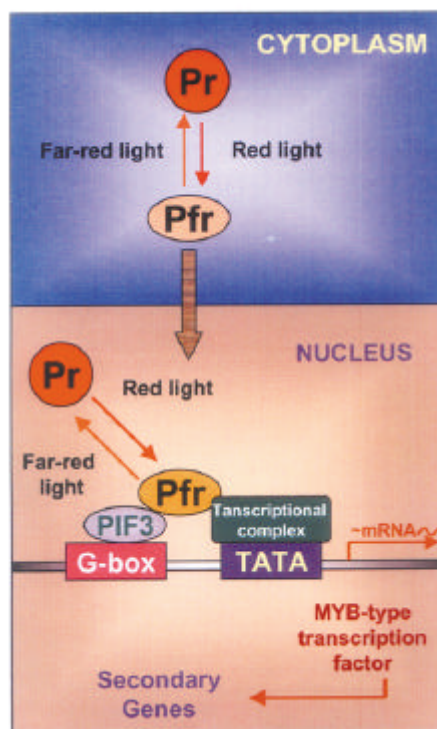


**Figure 2.** Transcriptional factors have various functional domains (violet, DNA-binding domain; hatched box, oligomerization domain; ladder box, regulation domain and NLS, nuclear localization signals) as well as regions (pink) of undefined function (after Liu *et al.*<sup>7</sup>).

In recent years, phosphorylation of regulatory proteins has emerged as a major event controlling the gene regulation in eukaryotes. Phosphorylation of plant regulatory factors may cause stimulation of binding/activation<sup>89,90</sup> or inactivation<sup>91,92</sup> of target genes. So far, only casein kinase II and serine kinases have been shown to be involved, although others may also be active. Further, it has been shown that phosphorylation may be effected by different cues like calcium<sup>92,93</sup> or circadian-clock<sup>91</sup>.

Multiplicity and redundancy of regulatory factors and their recognition sites raises the obvious 'Jig-saw puzzle' scenario and one wonders about the fundamental rules which establish regulation of gene expression. It is possible that concentration of a regulatory factor and its avidity to cognate DNA sequence establish a unique microenvironment for specific interactions. The combinatorial control by multiple DNA-binding regulatory proteins or association with coactivators can also provide specificity in action<sup>6,40</sup>. An interesting example of combinatorial light-dependent regulation has been revealed by the work of Martinez-Garcia *et al.*<sup>94</sup>. It has been shown that a helix-loop-helix transcription factor PIF3 binds to G-box motif of light-regulated MYB-type transcription factor. The photoreceptor, phytochrome binds to DNA-PIF3 complex photoreversibly, in its biologically active conformation. Since phytochrome is also known to have integral kinase activity, it is possible that phytochrome works as a coactivator to communicate with the transcriptional complex and to activate red-light-dependent expression (Figure 3). It is not beyond comprehension that light-activated MYB transcription factor can activate secondary genes in a light-dependent manner.

Since activity of regulatory factors controls downstream genes responsible to confer a particular phenotype on a plant or its response to extrinsic influences, it would be desirable to engineer regulatory factor genes for crop improvement<sup>95</sup>. Several primary and secondary metabolic pathways can be controlled by engineering key regulatory factors affecting transcription of genes for enzymes involved<sup>96,97</sup>. Already, expression of *LEAFY* under the control of CaMV 35S promoter has been shown to result in precocious flowering in aspen which takes 8–10 years to flower<sup>98</sup> and a new gene encoding a repressor of floral transition has been isolated recently<sup>99</sup>. Thus, an important trait like flowering time can be manipulated genetically. Manipulation of floral organs is possible by altering expression of homeotic genes in plants. One wonders if *Crocus* flowers can be made to produce more stigmatic tissue to obtain precious saffron or in rice the number of grains per spikelet could be increased by manipulating genes like *ZMM*<sup>100</sup>. Even seed production has been shown to increase by over expression of *HAT4*, a homeodomain zipper protein<sup>101</sup>. Dwarf varieties can be produced by manipulating



**Figure 3.** The photoreceptor, phytochrome, gets converted to Pfr form on absorbing red light and moves from the cytosol to the nucleus. In the nucleus, it interacts with G-box bound PIF3 (a basic helix-loop-helix transcription factor) on a light-regulated promoter. This results in activation of genes encoding MYB-type transcription factors (*CCA1*, *LHY*) which in turn can activate secondary light-regulated genes. Exposure to far-red light converts Pfr form to Pr form, which gets released from the transcriptional complex and shuts down gene expression (after Martinez-Garcia *et al.*<sup>94</sup>).

regulatory factors responding to the growth hormone gibberellin<sup>102</sup>. Recently, genes for regulatory factors have also been deployed to confer abiotic stress tolerance, with a view to stimulate expression of several downstream genes involved in conferring such a complex phenotype<sup>103,104</sup>. It, therefore, seems that genetic manipulation by deploying regulatory factor genes is becoming a reality.

### Conclusions and prospects

In last two decades, several promoters active in cells of dicot and monocot plants have been identified which are regulated by intrinsic and extrinsic stimuli. Functional genomics of *Arabidopsis* and rice, by way of promoter or enhancer traps, is likely to increase the repertoire of such elements to a great extent<sup>105,106</sup>. Simultaneous work on regulatory factors has led to identification of major structural features, including DNA binding and activation/repression domains. While the general theme of regulation across eukaryotes is represented by activation of genes, the signals and components for plants

might be specific, particularly due to the uniqueness of signals like light or plant hormones. Plants may also use mechanisms like genomic imprinting to regulate activity of genes at a genome-wide level as has been recently shown by the example of delayed activation of the paternal genome after fertilization in *Arabidopsis*<sup>107</sup>. It has been shown already that target-specific regulatory elements can be deployed for engineering traits like male-sterility. Genes for regulatory proteins seem to hold greater promise for crop improvement, particularly for traits involving multiple downstream genes. With the emerging scenario in the area of transgenics and genomics, there is need to focus on regulation biology, both for the sake of its novelty as well as application.

1. Jacob, F. and Monod, J., *J. Mol. Biol.*, 1961, **3**, 318–356.
2. Britten, R. J. and Davidson, E. H., *Science*, 1969, **165**, 349–357.
3. Struhl, K., *Cell*, 1999, **98**, 1–4.
4. Meinke, D. and Tanksley, S., *Curr. Opin. Plant Biol.*, 2000, **3**, 95–96.
5. Guilfoyle, T. J., in *Genetic Engineering* (ed. Setlow, J. K.), Plenum Press, New York, 1997, vol. 19, pp. 15–47.
6. Singh, K. B., *Plant Physiol.*, 1998, **118**, 1111–1120.
7. Liu, L., White, M. J. and MacRae, T. H., *Eur. J. Biochem.*, 1999, **262**, 247–257.
8. Weinzierl, R. O. J., *Mechanisms of Gene Expression*, Imperial College Press, London, 1999.
9. Wu, W. and Hampsey, M., *Curr. Biol.*, 1999, **9**, R606–R609.
10. Udvardy, A., *EMBO J.*, 1999, **18**, 1–8.
11. Hampsey, M. and Reinberg, D., *Curr. Opin. Genet. Develop.*, 1999, **9**, 132–139.
12. Ptashne, M. and Gann, A., *Nature*, 1997, **386**, 569–577.
13. Halle, J. P. and Meisterernest, M., *Trends Genet.*, 1996, **12**, 161–163.
14. Nayler, O. *et al.*, *Nucleic Acids Res.*, 1998, **26**, 3542–3549.
15. Okamura, J. K. and Goldberg, R. B., in *The Biochemistry of Plants: A Comprehensive Treatise* (ed. Marcus, A.), Academic Press, Inc., San Diego, 1989, vol. 15, pp. 1–82.
16. Timko, M. P. *et al.*, *Nature*, 1985, **318**, 579–582.
17. Tyagi, A. K., Mohanty, A., Bajaj, S., Chaudhury, A. and Maheshwari, S. C., *Crit. Rev. Biotechnol.*, 1999, **19**, 41–79.
18. Ellis, J. G., Lewellyn, D. J., Walker, J. C., Dennis, E. S. and Peacock, W. J., *EMBO J.*, 1987, **6**, 3203–3208.
19. Leung, J., Fukuda, H., Wing, D., Schell, J. and Masterson, R., *Mol. Gen. Genet.*, 1991, **230**, 463–474.
20. Zhang, W., McElroy, D. and Wu, R., *Plant Cell*, 1991, **3**, 1155–1165.
21. Toki, S. *et al.*, *Plant Physiol.*, 1992, **100**, 1503–1507.
22. Salgueiro, S., Pignocchi, C. and Parry, M. A. J., *Plant Mol. Biol.*, 2000, **42**, 615–622.
23. Benfy, P. N., Ren, L. and Chua, N.-H., *EMBO J.*, 1990, **9**, 1677–1684.
24. Benfy, P. N., Ren, L. and Chua, N.-H., *EMBO J.*, 1990, **9**, 1685–1696.
25. Comai, L., Moran, P. and Maslyar, D., *Plant Mol. Biol.*, 1990, **15**, 373–381.
26. Ni, M., Cui, D., Einstein, J., Narasimhulu, S., Vergara, C. E. and Gelvin, S. B., *Plant J.*, 1995, **7**, 661–676.
27. de Pater, S., Pham, K., Chua, N.-H., Memelink, J. and Kijne, J., *Plant Cell*, 1993, **5**, 877–886.
28. Yoshihara, T., Washida, H. and Takaiwa, F., *FEBS Lett.*, 1996, **383**, 213–218.
29. Wu, C., Suzuki, A., Washida, H. and Takaiwa, F., *Plant J.*, 1998, **14**, 673–683.
30. Wu, C., Washida, H., Onodera, Y., Harada, K. and Takaiwa, F., *Plant J.*, 2000, **23**, 415–421.
31. Tyagi, A. K., Dhingra, A. and Raghuvanshi, S., in *Probing Photosynthesis: Mechanism, Regulation and Adaptation* (eds Yunus, M., Pathre, U. and Mohanty, P.), Taylor and Francis Publishers Ltd, London, 2000, pp. 324–341.
32. Khurana, J. P., Kochhar, A. and Tyagi, A. K., *Crit. Rev. Plant Sci.*, 1998, **17**, 465–539.
33. Morelli, G., Nagy, F., Fraley, R. T., Rogers, S. G. and Chua, N.-H., *Nature*, 1985, **315**, 200–204.
34. Nagy, F., Morelli, G., Fraley, R. T., Rogers, S. G. and Chua, N.-H., *EMBO J.*, 1985, **4**, 3063–3068.
35. Cerdan, P. D., Staneloni, R. J., Casal, J. J. and Sanchez, R. A., *Plant Mol. Biol.*, 1997, **33**, 245–255.
36. Arguello-Astorga, G. R. and Herrera-Estrella, L. R., *Plant Physiol.*, 1996, **112**, 1151–1166.
37. Gilmartin, P. M., Sarokin, L., Memelink, J. and Chua, N. H., *Plant Cell*, 1990, **2**, 369–372.
38. Oelmuller, R., Bolle, C., Tyagi, A. K., Niekrawietz, N., Briet, S. and Herrmann, R. G., *Mol. Gen. Genet.*, 1993, **237**, 261–272.
39. Fliieger, K., Tyagi, A., Sopory, S. K., Cseplo, A. and Herrmann, R. G., *Plant J.*, 1993, **4**, 9–17.
40. Chattopadhyay, S., Puente, P., Deng, X.-W. and Wei, N., *Plant J.*, 1998, **15**, 69–77.
41. Gatz, C. and Lenk, I., *Trends Plant Sci.*, 1998, **3**, 352–358.
42. Reynolds, P. H. S. (ed.), *Inducible Gene Expression in Plants*, CABI Publishing, New York, 1999.
43. De Veylder, L., Van Montagu, M. and Inze, D., *Plant Cell Physiol.*, 1997, **38**, 568–577.
44. Gatz, C., in *Inducible Gene Expression in Plants* (ed. Reynolds, P. H. S.), CABI Publishing, New York, 1999, pp. 11–22.
45. Love, J., Scott, A. C. and Thompson, W. F., *Plant J.*, 2000, **21**, 579–598.
46. Llyod, A. M., Schena, M., Walbot, V. and Davis, R. W., *Science*, 1994, **266**, 436–439.
47. Aoyama, T. and Chua, N.-H., *Plant J.*, 1997, **11**, 605–612.
48. Martinez, A. and Jepson, I., in *Inducible Gene Expression in Plants* (ed. Reynolds, P. H. S.), CABI Publishing, New York, 1999, pp. 23–41.
49. Mett, V. L., Podivinsky, E., Tennant, A. M., Lockhead, L. P., Jones, W. T. and Reynolds, P. H. S., *Transgenic Res.*, 1996, **5**, 105–113.
50. Caddick, M. X. *et al.*, *Nature Biotechnol.*, 1998, **16**, 177–180.
51. Bonner, J., Huang, R.-C. and Maheshwari, N., *Biochem. Biophys. Res. Commun.*, 1960, **3**, 689–694.
52. Guilfoyle, T. J., in *Enzymes of Nucleic Acid Synthesis and Modification* (ed. Jacobs, S. T.), CRC Press, Boca Raton, 1983, vol. II, pp. 1–42.
53. Guilfoyle, T. J., Hagen, G. and Malcolm, S., *J. Biol. Chem.*, 1984, **259**, 640–648.
54. Guilfoyle, T. J. and Dietrich, M. A., in *Tailoring Genes for Crop Improvement* (eds Bruening *et al.*), Plenum Press, New York, 1986, pp. 87–100.
55. Connelly, S., Marshallsay, C., Leader, D., Brown, J. W. S. and Filipowicz, W., *Mol. Cell. Biol.*, 1994, **14**, 5910–5919.
56. Dietrich, M. A., Prenger, J. P. and Guilfoyle, T. J., *Plant Mol. Biol.*, 1990, **15**, 207–223.
57. Nawrath, C., Schell, J. and Koncz, C., *Mol. Gen. Genet.*, 1990, **223**, 65–75.
58. Saez-Vasquez, J. and Pikaard, C. S., *J. Biol. Chem.*, 2000, **275**, 37173–37178.
59. Gasch, A., Hoffman, A., Harikoshi, M., Roeder, R. G. and Chua, N.-H., *Nature*, 1990, **346**, 390–394.



60. Kim, J. L., Nikolov, D. B. and Burley, S. K., *Nature*, 1993, **365**, 520–527.
61. Washburn, K. B., Davis, E. A. and Ackerman, S., *Plant Mol. Biol.*, 1997, **35**, 1037–1043.
62. Baldwin, D. A. and Gurley, W. B., *Plant J.*, 1996, **10**, 561–568.
63. Pan, S., Czarnecha-Verner, E. and Gurley, W. B., *Plant Cell*, 2000, **12**, 125–136.
64. Burk, C., Yu, X. B., Marchitelli, I. Davis, E.-A. and Ackermann, S., *Nucleic Acids Res.*, 1990, **18**, 3611–3620.
65. Li, Y. F., Le Gourrierrec, J., Toriki, M., Kim, Y. J., Guerinneau, F. and Zhou, D. X., *Plant Mol. Biol.*, 1999, **39**, 515–525.
66. Sugiura, M., *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 1997, **48**, 383–398.
67. Bevan, M. *et al.*, *Nature*, 1998, **391**, 485–488.
68. Wu, K., Tian, L., Malik, K., Brown, D. and Miki, B., *Plant J.*, 2000, **22**, 19–27.
69. Katagiri, F., Lam, E. and Chua, N.-H., *Nature*, 1989, **340**, 727–730.
70. Schmidt, R. J., Burr, F. A., Aukermann, M. J. and Burr, B., *Proc. Natl. Acad. Sci., USA*, 1990, **87**, 46–50.
71. Sainz, M. B., Grotewold, E. and Chandler, V. L., *Plant Cell*, 1997, **9**, 611–625.
72. Takatsuji, H., *Cell. Mol. Life Sci.*, 1998, **54**, 582–596.
73. Schindler, U., Beckmann, H. and Cashmore, A. R., *Plant J.*, 1993, **4**, 137–150.
74. Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P. A. and Saedler, H., *EMBO J.*, 1987, **6**, 3553–3558.
75. Kranz, H., Scholz, K. and Weisshaar, B., *Plant J.*, 2000, **21**, 231–235.
76. Zhou, D.-X., *Trends Plant Sci.*, 1999, **4**, 210–214.
77. Yanofsky, M. F., Ma, H., Bowman, J. L., Drews, G. N., Feldmann, K. A. and Meyerowitz, E. M., *Nature*, 1990, **346**, 35–39.
78. Sommer, H., Beltran, J.-P., Huijser, P., Pape, H., Lönnig, W.-E., Saedler, H. and Schwarz-Sommer, Z., *EMBO J.*, 1990, **9**, 605–613.
79. Ma, J., Przibilla, E., Hu, J., Bogorad, L. and Ptashne, M., *Nature*, 1988, **334**, 631–633.
80. Goff, S. A., Cone, K. C. and Fromm, M. E., *Genes Dev.*, 1991, **5**, 289–309.
81. Sprenger-Haussels, M. and Weisshaar, B., *Plant J.*, 2000, **22**, 1–8.
82. Okanami, M., Meshi, T., Tamai, H. and Iwabuchi, M., *Genes Cells*, 1996, **1**, 87–99.
83. Döring, P., Treuter, E., Kistner, C., Lyck, R., Chen, A. and Nover, L., *Plant Cell*, 2000, **12**, 265–278.
84. Nakayama, T., Okanami, M., Meshi, T. and Iwabuchi, M., *Mol. Gen. Genet.*, 1997, **253**, 553–561.
85. Chern, M.-S., Bobb, A. J. and Bustos, M. M., *Plant Cell*, 1996, **8**, 305–321.
86. Le Gourrierrec, J., Li, Y.-F. and Zhou, D. X., *Plant J.*, 1999, **18**, 663–668.
87. Lyck, R., Harmening, U., Hohfeld, I., Treuter, E., Scharf, K.-D. and Nover, L., *Planta*, 1997, **202**, 117–125.
88. Jackson, D. and Hake, S., *Curr. Opin. Genet. Dev.*, 1997, **7**, 495–500.
89. Klimczak, L. J., Schindler, U. and Cashmore, A. R., *Plant Cell*, 1992, **4**, 87–98.
90. Dröge-Laser, W. *et al.*, *EMBO J.*, 1997, **16**, 726–738.
91. Ciceri, P. *et al.*, *Plant Cell*, 1997, **9**, 97–108.
92. Meshi, T., Moda, I., Minami, M. Okanami and Iwabuchi, M., *Plant Mol. Biol.*, 1998, **36**, 125–136.
93. Marechal, E., Hiratsuka, K., Delgado, J., Nairn, A., Qin, J., Chait, B. T. and Chua, N.-H., *Plant Mol. Biol.*, 1999, **40**, 373–386.
94. Martinez-Garcia, J. F., Huq, E. and Quail, P. H., *Science*, 2000, **288**, 859–863.
95. Martin, C., *Curr. Opin. Biotechnol.*, 1996, **7**, 130–138.
96. Van der Fits, L. and Memelink, J., *Science*, 2000, **289**, 295–297.
97. Yanagisawa, S., *Plant J.*, 2000, **21**, 281–288.
98. Mandel, M. A. and Yanofsky, M. F., *Nature*, 1995, **377**, 522–524.
99. Hartmann, U., Hohmann, S., Nettesheim, K., Wisman, E., Saedler, H. and Huijser, P., *Plant J.*, 2000, **21**, 351–360.
100. Cacharrn, J., Saedler, H. and Theissen, G., *Dev. Genes Evol.*, 1999, **209**, 411–420.
101. Schena, M., Lloyd, A. and Davis, R. W., *Genes Dev.*, 1993, **7**, 367–379.
102. Peng, J. *et al.*, *Nature*, 1999, **400**, 256–261.
103. Jaglo-Ottosen, K. R., Gilmour, S. J., Zarka, D. G., Schabenberger, O. and Thomashow, M. F., *Science*, 1998, **280**, 104–106.
104. Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K. and Shinozaki, K., *Nat. Biotechnol.*, 1999, **17**, 287–291.
105. Tyagi, A. K. and Mohanty, A., *Plant Sci.*, 2000, **158**, 1–18.
106. Parinov, S. and Sundaresan, V., *Curr. Opin. Biotechnol.*, 2000, **11**, 157–161.
107. Vielle-Calzada, J.-P., Baskar, R. and Grossniklaus, U., *Nature*, 2000, **404**, 91–94.

ACKNOWLEDGEMENT. Our work in the area of regulation biology is supported by the Department of Biotechnology, Government of India and the Rockefeller Foundation, USA.