Chapter 1

What Have We Learned About Synthetic Promoter Construction?

Paul J. Rushton

Abstract

The molecular components of transcriptional regulation are modular. Transcription factors have domains for specific functions such as DNA binding, dimerization, and protein–protein interactions associated with transcriptional activation and repression. Similarly, promoters are modular. They consist of combinations of *cis*-acting elements that are the binding sites for transcription factors. It is this promoter architecture that largely determines the expression pattern of a gene. The modular nature of promoters is supported by the observation that many *cis*-acting elements retain their activities when they are taken out of their native promoter context and used as building blocks in synthetic promoters. We therefore have a large collection of *cis*-acting elements to use in building synthetic promoters and many minimal promoters upon which to build them. This review discusses what we have learned concerning how to use these building blocks to make synthetic promoters. It has become clear that we can increase the strength of a promoter by adding increasing numbers of *cis*-acting elements. However, it appears that there may be a sweet spot with regard to inducibility as promoters with increasing numbers of copies of an element often show increased background expression. Spacing between elements appears important because if elements are placed too close together activity is lost, presumably due to reduced transcription factor binding due to steric hindrance. In many cases, promoters that contain combinations of *cis*-acting elements show better expression characteristics than promoters that contain a single type of element. This may be because multiple transcription factor binding sites in the promoter places it at the end of multiple signal transduction pathways. Finally, some *cis*-acting elements form functional units with other elements and are inactive on their own. In such cases, the complete unit is required for function in a synthetic promoter. Taken together, we have learned much about how to construct synthetic promoters and this knowledge will be crucial in both designing promoters to drive transgenes and also as components of defined regulatory networks in synthetic biology.

Key words Synthetic promoter, *Cis*-acting elements, Synthetic biology, Transgene expression, Plant biotechnology

1 Introduction

This review focuses mainly on the synthetic promoter projects that I have been involved with and serves as a guide to producing the best synthetic promoters. There are general trends, some of which we could not have predicted when we first started to construct

Reinhard Hehl (ed.), *Plant Synthetic Promoters: Methods and Protocols*, Methods in Molecular Biology, vol. 1482,
DOI 10.1007/978-1-4939-6396-6_1,

synthetic promoters, but there will always be exceptions to the rules. The reader is urged to use the observations presented here to help them in their own synthetic promoter projects but ultimately it is the activity of the constructed promoter that will decide whether a project is successful. One final note, in biotech projects a synthetic promoter will be used to drive a transgene and it is the best possible transgene expression that decides whether a project has ultimately been successful or not. A synthetic promoter can be used to optimize expression levels so that the transgene is expressed at the right time, in the right place, and to the optimum level. This potential optimization of expression (where, when, and how much) is where the advantage of synthetic promoters lies over native ones.

2 The Modular Nature of Transcriptional Regulation

The modular nature of transcription (and indeed signaling in general) has become apparent. Proteins have specific domains for certain function such as dimerization, ligand binding, nuclear localization, and so on. These domains can often retain their activities in domain swap or addition experiments. With transcription factors, this modularity is very clear. It includes nuclear localization domains, dimerization domains, calmodulin binding domains, protein–protein interaction domains associated with transcriptional activation or repression, and many others [1]. Building synthetic transcription factors with altered activities is therefore possible. For example, adding a repression domain such as an EAR domain from an ERF transcription factor can transform a transcription factor that normally functions as an activator into a dominant negative [2].

Similarly, promoters are modular as they typically contain combinations of *cis*-acting elements that are the binding sites for transcription factors. It is this promoter architecture that largely determines the expression pattern of a gene as it determines the specificity of transcription factor binding to the promoter. At the level of the promoter, binding of the transcription factors to the DNA is accompanied by protein–protein interactions between transcription factors themselves and also interactions with the general transcriptional machinery (general transcription factors, co-activators, and co-repressors) and other proteins that alter chromatin structure [1]. With each promoter containing multiple transcription factor binding sites and also with each transcription factor potentially forming multiple protein–protein interactions, it was originally unclear whether there would be any chance that a *cis*-acting element, when taken out of its native promoter context, could retain its activity [3]. This retention of activity would be a prerequisite for the construction of synthetic promoters.

In the late 1990s, I started a project on constructing pathogen-inducible synthetic promoters. There were many reports from the

literature of defined *cis*-acting elements retaining activity in synthetic promoters but it was unclear how widespread this phenomenon was. Several different types of known pathogen-responsive *cis*-acting elements were tested in synthetic promoters and strikingly, the majority of these elements retained their activity [3]. This included GCC-like boxes, W boxes, and Box D (which is still ill-defined). This showed that transcription at the promoter level is indeed modular and that many of these DNA modules can therefore be used to construct synthetic promoters. A synthetic promoter could therefore be build up from any number or combination of these modular building blocks in a similar way that someone builds something from Lego blocks.

3 Making a Synthetic Promoter

In its simplest form, a synthetic promoter will consist of a minimal promoter (the binding sites for general transcription factors including RNA polymerase II) and a defined *cis*-acting element [4]. The minimal promoter will typically contain a TATA box and a site at which transcription will start but little else as this may influence the expression characteristics of the promoter. Upstream of this minimal promoter are placed the *cis*-acting elements that will determine the expression characteristics of any transgene whose expression is driven by the promoter. These *cis*-acting elements can include any number of copies of an individual *cis*-acting element or combinations of different elements in any order and in any number. The possibilities are seemingly endless. With current advances in DNA technology, it is possible to simply synthesize any given synthetic promoter and this can speed up the process of building a promoter considerably. However, previously synthetic promoters were typically synthesized from ligating oligonucleotides containing the defined *cis*-acting element sequences upstream of a minimal promoter (Fig. 1).

Using technology based on two different restriction endonucleases with compatible sticky ends, this approach has the advantage that the resultant promoters can be used like Lego building blocks to optimize and test synthetic promoters. For this reason, this approach is still valuable today. Briefly, a defined *cis*-acting element is synthesized as two oligonucleotides, one for each strand of the DNA. When annealed together, the double-stranded DNA has sticky ends at both the 5 prime and 3 prime ends that are compatible (for example SpeI and XbaI or BamHI and BglII). The single copy of the *cis*-acting element is inserted into the corresponding restriction enzyme sites just upstream of a chosen minimal promoter to create a synthetic promoter with a single copy of the element (a 1 × construct). The beauty of this strategy becomes apparent when this 1 × construct is used to make other synthetic promoters. The 1 × construct is cut with a restriction enzyme that cuts the backbone

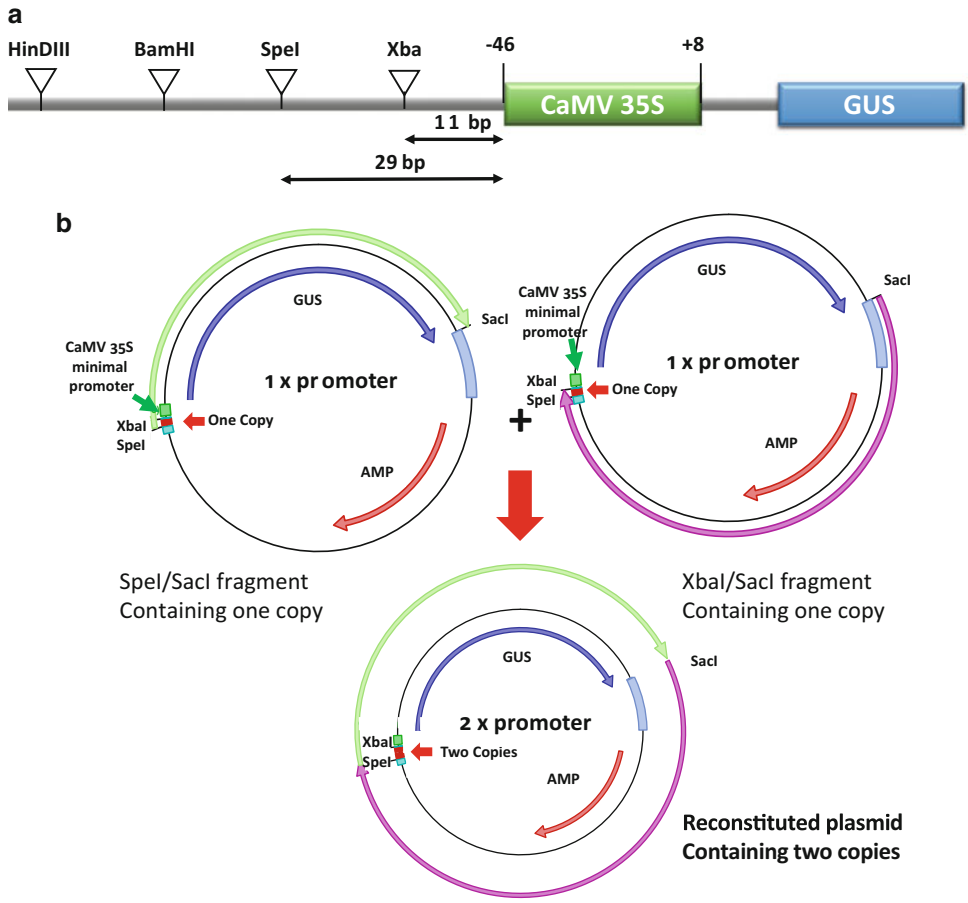

Fig. 1 A system to produce synthetic promoters with any number of *cis*-acting elements in any order. (**a**) The minimal promoter and restriction sites in MS23 (pBT10) [3]. Single-stranded oligonucleotides containing defined *cis*-acting elements with a SpeI sticky end at the 5′ end and an XbaI sticky end at the 3′ end are annealed and then inserted into SpeI/XbaI double-digested vector DNA 11 bp upstream of the CaMV35S -46 minimal promoter. (**b**) How to make a 2 × element promoter construct from 1 × element promoters. In two separate restriction digests, the 1 × promoter DNA is digested by either SacI and XbaI or SacI and SpeI. In each case the fragment containing the single copy of the *cis*-acting element is gel purified and the SpeI/SacI and XbaI/SacI fragments are then annealed to give a 2 × element promoter. The ligation of the SpeI and XbaI sticky ends destroys the restriction site in the middle of the 2 × element yielding a SpeI site at the 5′ end and an XbaI site at the 3′ end. This pattern of restriction sites is identical to the 1 × element construct and means that the process can be repeated to yield 4 × and then 8 × constructs and so on. The beauty of this system is that by using different promoter constructs as starting materials, promoters containing combinations of elements in any number and in any order can be produced

of the plasmid and then either of the two enzymes with compatible sticky ends (for example SpeI or XbaI). In each case, the fragment that contains the *cis*-acting element is then chosen and the two pieces are ligated together. Because each fragment contained a copy of the element, the resulting synthetic promoter contains two copies of the defined *cis*-acting element (a 2 × construct). The inventive

step of this approach is that where the two copies of the element come together, they ligate together as they have compatible sticky ends. However, the restriction site between the two elements is not recreated because the two restriction sites are different. The result is one piece of DNA with no internal restriction site but the same 5 prime and 3 prime sites that you started with. This means that the process can be repeated and two copies can become four and then eight. In addition, by choosing different *cis*-acting element monomers, promoters can be constructed with any number of elements in any combination and in any order. Once monomer constructs are available that contain different *cis*-acting elements, they can then truly act as Lego building blocks for building synthetic promoters to the design of the researcher [3, 4].

4 The Effect of *Cis*-acting Element Number on Strength and Inducibility

One of the first questions that I asked when constructing synthetic promoters was "What is the effect of increasing the number of copies of a single *cis*-acting element in a synthetic promoter?". Figure 2 shows that increasing the number of copies progressively from one to eight increases the strength of the promoter, presumably by providing more transcription factor binding sites. This suggests that an increasing number of transcription factors bound to the

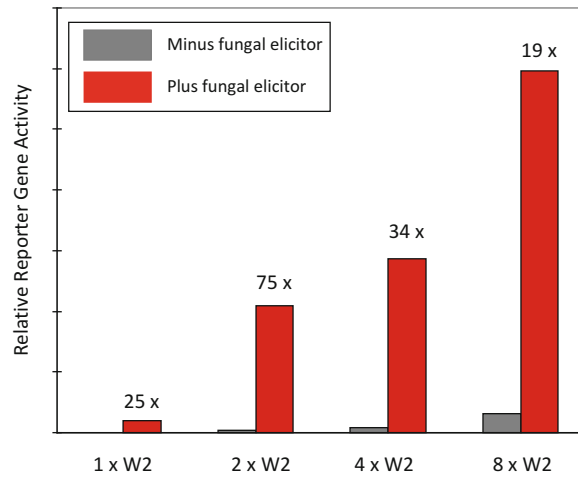

Fig. 2 Increasing the number of *cis*-acting elements in a synthetic promoter increases strength. 1 × W2, 2 × W2, 4 × W2, and 8 × W2 synthetic promoters were tested for induction by a fungal elicitor in a parsley transient expression system [3]. Increasing the number of copies progressively from one to eight increases the strength of the promoter, although the best signal-to-noise ratio is obtained with a 2 × W2 construct due to an increase in background expression with an increasing number of elements

promoter increases the rate of transcription from the synthetic promoter. This increase was observed not only in systems of reduced complexity but also in transgenic plants and has profound consequences for synthetic promoters in general. The fact that we can alter the strength of a synthetic promoter by varying the number of *cis*-acting element building blocks in the promoter means that we can modulate promoter strength by design. This immediately underlines an advantage of synthetic promoters over native ones—with synthetic promoters we can vary strength to find the optimum expression level of a transgene. This modulation is not possible when using a single native promoter.

One additional observation from Fig. 1, and this was also apparent in transgenic lines, is that there is one downside to increasing strength, namely that as the synthetic promoters get stronger the level of background expression in inducible promoters (such as pathogen-inducible promoters) often increases and therefore the fold inducibility is reduced [2]. The exact reason for this is unclear. It may be that more binding sites increase the level of basal transcription or alternatively allow increased binding of transcription factors that may have a lower affinity for the native promoter. Either of these two possibilities may increase the level of transcription in the absence of the signal and lead to increased background expression.

Again, the choice of synthetic promoter will be driven by the choice of transgene and how this transgene is best expressed. For some projects, reasonable levels of background expression could be tolerated (as the expression pattern is still considerably better than constitutive overexpression using, for example, the CaMV 35S promoter). For others, the best inducibility is required such as expression in infected plant tissues but not uninfected ones when using pathogen-inducible synthetic promoters. For the former, a strong promoter with eight copies of a *cis*-acting element might be best, whereas for the latter two copies may be preferred as it shows the best inducibility (signal:noise). The above examples provide a nice example of the value of synthetic promoters—we are designing promoters for specific purposes and different promoters will be suited to different projects.

5 The Effect of Spacing on Promoter Strength

Once the building blocks for a synthetic promoter have been chosen, how do we put them together to make a good synthetic promoter? Well one of the first considerations is spacing. This includes not only the spacing between multiple copies of an element but also spacing with respect to the minimal promoter. When I first started to construct a range of synthetic promoters, I suspected that spacing between elements might be crucial for promoter activity driven by the need for the cognate transcription factors to interact with other proteins in a productive way. However, although

spacing turns out to be important, results suggest that this is not in the way that I had envisaged. Elements such as GCC boxes and W boxes appear to function independently of each other and spacing between the elements themselves appears to have little or no effect. In fact, systematic rotating of *cis*-acting elements relative to each other by one base pair at a time through one complete turn of the DNA helix had a negligible effect on promoter activity. With *cis*-acting elements that function independently, it would appear that the exact distance between them has little or no effect.

However, spacing is crucial to synthetic promoter activity in one crucial respect—if you place the *cis*-acting elements too close together they lose activity. The exact distance will need to be determined experimentally, but in my experience if the core sequences of elements are less than 10–15 bp apart then activity is reduced. This makes sense if one considers the binding of proteins to the short promoter DNA sequence. A transcription factor will require a certain length of DNA to bind to and if this synthetic promoter puts two binding sites too close together then binding to one site will preclude binding of another transcription factor to the next site. This reduction in activity due to steric hindrance is also seen if the promoter puts a *cis*-acting element too close to the minimal/core promoter. In this case, general transcription factors will compete for binding to the promoter with the transcription factors that bind to the *cis*-acting elements that have been added upstream. As a rule of thumb, at least 15–20 bp should be allowed between multiple copies of a *cis*-acting element in a synthetic promoter so that their core sequences are separated by close to 50 bp. In addition, it is best to have at least 50 bp between the TATA Box and the core sequence of the first *cis*-acting element placed upstream of it (Fig. 3).

6 Combinations of *Cis*-acting Elements Appear Best

Some aspects of synthetic promoter technology were not necessarily predictable and only became apparent once a systematic approach was used to design a spectrum of different synthetic promoters [3, 4]. One of these observations is that synthetic promoters that contain more than one type of *cis*-acting element may be better than simpler promoters that contain multiple copies of only one type of element. With pathogen-inducible synthetic promoters, high expression at infection sites coupled to low background expression in non-infected tissues is preferred in order to reduce any negative effects of transgene expression in non-infected tissues. It was observed that promoters that contain more than one type of *cis*-acting element showed the best inducibility coupled with lower background, making them much better suited for transgene expression. It is likely that the reason for this is that multiple different *cis*-acting elements place a synthetic promoter at the end of more

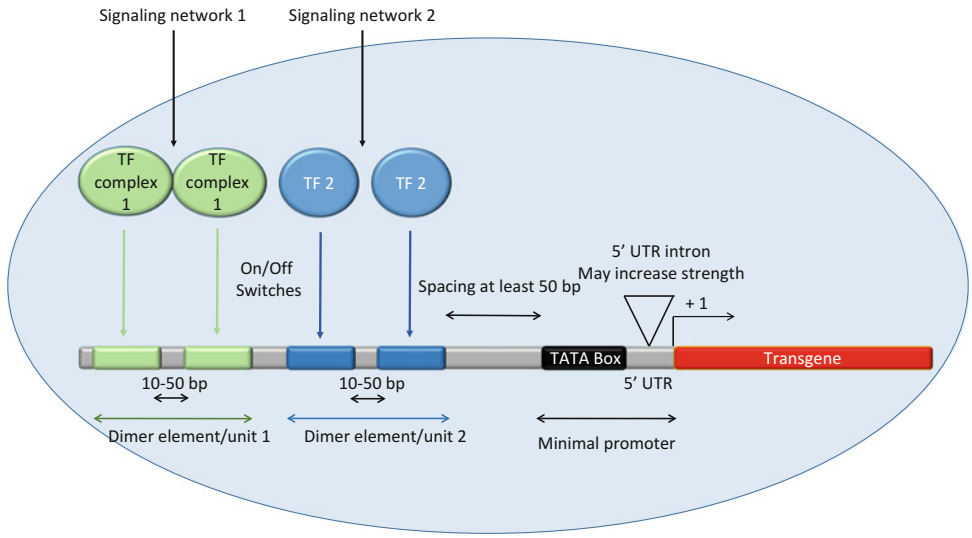

Fig. 3 What we have learned about synthetic promoter construction. Although promoters will need to be optimized for each transgene and each required transgenic plant line, a number of observations have been made concerning the best starting strategies and are summarized in this figure. The best signal:noise ratios appear to be produced from dimers of individual *cis*-acting elements/units. These dimers should be placed with spacing of 10–50 bp between the elements to avoid reduced activity. Synthetic promoters that contain more than one type of *cis*-acting element/unit appear to give better expression characteristics, presumably because this places the promoter at the end point of more than one signaling pathway so that it responds to multiple inputs. The minimal promoter should consist of the TATA Box, the 5′ UTR and the start of transcription only. This ensures that it is functional but does not affect promoter characteristics. The exception to this is that certain minimal promoters that contain introns in their 5′ UTRs may direct stronger expression. The individual *cis*-acting elements can contain a single element or consist of a functional unit of more than one element

than one signaling pathway. As the constructed promoter takes signals from multiple pathways and multiple transcription factor types, its activity is likely to be more tightly regulated resulting in a better expression pattern.

As we start to understand how to construct synthetic promoters for specific purposes, we can start to combine some of the observations when designing promoters. For example, two copies of an individual *cis*-acting element in a promoter probably give the best signal to noise ratio (Fig. 2) and multiple different *cis*-acting elements also seem to give better inducibility. It is therefore likely that a promoter that contains two copies of several *cis*-acting elements would be among the best promoters in terms of inducibility and that is exactly what was observed in the project reported by Rushton et al. [3]. The best synthetic pathogen-inducible promoter was $2 \times W2/2 \times S/2 \times D$. It combines three different types of *cis*-acting elements, two copies of each element, and at least two different families of transcription factors (WRKY and AP2/ERF) as end points in the signal transduction pathways. In addition, the elements are spaced far enough apart and from the minimal

promoter to avoid loss of activity due to steric hindrance. These observations give important pointers as to how we might use defined *cis*-acting elements to build the best synthetic promoters.

7 The Choice of Minimal Promoter

Many synthetic promoter projects have previously used the minimal CaMV 35S promoter as the start point for synthetic promoter construction. This was probably because the CaMV35S is the best characterized strong promoter and the -46 version shows minimal basal activity in the absence of added *cis*-acting elements. However, it appears that many minimal promoters can be used as the basis for synthetic promoter construction. Some work is however required as a minimal promoter needs to be defined that is active but that also does not affect expression characteristics (this normally contains the start of transcription and a TATA Box but little else). As a rule of thumb, the best minimal promoter to use is probably one from a gene whose expression characteristics already are closest to the desired expression characteristics of the final synthetic promoter. For a drought inducible promoter, this would be from a drought inducible gene, for a wound inducible promoter this would be from a wound inducible gene and so on.

One further observation is important in the production of synthetic promoters that are designed to direct strong expression levels because here the choice of minimal promoter may be more important. The promoters of several genes that show very high expression levels (for example several ubiquitin genes) contain introns in their 5 prime UTRs [5]. These introns appear to contribute to strength, and choosing a minimal promoter that contains such an intron may therefore be an important part of the design strategy when increasing strength.

8 Functional Units Need to Be Kept Together

More recently, we have gained more insights into the production of synthetic promoters. This can be illustrated by work using the GAG fragment from PMT promoters in tobacco [6]. The GAG fragment is so called because it consists of three parts. A G box followed by an AT-rich spacer region and then a GCC-like box. The GAG fragment, like the PMT promoters themselves directs jasmonate and wounding inducible expression (Fig. 4). The expression pattern for synthetic promoters with the GAG fragment is exquisite as wounding of a leaf or jasmonate treatment results in expression in the cortex of the root (tissue-specific expression at a distance). This is also the expression pattern of the native PMT promoters and illustrates that this small GAG fragment is sufficient to drive expression that is similar to the native promoter.

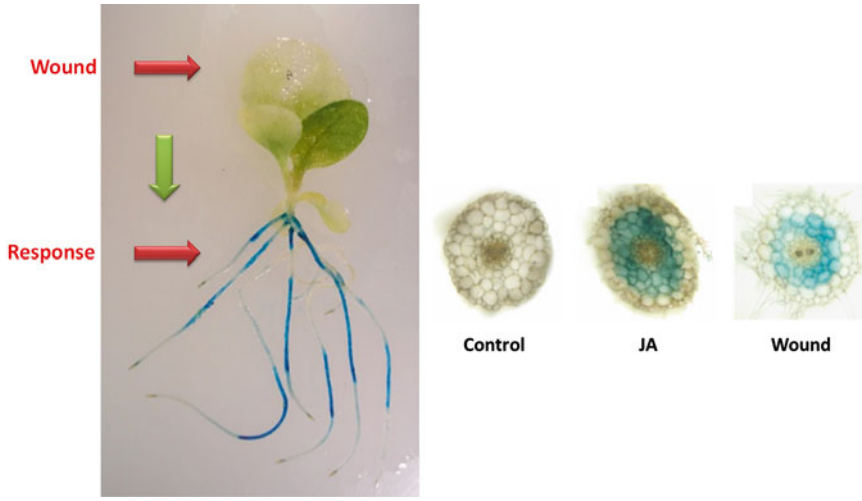

Fig. 4 Transgenic tobacco plants containing a 4 × GAG synthetic promoter show both wound and jasmonate inducible expression at a distance. Wounding or jasmonate treatment of the leaves leads to expression from the 4 × GAG synthetic promoter in the roots. This root-specific expression is concentrated in the cortex, the main site of nicotine biosynthesis

Synthetic promoters containing the GAG fragment and its three constituent parts present important new evidence concerning how to build synthetic promoters. Each one of the three constituent parts of the GAG fragment (the G box, the AT-rich region, and the GCC-like box) is inactive on its own. The G box and GCC-like box, although similar to known *cis*-acting elements that are bound by bHLH, bZIP, and AP2/ERF transcription factors, show no activity if used alone in synthetic promoters [6]. However, if the G box is combined with the GCC-like box, then jasmonate inducibility is restored. It is clear that the two *cis*-acting elements function together as a unit and that at least two transcription factors (a bHLH and an ERF) are required for function. However, the story does not end there because although the G box–GCC-like box unit is active, it is neither as strong nor as inducible as the G box–AT-rich region–GCC-like box unit (Fig. 5). It appears that the AT-rich region is required for full activity and that this activity is not dependent on the sequence of the region because an AT-rich region of different sequence but the same length appears similarly active. Taken together, it is clear that the GAG fragment is a unit consisting of three elements. Two of these elements appear to be binding sites for transcription factors and the third is most likely a spacer region. This suggests that the two transcription factors probably interact directly or indirectly for function of the GAG fragment.

The lesson for synthetic promoter construction is that some promoters consist of functional units with more than one constituent element. In such cases the entire unit needs to be used for activity. Importantly, if one uses each functional unit (for example

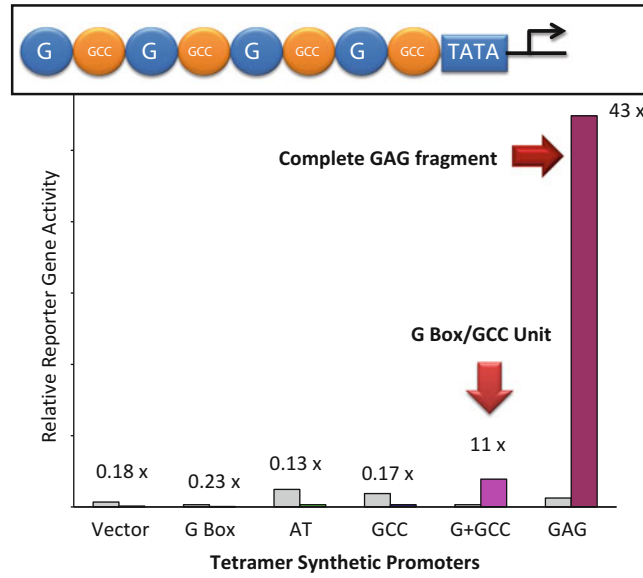

Fig. 5 Functional units consisting of more than one *cis*-acting element require all elements for full activity. Synthetic promoters containing tetramers of the GAG fragment and its constituent individual *cis*-acting elements (the G box, AT-rich region, and GCC-like element) were tested for activity (*gray bars*) and jasmonate inducibility (*colored bars*) in stably transformed BY-2 cells. None of the individual elements were functional alone. However, a tetramer of the G box–GCC-like element showed jasmonate inducibility suggesting that these elements are binding sites for transcription factors (bHLH and ERF) and together form a jasmonate response element in synthetic promoters. However, it is also clear that the GAG fragment is a unit consisting of three elements because synthetic promoters that also contain the AT-rich region (the complete GAG fragment) are both stronger and more inducible by jasmonate than the G box–GCC element. These data show that functional units that consist of more than one *cis*-acting element should be kept together when used in building synthetic promoters

the GAG fragment) in a similar way to *cis*-acting elements that function alone (such as W boxes) then the same rules apply. For example, 4 × GAG is stronger than 2 × GAG and so on. The modular nature of promoter technology still applies with some Lego bricks consisting of one *cis*-acting element whereas others consist of units of more than one element.

9 The Best Place to Start

Synthetic promoters have several potential advantages over native promoters. The main advantage is that the strength of the promoter can be altered to produce promoters that are stronger or weaker depending on the number of copies of each element/unit.

Additional potential advantages include the possibility of reducing unwanted expression characteristics by using a single *cis*-acting element from a promoter and eliminating other elements that may direct undesired expression characteristics. This has been a major theme in, for example, pathogen-inducible promoters where expression in uninfected tissues is undesirable. This strategy of eliminating other *cis*-acting elements has, however, met with only limited success. For example, a W box that directs pathogen-inducible expression (desired) often also directs wound-inducible expression (undesired). The likely cause of this is that the cognate transcription factors are involved in both pathogen and wound induced signaling and as a result, induction by the two stimuli cannot be separated.

Despite the potential advantages in using synthetic promoters to fine tune transgene expression in biotechnology projects, the best piece of advice is not to use one at all if a good native promoter is available that drives the desired expression characteristics! The other good piece of advice when choosing where to start, is to start with the native promoters that most closely fit the desired expression characteristics because they will be the best source of *cis*-acting element/unit building blocks to build an improved synthetic promoter. In the case of the GAG fragment, the source of the unit was the tobacco PMT promoters and the GAG fragment drives expression that has similarities to the full-length promoters. In the case of pathogen-inducible synthetic promoters, the best sources were the promoters of pathogenesis-related genes such as PR10s (the sources of various W boxes and Box D) [3].

10 Conclusions and Future Prospects

Many promising transgenes have failed not because of a poor choice of transgene but because of a poor choice of promoter with which to drive it with. I sometimes wonder how many transgenes have been discarded over the years as being unsuitable for improving crop plants based on results using high-level ectopic overexpression using the CaMV 35S promoter or other unsuitable promoters. The choice of promoter can make or break a project and over the years many plant scientists have been unimaginative in their choice of promoter. Synthetic promoters can change this and fine tuning promoter activity using synthetic promoters should be an increasingly important topic in plant biotechnology.

One area which should increase the use of synthetic plant promoters is synthetic biology. As we have seen above, signaling is modular, both at the protein domain level and the *cis*-acting element level with activity often residing in the individual protein domain or transcription factor binding site. Using these building blocks it should be possible to construct complete signaling pathways from the ground up using building blocks from different

proteins and promoters. With this approach, synthetic promoters will be a crucial part of synthetic biology as they represent synthetic end points for synthetic signaling pathways.

Acknowledgements

I would like to thank Peter Bowerman at BASF for critically reading this manuscript and Neal Stewart for inviting me to speak on this topic at the 2014 World Forum on Biology in Atlanta Georgia. This chapter is based on the seminar given there. I would like to thank Imre Somssich for his support and sound advice on the pathogen-inducible synthetic promoter projects. I would particularly like to thank Michael Timko for allowing me to include unpublished data on the GAG fragment in this chapter. I would also like to thank Reinhard Hehl for his patience while I was moving to a new position.

References

1. Liu W, Stewart CN (2016) Plant synthetic promoters and transcription factors. Curr Opin Biotechnol 37:36–44
2. Hiratsu K, Matsui K, Koyama T, Ohme-Takagi M (2003) Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in Arabidopsis. Plant J 34:733–739
3. Rushton PJ, Reinstädler A, Lipka V, Lippok B, Somssich IE (2002) Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen- and wound-induced signaling. Plant Cell 14:749–762
4. Gurr SJ, Rushton PJ (2005) Engineering plants with increased disease resistance: how are we going to express it? Trends Biotechnol 2005(23):283–290
5. Hernandez-Garcia CM, Bouchard RA, Rushton PJ, Jones ML, Chen X, Timko MP, Finer JJ (2010) High level transgenic expression of soybean (Glycine max) GmERF and Gmubi gene promoters isolated by a novel promoter analysis pipeline. BMC Plant Biol 10:237. doi:10.1186/1471-2229-10-237
6. Sears MT, Zhang H, Rushton PJ, Wu M, Han S, Spano AJ, Timko MP (2014) NtERF32: a non-NIC2 locus AP2/ERF transcription factor required in jasmonate-inducible nicotine biosynthesis in tobacco. Plant Mol Biol 84:49–66. doi:10.1007/s11103-013-0116-2