Plant Science 217–218 (2014) 109–119

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Review

Identification and validation of promoters and *cis*-acting regulatory elements☆

Carlos M. Hernandez-Garcia, John J. Finer*

Department of Horticulture and Crop Science, OARDC/The Ohio State University, 1680 Madison Avenue, Wooster, OH 44691, USA

ARTICLE INFO

Article history:
Received 4 September 2013
Received in revised form 3 December 2013
Accepted 5 December 2013
Available online 14 December 2013

Keywords:
Transcription
Genetic engineering
Synthetic promoters
Chromatin structure
Transcriptomics
Gene expression systems

ABSTRACT

Studies of promoters that largely regulate gene expression at the transcriptional level are crucial for improving our basic understanding of gene regulation and will expand the toolbox of available promoters for use in plant biotechnology. In this review, we present a comprehensive analysis of promoters and their underlying mechanisms in transcriptional regulation, including epigenetic marks and chromatin-based regulation. Large-scale prediction of promoter sequences and their contributing *cis*-acting elements has become routine due to recent advances in transcriptomic technologies and genome sequencing of several plants. However, predicted regulatory sequences may or may not be functional and demonstration of the contribution of the element to promoter activity is essential for confirmation of regulatory sequences. Synthetic promoters and introns provide useful approaches for functional validation of promoter sequences. The development and improvement of gene expression tools for rapid, efficient, predictable, and high-throughput analysis of promoter components will be critical for confirmation of the functional regulatory element sequences identified through transcriptomic and genomic analyses.

Contents

* Corresponding author. Tel.: +1 330 263 3880; fax: +1 330 263 3887.
E-mail address: finer.1@osu.edu (J.J. Finer).

0168-9452/$ – see front matter © 2013 The Authors. Published by Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.plantsci.2013.12.007

1. Introduction

Multicellular organisms maintain the same inherited genetic material in most of their cells throughout different cellular and developmental processes. However, organisms have acquired a diverse array of molecular mechanisms including coordinated expression of genes that tightly control metabolism during differentiation and development. Regulation of gene expression in different tissues and organs, during various growth and developmental stages, or as a consequence of external stimuli is mediated at the transcriptional, post-transcriptional, and post-translational level. Transcriptional regulation plays the greatest role in the activation and suppression of expression, and is largely controlled through gene promoters and their contributing *cis*-acting elements [1].

In the simplest terms, gene promoters are DNA sequences located upstream of gene coding regions and contain multiple *cis*-acting elements, which are specific binding sites for proteins involved in the initiation and regulation of transcription. DNA sequences located in the 3′-flanking or downstream region of the transcribed region, or even within the transcribed region, can also influence the initiation of transcription; however, these gene regulatory elements will not be discussed here. Promoters of protein-encoding genes often contain a "core promoter", which is a region located ~40 bp upstream of the transcriptional initiation site and comprises the TATA box [2]. The TATA box is the binding site for the transcription initiation factor TFIID TBP (TATA-box-Binding Protein) subunit. The core promoter also contains *cis*-elements that are binding sites for the basic transcriptional machinery, including RNA polymerase II and its corresponding subunits (Fig. 1) [3]. A protein complex, including general transcription factors such as TFIID and TFIIB is formed with RNA polymerase II prior to initiation of transcription [4].

Upstream of the core promoter region are the proximal and distal regions of promoters. Proximal and distal regions of the promoter contain different regulatory sequences such as enhancers, silencers, insulators, and *cis*-elements that contribute to the fine regulation of gene expression at the transcriptional level (Fig. 1) [4]. The physical demarcation of the upstream regions that contribute to the "full promoter" is more fluid than that of the core promoter, and the size of the active, fully functional promoter depends directly on the positional and combinatorial understanding of the *cis*-acting elements present in both the proximal and distal regions. During transcription, co-activators and transcription factors bind to specific DNA motifs and simultaneously interact with the transcriptional machinery attached to the core promoter [4]. This complex DNA/protein interaction leads to the activation, enhancement, or suppression of transcription. Thus, regulation of transcription depends on the: (a) availability and activity of transcription factors, and (b) the type, number, position, and combination of regulatory elements present in and around the promoter [1].

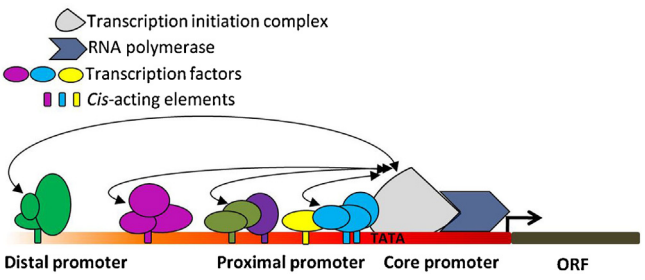

Fig. 1. Simplified model of transcriptional regulation of protein-encoding genes.

Regulation of gene expression at the promoter level is mainly controlled by the *cis*-acting elements localized upstream of the transcriptional start site. The physical interaction between regulatory proteins and the basic transcriptional machinery is straightforward during initiation of transcription due to the location of proximal elements to the core promoter. Distal promoter elements, located far away from the transcriptional start site can also impact gene expression. The mechanisms of how distal elements come into close proximity to the core promoter to modulate gene expression during transcription involve DNA folding mediated by conformational changes in the 3-dimensional structure of DNA and chromatin.

2. Chromatin structure and its role during transcription

Linear models depicting gene promoters (Fig. 1) are often simple representations of the contributing *cis*-acting elements positioned upstream of the transcriptional start site of a gene. However, *in vivo* regulation of transcription is more dynamic and portrays an augmented level of complexity. The 3-dimensional organization of DNA, influenced by folding and the association with chromatin appears to be highly organized and allows *cis*-acting elements located in distant regions to fold and spatially become proximal to the regulatory complex (Fig. 2) [5]. Under this dynamic definition of promoters, introns, 3′UTR, 5′UTR and even regulatory sequences positioned up to several Kb, and in extreme cases, more than 1 Mb away from the core promoter, can influence transcription rates [6].

Compacted assembly of the genomic DNA, wrapped by histones in a small nucleus is also a major constraint for transcriptional regulatory proteins and RNA polymerases limiting access to DNA and, thus, leverages gene transcription. Chromatin-based gene regulation includes replacement of common histones with specialized variant types and total or partial removal of histones from DNA [7]. The hypermobile animal nucleosomes containing the H3.3 and H2A.Z histone variants are relatively unstable, with these histones being easier to displace from DNA. These histones are predominantly associated with promoter regions, enhancers, and gene coding regions, where the nucleosomes are disrupted and reformed rapidly during transcription [8].

Post-translational modification of histones plays an important role in regulating transcription. Histones can be modified at their N-terminal tails through the addition of functional groups including methyl, acetyl, and phosphoryl. The addition of epigenetic marks to histones leads to activation or silencing of transcription as a result of either loosening or enhancing the association between histones and DNA [9]. For example, the trimethylation of the H3 histone protein at lysine 4 (H3K4me3) or lysine 27 (H3K27me3) are well-studied in animals and plants. The H3K4me3 histone variant is highly represented near the 5′ end of actively transcribed genes and associated with transcriptional initiation of the *Flowering Locus C* (FLC) in *Arabidopsis* [10]. Contrarily, the H3K27me3 variant is linked to gene silencing *via* chromatin condensation during plant development processes [9]. Transcription factors also have the ability to recruit coactivator proteins that acetylate histones and, thus, positively affect the activation of transcription.

Studies of chromatin-based regulation in plants are emerging and yet more research on epigenetic features and genome-wide mapping of histone modifications is required to predict active promoters and enhancers, and gain a better understanding of regulation of transcription at the promoter level. Transcription of transgenes in transformed organisms is also subject to chromatin-based regulation, albeit the degree of regulation depends on where the transgene is integrated. Most conventional DNA introduction methods result in somewhat random integration of transgenes in the host genome, including transcriptionally active and inactive