Biochimica et Biophysica Acta xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Towards genome-wide prediction and characterization of enhancers in plants☆

Alexandre P. Marand, Tao Zhang, Bo Zhu, Jiming Jiang *

Department of Horticulture, University of Wisconsin-Madison, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 12 April 2016
Received in revised form 13 June 2016
Accepted 14 June 2016
Available online xxxx

Keywords:
Enhancer
Promoter
Transcription co-factor
Open chromatin
Histone modification

ABSTRACT

Enhancers are important *cis*-regulatory DNA elements that regulate transcription programs by recruiting transcription factors and directing them to the promoters of target genes in a cell-type/tissue-specific manner. The expression of a gene can be regulated by one or multiple enhancers at different developmental stages and/or in different tissues. Enhancers are difficult to identify because of their unpredictable positions relative to their cognate promoters. Remarkably, only a handful of enhancers have been identified in plant species largely due to the lack of general approaches for enhancer identification. Extensive genomic and epigenomic research in mammalian species has revealed that the genomic locations of enhancers can be predicted based on the binding sites of transcriptional co-factors and several distinct features associated with open chromatin. Here we review the methodologies used in enhancer prediction in mammalian species. We also review the recent applications of these methodologies in *Arabidopsis thaliana* and discuss the future directions of enhancer identification in plants. This article is part of a Special Issue entitled: Plant Gene Regulatory Mechanisms and Networks, edited by Dr. Erich Grotewold and Dr. Nathan Springer.

1. Introduction

Plant development and responses to environmental and hormonal cues are carefully regulated by precise orchestration of gene transcription programs [1]. The spatiotemporal specificity of gene transcription can be largely attributed to the binding of transcription factors (TFs) to regulatory DNA elements [2]. Generally, the presence of regulatory elements located proximal to the transcription start site (TSS), also known as the core promoter, is sufficient for the assembly of transcriptional machinery, often leading to low or moderate gene expression. However, the participation of some promoter-distal regulatory elements can drastically alter the rate and quantity of mRNA biosynthesis through poorly understood mechanisms. These distant sequences contain clusters of short, 4–30 base pair (bp) DNA motifs which function as binding sites for sequence-specific TFs and collectively are referred to *cis*-regulatory elements (CREs) (Fig. 1). CREs can function in gene transcription, repression, chromatin organization and a multitude of other functions, however, this review will pay particular attention to the transcription activating CREs, termed enhancers.

Enhancer-bound TFs are capable of recruiting co-activators and other TFs that interact with different components of the mediator complex in order to recruit RNA polymerase II (RNAPII) to the TSS [3]. However, for distally activated elements to facilitate the interaction between TFs and RNAPII, the enhancer-associated chromatin must loop over the intervening sequences, and direct interaction between the enhancer-bound TF complex, the target gene promoter, and the transcriptional machinery, in a process that is typically mediated and stabilized by the chromosome-associated multi-subunit protein complex, cohesin [4] (Fig. 1). This demonstrates the ability of enhancers to act independently of distance, position and orientation with respect to the target gene(s). Surprisingly, in a few rare cases, enhancers have been implicated in interchromosomal regulation, by activating transcriptional programs of target genes found on different chromosomes, although the legitimacy of these claims remains somewhat contentious [5,6]. Additionally, enhancer-bound transcription factors are responsible for the recruitment of histone modifying enzymes and ATP-dependent chromatin remodeling complexes, cofactors which aid in the unraveling of tightly condensed chromatin fibers and thus increase the accessibility of DNA to other transcription regulatory proteins [7].

Substantial advancements in computational and molecular biology techniques have enabled the implementation of genome-wide tools for analysis of the genomic and functional characteristics of enhancers. These methods revealed several chromatin features associated with enhancers, such as hypersensitivity to DNase I, the presence of specific

☆ This article is part of a Special Issue entitled: Plant Gene Regulatory Mechanisms and Networks, edited by Dr. Erich Grotewold and Dr. Nathan Springer.

* Corresponding author.
E-mail address: jjiang1@wisc.edu (J. Jiang).

http://dx.doi.org/10.1016/j.bbagrm.2016.06.006
1874-9399/

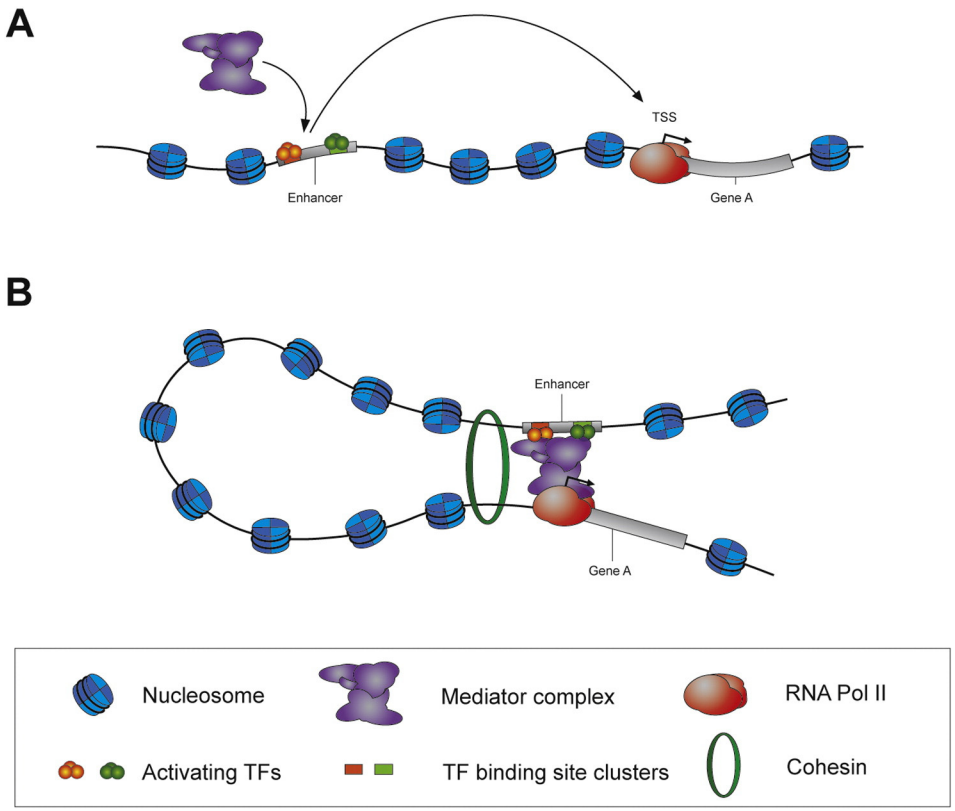

Fig. 1. A model for enhancer regulation of gene expression. Enhancers are distinct genomic regions that contain clusters of transcription factor binding sites which can enhance or up-regulate the transcription rate of a target gene(s), regardless of their location, orientation and distance with respect to the target(s). (A) Binding of transcription factors (TFs) to DNA motifs within the enhancer triggers the recruitment of the mediator complex and stimulates chromatin looping. (B) Cohesin stabilizes the enhancer-promoter interaction.

transcriptional cofactors, various histone modifications, chromatin looping, and nucleosome depletion and phasing. When analyzed congruently, enhancer-associated chromatin features lend an attractive and powerful avenue for the identification of novel enhancers. However, most of our current understanding of enhancer function stem from studies initiated by the Encyclopedia of DNA elements (ENCODE) project [8], which aimed to characterize and identify all functional elements, including enhancers, within the human genome. To date, no such effort has been initiated in any plant species, limiting our understanding of these regulatory elements in the context of the plant cells. Nevertheless, with the wide adoption of high-throughput sequencing methods coupled with enhancer-associated chromatin feature enrichment techniques, it has become more feasible than ever to answer basic questions in plant enhancer biology. This review presents the different methods pioneered by mammalian researchers for the identification of enhancers, as well as the genomic characteristics of enhancers that form the basis of these techniques. We also highlight studies, seminal and current, confirming enhancer activity through functional and genomic signature characterization, and discuss the future directions of enhancer research in plants.

2. Genome-wide features of enhancers

Recent advances in bioinformatics and molecular techniques have facilitated the adoption of genome-wide tools for enhancer prediction and identification. Several key chromatin features and enhancer-associated proteins have been thoroughly dissected in mammalian systems, enabling the translation of these approaches for use in the context of plants. A number of fundamental hallmarks have been used to locate enhancers, which include (*i*) TF-bound genomic regions; (*ii*) histone modifications of nucleosomes adjacent to enhancers; (*iii*) the presence of structural, transcriptional, and cofactor proteins, such as cohesin, RNAPII and the mediator complex distal to promoters; (*iv*) transcribed enhancer RNAs (eRNAs), and (*v*) open chromatin conformations. Here we present and discuss the roles of these chromatin features in enhancer biology as determined in mammalian studies as well as the methodologies used for their discovery. In addition, we evaluate the advantages and limitations of these methodologies in plant enhancer identification.

2.1. Transcriptional co-factors

One strategy to reveal genome-wide enhancers is by identifying their underlying chromatin features and associated proteins. To understand how these features are used to identify enhancers, we need to be familiar with the timing and specific roles of enhancer-associated proteins during enhancer-mediated transcription. Activation of an enhancer is strictly initiated by the binding of transcription factors to their cognate DNA motifs intrinsic to the enhancer sequence [9]. Upon DNA binding, TFs have been demonstrated to be responsible for recruiting the mediator complex, a major component of enhancer-directed transcription [10]. TFs additionally signal for the recruitment of chromatin remodeling complexes (Fig. 1), and histone modifying co-factors such as p300/CBP, proteins which directly regulate chromatin structure and accessibility [11]. Lastly, enhancers associate with transcriptional machinery such as general transcription factors and RNAPII. By mapping the locations of these transcriptional cofactors, several studies have demonstrated the effectiveness of this strategy for enhancer discovery.

A high-throughput method is required to reveal protein-DNA associations accurately and on a genome-wide scale. Chromatin immunoprecipitation (ChIP) with antibodies specific to enhancer-associated co-factors or transcription factors coupled with massively parallel sequencing (ChIP-seq) (Fig. 2A) has become the most common platform for the identification of the protein binding sites genome-wide [12,13]. It is

Please cite this article as: A.P. Marand, et al., Towards genome-wide prediction and characterization of enhancers in plants, Biochim. Biophys. Acta (2016), http://dx.doi.org/10.1016/j.bbagrm.2016.06.006