Plant Science 205–206 (2013) 55–62

Contents lists available at SciVerse ScienceDirect

Plant Science

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

Review

Genomic era analyses of RNA secondary structure and RNA-binding proteins reveal their significance to post-transcriptional regulation in plants

Ian M. Silverman[a,b,c,1], Fan Li[a,b,d,1], Brian D. Gregory[a,b,c,d,*]

[a] Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA
[b] PENN Genome Frontiers Institute, University of Pennsylvania, Philadelphia, PA 19104, USA
[c] Cell and Molecular Biology Graduate Group, University of Pennsylvania, Philadelphia, PA 19104, USA
[d] Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Received 1 December 2012
Received in revised form 22 January 2013
Accepted 23 January 2013
Available online 1 February 2013

Keywords:
RNA secondary structure
RNA-binding proteins
Post-transcriptional regulation
RNA genomics
RNA silencing
Small RNAs

ABSTRACT

The eukaryotic transcriptome is regulated both transcriptionally and post-transcriptionally. Transcriptional control was the major focus of early research efforts, while more recently post-transcriptional mechanisms have gained recognition for their significant regulatory importance. At the heart of post-transcriptional regulatory pathways are *cis*- and *trans*-acting features and factors including RNA secondary structure as well as RNA-binding proteins and their recognition sites on target RNAs. Recent advances in genomic methodologies have significantly improved our understanding of both RNA secondary structure and RNA-binding proteins and their regulatory effects within the eukaryotic transcriptome. In this review, we focus specifically on the collection of these regulatory moieties in plant transcriptomes. We describe the approaches for studying RNA secondary structure and RNA–protein interaction sites, with an emphasis on recent methodological advances that produce transcriptome-wide datasets. We discuss how these methods that include genome-wide RNA secondary structure determination and RNA–protein interaction site mapping are significantly improving our understanding of the functions of these two elements in post-transcriptional regulation. Finally, we delineate the need for additional genome-wide studies of RNA secondary structure and RNA–protein interactions in plants.

Contents

Abbreviations: dsRNase, double-stranded RNase; ssRNase, single-stranded RNase; miRNA, microRNA; PIP-seq, protein interaction profile sequencing; siRNA, small interfering RNA; PARS, parallel analysis of RNA structure.

* Corresponding author at: Department of Biology, University of Pennsylvania, 433 South University Avenue, Philadelphia, PA 19104, USA. Tel.: +1 215 746 4398; fax: +1 215 898 8780.
E-mail address: bdgregor@sas.upenn.edu (B.D. Gregory).
[1] These authors contributed equally to this work.

0168-9452/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.plantsci.2013.01.009

1. Introduction

Post-transcriptional regulation of the eukaryotic transcriptome can occur at any step of the RNA "life cycle" including maturation (e.g. splicing, polyadenylation, etc.); transport from the nucleus; localization within subcellular compartments; molecule stability; as well as the initiation, elongation, and termination of protein translation. Numerous *cis*- and *trans*-acting features and factors are integral to all of these regulatory processes.

The intrinsic secondary structure of RNA molecules is one such *cis*-acting feature. Secondary structure is the collection of intricate folding patterns that an RNA molecule forms through specific base pairing interactions encoded within its primary sequence [1–4]. Many RNA molecules cannot properly function without the formation of an extremely precise secondary structure [1–4]. For instance, ribosomal RNAs (rRNAs) must form structural folds that enable interactions with the correct ribosomal subunits at specific locations along their length, thereby allowing the formation of functional ribosomes [5]. Additionally, the structure of long non-coding RNAs (lncRNAs), not their primary sequence, drives their function in regulating gene expression [6,7]. Structural elements also affect the overall steady state abundance and stability of many eukaryotic mRNAs [8–10]. Thus, the secondary structure of RNAs is often required for their functionality in diverse cellular and regulatory processes.

RNA-binding proteins are a group of *trans*-acting regulatory factors that are integral to the post-transcriptional regulation of eukaryotic transcriptomes. A cellular RNA is involved in a multitude of complex interactions with numerous *trans*-acting RNA-binding proteins from the initial processing of a transcript in the nucleus to its final translation and decay in the cytoplasm [11–13]. These RNA-binding proteins interact with mRNAs and form dynamic multi-component ribonucleoprotein complexes, which can be the functional forms of mRNAs [14]. It is only through their proper formation that transcripts are correctly regulated and precisely produce the required amount of protein in a eukaryotic cell [2,11,13,14]. Thus, RNA–protein interactions are necessary for the functionality, processing, and regulation of many RNA molecules in plant cells. However, there is still much to be discovered about plant RNA-binding proteins and their interactions with target RNAs.

In this review, we explain the current understanding of the regulatory functions of RNA secondary structure and RNA-binding proteins in plant transcriptomes. We discuss advances in genomic technologies that allow a more comprehensive view of plant RNA secondary structure and RNA-binding proteins and their functional significance. Finally, we delineate experiments that will improve the understanding of these post-transcriptional regulatory elements and their effects on plant transcriptomes.

2. RNA secondary structure

RNA secondary structure is critical to post-transcriptional gene regulation. For example, riboswitches are a potent class of gene regulatory structural elements within mRNAs that directly bind to small metabolites. Other post-transcriptional processes such as protein translation [15,16] and RNA-mediated silencing [17,18] are also tightly controlled by structural features within RNA transcripts. In the next few sections, we summarize the role of RNA structure in numerous regulatory processes.

2.1. Riboswitches

Riboswitches are regulatory elements located within an mRNA that bind directly to small molecule ligands with no requirement for a protein partner. They mediate gene expression via ligand-induced conformational changes [19,20]. The thiamine pyrophospate riboswitch was first discovered in bacteria [21,22] and it can easily be distinguished by homology throughout the plant kingdom where it is located in the 3′ untranslated region of the *thiaminC* (*THIC* (*AT2G29630*)) gene that is required for thiamine biosynthesis [23,24]. The riboswitch controls the levels of mRNAs from this gene through conformational changes that either mask or unmask a 5′ splice site, and is one of the most thoroughly studied examples of gene regulation by RNA secondary structure in plants.

There are likely many other riboswitches encoded by plant transcriptomes that do not have bacterial orthologs (our unpublished results). This is of note because the search for additional riboswitches in plants has been primarily based on sequence homology to known bacterial riboswitch domains [25], and is limited by the lack of currently available data on RNA secondary structure in the plant lineage. Several novel genomics approaches (see below) could prove useful in detection of additional plant riboswitches by revealing the secondary structure of thousands of mRNAs in a single experiment.

2.2. A role for secondary structure in translation

RNA secondary structure also plays an important role in translational regulation. This was first demonstrated by studies in which stable stem-loop structures were shown to strongly modulate protein yield of plasmid-encoded preproinsulin [15,16]. A similar relationship between computationally predicted secondary structure and translational efficiency for both chloroplast and nuclear mRNAs was demonstrated in green algae and higher plants [26,27]. DNA microarrays have been used to assess translational regulation genome-wide in *Arabidopsis* under a variety of stress conditions including dehydration [28], sucrose starvation [29], salinity, and temperature [30]. These studies revealed a significant inverse correlation between computationally predicted thermodynamic stability and ribosome loading in the 5′ untranslated regions [31], suggesting that secondary structure could be important in mediating translational activity. Similarly, a strong correlation between genome-wide profiles of ribosome density and folding energy was identified in yeast [32], and a tendency for decreased secondary structure near the start codon of genes was observed for most cellular organisms and viruses [33]. In total, these results suggest that RNA secondary structure has a significant regulatory effect on overall protein translation from eukaryotic mRNAs.

2.3. RNA silencing pathways

Plant RNA silencing pathways are mediated by small RNAs, which consist primarily of microRNAs (miRNAs) and several classes of endogenous small interfering RNAs (siRNAs) [17,34]. miRNAs are short ~21–22 nucleotide RNAs that direct post-transcriptional or translational repression of specific mRNAs through direct base pairing interactions with complementary sites in the target transcript sequence. The miRNA-target interaction is thought to extend along the entire length of plant miRNAs [35]. However, in animals, this interaction mostly involves complementary base pairing only between nucleotides 2–8 of a miRNA (counted from its 5′ end) (seed region) and a binding site in a target transcript.

Various lines of evidence suggest that site accessibility mediates miRNA targeting efficiency [36–39]. The first study to incorporate target site structure in miRNA target prediction found 3-nucleotide accessible regions to be an important predictor of targeting efficiency in *Drosophila melanogaster* [38]. This observation was then extended to a more general trend of decreased structural complexity and increased accessibility in regions containing miRNA target sites [37]. A two-step nucleation/hybridization model based on Sfold-generated structure ensembles was used to identify