Plant Science 214 (2014) 1–12

Contents lists available at ScienceDirect

Plant Science

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

Review

Regulation of plant translation by upstream open reading frames

Albrecht G. von Arnim[a,b,*], Qidong Jia[b], Justin N. Vaughn[a,1]

[a] Department of Biochemistry, Cellular and Molecular Biology, The University of Tennessee, Knoxville, TN 37996-0840, USA
[b] Graduate School of Genome Science and Technology, The University of Tennessee, Knoxville, TN 37996-0840, USA

ARTICLE INFO

Article history:
Received 17 July 2013
Received in revised form 8 September 2013
Accepted 10 September 2013
Available online 18 September 2013

Keywords:
Protein synthesis
Translation initiation factor
Upstream open reading frame
Sensor
Development
Arabidopsis thaliana

ABSTRACT

We review the evidence that upstream open reading frames (uORFs) function as RNA sequence elements for post-transcriptional control of gene expression, specifically translation. uORFs are highly abundant in the genomes of angiosperms. Their negative effect on translation is often attenuated by ribosomal translation reinitiation, a process whose molecular biochemistry is still being investigated. Certain uORFs render translation responsive to small molecules, thus offering a path for metabolic control of gene expression in evolution and synthetic biology. In some cases, uORFs form modular logic gates in signal transduction. uORFs thus provide eukaryotes with a functionality analogous to, or comparable to, riboswitches and attenuators in prokaryotes. uORFs exist in many genes regulating development and point toward translational control of development. While many uORFs appear to be poorly conserved, and the number of genes with conserved-peptide uORFs is modest, many mRNAs have a conserved pattern of uORFs. Evolutionarily, the gain and loss of uORFs may be a widespread mechanism that diversifies gene expression patterns. Last but not least, this review includes a dedicated uORF database for Arabidopsis.

Contents

Abbreviations: AGI#, *Arabidopsis* gene identifier; CaMV, cauliflower mosaic virus; eIF, eukaryotic translation initiation factor; mORF, major open reading frame; NMD, nonsense mediated decay; RPL, ribosomal protein of the large subunit; RPS, ribosomal protein of the small subunit; uORF, upstream open reading frame; CPuORF, conserved peptide uORF; UTR, untranslated region.

* Corresponding author at: Department of Biochemistry, Cellular and Molecular Biology, The University of Tennessee, Knoxville, TN 37996-0840, United States. Tel.: +1 865 974 6206.
E-mail addresses: vonarnim@utk.edu (A.G. von Arnim), qjia2@utk.edu (Q. Jia), jnvaughn@uga.edu (J.N. Vaughn).
[1] Current address: Department of Genetics, University of Georgia, Athens, GA 30602-7223, USA.

0168-9452/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.plantsci.2013.09.006

1. Introduction

1.1. Definitions and early cases

Upstream open reading frames (uORFs) are protein coding regions in mRNAs that lie upstream of the main protein coding region, i.e. in the 5′ untranslated region of the mRNA (Fig. 1). Although counterintuitive, the so-called 5′ untranslated regions (5′ UTR) of mRNAs are often partially translated. According to Kozak's scanning model of translation initiation, the ribosome scans the mRNA from the 5′ cap and engages at the first AUG triplet that it encounters. If the first AUG is the start codon of a uORF, it typically reduces the efficiency of translation of the main coding region of the mRNA (major open reading frame or mORF) [1,2]. However, uORFs usually do not eliminate translation altogether. This was shown *in planta* with the maize transcription factors, *R/Lc* and *Opaque-2* [3,4].

One of the earliest and most unconventional uORF-based translational control systems was discovered in the pararetrovirus, cauliflower mosaic virus (CaMV)[5,6]. Briefly, the long CaMV 5′ leader contains six uORFs (Supplemental Fig. 1) and a long hairpin-loop structure downstream from the first uORF. A ribosome that has translated the first short uORF is competent to scan past the hairpin without unfolding it, an event called shunting. The shunting mechanism ensures that a specific region of the 5′ UTR remains free of ribosomes. This region contains an RNA encapsidation signal that binds to the CaMV coat protein [6,7]. It is notable that these viruses do not use an internal ribosome entry mechanism that maintains their 5′ UTR free of ribosomes. Internal ribosome entry sites are RNA sequence elements commonly used by metazoan viruses that direct ribosomes to specific sites adjacent to their translation start codons and circumvent cap-dependent translation initiation.

Until recently, shunting was thought to be a peculiarity of the pararetroviruses such as CaMV and the related pararetrovirus, rice tungro bacilliform virus. However, uORF-stimulated shunting was recently discovered in a picorna-like RNA virus (rice tungro spherical virus). Because the two rice viruses coexist together in the same host, it seems very likely that the RNA virus may have acquired the shunting mechanism by cohabitation with rice tungro bacilliform virus [8].

1.2. Types of uORFs and their distribution

The preponderance of uORFs is clearly biased with respect to gene function. Highly expressed mRNAs such as those of many housekeeping genes tend to have short 5′ UTRs that are devoid of uORFs. Poorly expressed mRNAs such as mRNAs for transcription factors and kinases often have longer 5′ UTRs and are rich in uORFs [9]. This feature is pan-eukaryotic and was first observed in 1987 [10]. These results suggest that the presence or absence of uORFs is most likely adaptive. However, not many uORFs have been directly examined for their functional significance at the whole plant level. Only in a handful of cases has the mutation of the uORF revealed significant growth defects [11–13]. In addition, very few uORFs were discovered by classical forward genetic analysis, that is, because a mutation in the uORF altered the phenotype of the plant [12,14].

Some uORFs overlap the major open reading frame of the mRNA (major ORF). A recent study in yeast concluded that, among all the possible mutations in a gene's 5′ upstream region, mutations that cause uORFs to overlap with the major ORF have the most dramatic inhibitory effect on gene expression [15].

uORFs are classified along evolutionary lines. For a fairly small fraction of uORFs the peptide sequence is noticeably conserved in evolution (Conserved peptide uORFs or CPuORFs). In these cases, the peptide sequence is key for translational repression. Several surveys on CPuORFs have been published [16–19]. CPuORFs have rightfully been assigned their own gene identifiers (AGI numbers) in Arabidopsis. In some cases, and in keeping with similar CPuORF peptides in fungi [20], the conserved peptide is hypothesized to stall the ribosome as a nascent peptide while located in the ribosome exit tunnel, thus blocking the progression of upstream ribosomes or suppressing reinitiation [11,21]. In other cases, the uORF peptide may exert its function after it has been released from the ribosome. Only two such cases are known. In one case, the peptide binds to the mRNA and destabilizes it [22]. In another case, the synthesized uORF peptide can inhibit translation when added to an in vitro translation system [23]. CPuORFs are currently sorted into more than 30 homology groups that are spread over more than 79 Arabidopsis genes [17,19,24], while up to 150 cereal genes are now estimated to have CPuORFs [18]. Most CPuORFs are conserved between monocots and dicots [16–18,25]. Lineage specific gain or loss of CPuORFs is uncommon [19], but it does occur. Arabidopsis often has uORF features different from those of other dicots [25]. For example, the highly conserved CPuORF in the mRNA for ribosomal protein S6 kinase has lost its AUG in the Brassicaceae lineage, and has been replaced by a different uORF in a different frame.

About 35% of Arabidopsis genes give rise to a uORF-containing mRNA, and about half of these have multiple uORFs [9]. Other plant species have similar fractions of transcripts with uORFs (Fig. 2A). The Arabidopsis genome encodes more than 20,000 uORFs, almost as many as major ORFs (Supplemental Data File S1). Because the AUG triplet is only slightly underrepresented in Arabidopsis 5′ UTRs [9], the number of uORFs is only slightly lower than predicted by chance alone. However, longer uORFs are more underrepresented than shorter ones (Fig. 2B).

The CPuORFs represent only a small fraction of all uORFs. However, the uORFs that do not qualify as CPuORFs (nonCPuORFs) commonly influence the level of gene expression of the major ORF (Table 1). Therefore, a large number of nonCPuORFs are functional. Then, what fraction of the nonCPuORFs was, or still is, subject to selection? What fraction of them has adaptive significance? There is evidence that uORFs other than the known CPuORFs are subject to selection. First, the presence of the nonCPuORF is sometimes conserved, even if their amino acid sequence is not [25]. Similar to the situation in mouse and human [26], the AUG triplet is the most frequently conserved triplet when 5′ UTRs from two plant families are aligned (Fig. 2C), consistent with the notion that many uAUGs are under stabilizing selection. Furthermore, certain gene networks that possess CPuORFs also include other genes with nonCPuORFs, for example the polyamine gene network (Fig. 3) and the auxin response transcription factors [27]. Finally, it should be recognized that the definition of a uORF as a CPuORF is constrained by our statistical power. All of these nonrandom patterns suggest that many nonCPuORFs are biologically significant.

Such observations notwithstanding, it is also evident that many uORFs are unconserved. It was speculated that uORFs might counterbalance evolutionary changes in transcriptional control [4]; this interesting idea of coevolution between transcriptional and post-transcriptional regulation has not been scrutinized.

1.3. How does the ribosome get past uORFs: Leaky scanning, shunting, and reinitiation

In vitro, uORFs suppress translation in a length dependent manner [2,28]. One of the more detailed in vivo surveys was performed in human cells [26] with >25 uORFs that inhibited protein expression between less than 5% and 100%. Although this study did not detect a correlation between uORF length and protein expression, evidence from plants (Fig. 2D) shows a moderate correlation between inhibition of gene expression and uORF length. uORFs of more than 16 codons can be expected to inhibit translation. In contrast,