The Plant Cell, Vol. 22: 3764–3777, November 2010, www.plantcell.org © 2010 American Society of Plant Biologists

Noncanonical Translation Initiation of the *Arabidopsis* Flowering Time and Alternative Polyadenylation Regulator FCA [C] [W]

Gordon G. Simpson,[a,1,2] Rebecca E. Laurie,[a,b,1] Paul P. Dijkwel,[a,3] Victor Quesada,[a,4] Peter A. Stockwell,[b] Caroline Dean,[a] and Richard C. Macknight[b,5]

[a] Department of Cell and Developmental Biology, John Innes Centre, Norfolk NR4 7UH, United Kingdom
[b] Department of Biochemistry, University of Otago, Dunedin 9054, New Zealand

The RNA binding protein FCA regulates the floral transition and is required for silencing RNAs corresponding to specific noncoding sequences in the *Arabidopsis thaliana* genome. Through interaction with the canonical RNA 3′ processing machinery, FCA affects alternative polyadenylation of many transcripts, including antisense RNAs at the locus encoding the floral repressor *FLC*. This potential for widespread alteration of gene regulation clearly needs to be tightly regulated, and we have previously shown that *FCA* expression is autoregulated through poly(A) site choice. Here, we show distinct layers of FCA regulation that involve sequences within the 5′ region that regulate noncanonical translation initiation and alter the expression profile. FCA translation in vivo occurs exclusively at a noncanonical CUG codon upstream of the first in-frame AUG. We fully define the upstream flanking sequences essential for its selection, revealing features that distinguish this from other non-AUG start site mechanisms. Bioinformatic analysis identified 10 additional *Arabidopsis* genes that likely initiate translation at a CUG codon. Our findings reveal further unexpected complexity in the regulation of *FCA* expression with implications for its roles in regulating flowering time and gene expression and more generally show plant mRNA exceptions to AUG translation initiation.

INTRODUCTION

The *Arabidopsis thaliana* RNA binding protein FCA was first identified through its function in flowering time regulation (Macknight et al., 1997). Loss-of-function *fca* mutants flower late, and FCA forms part of the genetically defined autonomous pathway that prevents the accumulation of mRNA encoding the floral repressor FLOWERING LOCUS C (FLC) (Koornneef et al., 1991; Michaels and Amasino, 1999).

FCA physically interacts with a highly conserved RNA 3′-end–processing factor FY, and this interaction is absolutely required for the function FCA performs in regulating flowering time (Simpson et al., 2003). FY is the *Arabidopsis* homolog of *Saccharomyces cerevisiae* Pfs2p (Ohnacker et al., 2000) and human cleavage and polyadenylation specificity factor WDR33 (Shi et al., 2009), and each of these proteins is required for the cleavage and polyadenylation of mRNA. FCA regulates *FLC* expression by affecting poly(A) site selection in noncoding antisense RNAs, which initiate in the 3′ region of the *FLC* locus (Hornyik et al., 2010; Liu et al., 2010). FCA activity correlates with alternative poly(A) site selection in the *FLC* antisense transcripts and transcriptional regulation of *FLC* expression. This function of FCA is dependent on FLOWERING LOCUS D, an *Arabidopsis* protein related to human lysine-specific demethylase (LSD1 or KDM1 under new nomenclature) (Liu et al., 2007; Baurle and Dean, 2008).

The regulation of *FCA* expression itself is unusually complex (Macknight et al., 2002). *FCA* pre-mRNA is alternatively spliced and subject to alternative polyadenylation (Macknight et al., 1997). Some of this alternative processing results from autoregulation of expression mediated by FCA (Quesada et al., 2003): FCA promotes promoter-proximal cleavage and polyadenylation within intron 3 of its own pre-mRNA, resulting in the formation of a prematurely truncated, inactive isoform (*FCA-β*) at the expense of full-length active *FCA-γ+* mRNA transcript (Quesada et al., 2003). This tight autoregulation suggests that the level of *FCA* expression is particularly important. This process is under developmental regulation and has a functional impact on the timing of flowering (Quesada et al., 2003).

Genetic evidence indicates that FCA regulates the floral transition by repressing *FLC* (Michaels and Amasino, 2001). However, FCA was found to play much wider roles in the *Arabidopsis*

[1] These authors contributed equally to this work.
[2] Current address: Division of Plant Sciences, College of Life Sciences, Dundee University at Scottish Crop Research Institute, Invergowrie, Scotland DD2 5DA, UK.
[3] Current address: Institute of Molecular BioSciences, Massey University, Private Bag 11222, Palmerston North, New Zealand.
[4] Current address: División de Genética, Departamento de Biología Aplicada e Instituto de Bioingenería, Edificio Vinalopó, Universidad Miguel Hernández, Avenida de la Universidad s/n, 03202 Elche (Alicante), Spain.
[5] Address correspondence to richard.macknight@otago.ac.nz.
The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) are: Caroline Dean (caroline.dean@bbsrc.ac.uk) and Richard C. Macknight (richard.macknight@otago.ac.nz).
[C] Some figures in this article are displayed in color online but in black and white in the print edition.
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.110.077990

genome. For example, FCA is required for mediating RNA silencing of endogenous loci in response to a signal from an inverted repeat transgene (Baurle et al., 2007). In addition, RNA sequences related to transposons, retrotransposons, and dispersed repeats that are normally silenced by the RNA-directed DNA methylation pathway are misregulated in *fca* mutants (Baurle et al., 2007). Loss of FCA also leads to defects in root development (Macknight et al., 2002) and gametophytic development (Baurle et al., 2007).

In experiments designed to examine whether increased expression of *FCA* could alter flowering time, we exchanged the *FCA* promoter and a large fraction of 5′ leader sequence for the cauliflower mosaic virus (CaMV) 35S promoter, thereby fusing it upstream of the first in-frame ATG codon of *FCA*. In the first transgenes analyzed, native *FCA* introns were retained, but we were unable to detect overexpressed FCA protein in the corresponding transgenic *Arabidopsis* plants (Quesada et al., 2003). This led us to discover FCA autoregulation through alternative poly(A) site selection within the *FCA* transcript (Quesada et al., 2003). Subsequently, by removing the introns from the transgenes, we were able to overexpress FCA protein. However, the protein produced in vivo was shorter than the endogenous wild-type FCA protein (Macknight et al., 2002). Here, we investigate this size difference and show that translation of FCA normally initiates upstream of the first in-frame AUG codon through use of a CUG triplet. This noncanonical initiation of translation appears to be conserved in *FCA* found in other plants. We identify *cis*-elements within the *FCA* 5′ untranslated region (UTR) that are required for CUG selection and that are not required for translation initiation when this codon is mutated to the canonical AUG triplet. We also show that the 5′ UTR region of *FCA* contains sequences that influence transcription. Bioinformatic analysis of the *Arabidopsis* genome identified 10 other genes that are likely to initiate translation at a CUG codon. These findings reveal an added complexity of *FCA* regulation and have implications for the roles that FCA might play in plant biology as well as for the study and annotation of plant proteins in general.

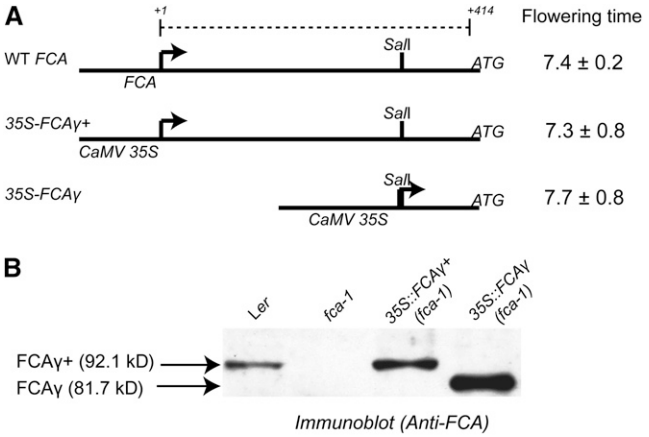

Figure 1. Fusion of the CaMV 35S Promoter at a *Sal*I Site Upstream of the First in-Frame ATG of *FCA* Results in Expression of a Truncated FCA Protein.

(A) Scheme of constructs designed to express FCA from the CaMV 35S promoter and the flowering time, determined as total leaf number at flowering, of wild-type (WT; Landsberg *erecta* [L*er*]) and the two constructs in an *fca-1* mutant background (mean ± SE calculated using at least 15 individuals for each genotype; *fca-1* typically flowers with ~30 leaves when grown in similar conditions).

(B) Immunoblot of FCA protein, detected by anti-FCA KL4 antibodies, in wild-type (L*er*) and *fca-1* mutant backgrounds compared with transgenic *FCA* expressed from the CaMV 35S promoter. Full-length FCA-γ+ and truncated FCAγ proteins are indicated by arrows.

RESULTS

Full-Length FCA Protein Expression Requires Sequences Upstream of the First in-Frame AUG

As part of our initial characterization of *FCA*, we generated a transgene designed to increase levels of FCA. The CaMV 35S promoter was fused at a convenient restriction site (*Sal*I, position +349 with respect to the *FCA* transcription start site) 65 bp upstream of the first in-frame +414ATG (+414 refers to the position of A with respect to the *FCA* transcription start site). The transgenic FCA protein expressed in these lines was ~10 kD shorter than the endogenous wild-type protein (Figure 1) (Macknight et al., 2002). We refer to the truncated protein as FCA-γ and the wild-type protein as FCA-γ+. An explanation for this discrepancy may relate to ectopic overexpression of *FCA* from the CaMV 35S promoter. Alternatively, this distinction may be related to the deletion of 349 nucleotides of upstream transcribed sequence in the course of constructing this transgene. To distinguish between these possibilities, a second transgene was made: in this case, the CaMV 35S promoter was fused to the *FCA* transcription start site and the constructs were introduced into mutant *fca-1* plants. FCA protein produced from this transgene complemented the late-flowering phenotype of these plants and was full length (FCA-γ+) as judged by its similar migration to the wild type when separated by SDS-PAGE (Figure 1). We therefore concluded that sequences upstream of the first in-frame ATG codon (414ATG) were necessary for full-length FCA-γ+ protein expression.

A Conserved Open Reading Frame Extends Upstream of the First in-Frame ATG to the FCA Transcription Start Site

Inspection of the *FCA* nucleotide sequence revealed its potential to encode an open reading frame (ORF) that extended throughout the entire 5′ region upstream of 414ATG to the transcription start site (Macknight et al., 1997) (see Supplemental Figure 1 online). Much of the nucleotide sequence in this region was relatively GC rich, compared with the typically AU-rich nature of higher plant 5′ UTRs (Gallie, 1996). Together, these findings raised the possibility that not all of the sequence upstream of 414ATG might constitute the *FCA* 5′ UTR. Instead, at least part of this region might be translated, and unusually, this would have to involve the selection of a noncanonical or non-AUG codon. In eukaryotes, the initiation of translation from non-AUG codons is apparently rare but is often associated with highly regulated genes (van der Velden and Thomas, 1999).

To begin to address the possibility that translation of *FCA* might initiate from a non-AUG codon, we inspected the *FCA*

sequence for codons that match criteria important for their selection as suggested by Kozak (1990, 1997). Only one in-frame codon, 141CTG, was identified that differed from ATG at only one position and had an adjacent purine at –3 and G at +4 (notably, the first in-frame ATG is in a poor Kozak consensus). Importantly, conceptual translation of *FCA* from 141CTG to 414ATG resulted in a predicted molecular mass of 10.4 kD, close, therefore, to the 10-kD size difference identified between the transgenic FCA proteins expressed from 414ATG (FCA-γ) and the transgenic protein expressed from the *FCA* transcription start site (FCA-γ+) (Figure 1). Transient expression studies of translation initiation from synthetic non-AUG start codons in plant cells previously found that, in plant protoplasts, CUG was the most effective triplet, exhibiting 30% of the AUG activity (Gordon et al., 1992). We therefore considered it possible that *FCA* translation initiated from 141CUG and set out to test this idea.

FCA Sequences Upstream of the First in-Frame AUG Codon Affect the Translation of a Reporter Protein in Vitro

To determine whether *FCA* sequences upstream of the first in-frame 414ATG could influence translation, their effect on a reporter protein (the spliceosomal protein, U1A) was examined through in vitro translation experiments. First, the sequence of *Arabidopsis FCA*, from transcription start site to 414ATG, was fused upstream of the U1A ATG start codon. Proteins were translated in vitro from RNA transcribed from this construct in rabbit reticulocyte extract in the presence of [^{35}S]Met. The fusion of the upstream *FCA* sequence to U1A resulted in the translation of a protein of greater apparent molecular mass than the U1A control (we refer to the larger protein as U1A+) (Figure 2). Consistent with this effect being related to translation, the introduction of a stop codon midway between 141CUG and 414AUG at position +249 in the *FCA* transcript (by point mutation of the triplet 249GAA to UAA [141CTG//STOP]) inhibited U1A protein production (Figure 2). This result is consistent with translation starting upstream of 414ATG. We next asked whether mutation of 141CTG would also affect the translation of the U1A reporter protein. Two different point mutations were introduced: first, a 141CTG to CTC change that results in a triplet differing from the canonical ATG at two positions was introduced (CTC, like CTG codes for the amino acid Leu). Second, a 141CTG-to-ATG change that results in the reconstruction of a canonical translation start codon was made. Consistent with 141CUG being used as the initiating codon in vitro, the 141CTG-to-CTC mutation blocked the formation of full-length U1A+ protein (Figure 2). However, a protein matching the size of U1A+ produced from the fusion of wild-type FCA sequence was produced from RNA bearing the 141CUG-to-AUG mutation (Figure 2).

In summary, these data support the idea that *FCA* sequences upstream of 414AUG harbor a noncanonical translation initiation codon and specifically indicate that this codon is 141CUG. Since this noncanonical translation initiation could be conferred upon a heterologous protein by fusion of the sequence upstream of *FCA* 414AUG, it indicates that sequences necessary and sufficient for noncanonical translation initiation are coded within this region.

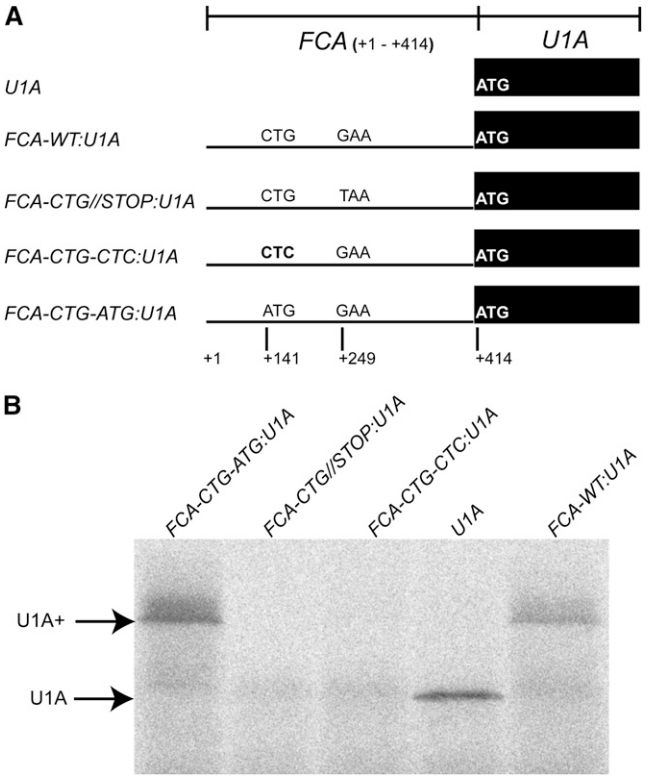

Figure 2. In Vitro Translation of a Reporter Protein, U1A, Expressed from RNA Derived from *FCA-U1A* Fusions in Rabbit Reticulocyte Lysates.

(A) Schematic representation of DNA constructs used: Wild-type *FCA* sequence from +1 to +414 and related sequences bearing point mutations were fused upstream of *U1A*. The position of +141CTG, +249GAA (which was mutated to a stop codon in one construct), and the first in-frame +414ATG are shown. Expression was driven by T7 RNA polymerase.

(B) Autoradiograph of proteins expressed in vitro in the presence of [^{35}S]Met from constructs depicted in **(A)**. The position of full-length U1A protein and the size-shifted U1A+ protein detected with FCA-CTG-ATG:U1A and FCA-WT:U1A fusions are indicated with an arrow.

Stable Expression of Transgenes Bearing Point Mutations in *FCA* Provides Evidence for Noncanonical Translation Initiation in Vivo

We next asked whether the results found in vitro using a heterologous host extract were relevant to what might happen in *Arabidopsis* itself. We therefore expressed *FCA* transgenes carrying the same point mutations in the sequence upstream of 414ATG as had been tested in vitro (Figure 2). Transgene expression was driven by the native *FCA* promoter. These mutant *FCA* transgenes were introduced into *fca-1* loss-of-function mutant backgrounds, and at least three independent transgenic lines were analyzed for each construct (Figures 3A to 3C). The *fca-1* mutant accumulates reduced levels of truncated FCA, as shown in Figure 3C (Quesada et al., 2003). This enables the detection of full-length FCA expression from the test transgenes using anti-FCA antibodies without an interfering signal from endogenous FCA. Simultaneously, it enables the

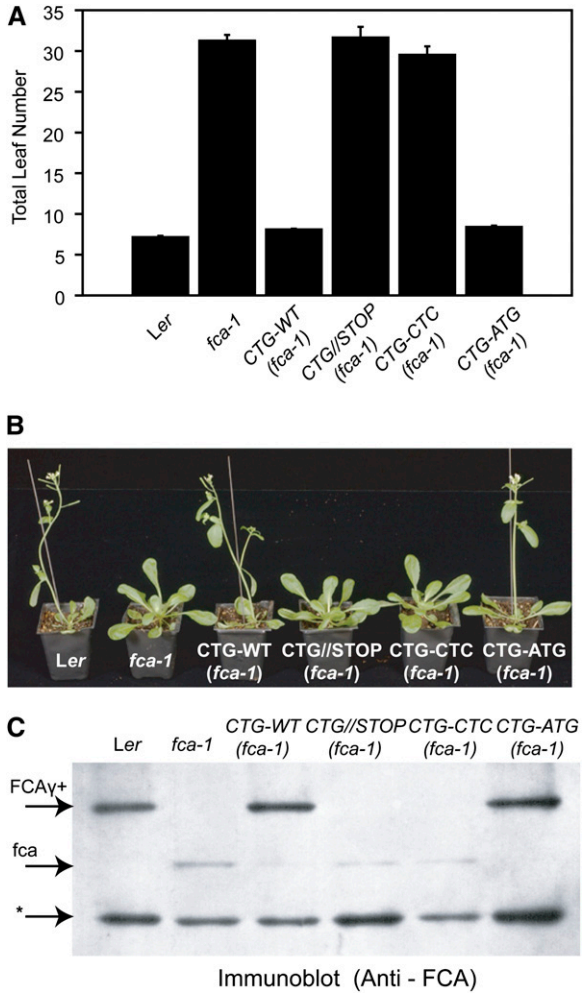

Figure 3. Flowering Time and Expression Analysis of *FCA* Transgenes Bearing Single Point Mutations Upstream of +414ATG When Introduced into a Mutant *fca-1* Background.

(A) Quantification of flowering time, determined as total leaf number after flowering (mean ± SE calculated using at least 20 individuals for each genotype).

(B) Representative plants illustrating the phenotypic difference between early and late flowering genotypes. Early flowering plants, such as wild-type L*er* have already flowered, while other genotypes grown alongside in identical conditions for the same period of time remain in the vegetative state and continue to produce rosette leaves.

(C) Immunoblot showing the expression of FCA in representative genotypes of wild-type, mutant *fca-1*, and *fca-1* plants expressing mutated and wild-type *FCA* transgenes. The position of full-length FCAγ+ and truncated mutant protein (fca) expressed in *fca-1* is indicated by arrows. Asterisk denotes a non-FCA-derived cross-reacting product that serves as an internal control.

[See online article for color version of this figure.]

assessment of the functional significance of this unusual translation by measuring the degree to which this late flowering phenotype is complemented by the test *FCA* transgenes (Figures 3A and 3B). Wild-type *FCA* (lacking introns) expressed from the native *FCA* promoter complements the *fca-1* mutation (Macknight et al., 2002). By contrast, a transgene bearing the point mutation tested in vitro that introduces a stop codon midway between 141CTG and 414ATG (141CTG//STOP) failed to complement the late flowering phenotype of *fca-1* (Figures 3A and 3B). Consistent with this, no full-length FCA-γ+ protein could be detected in these transgenic lines (Figure 3C). We next introduced the transgenes bearing point mutations in 141CTG itself: 141CTG to CTC and 141CTG to ATG. Transgenic lines harboring the 141CTG-to-CTC mutant *FCA* transgene flowered late and produced no full-length FCA-γ+ protein (Figures 3A to 3C). By contrast, the 141CTG-to-ATG lines flowered early and expressed FCA protein that migrated with the same apparent mobility as wild-type FCA-γ+ (Figures 3A to 3C).

FCA negatively regulates its own expression by modulating the site of 3′ end formation in its own pre-mRNA (Quesada et al., 2003). Consistent with this negative feedback (and the activity of the transgenic FCA proteins in flowering time regulation), those backgrounds expressing full-length active FCA protein show reduced levels of the endogenous truncated FCA mutant protein (Figure 3C). The genotype of these lines was confirmed by derived cleaved-amplified polymorphic sequence analysis (Neff et al., 1998; Quesada et al., 2003).

In summary, the results obtained in vivo with the transgenic lines expressing mutant *FCA* transgenes closely resemble those obtained in vitro with constructs fusing mutated *FCA* sequence to a heterologous reporter protein. Together, the results are consistent with *FCA* translation initiating from a noncanonical CUG codon.

141CTG and the Deduced ORF Upstream of the First in-Frame ATG Are Conserved in Homologs of *FCA* from Other Plant Species

The availability of sequence data for *FCA* homologs from other plant species indicates that the selection of a noncanonical codon may not be restricted to *Arabidopsis*. For example, an ORF extends upstream of the first in-frame ATG codon of rapeseed (*Brassica napus*), cotton (*Gossypium hirsutum*), and *Medicago truncatula FCA.* Comparison of these potential ORFs with that encoded by *Arabidopsis FCA* sequence +1 to +414 reveals significant stretches of homology (Figure 4). In each case, a CTG at a position equivalent to 141CTG is found that lies within a good Kozak consensus. Thus, in contrast with the situation with *Arabidopsis AGAMOUS*, where noncanonical ACG usage is not conserved in homologous genes from other species (Riechmann et al., 1999), this sequence data raises the possibility that this unusual form of translation is conserved in *FCA* homologs. Further consideration of related *FCA* sequences from other species reveals stretches of conservation of the predicted ORF encoded upstream of the first in-frame ATG of *Arabidopsis FCA* as well. However, in the case of tomato (*Solanum lycopersicum*) and grape (*Vitis vinifera*), for example, an ATG codon is found at the position corresponding to 141CTG. The conservation of this ORF and the presence of a conventional ATG codon at the position corresponding to 141CTG is consistent with our experimental data, revealing that *Arabidopsis* FCA protein translation begins upstream of the first in-frame AUG codon. Initiation of translation at a CUG codon is likely to be a conserved, but not obligatory, feature of FCA protein expression.

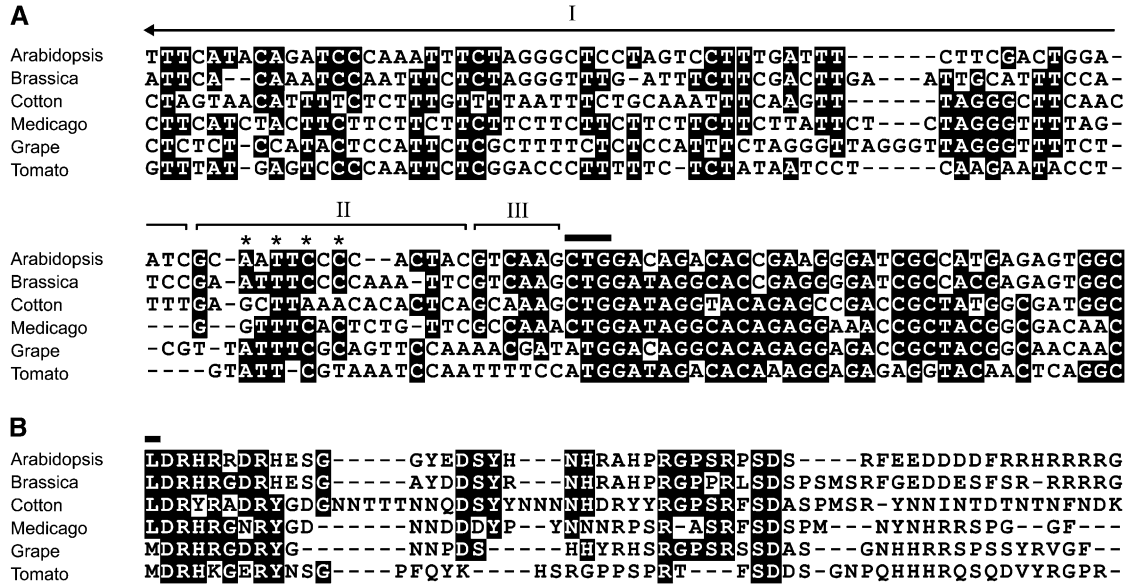

Figure 4. *FCA* DNA and Peptide Sequence Are Highly Conserved at the *CUG* Start Codon.

DNA **(A)** and protein **(B)** sequence alignments of various plant *FCAs* (*Arabidopsis*; Brassica, *Brassica napus*; cotton, *Gossypium*; Medicago, *Medicago truncatula*; grape, *Vitus vinifera*; tomato, *Solanum lycopersicon*). DNA nucleotides and amino acids with >60% identity are shaded black. The position of 141CTG is marked by a thick black line **(A)** and corresponds to the first amino acid in **(B)**. Nucleotides mutated to create the *CTG region II* mutant are indicated by an asterisk. I, II, and III (nucleotides 1 to 60 are not shown) represent three regulatory regions within the *Arabidopsis FCA* 5′ UTR.

Sequences Downstream of 141CUG Are Not Required for Translation Initiation at *FCA* 141CUG

To define the sequence features required for noncanonical CUG translation initiation, we established a transient assay experimental system. The sequence of *FCA* between the transcription start site and 450ATG (the second in-frame ATG) was fused to a firefly luciferase reporter gene and expressed with the CaMV 35S promoter. This construct also contained a 35S promoter-*Renilla* luciferase gene as an internal control (Hellens et al., 2005). To determine whether any effect was specific to noncanonical initiation codons, we prepared an identical construct carrying the 141CTG-to-ATG mutation used in the experiments described in Figure 3. Both the 141CTG- and 141ATG-containing constructs were transiently expressed in *Nicotiana benthamiana* leaves. The strong CaMV 35S promoter and heterologous expression system ensured similar levels of transcription between different constructs. This was confirmed by measuring the reporter RNA levels with quantitative RT-PCR, and the results are expressed as a ratio of luciferase activity versus relative luciferase mRNA levels. Surprisingly, we found that there was only a modest (1.3-fold) increase in translation efficiency when an AUG rather than a CUG codon was present at position +141 (Figure 5, i and ii). This result was confirmed in stable transgenic *Arabidopsis* lines where the CaMV 35S promoter was used to drive an in-frame fusion between the FCAs 5′ region (+1 to 450), containing either CUG or AUG at position 141, and β-glucuronidase (GUS) (see Supplemental Figure 2 online). We then examined whether translation initiated at the 141CUG or AUG codon of RNA expressed from these reporter gene constructs by deleting the downstream AUG codons (306 to 450) and mutating the 141CUG to CUC. Deletion of the downstream AUG codons had little effect on the level of luciferase activity, whereas mutating 141CUG to CUC abolished activity (Figure 5, iii to v). Therefore, as with stable transgenic *Arabidopsis* lines and in vitro translation experiments, this experimental system confirms translation initiation at a Leu-coding CUG, but not a CUC codon, within *FCA* 5′ sequences. Again, there was little difference in the levels of luciferase produced from the CUG- and AUG-containing constructs (Figure 5, iii and iv). Thus, in contrast with examples where translation initiation from synthetic constructs was shown to occur ~3-fold less efficiently at CUG compared with AUG codons (Gordon et al., 1992), translation initiation at *FCA* 141CUG was similar to an AUG codon at the corresponding positions.

Noncanonical translation initiation can be enhanced when an RNA hairpin is located 14 nucleotides from the initiating codon, as this corresponds to the distance from the leading edge of the ribosome to the AUG recognition center (Kozak, 1990). To investigate if such sequences were required for FCA translation initiation, the region immediately downstream of 141CUG (from nucleotides 147 to 188) was deleted. We found that this deletion had no effect on translation initiation as measured by luciferase activity (Figure 5, vi). Furthermore, deletion of two downstream regions predicted to form RNA secondary structures (region 195 to 239 and region 249 to 287; Figure 5, vii and viii, respectively) had little effect on translation initiation either. In fact, deletion of all but the next codon after 141CUG (region 147 to 450; Figure 5, ix) did not prevent 141CUG translation initiation as measured by luciferase activity. We therefore conclude that sequences

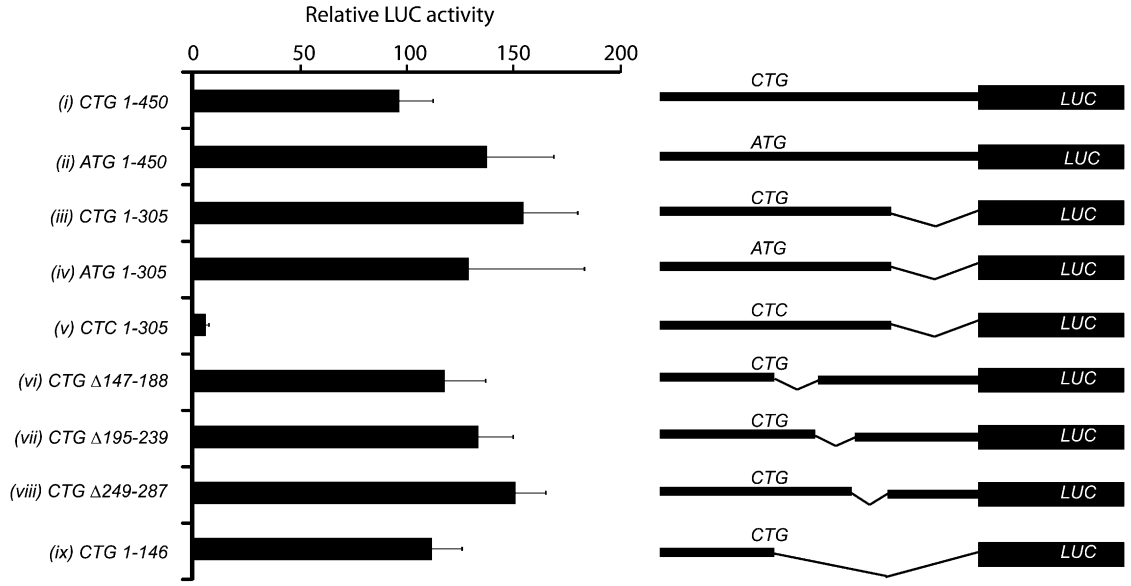

Figure 5. Downstream Sequences Are Not Required for Translation of *FCA* from 141CUG.

Relative luciferase (LUC) activities produced from various constructs containing 141CUG, 141AUG, or 141CUC initiation codons fused to a firefly luciferase reporter gene and expressed using an *N. benthamiana* transient assay system. Relative LUC activity represents the ratio of LUC activity versus relative *LUC* mRNA levels to ensure any differences are due to altered translational efficiency rather than RNA levels (data represent the mean ± SE from at least 10 independent assays**)**. Variation in transformation efficiencies between *Agrobacterium* infiltrations of *N. benthamiana* leaves were corrected using the ratio of *Renilla* luciferase activity. The name of each construct indicates the codon at position 141 and the amount of *FCA* sequence, numbered from the transcription start, that was fused in frame with luciferase. The constructs are shown with *FCA* sequence as thick lines, deletions as thin lines, and luciferase sequence as rectangles.

downstream of 141CUG are not required for its selection as an initiation codon. Interestingly, deletion of downstream sequences from constructs bearing the corresponding 141CUG-AUG mutation resulted in increased translation of mRNA by about 2-fold (see Supplemental Figure 3 online). Therefore, although RNA sequences downstream of 141CUG are not required for its selection, they contribute to the relative efficiency of translation initiation of an AUG codon in the same context, revealing contrasting features of CUG and AUG initiation.

Overall, these data reveal that sequences downstream of 141CUG are not required for its selection. Thus, unlike other examples of noncanonical start codon usage (Kozak, 2002), *FCA* translation initiation is not explained by downstream secondary structure slowing ribosome scanning and enabling CUG codon recognition.

Multiple Upstream Sequences within the 5′ UTR Are Required for Efficient CUG Translation Initiation of *FCA*

Given that sequences downstream of 141CUG were not required for its selection as an initiation codon, we next examined the potential role upstream sequences might play in this same experimental system: First, in the same pair of firefly and *Renilla* luciferase reporter constructs used in Figure 5, i and ii (and shown again in Figure 6A, i and ii), we deleted all but six base pairs (deletion 1-134) upstream of 141CTG and 141ATG (Figure 6A, iii and iv). This deletion almost abolished translation of mRNA expressing 141CUG but had a minor effect on translation of mRNA carrying the 141CUG-AUG mutation (Figure 6A, iii and iv). Thus, sequences upstream of position 141 are specifically required for CUG-dependent, but not AUG-dependent, translation initiation.

To determine what features of the 5′ UTR were required for 141CUG translation initiation, we first asked whether the impact of this deletion was derived from the loss of specific *cis*-acting elements or simply the loss of upstream sequence, as the length of mRNAs we were comparing was now different. To address this, we synthesized a construct designed to express an mRNA with a 5′ UTR of the same length and GC content upstream of 141CUG as in *FCA*, but comprising a randomized sequence that differed from wild-type *FCA* in region 1-134. Again, almost no luciferase activity was detected upon expression of this mRNA (Figure 6A, v). This finding reveals that specific *cis*-elements, rather than a particular 5′ UTR length, are required for 141CUG translation initiation.

What might these *cis*-elements be? Alignment of the sequence of *FCA* 5′ UTR from a range of plants revealed little similarity, apart from a weakly conserved region between −6 and −21 nucleotides upstream of 141CUG (Figure 4, region II). To investigate whether this conserved region affected noncanonical translation initiation, we first asked whether this sequence was sufficient to promote CUG translation initiation. Constructs were generated with all but 21 nucleotides of the 5′ UTR either deleted or comprised of randomized nucleotides. When compared with

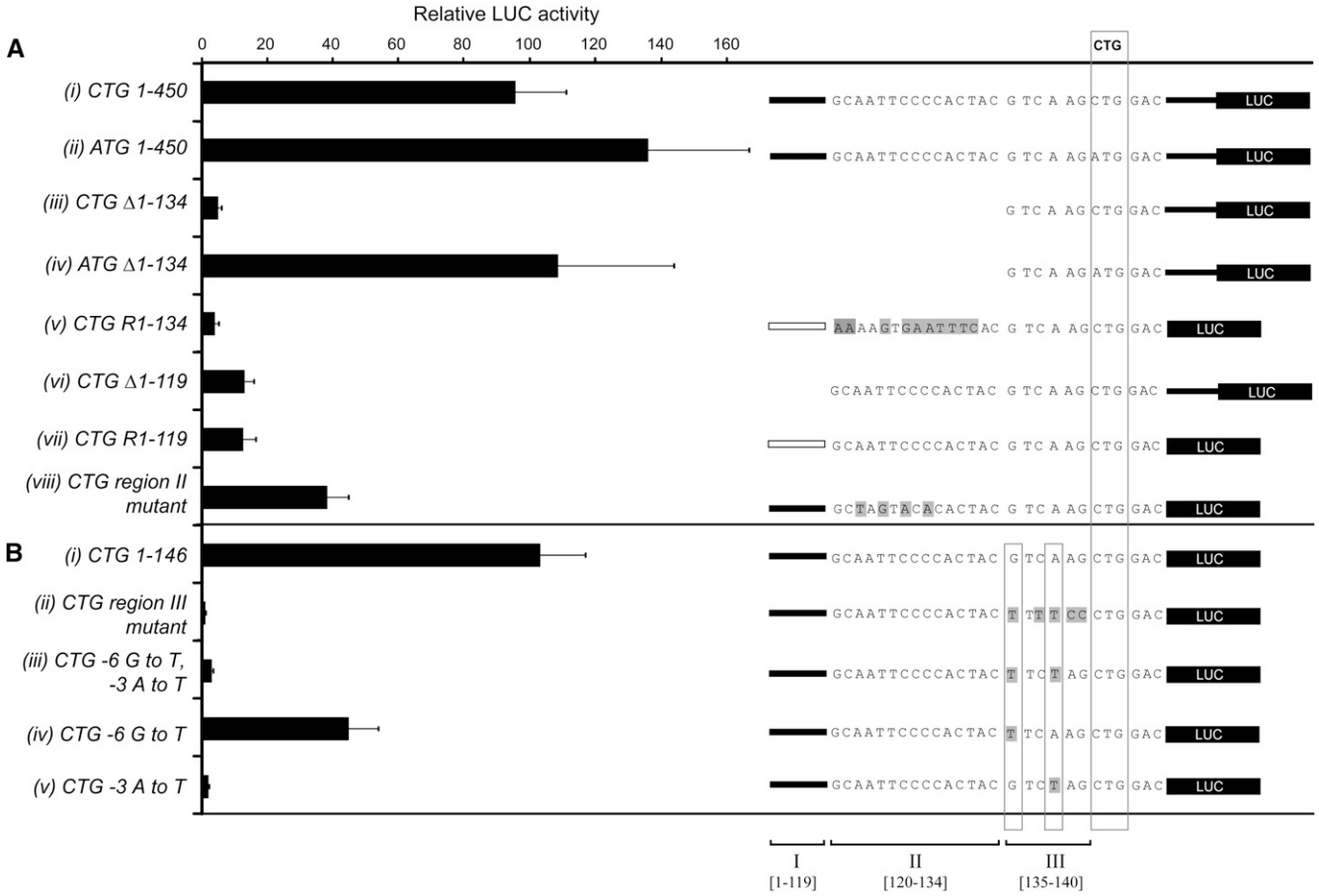

Figure 6. Sequences within the *FCA* 5′ UTR Are Required for *141CUG* Translation Initiation.

Relative luciferase (LUC) activities produced from constructs containing deletions or mutations within the 5′ UTR **(A)** and mutations that alter the context of the 141CUG initiation codon **(B)**. Relative LUC activities are given as a ratio of LUC activity versus relative *LUC* mRNA levels (data represent the mean ± SE from at least 10 independent assays). The constructs were transiently expressed in *N. benthamiana* leaves, and variations in transformation efficiencies between *Agrobacterium* infiltrations of leaves were corrected for using *Renilla* luciferase activity. The *FCA* sequence within each construct is shown numerically (i.e., *CTG 1-450* incorporates the *FCA* sequence from transcription start to position +450) and diagrammatically. Individual point mutations are shaded gray, and deleted bases are not shown. Black bars represent 119 bp of the *FCA* sequence upstream of 141NTG or 300 bp of downstream sequence (not to scale). Unfilled bars indicate 119 bp of randomly generated sequence (www.molbiol.ru/) containing 37% GC content, with no introduced translation initiation codons.

translation measured from constructs containing only six nucleotides of wild-type *FCA* 5′ UTR (nucleotides 135 to 140) (Figure 6A, iii and v), the constructs containing a 21-nucleotide 5′ UTR (nucleotides 120 to 140) either alone or with extra randomized 5′ UTR sequences resulted in about a 3-fold increase in luciferase activity (Figure 6A, vi and vii). Therefore, this conserved region II is sufficient to enhance CUG translation initiation efficiency.

Four nucleotides in region II are conserved between *FCA* sequences from other species (nucleotides marked with an asterisk in Figure 4). To test if these nucleotides were required for CUG translation initiation, we mutated them in a construct that contains the entire 5′ UTR. This specific disruption of region II resulted in 60% less detectable luciferase activity (Figure 6A, viii) compared with wild-type FCA 5′ sequences (Figure 6B, i), revealing that not only was region II sufficient for CUG translation initiation, but that it was also required for complete wild-type levels of recognition.

Other upstream sequences, in addition to region II, contribute to the efficiency of noncanonical translation efficiency: we found that constructs expressing the entire 5′ UTR (e.g., Figure 6A, i, and 6B, i) result in ~8-fold higher luciferase activity than constructs in which nucleotides 1 to 120 of the *FCA* 5′ UTR (which we refer to as region I) have been either deleted or replaced by randomized sequence of the same length and GC content (Figure 6A, vi and vii). Together, these experiments demonstrate that the 5′ UTR of *FCA* contains at least two *cis*-elements (present within regions I and II) required to promote efficient translation initiation at 141CUG, but which are dispensable for AUG translation initiation in the same context.

We next attempted to establish whether noncanonical translation might play role in regulating FCA expression. However, while the *FCA* promoter and 5′ *FCA* sequence (either just the 5′ UTR [*FCA* 1-141] or 5′ UTR and ORF to 483ATG) was capable of

driving comparatively high levels of GUS expression, the *FCA* promoter alone resulted in almost no detectable GUS activity or mRNA (see Supplemental Figure 4 online). Thus, the *FCA* 5′ UTR, which is required for CUG translation initiation, also contains elements that affect *FCA* expression. In support of this conclusion, the *FCA* 5′ UTR was able to drive GUS expression when fused downstream of a minimal TATA box CaMV 35S promoter (see Supplemental Figure 4C online). The fact that the FCA 5′ region contains both transcriptional and translational control elements as well as the FCA ORF will make dissecting the biological role of these sequences more difficult.

FCA Translation Initiation Depends on 141CUG Context

Apart from FCA, the only other *Arabidopsis* mRNA that has been shown to initiate translation at a CUG codon is that expressed from *Polγ2*. Using a wheat germ in vitro translation system, Wamboldt et al. (2009) provided evidence that an AAG purine triplet preceding CUG is required for Polγ2 CUG translation. Mutating either the -3A or -2A to C did not prevent CUG translation initiation; however, mutating all three purines to pyrimidines (AAG to TTC) abolished translation. Wamboldt et al. (2009) hypothesized that the minimal requirement for CUG translation initiation is the context A/G A/G A/G $\underline{C^{+1}UG}$ G^{+4} and demonstrated that the addition of the sequence AAG $\underline{CUG}$ G (the introduced in-frame CUG is underlined) upstream of the canonical AUG within At2g15790 (cyclophilin 40) mRNA, was sufficient to promote CUG translation initiation in wheat germ in vitro translation assays. *Arabidopsis FCA* also has a purine triplet at the 5′ flank of *FCA* 141CTG. However, we showed that a purine triplet was not sufficient to allow *FCA* translation initiation in planta and that specific *cis*-elements within the 5′ UTR are required (Figure 6A), indicating that features of noncanonical *FCA* translation initiation differ from the leaky scanning mechanism established for *Polγ2*. We therefore asked what contribution the sequence context of 141CUG made to FCA noncanonical translation initiation.

A purine triplet upstream of 141CUG, along with a G at −6 and +4, is a conserved feature of *FCA* mRNAs from species other than *Arabidopsis* that have a predicted CUG initiation codon. However, they are not a conserved feature of *FCA* mRNAs predicted to initiate translation at an AUG codon (these sequences are annotated as region III in Figures 4 and 6). To determine if this conserved sequence is required for noncanonical translation initiation, we mutated five of the six nucleotides immediately upstream of 141CUG, converting purines to pyrimidines and matching the corresponding sequence in tomato *FCA*, which has an AUG initiation codon (Figure 6B, ii). These mutations resulted in an ~100-fold decrease in luciferase activity, indicating that the upstream purines are essential for 141CUG translation initiation in *FCA* mRNA. The importance of the −6 and −3 purines was then examined by creating a construct in which these two nucleotides were converted to thymines (G135 to T and A138 to T). These mutations almost abolished luciferase activity (Figure 6B, iii). We then examined the role of the individual purines. The point mutation at −6 (G135 to T) resulted in an ~2-fold reduction in luciferase activity (Figure 6B, iv), but the point mutation converting the −3 purine to a pyrimidine (A138 to T) almost abolished activity (Figure 6B, v). These results demonstrate that sequence immediately 5′ of the CUG (within region III of Figure 6) is important for noncanonical translation initiation of FCA and that a purine (adenine) at −3 is essential.

Our results support the conclusion of Wamboldt et al. (2009) that upstream purines are necessary for FCA CUG translation initiation but, in contrast with Wamboldt et al. (2009), we find that they are not sufficient for noncanonical FCA translation initiation.

CUG Translation Initiation Is a Conserved Feature of a Small Proportion of Plant Genes

A previous computational search for CUG codons (with the consensus Axx $\underline{CUG}$ G) upstream of in-frame AUG codons in *Arabidopsis* 5′ UTRs identified 76 candidate mRNAs that might initiate translation from a CUG codon (Wamboldt et al., 2009). Our in planta data suggest that while such a context is required, it is not sufficient for CUG translation initiation to occur. If CUG translation initiation were biologically significant, it is likely to be conserved between orthologous genes in other plants. Therefore, to investigate if other *Arabidopsis* mRNAs might initiate at a CUG, we used an alternative approach that incorporated diverse plant EST and genomic sequence databases to identify mRNAs with conservation between an upstream CUG and the annotated AUG start codon.

Our experimental data define the following criteria for CUG translation initiation of *Arabidopsis FCA* mRNA: first, it contains an upstream CUG codon, with G at +4, that is in frame with the annotated AUG of the *FCA* ORF; second, the deduced translated sequence between CUG to AUG sequences is conserved in *FCA* orthologs from other plant species; third, this amino acid conservation is not observed in sequence upstream of Leu/CUG. We undertook a bioinformatic analysis using these criteria to identify other mRNAs that might initiate translation at a CUG codon. A flow diagram of the steps involved is shown in Figure 7. We identified 856 genes with upstream in-frame CTG codons (with a G at +4), but only 10 (in addition to *FCA* and *Polγ2*) in which the putative CUG translation initiation codon and translated sequence between the CUG and the downstream annotated AUG start codon were conserved in putative orthologous sequences from other plants (Table 1; see Supplemental Figure 5 online for further details). Interestingly, for two of the genes, a GUG rather than a CUG codon was present in some species. GUG has also been shown to function as a translation initiation codon, although at lower efficiency (Gordon et al., 1992; Depeiges et al., 2006). Next, we examined if the 10 genes identified in our bioinformatic search share any common sequence elements that might be required for CUG translation initiation and found that a purine was present at −6 in 80% of the genes and present at −3 in all genes, and that a guanine was always at −4 (Figure 8; see Supplemental Figure 6 online). Analyzed in this way, our findings suggest that CUG translation initiation is relatively rare, although our approach is quite conservative and some genes might be missed, for example, if orthologs have not been sequenced or if there is insufficient translated sequence between CUG and the annotated AUG for BLAST matching.

To validate these candidate examples of noncanonical CUG translation initiation, the deduced 5′ UTR of each gene was fused

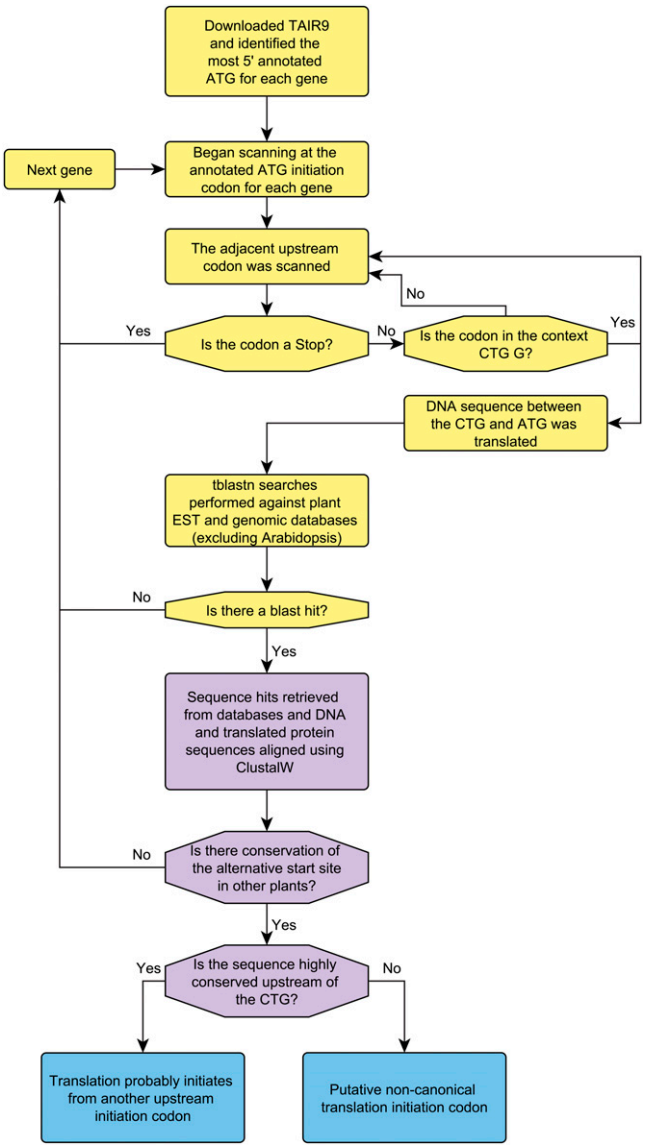

Figure 7. Flow Diagram Explaining the Bioinformatic Search for Genes That Initiate Translation at CUG.

Processes described in orange boxes were performed using computer algorithms, whereas purple boxes show manual processes.
[See online article for color version of this figure.]

in frame with the luciferase reporter gene under the control of the CaMV 35S promoter and translation examined using infiltrated *N. benthamiana* leaves. As with our previous analysis using this experimental system, luciferase activity should only be observed if translation is initiated at a noncanonical codon. We examined four genes with a conserved upstream CUG (and for comparison *FCA*) and four genes without a conserved upstream CUG (i.e., the upstream CUGs were only found in *Arabidopsis*). The upstream CUG codon of all the genes examined was in a good context (i.e., a purine at −3 and a G at +4). Luciferase activity was detected from FCA and the other four constructs derived from genes with conserved CUG initiation codons (Table 2) but barely detectable from constructs derived from two genes lacking conserved CUG codons (At3g05580 and At3g58690), while another (At4g21120) had luciferase activity lower than those constructs derived from the conserved CUG set, and At3g50830 had luciferase activity that was readily detectable (Table 2). These data also reinforce our findings with *FCA* that an in-frame CUG in a good context is not sufficient for efficient translation initiation and that other sequences contribute to CUG translation initiation.

Leaky scanning translation initiation in *Polγ2* mRNA at both a CUG and a downstream AUG allows dual targeting of the protein to chloroplasts and mitochondria, respectively (Christensen et al., 2005). For five of the genes identified with our search criteria, dual targeting to different organelles was also predicted, dependent on whether initiation occurred at CUG or AUG (Table 1; see Supplemental Table 1 online for prediction details). However, for the other five deduced proteins, there was no indication of dual targeting. Thus, while noncanonical translation and leaky scanning can facilitate dual targeting for protein isoforms, this is not the sole purpose of noncanonical translation initiation as exemplified by FCA and nor is leaky scanning the universal mechanism by which such codons are selected.

DISCUSSION

We report here that translation of *FCA* initiates upstream of the first in-frame AUG codon. By combining in vitro translation and in vivo genetic approaches, we obtained evidence consistent with translation of *FCA* initiating from a CUG triplet. This is not a feature unique to *Arabidopsis FCA*, as conservation of sequence strongly indicates that *FCA* from other plants also initiates translation at a CUG codon. Specific *cis*-elements upstream of this CUG codon are required for its selection but not for initiation at an AUG codon at the same position. This implies that FCA translation has the potential to be regulated in a specific manner. Our findings have wider implications for understanding how plant proteins in general are translated and determining how plant genome sequences should be annotated.

Noncanonical translation initiation is unexpected, and alternative explanations for our findings could invoke a role for splicing of the 5′ region of *FCA*. For example, it is possible that a splicing event within the 5′ region of *FCA* creates an in-frame AUG codon that is the site of translation initiation. Alternatively, translation initiation at an upstream AUG codon in a different reading frame coupled with a splicing event within the 5′ leader could recreate the FCA ORF of full-length FCAγ+. However, our experimental approaches rule out a role for splicing. First, in a detailed set of analyses of alternative processing of *FCA* pre-mRNA (Macknight et al., 1997, 2002; Quesada et al., 2003), we never detected splicing events in the 5′ region of *FCA* pre-mRNA. Second, the absolute sensitivity of *FCA* translation to the 141CUG-to-CUC but not the 141CUG-to-AUG mutation is inconsistent with a role for splicing of this RNA. Third, we were able to recapitulate the in vivo effects of specific mutations on *FCA* translation in vitro using rabbit reticulocyte extracts that are unable to splice pre-mRNA.

Table 1. List of Genes Identified as Initiating Translation at CUG and the Predicted Cellular Localization if Initiation Occurs at the CUG or Downstream AUG

At No.	Description	Predicted Targeting	
		CUG Protein	AUG Protein
At1g03260	Similar to unknown protein	–	ER
At1g21000	Zinc binding family protein	–	–
At1g55760	BTB/POZ domain-containing protein	–	Mitochondria
At2g20680	Glycosyl hydrolase family 5 protein	Mitochondria	ER
At3g02750	Protein phosphatase 2C	Plastid	Plastid
At3g04080	APY1, apyrase	Plastid	Mitochondria
At3g10985	SAG20, wound-induced protein 12	–	–
At3g62660	GATL7, galacturonosyltransferase-like 7	–	ER
At4g02680	EOL1, ETO1-like 1	–	–
At5g14500	Aldose 1-epimerase family protein	–	–

ER, endoplasmic reticulum.

How is the noncanonical codon selected in *FCA* mRNA? Two principal mechanisms, referred to as linear scanning and internal ribosome entry, explain translation initiation, and noncanonical initiation codons can be selected in both cases. The linear scanning model explains eukaryotic translation initiation in most cases: this process is dependent on the interaction of eukaryotic initiation factor (eIF) 4E with the 7-methyl guanosine cap structure at the 5′ end of the mRNA. This provides a landing platform for the 43S preinitiation complex of 40S ribosomes, eIF-2, eIF-3, Met-tRNA^{i}, and GTP, which migrates from the 5′ end of the mRNA to the first AUG triplet in a favorable sequence context, from where protein synthesis begins (Kozak, 1999). Translation initiation via internal ribosome entry involves *cis*-elements, known as internal ribosome entry sites (IRES), in a cap-independent process (Hellen and Sarnow, 2001). No single mechanism or sequence accounts for IRES-mediated translation initiation. Viral and cellular IRESs are typically characterized by long leader sequences and complex RNA structures, but this is not true of all cellular IRESs (Komar and Hatzoglou, 2005). Our findings do not distinguish between scanning and internal ribosome entry-mediated translation of *FCA* mRNA, but we found no evidence of leaky scanning (see below) or of downstream sequences being required to slow progression of scanning ribosomes to allow recognition of a noncanonical codon. We did, however, discover that upstream sequences were essential for CUG recognition, but these were dispensable for AUG recognition in the same context.

Where non-AUG translation initiation has been documented, it has sometimes been associated with additional sites of initiation resulting from leaky scanning (Acland et al., 1990; Vagner et al., 1995; Nanbru et al., 1997; Stoneley et al., 1998; Kozak, 2002). Leaky scanning involves inefficient recognition of an initiator codon that results in a proportion of 43S preinitiation complexes continuing to scan and initiate translation at a downstream site. This leads to the formation of different protein isoforms whose production can be differentially regulated (Vagner et al., 1996), which may be localized differently (Acland et al., 1990) or are functionally distinct. Certain viruses that infect plants exploit leaky scanning as a mechanism to translate distinct proteins from their compact genomes (Futterer et al., 1997). It has been proposed that leaky scanning of *Arabidopsis Polγ2* enables the formation of distinct peptide target sequences that allows dual targeting of Polγ2 to plastids and mitochondria (Wamboldt et al., 2009). This raises the question of whether leaky scanning from 141CUG might normally produce FCA-γ from 414AUG in vivo and whether this isoform may behave differently to FCA-γ+. However, we have no positive evidence for either of these possibilities: first, these two FCA isoforms behave similarly, in that both FCA-γ and FCA-γ+ associate with *FLC* chromatin (Liu et al., 2007) and can complement the *fca-1* mutation. Second, we have been unable to detect endogenous FCA-γ: when we use highly denaturing protein extraction buffers (that include guanidine hydrochloride or urea), we can detect transgenic FCA-γ and FCA-γ+ following immunoblot analysis, but only FCA-γ+ is found in wild-type *Arabidopsis* extracts (Figure 1). Therefore, *Arabidopsis FCA* appears to be translated exclusively from a noncanonical initiation codon without detectable alternative isoforms being produced by leaky scanning. Of the 10 genes we identified through a computational approach as likely to initiate translation at a CUG codon, five are predicted to encode proteins that are differentially targeted depending on whether

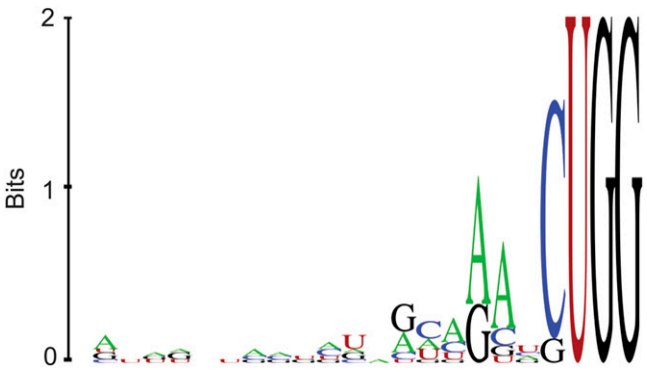

Figure 8. Logo Plot Showing the Consensus Surrounding the CUG Codon from the 10 *Arabidopsis* Genes Identified in Our Bioinformatic Search and Their Homologs.

[See online article for color version of this figure.]

Table 2. List of Genes with an Upstream CUG Codon and Their Translation Initiation Efficiency

Genes	CUG Context	Luciferase Activity
Genes with Conserved CUG		
At5g14500	**A**TC **A**AG CTG G	2.4 ± 0.2
At2g20680	**A**AA **G**AA CTG G	3.7 ± 0.2
At1g03260	**A**CA **A**AT CTG G	5.9 ± 0.5
At3g04080	**G**CC **G**AT CTG G	44.7 ± 0.9
FCA	**G**TC **A**AG CTG G	81.5 ± 0.5
Genes without Conserved CUG		
At3g05580	**T**GG **A**TT CTG G	0.1 ± 0.1
At3g58690	**T**GG **A**AA CTG G	0.2 ± 0.0
At4g21120	**C**AG **A**AT CTG G	1.7 ± 0.1
At3g50830	**T**GG **A**TT CTG G	11.8 ± 0.2

they initiate at CUG or a downstream AUG (Table 1). For the remaining five genes, the predicted subcellular localization of the translated protein was unaffected by selection of distinct initiation codons. Therefore, our work with *FCA* demonstrates that dual targeting of proteins to different subcellular locations is not the only reason that plants use noncanonical CUG translation initiation, and leaky scanning does not necessarily result from noncanonical codon initiation in plants.

A CUG initiation codon can form only two base pairs, instead of three, with the anticodon of the Met-tRNA^i (polypeptides with N termini Met residues are produced from noncanonical translation initiation; Peabody, 1989). The sensitivity of *FCA* translation initiation at 141CUG-to-the CUC mutation is consistent with its recognition being dependent on Met-tRNA^i. Non-AUG initiation requires that the weakened codon–anticodon interaction is compensated for by contacts with nearby nucleotides: a purine at position –3 and a G at position +4 augments recognition of non-AUG codons (Kozak, 1997). Significantly, *FCA* 141CUG lies in such a context, and we showed that mutation of the −3 purine to a pyrimidine results in a dramatic reduction in translation efficiency. CUG context, comprising a purine triplet 5′ of a CUG initiation codon, is an important element of the mechanism that mediates alternative translation initiation of Polγ2 dual targeted proteins (Wamboldt et al., 2009). A purine triplet is also found upstream of *Arabidopsis FCA* 141CUG. Furthermore, a purine at −3 and guanine at +4 was a conserved feature of the additional genes with putative CUG translation initiation codons that we identified in our bioinformatic analysis (Figure 8). Therefore, our analysis is consistent with local sequence context of the initiation codon being crucial to their recognition.

In addition to local CUG context, however, an important discovery was that deletion of sequence upstream of 141CUG (which did not affect the context of the CUG) substantially reduced (~100-fold) translation initiation at this codon but had no effect when 141CUG was mutated to AUG. By contrast, sequences downstream of 141CUG were not required for its selection. These findings demonstrate that upstream sequences are essential specifically for noncanonical translation initiation. Two upstream regions were found to be required for efficient *FCA* CUG translation initiation: a short 15-nucleotide region (beginning at −7 from the CUG) and the remaining 120 nucleotides of the 5′ UTR.

In addition to noncanonical translation initiation, our analysis of the 5′ end of *FCA* upstream of the first in-frame AUG revealed roles for the region in regulating RNA accumulation and encoding the *FCA* ORF. We discovered that the *FCA* promoter confers almost no detectable level of transcription, beyond expression in stipules. Instead, sequences corresponding to the 5′ transcribed region significantly enhance the pattern and level of *FCA* RNA expression. The multifunctionality of this region has practical implications for our study of *FCA* translation. We have been unable to test the biological consequences in *Arabidopsis* of deleting large segments of this 5′ region because both translation and RNA expression are affected. For example, although we demonstrated that a truncated version of FCA initiating translation at +414AUG and expressed from the CaMV 35S promoter complements the late flowering phenotype of loss-of-function *fca* mutants (Macknight et al., 2002), we cannot conclude that the additional *FCA* ORF we have discovered here is not necessary for FCA function because CaMV 35S drives ectopic and overexpression of FCA protein. When we drive expression of the same construct with the *FCA* promoter, the late flowering phenotype of *fca-1* is not complemented, but as we have deleted both the 5′ UTR and the ORF to the first AUG, RNA expression from such a construct is not detectable either.

Since the complexity and diversity of the functions of this 5′ *FCA* sequence make the precise examination of the biological significance of *FCA* initiating translation at a CUG difficult, we can only speculate as to its significance. The fact that additional upstream sequences are required suggests that specific factors could potentially regulate access of the translation machinery to the CUG codon of *FCA* mRNA. Translational control of *Arabidopsis* protein expression has received relatively little attention. IRES-mediated regulation of translation has been described in eukaryotic cellular mRNAs that are active during mitosis or under stress conditions when cap-dependent translation is inhibited (Pyronnet et al., 2000; Gilbert et al., 2007; Sonenberg and Hinnebusch, 2007), thereby providing a mechanism for continued synthesis of specific proteins when global translation is prevented. Consistent with this, the first IRES identified in plant cells was in mRNA encoding the maize (*Zea mays*) heat shock protein, Hsp101, and this accounts for its cap-independent translation during heat stress (Dinkova et al., 2005). FCA functions to regulate flowering time by repressing *FLC*, but we also know it plays other roles in regulating the expression of genome sequences that are also targets of RNA-mediated chromatin silencing (Baurle et al., 2007) and likely acts somewhat redundantly with other autonomous pathway components to regulate a wider set of gene targets (Veley and Michaels, 2008). Therefore, continued translation of FCA through mitosis may be required to prevent changes in otherwise epigenetically silenced loci, or translation regulation may affect FCA function in stress-dependent conditions, distinct from its role in flowering.

Protein sequences are mostly indirectly deduced from DNA sequence. Where computational programs are used in genome annotation, they include obligatory rules to search for AUG codons as translational start sites. However, this does not take into account the possibility that some genes initiate translation at a noncanonical start codon. A recent computational search for CUG codons upstream of in-frame AUG codons in *Arabidopsis* 5′ UTRs

(Wamboldt et al., 2009) suggested that CUG translation initiation might be relatively common, while our more conservative analysis based on what we had learned from FCA translation identified only 10 *Arabidopsis* genes, in addition to *FCA* and *Pol*γ*2*, in which translation initiation from a CUG codon is likely. As we included sequence conservation among putative orthologs as a criterion in this search, our findings may be of biological relevance.

Overall, our results suggest that a small number of genes have evolved the ability to initiate translation at the noncanonical codon CUG. For CUG translation initiation to occur, it appears that both CUG context and additional sequences are needed. In the case of *FCA*, specific sequences upstream of the CUG codon are required for efficient translation initiation. Our work on *FCA* also demonstrates that alternative initiation codons are not always used to target proteins from a single gene to different subcellular organelles. It will now be important to understand the novel features of the mechanism of *FCA* translation initiation and its significance in the regulation of FCA function.

METHODS

Plant Material and Growth Conditions

The mutant *fca-1* was provided by M. Koornneef (Koornneef et al., 1991). *Arabidopsis thaliana* seeds were sown aseptically in Petri dishes containing GM medium as described previously (Macknight et al., 2002), stratified for 2 d at 4°C, and planted in soil (6:1 mixture of Levingtons M3 compost with grit) at the four-leaf stage. Plants were grown in controlled environment rooms at 20°C as previously described (Macknight et al., 2002). Flowering time was measured by counting the number of rosette and cauline leaves on the main stem. *Nicotiana benthamiana*, provided by R. Hellens, was grown as described (Hellens et al., 2005).

In Vitro Translation

In vitro translation was performed with a coupled transcription and translation procedure (TnT) using rabbit reticulocyte extract in the presence of [^{35}S]Met according to the manufacturer's instructions (Promega). Translated proteins were separated by SDS-PAGE (15% polyacrylamide), fixed, dried, and revealed by exposure to a phosphor imager.

Plasmid Constructions

All primer sequences are listed in Supplemental Table 2 online. The CTG-CTC, CTG//STOP, and CTG-ATG mutations were introduced by site-directed mutagenesis of the 5′ coding sequence of plasmid pFCA-cDNA (Macknight et al., 2002) using the Pharmacia Biotech U.S.E. mutagenesis kit and the primers CTG-CTC, CTG//STOP, and CTG-ATG, respectively. The XHOX primer was used as a selection primer to remove the unique *Xho*I site. The mutated plasmids were cut with *Cla*I and *Xho*I and ligated into *pFCA-cDNA* in SLJ755I5, cut with *Cla*I and *Xho*I, to restore the *Xho*I site. FCA sequences were fused upstream of the U1A coding sequence in pGem3Zf+ (Simpson et al., 1995). The 5′ sequence of *FCA* (+1 to +416) was PCR-amplified from pFCA-cDNA with *Pfu* DNA polymerase using GSO256 and GSO372. The PCR product was cut with *Eco*RI and *Nco*I and cloned into the U1A plasmid cut similarly. The mutated CTG//STOP, CTG-CTC, and CTG-ATG sequences were prepared in the same way, except that the corresponding mutations in pFCA-cDNA were used as templates in the PCR reaction. For transient gene expression in *N. benthamiana* leaves, FCA fragments amplified using *Pfu* or Expand Long-template DNA polymerases were cloned into the pGreenII 0800-LUC vector (Hellens et al., 2005). CTG 1-448 was amplified from *pFCA-cDNA* with RCM505 and RCM506 and cloned into pGreenII 0800-LUC using *Hin*dIII and *Nco*I restriction sites. To generate *ATG 1-448*, primers RCM547 and RCM687 were used to amplify the entire *CTG 1-448* plasmid and the ends ligated to replace the CUG translation start codon with AUG. This strategy was used to prepare *CTG 1-146* and *ATG 1-146*, using primers RCM546(CTG) or RCM547(ATG) with RCM545. *CTG Δ147-188* and *ATG Δ147-188* were prepared using the same method with PCR products amplified from *CTG 1-448* or *ATG 1-448* using RCM546/RCM547 and RCM688. *CTG/ATG* and *CTC 1-305* were created using primers RCM737 and RCM545. *CUG/AUG Δ147-188*, *Δ195-239*, *Δ249-287*, and *Δ147-287* were generated using the following primers from *CTG 1-448* or *AUG1-448* template: RCM546/RCM547 and RCM673, RCM674 and RCM675, RCM676 and RCM688, and RCM546/RCM547 and RCM688. *CTG/ATG Δ1-119* and *CTG/ATG Δ1-134* were amplified using RCM727 or RCM739 (CTG)/RCM740(ATG) with RCM506, from *CTG 1-448* or *ATG 1-448* templates. *CTG/ATG Δ1-119* and *CTG/ATG Δ1-134* PCR fragments were cloned into pGreenII0800-LUC using *Hin*dIII and *Nco*I restriction sites. The following fragments were synthesized by Genescript and subcloned into modified pGreenII 0800-LUC, in which the *LUC AUG* initiation codon is removed (pGreenII 0800-LUCm), using *HindIII* and *Bam*HI restriction sites*: CTG R1-119*, *CTG R1-1134*, *CTG region II mutant*, *CTG region III mutant*, *CTG -6 GtoT -3 AtoT*, *CTG -6 GtoT*, and *CTG -3 AtoT*. To investigate other noncanonical start codons, genomic DNA sequences amplified from *Arabidopsis* Columbia seedlings using Phusion DNA polymerase (Finnzymes) were cloned into pGreenII 0800-LUCm vector. For overexpression of *CTG1-448:GUS* and *ATG1-448:GUS* in *Arabidopsis*, *FCA* fragments were amplified from *CTG1-448* or *ATG1-448* using RCM505 and RCM506 and cloned into pCAMBIA 1391x using *Hin*dIII and *Nco*I restriction sites. The *FCA*$_{pro}$*-FCA* 1-450:GUS transgene was described previously (Macknight et al., 2002; *P*$_{FCA}$*–FCA*$_{to\ ATG}$*:GUS*). To generate *FCA*$_{pro}$*-*GUS, the *FCA* promoter was isolated from P$_{FCA}$-γ KSII (Macknight et al., 1997) as an *Eco*RI-*Sph*I fragment. An *Sph*I site 7 bp 3′ of the transcription start was introduced using FCAmut1. The *GUS* gene and *FCA* terminator (as described in Macknight et al., 2002) were cloned downstream of *FCA*$_{pro}$ as an *Sph*I-*Xho*I fragment. *FCA*$_{pro}$:GUS was ligated into pSLJ1714 (Jones et al., 1992).

To make TATA-*FCA* 1-450:GUS, first a minimal 35S promoter (−46 to +9) (Odell et al., 1985), TATA, was amplified using RL01 and RL02. The TATA fragment was cloned into pBluescript KSII (Stratagene) as an *Eco*RI-*Xho*I fragment to make TATA-KSII. The *GUS* gene and *FCA* termination sequences were then ligated into TATA-KSII using *Sph*I-*Xho*I, and the entire TATA-GUS-FCA 3′ UTR was subcloned into binary vector pSLJ755I6. Next, the FCA 5′ region was amplified using RL05 and RL06 from *P*$_{FCA}$*–FCA*$_{to\ ATG}$*:GUS* and ligated into TATA-GUS-FCA 3′ using *Hin*dIII restriction sites to produce the final binary vector, TATA-*FCA* 1-450:GUS. To generate the *FCA*$_{pro}$*-FCA* 1-141:GUS construct, the *FCA*$_{pro}$ *FCA* 1-141 fragment was amplified from *FCA*$_{pro}$*-FCA* 1-450:GUS using RLFCA11 and RLFCA12. The TATA fragment in TATA-GUS was replaced with *FCA*$_{pro}$*-FCA* 1-141 using *Eco*RI and *Sph*I restriction sites.

Plant Transformation

Transgenes were introduced into *Agrobacterium tumefaciens* strain C58C1 pGV2260 by triparental mating and transformed into Landsberg *erecta* or *fca-1* by floral dip (Clough and Bent, 1998). For transient expression in *N. benthamiana* leaves, transgenes were introduced into *Agrobacterium* strain GV3101 (MP90) using electroporation and infiltrations performed as described by Hellens et al. (2005). Transient expression was assayed 4 d after infiltration.

Immunodetection of FCA Protein

Protein extraction and immunoblotting were performed as before (Quesada et al., 2003).

Protein, Luciferase, and RNA Analysis

RNA and protein extraction plus RNA gel blot and immunoblot procedures were as described previously (Quesada et al., 2003). Firefly luciferase and *Renilla* luciferase were assayed using the Dual-Luciferase Reporter Assay System (Promega) as described by Hellens et al. (2005). Histochemical GUS staining of transgenic *Arabidopsis* plants was performed as described by Macknight et al. (2002). Quantitative GUS assays were performed as described (Jefferson et al., 1987) and total protein quantified with Bradford reagent as per the manufacturer's instructions (Bio-Rad). Total RNA was obtained from 10-d-old *Arabidopsis* seedlings and *N. benthamiana* leaf discs using Invitrogen Plant RNA Purification reagent and cDNA synthesized with the Roche Transcriptor reverse transcriptase. Relative quantification of *firefly* luciferase expression was determined using quantitative RT-PCR (Roche Lightcycler480 and reagents) using LUC RT F and LUC RT R primers and data normalized to *Renilla* luciferase using REN RT F and REN RT R (see Supplemental Table 2 online).

Bioinformatics Analysis

TAIR9 REFSEQ *Arabidopsis* genomic sequences were downloaded from The Arabidopsis Information Resource website (http://www.Arabidopsis.org). Noncanonical in-frame CTG start codons, with a +4 guanine, and not more than 500 bases upstream of the annotated ATG initiation codon were identified using a program (provided in Supplemental Data Set 1 online) that made extensive use of code developed for reading gene termination regions for construction of the TransTerm translation terminator database (http://transterm.otago.ac.nz). This analysis used genomic sequence to avoid missing genes that lacked or had truncated mRNA sequences. For splice variants and genes with multiple start sites, only the most 5′ transcript was scrutinized. A program (provided in Supplemental Data Set 1 online) was also developed to extract and translate sequences between the noncanonical CTG and annotated ATG and carry out tBLASTn searches against plant EST (derived from ESTs deposited at http://compbio.dfci.harvard.edu/tgi/, excluding *Arabidopsis*) and genomic databases. Multiple sequence alignments were prepared using the Geneious Pro ClustalW plug-in (http://www.geneious.com) for all genes with sequence conservation extending upstream of the annotated ATG. Logo plots were also created using Geneious Pro.

Accession Numbers

Sequence data from this article can be found in the Arabidopsis Genome Initiative or GenBank/EMBL databases under the following accession numbers: *Arabidopsis FCA*, At4g16280; *Brassica napus FCA* (AF414188), *Gossypium raimondii FCA* (CO097888), *Medicago truncatula FCA* (BG449291), *Vitus vinifera FCA* (AM461090), *Solanum lycopersicon FCA* (BI929080), and *Arabidopsis Polγ2* (At1g50840). The 10 *Arabidopsis* genes identified in our bioinformatic analysis are At1g03260, At1g21000, At1g55760, At2g20680, At3g02750, At3g04080, At3g10985, At3g62660, At4g02680, and At5g14500.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Nucleotide Sequence and Deduced Open Reading Frame Upstream of the First in-Frame AUG Codon in *Arabidopsis FCA* mRNA.

Supplemental Figure 2. *FCA* Translation Initiation Is Equally Efficient from Either a CUG or AUG Start Codon in Transgenic *Arabidopsis* Plants.

Supplemental Figure 3. *FCA* Sequences Downstream of a 141AUG Start Codon Influence Translation Efficiency.

Supplemental Figure 4. The *FCA* 5′ Region Contains Elements That Enhance Expression.

Supplemental Figure 5. Nucleotide and Protein Alignments of Sequences between the Putative CUG Initiation Codon and the Annotated AUG for Each of the 10 Genes and Their Homologs Identified in Our Bioinformatic Search.

Supplemental Figure 6. Nucleotide Alignments of Sequences Surrounding the Putative CUG Initiation Codon for Each of the 10 Genes and Their Homologs Identified in Our Bioinformatic Search.

Supplemental Table 1. Subcellular Localization Predictions for Proteins Produced from Either the Upstream CUG or Annotated AUG.

Supplemental Table 2. Sequence of Primers Used in This Study.

Supplemental Data Set 1. Programs Used in Bioinformatic Analysis.

ACKNOWLEDGMENTS

We thank L. Clissold, C. Nicholls, M. Smith, and J. Campbell for technical assistance and Philip Smith for proofreading the manuscript. This work was supported by European Union Marie Curie fellowships to V.Q. and P.P.D., a Ministerio de Educacion y Cultura of Spain fellowship to V.Q., an AGMARDT fellowship to R.E.L., the New Zealand Marsden Fund, and a Core Strategic Grant from the Biotechnology and Biological Sciences Research Council.

Received July 26, 2010; revised September 13, 2010; accepted October 28, 2010; published November 12, 2010.

REFERENCES

Acland, P., Dixon, M., Peters, G., and Dickson, C. (1990). Subcellular fate of the int-2 oncoprotein is determined by choice of initiation codon. Nature **343:** 662–665.

Baurle, I., and Dean, C. (2008). Differential interactions of the autonomous pathway RRM proteins and chromatin regulators in the silencing of *Arabidopsis* targets. PLoS ONE **3:** e2733.

Baurle, I., Smith, L., Baulcombe, D.C., and Dean, C. (2007). Widespread role for the flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing. Science **318:** 109–112.

Christensen, A.C., Lyznik, A., Mohammed, S., Elowsky, C.G., Elo, A., Yule, R., and Mackenzie, S.A. (2005). Dual-domain, dual-targeting organellar protein presequences in *Arabidopsis* can use non-AUG start codons. Plant Cell **17:** 2805–2816.

Clough, S.J., and Bent, A.F. (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. **16:** 735–743.

Depeiges, A., Degroote, F., Espagnol, M.C., and Picard, G. (2006). Translation initiation by non-AUG codons in *Arabidopsis thaliana* transgenic plants. Plant Cell Rep. **25:** 55–61.

Dinkova, T.D., Zepeda, H., Martinez-Salas, E., Martinez, L.M., Nieto-Sotelo, J., and de Jimenez, E.S. (2005). Cap-independent translation of maize Hsp101. Plant J. **41:** 722–731.

Futterer, J., Rothnie, H.M., Hohn, T., and Potrykus, I. (1997). Rice tungro bacilliform virus open reading frames II and III are translated from polycistronic pregenomic RNA by leaky scanning. J. Virol. **71:** 7984–7989.

Gallie, D.R. (1996). Translational control of cellular and viral mRNAs. Plant Mol. Biol. **32:** 145–158.

Gilbert, W.V., Zhou, K., Butler, T.K., and Doudna, J.A. (2007).

Cap-independent translation is required for starvation-induced differentiation in yeast. Science **317:** 1224–1227.

Gordon, K., Futterer, J., and Hohn, T. (1992). Efficient initiation of translation at non-AUG triplets in plant cells. Plant J. **2:** 809–813.

Hellen, C.U., and Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. Genes Dev. **15:** 1593–1612.

Hellens, R.P., Allan, A.C., Friel, E.N., Bolitho, K., Grafton, K., Templeton, M.D., Karunairetnam, S., Gleave, A.P., and Laing, W.A. (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. Plant Methods **1:** 13.

Hornyik, C., Terzi, L.C., and Simpson, G.G. (2010). The spen family protein FPA controls alternative cleavage and polyadenylation of RNA. Dev. Cell **18:** 203–213.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: Beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. **6:** 3901–3907.

Jones, J.D.G., Shlumukov, L., Carland, F., English, J., Scofield, S.R., Bishop, G.J., and Harrison, K. (1992). Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants. Transgenic Res. **1:** 285–297.

Komar, A.A., and Hatzoglou, M. (2005). Internal ribosome entry sites in cellular mRNAs: Mystery of their existence. J. Biol. Chem. **280:** 23425–23428.

Koornneef, M., Hanhart, C.J., and van der Veen, J.H. (1991). A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. Mol. Gen. Genet. **229:** 57–66.

Kozak, M. (1990). Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA **87:** 8301–8305.

Kozak, M. (1997). Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. EMBO J. **16:** 2482–2492.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. Gene **234:** 187–208.

Kozak, M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. Gene **299:** 1–34.

Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C. (2010). Targeted 3′ processing of antisense transcripts triggers *Arabidopsis FLC* chromatin silencing. Science **327:** 94–97.

Liu, F., Quesada, V., Crevillen, P., Baurle, I., Swiezewski, S., and Dean, C. (2007). The *Arabidopsis* RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate *FLC*. Mol. Cell **28:** 398–407.

Macknight, R., Bancroft, I., Page, T., Lister, C., Schmidt, R., Love, K., Westphal, L., Murphy, G., Sherson, S., Cobbett, C., and Dean, C. (1997). *FCA*, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. Cell **89:** 737–745.

Macknight, R., Duroux, M., Laurie, R., Dijkwel, P., Simpson, G., and Dean, C. (2002). Functional significance of the alternative transcript processing of the Arabidopsis floral promoter *FCA*. Plant Cell **14:** 877–888.

Michaels, S.D., and Amasino, R.M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell **11:** 949–956.

Michaels, S.D., and Amasino, R.M. (2001). Loss of *FLOWERING LOCUS C* activity eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutations but not responsiveness to vernalization. Plant Cell **13:** 935–941.

Nanbru, C., Lafon, I., Audigier, S., Gensac, M.C., Vagner, S., Huez, G., and Prats, A.C. (1997). Alternative translation of the proto-oncogene c-myc by an internal ribosome entry site. J. Biol. Chem. **272:** 32061–32066.

Neff, M.M., Neff, J.D., Chory, J., and Pepper, A.E. (1998). dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. Plant J. **14:** 387–392.

Odell, J.T., Nagy, F., and Chua, N.-H. (1985). Identification of DNA sequences required for activity of the cauliflower mosaic visrus 35S promoter. Nature **313:** 810–812.

Ohnacker, M., Barabino, S.M., Preker, P.J., and Keller, W. (2000). The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3′-end-processing complex. EMBO J. **19:** 37–47.

Peabody, D.S. (1989). Translation initiation at non-AUG triplets in mammalian cells. J. Biol. Chem. **264:** 5031–5035.

Pyronnet, S., Pradayrol, L., and Sonenberg, N. (2000). A cell cycle-dependent internal ribosome entry site. Mol. Cell **5:** 607–616.

Quesada, V., Macknight, R., Dean, C., and Simpson, G.G. (2003). Autoregulation of *FCA* pre-mRNA processing controls *Arabidopsis* flowering time. EMBO J. **22:** 3142–3152.

Riechmann, J.L., Ito, T., and Meyerowitz, E.M. (1999). Non-AUG initiation of AGAMOUS mRNA translation in *Arabidopsis thaliana*. Mol. Cell. Biol. **19:** 8505–8512.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates III, J.R., Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3′ processing complex. Mol. Cell **33:** 365–376.

Simpson, G.G., Clark, G.P., Rothnie, H.M., Boelens, W., van Venrooij, W., and Brown, J.W. (1995). Molecular characterization of the spliceosomal proteins U1A and U2B" from higher plants. EMBO J. **14:** 4540–4550.

Simpson, G.G., Dijkwel, P.P., Quesada, V., Henderson, I., and Dean, C. (2003). FY is an RNA 3′ end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition. Cell **113:** 777–787.

Sonenberg, N., and Hinnebusch, A.G. (2007). New modes of translational control in development, behavior, and disease. Mol. Cell **28:** 721–729.

Stoneley, M., Paulin, F.E., Le Quesne, J.P., Chappell, S.A., and Willis, A.E. (1998). C-Myc 5′ untranslated region contains an internal ribosome entry segment. Oncogene **16:** 423–428.

Vagner, S., Gensac, M.C., Maret, A., Bayard, F., Amalric, F., Prats, H., and Prats, A.C. (1995). Alternative translation of human fibroblast growth factor 2 mRNA occurs by internal entry of ribosomes. Mol. Cell. Biol. **15:** 35–44.

Vagner, S., Touriol, C., Galy, B., Audigier, S., Gensac, M.C., Amalric, F., Bayard, F., Prats, H., and Prats, A.C. (1996). Translation of CUG- but not AUG-initiated forms of human fibroblast growth factor 2 is activated in transformed and stressed cells. J. Cell Biol. **135:** 1391–1402.

van der Velden, A.W., and Thomas, A.A. (1999). The role of the 5′ untranslated region of an mRNA in translation regulation during development. Int. J. Biochem. Cell Biol. **31:** 87–106.

Veley, K.M., and Michaels, S.D. (2008). Functional redundancy and new roles for genes of the autonomous floral-promotion pathway. Plant Physiol. **147:** 682–695.

Wamboldt, Y., Mohammed, S., Elowsky, C., Wittgren, C., de Paula, W.B., and Mackenzie, S.A. (2009). Participation of leaky ribosome scanning in protein dual targeting by alternative translation initiation in higher plants. Plant Cell **21:** 157–167.