A link between RNA metabolism and silencing affecting *Arabidopsis* development

Brian D. Gregory 1,2, Ronan C. O’Malley 1,2, Ryan Lister 1,2, Mark A. Urich 1, Julian Tonti-Filippini 3, Huaming Chen 2, A. Harvey Millar 3, and Joseph R. Ecker 1,2,*

1 Plant Biology Laboratory and 2 Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA; 3 ARC Centre of Excellence in Plant Energy Biology, The University of Western Australia, Crawley, WA 6009, Australia

* To whom correspondence should be addressed:

Joseph R. Ecker
Plant Biology Laboratory and Genomic Analysis Laboratory
The Salk Institute for Biological Studies
10010 N. Torrey Pines Rd., La Jolla, CA 92037
Telephone: (858) 453-4100 x1795
Fax: (858) 558-6379
E-mail: ecker@salk.edu

Abbreviations: miRNA, microRNA; siRNA, small-interfering RNA; tasiRNA, trans-acting siRNA; CBC, nuclear mRNA cap-binding complex; RDR, RNA-dependent RNA Polymerase; PPM, parts per million; RACE, Rapid Amplification of cDNA Ends; GMUCT, Genome-wide Mapping of Uncapped and Cleaved Transcripts; ACC, 1-aminocyclopropane-1-carboxylic acid.

Notes to reviewers: We have created an interactive public web browser to house all of the smRNA and GMUCT sequencing data ([http://neomorph.salk.edu/aj_salk/smRNAome.html](http://neomorph.salk.edu/aj_salk/smRNAome.html)). For browsing at this website, please ensure to use Firefox. In order to protect the anonymity of the reviewers, we suggest using a proxy server such as the free anonymous browsing provided by [http://www.anonymizer.com/](http://www.anonymizer.com/). Upon acceptance and publication of our paper, everything that is available on our website will be made publicly available. As indicated in this manuscript, all of the sequencing data herein will be available via this interface, as well as through GenBank (GEO accession numbers are currently being obtained). Data release in both locations will coincide with publication of the paper.
SUMMARY

Micro (mi)RNAs and small-interfering (si)RNAs are abundant endogenous small (sm)RNAs that control transcript expression through post-transcriptional gene silencing. Here, we show that concomitant loss of XRN4/EIN5, a 5’-3’ exoribonuclease, and CBP80/ABH1, a subunit of the mRNA cap binding complex, results in Arabidopsis plants manifesting myriad developmental defects. Through the analysis of ein5 abh1 double mutant plants, we find that ABH1/CBP80 is necessary to obtain proper mature miRNA levels, which suggests this protein affects the miRNA-mediated RNA silencing pathway. Additionally, we show that XRN4/EIN5 affects the levels of a smRNA class that is processed from both sense and anti-sense strands of ~130 endogenous transcripts that apparently are converted to double-stranded RNA and subsequently processed. We find that the parent transcripts of these smRNAs accumulate in an uncapped form upon loss of XRN4/EIN5, which suggests that uncapped endogenous transcripts can become smRNA biogenesis substrates. Overall, our results reveal unexpected connections between RNA metabolism and silencing pathways.
INTRODUCTION

RNA silencing represents a pathway that controls gene expression transcriptionally and post-transcriptionally (Baulcombe, 2004). In RNA silencing, production of double-stranded RNA (dsRNA) or self-complementary fold-back structures give rise to small RNAs (smRNAs) through the activity of DICER or DICER-LIKE (DCL) RNase III-type ribonucleases (Jones-Rhoades et al., 2006). These smRNAs comprise the sequence-specific effectors of RNA silencing pathways that direct the negative regulation or control of genes, repetitive sequences, viruses, and mobile elements (Almeida and Allshire, 2005; Tomari and Zamore, 2005). In plants, these smRNAs are comprised of microRNAs (miRNAs) and several classes of endogenous small interfering RNAs (siRNAs), which are differentiated from one another by their distinct biogenesis pathways and the classes of genomic loci from which they arise (Baulcombe, 2004).

miRNAs are a class of smRNAs that are 20-24 nucleotides (nt) in length, which arise from much longer primary transcripts that form characteristic stem-loop structures (Bartel, 2004). In Arabidopsis, the stem-loop precursor of miRNAs is processed by DCL1 ribonuclease to generate a miRNA/miRNA* duplex with a 2-nucleotide (nt) 3’ overhang. The miRNA* is derived from the opposite strand of the stem-loop structure and pairs imperfectly to the miRNA (Bartel, 2004; Jones-Rhoades et al., 2006). The miRNA is then incorporated into an RNA-induced silencing complex (RISC) that has the ARGONAUTE1 (AGO1) protein at its core (Baumberger and Baulcombe, 2005; Qi et al., 2005).

Plant miRNAs have imperfect but extensive complimentarity to their mRNA targets, and typically direct cleavage of these transcripts (Jones-Rhoades and Bartel, 2004; Llave et al., 2002). The targets of many plant miRNAs are mRNAs encoding transcription factors (Jones-Rhoades and Bartel, 2004), and the importance of miRNA-mediated regulation of a number of these target transcripts for proper development is well established (Mallory and Vaucheret, 2006; Willmann and Poethig, 2007). Additionally, plants containing mutations in genes encoding proteins involved in miRNA biogenesis or function (AGO1, DCL1, HEN1, HYL1, SERRATE, and HST) exhibit myriad dramatic developmental abnormalities, exemplifying the importance of this class of smRNAs in growth and differentiation (Boutet et al., 2003; Lobbes et al., 2006; Park
et al., 2002; Prigge and Wagner, 2001; Vaucheret et al., 2004; Vazquez et al., 2004a; Yang et al., 2006).

Endogenous siRNAs are a class of smRNAs that arise from long dsRNA that are formed as a product of an RNA-dependent RNA polymerase (RDR), by convergent transcription, or transcription of repetitive elements. siRNAs typically perform autosilencing, in that they target DNA or transcripts corresponding to (or homologous in sequence to) the loci from which they are processed (Baulcombe, 2004). However, the trans-acting siRNAs (tasiRNAs), which are processed from non-coding RNAs known as TRANS-ACTING siRNA (TAS) genes, are the exception in that they post-transcriptionally down-regulate protein-coding transcripts from unrelated loci in a fashion reminiscent of the miRNA-directed RNA silencing pathway (Allen et al., 2005; Peragine et al., 2004; Vazquez et al., 2004b; Yoshikawa et al., 2005). During the biogenesis of tasiRNAs, a segment of the TAS transcript that is defined by miRNA-mediated RISC cleavage is converted by RDR6 to dsRNA, which is successively cleaved by DCL4 into 21 nt siRNAs (Adenot et al., 2006; Allen et al., 2005; Fahlgren et al., 2006; Garcia et al., 2006).

Another class of plant endogenous siRNAs is the heterochromatic siRNAs, which are smRNAs that are mostly 24 nt in length and associated with DNA methylation. The concerted activity of plant-specific DNA-dependent RNA polymerases, PolIVa and PolIVb, correlates with the accumulation of 24 nt heterochromatic siRNAs via dsRNA formation by RDR2 and DCL3-mediated processing (Chan et al., 2005; Herr et al., 2005; Onodera et al., 2005; Xie et al., 2004). A fraction of these siRNAs associate with AGO4 to form a silencing complex thought to direct sequence-specific methylation events (Chan et al., 2005; Qi et al., 2006). Subsequently, this siRNA-directed DNA methylation can result in maintained transcriptional gene silencing at loci from which the smRNAs arise (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005).

The 5’ cap structure and the 3’ poly(A) tail are the two boundary marks that define the extreme borders of a eukaryotic mRNA. In the eukaryotic cell nucleus, the 5’ cap is recognized by the nuclear mRNA cap binding complex (CBC). CBC is a heterodimeric complex that consists of a small (CBP20) and a large (CBP80) protein
subunit (Mazza et al., 2001) that plays numerous roles in RNA metabolism (Aguilera, 2005). The Arabidopsis homologue of CBP80 is encoded by the ABH1 gene. Plants harboring a genetic lesion in ABH1 (abh1-1 mutant plants) manifest an ABA-hypersensitive regulation of seed germination phenotype that suggests a link between mRNA metabolism and ABA signaling. Interestingly, the characterization of abh1-1 mutant plants resulted in the discovery that Arabidopsis CBC is entirely composed of the ABH1/CBP80-AtCBP20 heteroduplex (Hugouvieux et al., 2001).

In Arabidopsis, the 5′-3′ exoribonuclease XRN4/EIN5 (henceforth referred to as EIN5) acts as an mRNA-degrading enzyme that is involved in the decay of specific transcripts that include the 3′ products of miRNA-mediated cleavage (Kastenmayer and Green, 2000; Souret et al., 2004). Two of its other mRNA targets, EBF1 and EBF2, encode F-box proteins that target the ubiquitin/proteasome-mediated turnover of EIN3, a key transcription factor mediating gene expression regulated by the phytohormone ethylene gas. These specific EIN5 targets account for the ethylene insensitive (Ein−) phenotype exhibited by ein5 mutant plants (Olmedo et al., 2006; Potuschak et al., 2006).

Here, through analysis of the developmental defects manifested on ein5-6 abh1-1 double mutant plants, we demonstrate unexpected roles in RNA silencing pathways for two proteins involved in general RNA metabolism, ABH1/CBP80 (henceforth referred to ABH1) and EIN5. First, we find that loss of ABH1 decreases the levels of mature miRNAs, thus suggesting this protein functions in the miRNA-mediated RNA silencing pathway. Additionally, we show that EIN5 affects the abundance of a distinct class of mostly 21 nt smRNAs that in many cases emanate from the entire length of endogenous functionally-annotated transcripts, many of which accumulate in an uncapped form in ein5 mutant plants. Taken together, our results suggest that an additional fate for endogenous uncapped transcripts is shuttling into an RNA silencing pathway where they become smRNA-biogenesis substrates.

RESULTS

Ethylene insensitivity of ein5 mutant plants is suppressed by the abh1 mutation
The position that ABH1 occupies on mRNA molecules suggests that it may counteract the function of 5’-3’ exoribonucleases, such as EIN5 (Aguilera, 2005). Therefore, in order to determine if mutation of ABH1 can suppress the ethylene insensitivity of ein5 mutant plants, we generated double mutants of ein5 and abh1. We found that loss of ABH1 (abh1-1) almost completely suppresses the Ein- phenotype of ein5-6 mutant plants (Figures 1A-E). Previously, we and others have demonstrated that the ethylene insensitivity of ein5 mutant plants is a consequence of EBF1 and EBF2 mRNA accumulation (Olmedo et al., 2006; Potuschak et al., 2006). To characterize the effect of ABH1 on EBF1 and EBF2 mRNA levels, we performed northern blot analysis using total RNA from 3-day-old etiolated seedlings of wild-type Col-0, abh1-1, ein5-6, and ein5-6 abh1-1 grown in hydrocarbon-free air or 10 parts per million (ppm) ethylene for various time periods (Figure 1F). We found that abh1-1 seedlings maintained similar levels of EBF1 and EBF2 mRNA to wild-type Col-0 (Figure 1F). As expected, ein5-6 seedlings accumulated a significantly increased level of both mRNAs in air and upon ethylene treatment (Figure 1F). Conversely, ein5-6 abh1-1 seedlings exhibited levels of both EBF1 and EBF2 mRNA similar to wild-type Col-0 (Figure 1F), suggesting that loss of ABH1 restores proper maintenance of these two mRNAs in ein5 mutant plants. Taken together, these results reveal that ABH1 is required for the accumulation of increased levels of EBF1 and EBF2 mRNAs in the absence of EIN5, and abh1 mutation can act as a genetic suppressor of the hormone response phenotype manifested by ein5 mutant plants.

**ein5-6 abh1-1 plants manifest developmental defects similar to those observed for miRNA pathway mutants**

During analysis of the ethylene response of ein5 abh1 mutant plants, we noticed they manifested severe developmental defects. Previously, abh1-1 and ein5-6 single mutant plants had been found to present a serrated leaf margin phenotype (Hugouvieux et al., 2002; Olmedo et al., 2006), which we have determined is strongly enhanced in ein5-6 abh1-1 double mutant plants (Figure 2Q). In addition, a small percentage of abh1-1 (~7%) and ein5-6 (~1%) single mutant seedlings demonstrated fused cotyledons when compared to wild-type Col-0 plants, which never displayed this phenotype. This
phenotype was also greatly enhanced in *ein5-6 abh1-1* double mutant seedlings, where we observed over 16% of seedlings with fused cotyledons (Figures 2A-P). *ein5-6 abh1-1* double mutant flowers also manifested defects not observed for the other three genotypes, including extranumerary petals and bending of the gynoecium (Figure S1). Finally, *abh1-1* and *ein5-6* single mutant plants were also found to manifest an altered phyllotactic pattern, where the internode length is severely decreased resulting in multiple fruits emanating from the same node, which was observed at a far greater frequency in *ein5-6 abh1-1* double mutant plants (Figures 2R-V). In fact, we observed numerous nodes from which more than three fruits emanated on double mutant plants, while this phenotype was never manifested by plants of the other genotypes (Figure 2Y). Interestingly, Prigge and Wagner (Prigge and Wagner, 2001) previously determined that mutation of the *SERRATE* gene in *Arabidopsis* resulted in plants that presented similar phenotypes to those observed for *ein5 abh1*. Subsequently, it was determined that SERRATE is required for proper miRNA biogenesis (Lobbes et al., 2006; Yang et al., 2006). Therefore, the developmental defects manifested by *ein5-6 abh1-1* (and to a lesser extent *abh1-1* and *ein5-6*) suggested that loss of ABH1 and EIN5 might affect the miRNA-mediated RNA silencing pathway.

**Loss of ABH1 and EIN5 disrupts the miRNA-mediated RNA silencing pathway**

To determine if EIN5 and/or ABH1 have effects on the miRNA-mediated RNA silencing pathway, we carried out an unbiased analysis of the transcriptome using whole-genome tiling microarrays (Chekanova et al., 2007; Kapranov et al., 2002) (Figure S2 and Tables S1-6). Interestingly, tiling microarray analysis revealed that the levels of a number of primary *MIRNA* transcripts were significantly increased in *ein5-6 abh1-1* (18 increased) and *abh1-1* (19 increased) compared to *ein5-6* and wild-type Col-0 plants (Figure 3A and Table S7). To determine that the stem loop structure containing the miRNA/miRNA* duplex was part of the primary *MIRNA* transcript up-regulated in our tiling array analysis and to provide validation for this class of mRNAs, we performed reverse transcription (RT) quantitative polymerase chain reaction (RT qPCR) with primer sets homologous to sequences just upstream and downstream of this structural moiety for a subset of the primary *MIRNA* transcripts that were statistically up-regulated.
from our microarray studies (Table S7). We found that ein5-6 abh1-1 and abh1-1 mutant plants accumulated significantly more MIRNA158a, MIRNA164b, MIRNA167a, and MIR168a transcripts that contained the stem loop structural moiety than did wild-type Col-0 or ein5-6 plants (Figure 3B), suggesting that it is the loss of ABH1 that results in accumulation of this class of mRNAs (Figures 3A-B and Table S7).

Furthermore, tiling array analysis revealed that a number of miRNA-target mRNAs accumulated in ein5-6 abh1-1 double mutant plants to levels far exceeding those observed for the three other genotypes, two examples of which can be seen in Figure 3C (Figure 3C and Table S5). We validated these two examples, CUC1 and HAP2C, using RT qPCR (Figure 3D). As expected based on the tiling microarray results, we determined that ein5-6 abh1-1 double mutant plants accumulated increased levels of the miRNA-target transcripts CUC1 and HAP2C compared to the other three genotypes (Figure 3D). These results provided further validation of our tiling microarray analysis and suggested that the concomitant loss of ABH1 and EIN5 results in the accumulation of miRNA-target mRNAs. Previously, it was determined that these two classes of mRNAs (primary MIRNA transcripts and miRNA-target mRNAs) were significantly increased in Arabidopsis plants lacking proteins involved in miRNA biogenesis and function (Llave et al., 2002; Lobbes et al., 2006). Overall, these results suggested that concomitant loss of ABH1 and EIN5 disrupts miRNA-mediated RNA silencing.

**ABH1 affects the miRNA-mediated RNA silencing pathway.**

Next, we wanted to obtain a genome-wide view of the effects of ABH1 and/or EIN5 on the smRNA populations of Arabidopsis. In order to accomplish this, we employed deep sequencing of smRNA samples from wild-type Col-0, abh1-1, ein5-6, and ein5-6 abh1-1 immature flower buds using an Illumina Genetic Analyzer (GA) (Figures S3-4). A total of 3,264,170 (1,036,593 unique), 4,446,687 (1,400,713 unique), 3,173,518 (1,030,343 unique), and 2,910,019 (959,915 unique) smRNAs were identified from the wild-type Col-0, abh1-1, ein5-6, and ein5-6 abh1-1 sequencing libraries, respectively (Figure S4). We found that the majority (~86%) of smRNAs in all genotypes sequenced were 21-24 nt in size (data not shown). Upon focusing our analysis on the
sequencing data encompassing the 21 nt size class of smRNAs, we noticed that this smRNA size class was overall underrepresented (~20%) in abh1-1 single mutant plants compared to wild-type Col-0 (Figure S5A). These findings indicated that ABH1 is required to obtain proper levels of a class of smRNAs that are 21 nt in length. Conversely, we found that ein5-6 abh1-1 double mutant plants accumulated higher levels of 21 nt smRNAs than the other three genotypes (Figure S5A), which suggests that EIN5 negatively regulates a class of 21 nt smRNAs. To obtain a genome-wide view of the 21 nt smRNA population, we first normalized the number of sequence reads at all genomic locations corresponding to 21 nt smRNAs by the total reads sequenced for the matching library, then parsed the values into 250 kilobase (kb) bins, and plotted these locations across the five Arabidopsis nuclear chromosomes. We found that for abh1-1 and ein5-6 abh1-1 plants there was a reduction on average of ~60% in the levels of 21 nt smRNAs corresponding to the highly abundant members of this smRNA class, which consist of mostly miRNAs (Figure 4A). These results suggested that ABH1 is required to obtain proper miRNA levels. Interestingly, we also observed that ein5-6 abh1-1 and to a lesser extent ein5-6 mutant plants accumulated 21 nt smRNA clusters that mapped to a number of different gene-rich locations throughout the Arabidopsis genome compared to the other two genotypes (Figure 4A, red arrows), suggesting that EIN5 negatively affects a population of 21 nt smRNAs (see below). To determine if these two proteins regulated the levels of specific classes of 21 nt smRNAs, we used our sequencing data encompassing another size class of smRNAs, 24 nts, as a control. This class of smRNAs has been demonstrated to consist mostly of heterochromatic siRNAs (Kasschau et al., 2007; Rajagopalan et al., 2006). We found that on a genome-wide scale the levels of 24 nt smRNAs were not significantly different between the four sequenced genotypes (Figures S5B), indicating that EIN5 and ABH1 specifically regulated certain classes of 21 nt smRNAs.

To validate our sequencing data and further characterize ABH1’s effect on miRNA levels, we performed smRNA-enriched northern blot analysis on samples from wild-type Col-0, abh1-1, ein5-6, and ein5-6 abh1-1 immature flower buds. These experiments revealed that abh1-1 and ein5-6 abh1-1 plants accumulated from 62-83% less miRNA for all those interrogated (miRNA156, 158, 159, 164, 167, 169, and 390)
compared to wild-type Col-0 and ein5-6 (Figures 4B-C, S6). Furthermore, this reduction in the levels of miRNAs upon loss of ABH1 function was corroborated by our smRNA sequencing data when the number of reads for specific miRNAs was normalized to the total sequenced read numbers (Figure S7A). Taken together, these results suggest that ABH1 is required to obtain proper miRNA levels. Notably, the smRNA sequencing data demonstrated that the normalized levels of some well-characterized heterochromatin-associated siRNAs (24 nt) were not reduced by the loss of ABH1 function (Figure S7B), which suggested this protein is specific in its effect on miRNAs levels.

To test if the reduction in miRNA levels observed for abh1-1 and ein5-6 abh1-1 plants affected miRNA-mediated target mRNA cleavage, we employed a modified 5'-rapid amplification of cDNA ends (5'-RACE) protocol (Llave et al., 2002). Using this methodology we found that for abh1-1 and ein5-6 abh1-1 plants there was decreased miRNA-mediated cleavage of the target mRNAs interrogated compared to wild-type Col-0 and ein5-6 (Figures 4D-E). Overall, these results demonstrated that ABH1 is required to obtain wild-type levels of miRNAs, and suggested it has effects on the miRNA-mediated RNA silencing pathway through the reduction of this class of smRNA.

Next, we wanted to determine if mutation of ABH1 could enhance the developmental defects manifested by plants containing a hypomorphic allele of AGO1, since AGO1 is known to function in the miRNA-mediated RNA silencing pathway. To do this, we generated double mutants between abh1-8 (a null allele) and plants containing a hypomorphic allele of AGO1, ago1-38. A hypomorphic allele of AGO1 was used because plants harboring null mutations in this gene do not survive. We then analyzed the abh1-8 ago1-38 double mutant plants for novel growth defects or enhancement of developmental abnormalities manifested in abh1-8 and/or ago1-38 single mutant plants (Figure 5). We found that approximately 15% of abh1-8 ago1-38 double mutant seedlings exhibited fused cotyledons, which was more than double the amount manifested by either abh1-8 or ago1-38 single mutant plants (Figures 5B-E, 5G-J, 5L-O and 5P). Furthermore, none of the wild-type Col-0 seedlings analyzed exhibited this developmental abnormality (Figures 5A and 5P). Additionally, we found that approximately 12% of abh1-8 ago1-38 double mutant seedlings never developed a root, even after two weeks of growth (Figures 5E, 5J, 5O, and 5Q). Interestingly, this
developmental defect was never observed for wild-type Col-0, abh1-8, nor ago1-38 single mutant plants (Figures 5A-C, 5F-H, 5K-M, and 5Q). We also found that abh1-8 ago1-38 double mutant plants presented an enhancement of the altered phyllotaxy where the internode length is severely decreased, resulting in multiple fruits emanating from the same node also demonstrated by both abh1-8 and ago1-38 single mutant plants. Specifically, ~16% of all abh1-8 ago1-38 double mutant plants manifested more than three fruits emanating from the same node, which is more than double the amount witnessed for wild-type Col-0, abh1-8, or ago1-38 single mutant plants (Figures 5R-S). These results demonstrated that the abh1-8 mutation is able to enhance the developmental defects manifested by Arabidopsis plants harboring a hypomorphic mutant allele of AGO1, which suggested that like AGO1, ABH1 also affects the miRNA-mediated RNA silencing pathway.

EIN5 affects the levels of a class of smRNAs processed from functionally-annotated mRNAs

As noted previously, we observed that in the absence of EIN5 function clusters of 21 nt smRNAs accumulate in euchromatic regions of the genome (Figure 4A). Therefore, we filtered the sequenced smRNAs for 21 nt smRNAs present at least twice in abh1-1, ein5-6, or ein5-6 abh1-1, but not found in Col-0. These three lists were further filtered for functionally-annotated mRNAs from which at least two independent smRNAs were processed, which identified 67, 253, and 235 such transcripts from abh1-1, ein5-6, and ein5-6 abh1-1 plants, respectively (Figure 6A and Tables S8-10). We found that 156 of these transcripts overlap between the ein5-6 and ein5-6 abh1-1 lists, with 23 of them also overlapping with those found for abh1-1 (Table S11). Since the smRNAs observed (Figure 4A) were only present in the absence of EIN5, we focused on the 133 transcripts that overlapped between the lists from the ein5-6 and ein5-6 abh1-1 smRNA libraries (Figure 6A, Tables S11-12). We found that clusters of mostly 21 nt smRNAs were generated from both sense and anti-sense strands of these 133 functionally-annotated transcripts (Figures 6B-C), suggesting that these mRNAs had been converted to double-stranded RNA and then processed in 21 nt increments in the absence of functional EIN5 (Figures 6B-C). In many cases, these EIN5-affected
smRNAs are processed from along the entire length of the parental transcript in multiple 21 nt registers, which suggests there is not an ordered processing event or “common” mRNA terminus that is consistently required for biogenesis of these smRNAs (Figures 6B, S8). There were also numerous examples where the EIN5-affected smRNAs accumulated to significantly higher levels in ein5-6 abh1-1 compared to ein5-6 plants (Figures 6B, 6D, S9 and Table S12). In fact, these smRNAs compose ~19% of the entire 21 nt smRNA population of ein5-6 abh1-1 double mutant plants (Figure S10), which suggested that concomitant loss of ABH1 and EIN5 results in the enhanced processing and/or stability of these smRNAs. Interestingly, we observed that often low levels of EIN5-affected smRNAs were present in wild-type plants (Figures 6B, 6D, S9 and Table S13), which indicates that this class of smRNAs are indeed processed under normal conditions and are not just a byproduct of unnatural consequences (ein5 mutation). Overall, these results suggest that the 5’-3’ ribonuclease EIN5 negatively regulates the levels of a class of endogenous siRNAs that are processed from functionally-annotated Arabidopsis gene transcripts.

**Endogenous uncapped transcripts likely act as smRNA biogenesis substrates**

It had been previously reported that EIN5 antagonizes RDR-dependent RNA silencing of exogenously introduced transgenes likely through degradation of “aberrant” mRNAs that may lack a 5’ terminal cap moiety and serve as biogenesis substrates for smRNA production (Gazzani et al., 2004). Many of the features exhibited by our endogenous EIN5-affected smRNA-generating transcripts (Figure 6) suggest that they may also lack a 5’ terminal cap moiety, and therefore are regulated in a similar manner to what was previously witnessed for transgene mRNAs. To detect uncapped 5’ RNA ends, we used a modified 5’-RACE protocol (Llave et al., 2002). Full-length, decapped mRNA corresponding to the EIN5-affected smRNA-generating transcripts (verified by cloning and sequencing of the cDNA (Figure S11)) consistently accumulated in ein5-6 and ein5-6 abh1-1 plants (Figure 7A). In order to generate a more comprehensive, genome-wide view of all 133 smRNA-generating transcripts, we designed an Illumina GA sequencing-based 5’-RACE assay termed Genome-wide Mapping of Uncapped and Cleaved Transcripts (GMUCT) (see Figure S12, and Materials and Methods for details).
We then applied GMUCT to unmodified, poly-adenylated RNA samples from wild-type Col-0 and ein5-6 immature flower buds. As expected for this modified 5’-RACE protocol, we were able to identify known miRNA-mediated target mRNA cleavage sites from the obtained sequencing data (Figure S13), thereby providing validation of this methodology. Using this novel assay we determined that at least 46% of the EIN5-affected siRNA-generating transcripts accumulate in an uncapped form in the absence of EIN5 function (Figures 7B-C). Overall, we validated ~50% of the total population of EIN5-affected smRNA-generating transcripts as accumulating in an uncapped form in ein5-6 mutant plants through the combination of 5’-RACE and GMUCT (Figure 7B), with two of these mRNAs being validated by both methodologies. These results demonstrated that EIN5 is responsible for the removal of the uncapped form of these 133 transcripts. Thus, we have identified a class of endogenous EIN5-affected mRNAs that are processed into smRNAs by an RNA silencing pathway similar to one previously observed to regulate “aberrant” transgenic RNAs ((Gazzani et al., 2004), Figure 7). More specifically, the endogenous mRNAs also accumulate in an uncapped form, which likely is the defining feature that shunts them into an RNA silencing pathway where they are processed into mostly 21 nt smRNAs.

**DISCUSSION**

The mRNA cap-binding complex (CBC) is known to increase the efficiency of pre-mRNA splicing and non-sense mediated decay (NMD) (an mRNA surveillance pathway that generally eliminates messenger RNAs that prematurely terminate translation) through direct interaction with proteins involved in these two important processes (Aguilera, 2005). Therefore, we hypothesize that as ABH1 is a subunit of the CBC, its role in the miRNA-mediated mRNA silencing pathway could be to increase the efficiency of processing of these smRNAs from the initial primary MIRNA transcript. Specifically, we suggest that CBC bound to the 5’ end of primary MIRNA transcripts interacts with one of the proteins of the miRNA biogenesis complex (SERRATE, DCL1, and/or HYL1), thus increasing the likelihood of proper processing of these important smRNAs. The overlapping phenotypic similarities between abh1 and se mutant plants ((Bezerra et al., 2004) and Figures 2 and 5) offer a tantalizing suggestion that ABH1...
may interact directly with SERRATE to ensure proper and efficient processing of primary MIRNA transcripts into the biologically-active miRNA products.

Interestingly, consistent with our findings and providing evidence that a role for ABH1 in the miRNA-mediated RNA silencing pathway may also be conserved in animals, Parry et al. (Parry et al., 2007) recently performed a whole-genome RNA interference (RNAi) screen in C. elegans that suggested a function for the worm homologue of ABH1 (F37E3.1) specifically in the let-7 miRNA-mediated RNA silencing pathway. Upon performing smRNA northern blots using RNA isolated from young adult worms that were fed an RNAi knockdown construct targeting F37E3.1, only a slight reduction in mature let-7 levels compared to wild-type animals was detected. Based on these results, this group concluded that the miRNA-debilitating gene inactivation of F37E3.1 likely abrogated gene function downstream of the expression and processing of let-7. Here, we demonstrate that Arabidopsis ABH1 is intimately required for accumulation of mature miRNA levels, which suggests a role for this protein either in or upstream of miRNA processing. These discrepancies may result from differences in the miRNA biogenesis pathways between the two organisms (C. elegans and Arabidopsis) examined in these studies, or more likely are due to differences in methodologies employed; RNAi knockdown (Parry et al., 2007) versus stable genetic mutation (abh1 mutant plants (this study)). The latter suggestion is supported by the fact that even complete and stable loss of ABH1 in Arabidopsis caused an ~66% reduction in mature miRNA levels not a complete loss (Figures 4 and S6-7). Therefore, transient and incomplete reduction of F37E3.1 levels by RNAi likely only causes a slight decrease in mature miRNA levels, which would result in the differing interpretation of results. Nevertheless, in combination these studies suggest an unexpected role for ABH1 (CBP80) in the miRNA-mediated RNA silencing pathway of eukaryotic organisms.

Additionally, we demonstrate that EIN5 affects the levels of a class of mostly 21 nt smRNAs that are processed from functionally-annotated Arabidopsis gene transcripts (Figure 6). We find that the smRNA-generating mRNAs accumulate in an uncapped form in the absence of EIN5 function (Figure 7), which suggests that this may be the defining feature of these transcripts that shuttles them into an RNA silencing pathway, much like was previously demonstrated to occur for capless mRNAs derived from
introduced transgenes (Gazzani et al., 2004). An alternative hypothesis would be that loss of EIN5 function allows cryptic anti-sense transcripts specific to the 133 EIN5-affected smRNA-generating loci to accumulate, thereby resulting in dsRNA formation and subsequent DICER processing into smRNAs. We strongly disfavor this latter hypothesis because our tiling microarray analysis did not reveal upregulated anti-sense transcripts corresponding to any of the 133 EIN5-affected smRNA-generating loci. Additionally, none of the EIN5-affected smRNAs map to the intronic regions of transcripts from which they are processed (Figure 6B), which strongly suggests that the mature mRNA molecule itself is acting as the template for synthesis of the anti-sense strand likely in an RDR-dependent manner. Furthermore, we observe an enhancement in the levels of many of the EIN5-affected smRNAs in ein5-6 abh1-1 double mutant plants compared to any of the other three genotypes (Figures 6B, 6D S9 and Table S13), which suggests that the processing and/or stability of these smRNAs is enhanced by the absence of both ABH1 and EIN5 function. While these results led us to prefer the latter explanation where enhanced processing of the EIN5-affected smRNAs results from an increased availability of their biogenesis substrate, these models are not mutually exclusive. Specifically, based on the data from the ein5-6 abh1-1 double mutant, we hypothesize that the loss of ABH1 function likely reduces protection of the 5' cap structure from removal by decapping enzymes, while at the same time the lack of functional EIN5 stabilizes these uncapped RNA intermediates. Overall, these results support a model in which the 133 transcripts that accumulate in an uncapped form in the absence of EIN5 function are shunted into an RDR-dependent RNA silencing pathway where they are processed into smRNAs. Furthermore, the findings presented herein for the single and double mutants suggest an integration of EIN5 and ABH1 functions in protecting these 133 endogenous transcripts from this proposed fate.

Recently, it was demonstrated that the terminal 5' nucleotide of Arabidopsis smRNAs specifies the AGO protein with which they associate. Specifically, it was found that smRNAs with a 5' terminal adenosine preferentially associate with AGO2 and AGO4, smRNAs with a 5' terminal cytosine preferentially associate with AGO5, and smRNAs with a 5' terminal uridine preferentially associate with AGO1 (Mi et al., 2008). From deep smRNA sequencing, we find that the majority of 21 nt smRNAs from
immature flower buds begin with a 5’ adenosine (Figure S14 and Table S13), and this is also observed for the entire population of sequenced smRNAs (data not shown). Therefore, the majority of smRNAs in these plant organs associate with AGO2 and AGO4. Guanine is the next most abundant 5’ terminal nucleotide for 21 nt smRNAs, while T and C are the least favored 5’ terminal nucleotides for 21 nt smRNAs (Figure S14 and Table S13). Interestingly, the distribution of 5’ terminal nucleotides amongst the 21 nt EIN5-affected smRNAs is not significantly different from the total population of 21 nt smRNAs (Figure S14 and Table S13). Overall, these results suggest that EIN5-affected smRNAs are quite diversified at the 5’ terminal nucleotide position, which likely allows them to associate with a number of different *Arabidopsis* AGO proteins.

As noted above, we observed numerous cases of EIN5-affected smRNAs that correspond to gene transcripts also detectable in wild-type plants (Figures 6B, 6D, S9 and Table S13), which indicates that the processing of this class of smRNAs is not simply a byproduct of genetic mutation of *EIN5* and *ABH1*. An intriguing possibility is that that these smRNAs may have regulatory function, and that they might be processed more readily and/or accumulate in response to specific physiological or environmental conditions in which the function of EIN5 is negatively regulated. In this model, the expansion of EIN5-affected smRNAs could thus provide an additional population of smRNAs to regulate target mRNAs with sites of sequence complimentarity both in *cis* and *trans*. Although evidence that EIN5-affected smRNAs are able to regulate mRNAs posttranscriptionally is still lacking, this tantalizing possibility warrants further examination. Overall, these findings suggest that there is still much to learn concerning RNA silencing and the regulation of these posttranscriptional regulatory pathways.

**EXPERIMENTAL PROCEDURES**

**Plant materials**

The Columbia (Col-0) ecotype of *Arabidopsis* was used in this study. The *abh1-1* and *ein5-6* mutants were previously identified (Hugouvieux et al., 2001; Hugouvieux et al., 2002; Olmedo et al., 2006). *abh1-8* (this study) was recovered from the Salk T-DNA collection (Alonso et al., 2003). *ago1-38* comes from an ethyl methanesulfonate (EMS)-
mutagenized population, and contains a Gly to Arg substitution at position 186 (seeds kindly provided by E. Meyerowitz, California Institute of Technology).

**Genetic Analysis of Double Mutants**

Double mutant plants were generated by genetic crosses, and homozygous lines were identified by PCR-based genotyping. All primers used for PCR-based genotyping of mutant plants used in this study can be found in Table S14. In the case of ein5-6 genotyping, the amplified PCR products were subsequently cleaved with MboII. In the case of ago1-38, genotyping the amplified PCR products were subsequently cleaved with EcoNI.

**RNA Analyses**

Immature flower bud clusters were collected for RNA isolation using the RNeasy Plant Mini Kit (Qiagen, Valencia, CA) for all experiments performed herein. Low-molecular-weight RNA was purified from total RNA and northern blots for mature miRNAs were performed as previously described (Peragine et al., 2004). Northern blot analysis of EBF1 and EBF2 mRNA levels was performed as previously described (Olmedo et al., 2006). Transcripts were quantified by RT qPCR using the comparative threshold cycle method (\(\Delta\Delta C_t\), primers listed in Table S14), using Actin 2 (At3g18780) as the endogenous reference.

**5' Rapid Amplification of cDNA ends (5' RACE)**

5' RACE and cloning of 5' RACE products was carried out using the GeneRacer Kit (Invitrogen, Carlsbad, CA) as previously described (Kasschau et al., 2003). Briefly, there were no modifications made to the total RNA used for the 5'-RACE experiments. The 5'-adapter molecule was immediately ligated onto the total RNA population, thereby isolating only mRNAs or mRNA fragments with a 5' end with a free 5' phosphate. 5'-RACE products of miRNA-directed cleavage (Figure S7) were quantified by RT qPCR using the comparative threshold cycle method (\(\Delta\Delta C_t\), primers listed in Table S14), using Actin 2 (At3g18780) as the endogenous reference.
Microarray Analysis

Tiling microarray analysis was carried out as previously described (Chekanova et al., 2007). For more detailed methodology see Supplemental Methods.

smRNA library construction

Small RNA libraries were constructed as per manufacturer’s instructions (Illumina Inc., San Diego, CA). For more detailed methodology see Supplemental Methods and Figure S3.

GMUCT sequencing library construction

Briefly, GMUCT libraries were constructed by adapting a modified 5’-RACE protocol (see 5’-RACE section above) for sequencing on an Illumina Genetic Analyzer. For more detailed methodology see Supplemental Methods and Figure S12.

High-throughput sequencing

Small RNA and GMUCT libraries were sequenced using the Illumina Genetic Analyzer (GA) as per manufacturer’s instructions (Illumina Inc., San Diego, CA), except sequencing of GMUCT libraries was performed for 50 cycles to yield longer sequences that are more amenable to unambiguous mapping to the Arabidopsis genome sequence.

Mapping small RNA reads

Sequence information was extracted from the image files with the Illumina Firecrest and Bustard software applications. Prior to alignment of the smRNA reads, a custom Perl script was used to identify the first seven bases of the 3’ adapter sequence, and the read was truncated up to the junction with the adaptor sequence. No further analysis was performed on reads that do not contain adapter sequence, as those reads lacking an adapter cannot be precisely sized. The reads were then mapped to the Arabidopsis Col-0 reference genome (TAIR v.7) with the Illumina Eland application. Eland aligns 32 base pair or shorter reads with up to two mismatches to a reference genome. However, only perfect matches were excepted since shorter reads will have a
higher tendency to falsely map than longer reads. Since Eland does not return positions for reads that map to multiple positions in the reference genome, we utilized the Basic Local Alignment Search Tool (BLAST) to map these non-unique reads, run on using a word size of 10 and expectation value of 10. Once again, only perfect matches were accepted, as these shorter reads will have a higher tendency to falsely map than longer reads.

**Mapping GMUCT reads**

In order to avoid omitting reads mapping within unannotated transcripts, 36 base sequence reads were aligned to the Arabidopsis reference genome sequence (TAIR v.7) with the ELAND algorithm.

**Accession Numbers**

All raw microarray data (CEL files) for expression analyses (wild typeCol-0, abh1-1, ein5-6, and ein5-6 abh1-1), as well as smRNA (Col-0, abh1-1, ein5-6, and ein5-6 abh1-1 libraries), and GMUCT (wild-type Col-0 and ein5-6 libraries) sequence data from our analyses were deposited in GEO under the accession GSE11070.

**REFERENCES**


ACKNOWLEDGMENTS
The authors thank Angelita C. Garcia for technical assistance; Kim Emerson for assistance with preparing digital artwork; and Jennifer C. Fletcher, Eva C. Ziegelhoffer, and Elliot M. Meyerowitz for ago1-38 seeds. B.D.G. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-1909-06). R.L. is supported by a Human Frontier Science Program Long-term Fellowship. J. T-F. is supported by Hackett and Shacklock Scholarships from the University of Western Australia. The development of the visualization and genome profiling browser AnnoJ was supported by grants from the Australian Research Council to AHM (CE0561495, DP0771156). This work was supported by grants from the NSF (MCB-0516888) and DOE (DE-FG02-04ER15517) to J.R.E.

FIGURE LEGENDS

Figure 1. The ethylene insensitivity of ein5 mutant plants is suppressed by abh1 mutation. (A-D) Triple-response phenotypes of wild-type Col-0, abh1-1, ein5-6, or ein5-6 abh1-1 3-day-old etiolated seedlings grown on LS medium without the ethylene precursor ACC (A), or with 0.1 µM (B), 1 µM (C), 10 µM ACC (D). (E) Dosage response of 3-day-old etiolated seedlings of the indicated genotypes to various concentrations of ACC. (F) Levels of EBF1 and EBF2 mRNA determined by northern analyses of total RNA from 3-day-old etiolated seedlings from the indicated genotypes grown in hydrocarbon-free air or treated for various times (indicated above wells) with 10 ppm ethylene (C2H4) gas using a 3' probe to EBF1 or EBF2. 25S rRNA shown as a loading control. Normalized values of EBF1 and EBF2 mRNA to rRNA control (with Col-0 without ethylene treatment set at 1.0) are indicated below the respective blot.

Figure 2. ein5-6 abh1-1 double mutant plants manifest developmental defects similar to those observed for miRNA pathway mutants. (A-E) 1-week-old seedlings of wild-type Col-0 (A), abh1-1 (B), ein5-6 (C), and ein5-6 abh1-1, one of which manifests partial (D) and the other full (E) cotyledon fusion phenotypes. (F-J) 10-day old seedlings of the same genotypes. (K-O) 2-week-old seedlings of the same genotypes. (P) The percentage of the indicated genotype seedlings that manifest the cotyledon fusion defect, scored using 1-week old plants. (Q) Leaves of 5-week-old plants of the indicated
genotypes. Both *ein5-6* and *abh1-1* single mutant plants have serrated leaf edges, which is strongly enhanced on *ein5-6 abh1-1* double mutant plants. The leaf margins of *ein5-6 abh1-1* double mutant plants are highly dentate and also often become asymmetric (left leaf) in their growth pattern. (R-U) Inflorescence stems of the indicated genotypes. (V) The percentage of the indicated genotypes inflorescence stems that manifest the developmental defect where multiple fruits (2, 3, or more than 3) emanate from the same node.

Figure 3. Tiling microarray analysis suggests a role for ABH1 and EIN5 in the miRNA-mediated RNA silencing pathway. (A) The levels of primary *MIRNA158a*, *MIRNA164b*, *MIRNA167a*, and *MIRNA168a* mRNA were determined by *Arabidopsis* whole genome tiling microarray expression analysis. The top four tracks display the level of primary *MIRNA* mRNAs in the indicated genotypes. The bottom track is the annotated gene models for the primary *MIRNA* loci. (B) Oligo(dT)-primed RT qPCR specific for the stem-loop region of primary *MIRNA158a*, *MIRNA164b*, *MIRNA167a*, and *MIRNA168a* mRNAs for wild-type Col-0 (blue bar), *abh1-1* (red bar), *ein5-6* (yellow bar), and *ein5-6 abh1-1* (light green bar) plants. Error bars, ±SD. (C) The levels of specific miRNA-target mRNAs *CUC1* (*At3g15170*) and *HAP2C* (*At1g72830*), were determined by *Arabidopsis* whole genome tiling microarray expression analysis. The top four tracks display the level of *CUC1* and *HAP2C* mRNAs in the indicated genotypes. The bottom track is the annotated gene models for the *CUC1* and *HAP2C* loci. (D) Oligo(dT)-primed RT qPCR analysis of *CUC1* and *HAP2C* mRNA levels. Error bars, ±SD.

Figure 4. ABH1 affects the miRNA-mediated RNA silencing pathway. (A) Panels demonstrate the normalized abundance of 21 nt smRNAs (y axis, left-side scale) in a sliding 250 kb window in wild-type Col-0 (blue line), *abh1-1* (pink line), *ein5-6* (yellow line), and *ein5-6 abh1-1* (green line) double mutant plants. Red arrows indicate peaks of EIN5-affected smRNAs sequenced from *ein5-6* and *ein5-6 abh1-1* libraries. (B) smRNA northern blot analysis of samples from indicated genotypes with DNA probes complementary to miRNA158, 164, 167, and 390. U6 is shown as a loading control. (C) Normalized values of miRNA to the U6 control for 3 biological replicates for the
indicated genotypes. The Col-0 value is set as 1.0. (D) Agarose gel analyses of 5’-RACE products for At5g07680, ARF8 (At5g37020), and APETALA2 (At4g36920) miRNA-directed cleavage from the indicated genotypes. Actin was used as a loading control for the 5’-RACE experiments. (E) Oligo(dT)-primed RT qPCR analysis of 5’-RACE products for At5g07680, ARF8 (At5g37020), and APETALA2 (At4g36920) miRNA-directed cleavage from the indicated genotypes. Error bars, ±SD.

Figure 5. The abh1 mutation enhances developmental defects associated with plants containing a hypomorphic genetic lesion in AGO1, ago1-38. (A-E) 1-week-old seedlings of wild type (Col-0) (A), abh1-8 (B), ago1-38 (C), and abh1-8 ago1-38 (D-E) double mutant plants, one of which manifests a cotyledon fusion phenotype and has no root (E). (F-J) 10-day old seedlings of the same genotypes (K-O) 2-week-old seedlings of the same genotypes (P) The percentage of wild-type (Col-0), abh1-8, ago1-38, and abh1-8 ago1-38 double mutant seedlings that manifest the cotyledon fusion defect, which was scored using 1-week old plants. (Q) The percentage of wild-type (Col-0), abh1-8, ago1-38, and abh1-8 ago1-38 double mutant seedlings that do not develop a root, which was scored using 2-week old plants. (R) Inflorescence stems of plants of the indicated genotypes. (S) The percentage of wild-type (Col-0), abh1-8, ago1-38, and abh1-8 ago1-38 double mutant inflorescence stems that manifest the developmental defect where multiple fruits (2, 3, or more than 3) emanate from the same node.

Figure 6. Clusters of EIN5-affected smRNAs that emanate from functionally-annotated endogenous transcripts. (A) Venn diagram representation of annotated transcripts from which multiple 21 nt smRNAs were sequenced at least twice for abh1-1, ein5-6, or ein5-6 abh1-1 double mutant plants but not for wild-type Col-0. (B) Two examples of EIN5-affected 21 nt smRNA-generating transcripts (screenshots from the smRNAome database, http://neomorph.salk.edu/aj_salk/smRNAome.html). W (red bars) and C (green bars) indicate signal from Watson and Crick strands, respectively. (C) Table of Gene Ontology (GO) functional annotations that are over-represented by at least 1.5-fold compared to the Col-0 reference genome in the 133 transcripts from which the EIN5-affected smRNAs are processed. (D) The total number of smRNA sequence reads.
that correspond to EIN5-affected smRNAs from the wild-type Col-0, abh1-1, ein5-6, and ein5-6 abh1-1 libraries, which was determined using the data in Table S13.

Figure 7. Transcripts from which EIN5-affected smRNAs are processed accumulate in an uncapped form. (A) 5'-RACE products of SPL4, At1g75240, At2g35945, and At3g46550 that map to the very 5' end (free 5' terminal phosphate) of the transcripts, which can only be detected in ein5-6 and ein5-6 abh1-1 samples. Actin was used as a loading control. (B) The percent of EIN5-affected smRNA-generating transcripts validated as accumulating in an uncapped form by two methods, 5'-RACE and/or GMUCT. (C) Two examples of uncapped EIN5-affected smRNA-generating transcripts (screenshots from the smRNAome database) that were validated by the GMUCT method. W and C indicate signal from Watson and Crick strands, respectively. Gray bars represent independent Illumina sequencing reads.