Published online 11 July 2020

Nucleic Acids Research, 2020, Vol. 48, No. 15 ***8767–8781***
doi: 10.1093/nar/gkaa577

Intact RNA structurome reveals mRNA structure-mediated regulation of miRNA cleavage *in vivo*

Minglei Yang[1,†], Hugh C. Woolfenden[1,†], Yueying Zhang[1,†], Xiaofeng Fang[1,†], Qi Liu[1], Maria L. Vigh[2], Jitender Cheema[1], Xiaofei Yang[1], Matthew Norris[1], Sha Yu[1,3], Alberto Carbonell[4], Peter Brodersen[2], Jiawei Wang[3,5] and Yiliang Ding[1,*]

[1]Department of Cell and Developmental Biology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK, [2]Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200, Copenhagen N, Denmark, [3]National Key Laboratory of Plant Molecular Genetics (NKLPMG), CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology (SIPPE), Shanghai Institutes for Biological Sciences (SIBS), Shanghai 200032, People's Republic of China, [4]Instituto de Biología Molecular y Celular de Plantas (Consejo Superior de Investigaciones Científicas-Universidad Politécnica de Valencia), Valencia, 46022, Spain and [5]ShanghaiTech University, Shanghai 200031, People's Republic of China

Received February 27, 2020; Revised June 11, 2020; Editorial Decision June 23, 2020; Accepted June 24, 2020

ABSTRACT

MicroRNA (miRNA)-mediated cleavage is involved in numerous essential cellular pathways. miRNAs recognize target RNAs via sequence complementarity. In addition to complementarity, *in vitro* and *in silico* studies have suggested that RNA structure may influence the accessibility of mRNAs to miRNA-induced silencing complexes (miRISCs), thereby affecting RNA silencing. However, the regulatory mechanism of mRNA structure in miRNA cleavage remains elusive. We investigated the role of *in vivo* RNA secondary structure in miRNA cleavage by developing the new CAP-STRUCTURE-seq method to capture the intact mRNA structurome in *Arabidopsis thaliana*. This approach revealed that miRNA target sites were not structurally accessible for miRISC binding prior to cleavage *in vivo*. Instead, we found that the unfolding of the target site structure plays a key role in miRISC activity *in vivo*. We found that the single-strandedness of the two nucleotides immediately downstream of the target site, named Target Adjacent nucleotide Motif, can promote miRNA cleavage but not miRNA binding, thus decoupling target site binding from cleavage. Our findings demonstrate that mRNA structure *in vivo* can modulate miRNA cleavage, providing evidence of mRNA structure-dependent regulation of biological processes.

INTRODUCTION

MicroRNAs (MiRNAs) are ∼21 nt RNAs that are involved in various aspects of development and stress responses by post-transcriptionally regulating gene expression (1). MiRNAs are loaded onto ARGONAUTE proteins (AGO) to form functional post-transcriptional gene silencing effector complexes, miRNA-induced silencing complexes (miRISCs) (2). miRISC is guided by the miRNA to bind to target RNAs through sequence complementarity and then to cleave the target RNAs (3,4). However, previous studies found that sequence complementarity is not the sole factor dictating miRNA cleavage (2), with RNA structure suggested to influence the silencing efficiency (5–7). However, these studies were unable to reveal native RNA structure features for several reasons. First, these *in vitro* and *in silico* studies do not reflect the RNA structure folding status in living cells (8–10). Second, the structure was assessed by introducing an artificial long sequence (predicted to form a strong structure, e.g. a hairpin). Finally, the target site together with the flank regions were assessed as one, thus making it difficult to separate the contribution from each region (5–7). This confounded the identification of a specific RNA structure motif that regulated miRNA cleavage.

To assess whether RNA structure influences miRNA cleavage *in vivo*, a method is required that obtains RNA structures before cleavage. Recently, several transcriptome-wide *in vivo* structure probing methods for RNA have been established (8–10), and these provide powerful tools for acquiring RNA structures under physiological conditions. These methods are based on the quantification

*To whom correspondence should be addressed. Tel: +44 1603 450266; Fax: +44 1603 450045; Email: yiliang.ding@jic.ac.uk
†The authors wish it to be known that, in their opinion, the first four authors should be regarded as Joint First Authors.

© The Author(s) 2020. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

of sequencing reads and in particular the reverse transcriptase (RT) stalling position on an RNA in a sample treated with a 'structure probing' chemical and a control sample. The DMS-based methods use **dim**ethyl **s**ulphate to probe for single-stranded A and C nucleotides (11), while the SHAPE (**S**elective 2′-**H**ydroxyl **A**cylation analyzed by **P**rimer **E**xtension)-based methods utilize a variety of chemicals, e.g. 2-methylnicotinic acid (NAI), to probe the single-strandedness of all four RNA nucleotides (12). The RT stop counts are used to generate a per nucleotide chemical reactivity with a high value indicative of single-strandedness. However, the RNA structure methods that detect RT stalling (9–10,13–14) or alternative methods that detect nucleotide mutation (15,16) are unable to discern whether the chemical reactivity represents the RNA structure information for the endogenous degraded RNAs or for the intact RNAs (Supplementary Figure S1). Additionally, degraded mRNAs are capable of introducing false positive signals in the RT stalling methods because the 5′ end of the degraded mRNA will have an extremely high RT stalling signal (Supplementary Figure S1). Therefore, these methods are not able to reveal the causal relationship between RNA structure and miRNA cleavage.

To decipher the *in vivo* relationship between mRNA structure and miRNA cleavage, we developed the novel method, CAP-STRUCTURE-seq, to obtain *in vivo* structures of target mRNAs before cleavage. We found that miRNA target sites were not structurally accessible *in vivo*. Instead, our analysis suggests that the unfolding of the target site structure could be the primary determinant of miRISC binding prior to cleavage *in vivo*. Furthermore, by assessing the structure features flanking the miRNA target sites, we find that the single-strandedness of the two nucleotides immediately downstream of the target site, which we named Target Adjacent nucleotide Motif (TAM), can promote miRNA cleavage but not miRNA binding. Thus, TAM decouples target site binding from cleavage. Our study revealed the role of *in vivo* mRNA structure in the regulation of miRNA cleavage, providing evidence of mRNA structure-dependent regulation of biological processes.

MATERIALS AND METHODS

Plant materials and growth conditions

Arabidopsis thaliana seeds of the Columbia (Col-0) and the *xrn4* mutant accession (17,18) were sterilized with 70% (v/v) ethanol and plated on half-strength Murashige and Skoog medium (1/2 MS). The plates were wrapped in foil and stratified at 4°C for 3–4 days and then grown in a 22–24°C growth chamber for 5 days.

Gel-based 18S rRNA structure probing

The gel-based method of structure probing used the same *in vivo* total RNA pools as for CAP-STRUCTURE-seq. To accomplish gel-based structure probing, reverse transcription was performed using 18S rRNA gene-specific DNA primers with 5′ end labeled Cy5 (TAGAATTACTACGG TTATCCGAGTA). The whole procedure was performed according to Ding *et al*. (8). Each gel was detected by Typhoon FLA 9500 (GE Healthcare).

(+)SHAPE and (−)SHAPE CAP-STRUCTURE-seq library construction

We modified the *in vivo* chemical probing protocol (8) by changing the reagent from dimethyl sulphate (DMS) to the SHAPE reagent, 2-methylnicotinic acid (NAI). NAI was prepared as reported previously (19). Briefly, 5-day-old *A. thaliana* etiolated seedlings were suspended and completely covered in 20 ml 1× SHAPE reaction buffer (100 mM KCl, 40 mM HEPES (pH7.5) and 0.5 mM $MgCl_2$) in a 50 ml Falcon tube. NAI was added to a final concentration of 150 mM and the tube swirled on a shaker (1000 rpm) for 15 min at room temperature (22°C). This NAI concentration and reaction time had been optimized to allow NAI to penetrate plant cells and modify the RNA *in vivo* under single-hit kinetics conditions (Supplementary Figure S2A). After quenching the reaction with freshly prepared dithiothreitol, the seedlings were washed with deionized water and immediately frozen with liquid nitrogen and ground into powder. Total RNA was extracted using RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions, followed by on-column DNaseI treatment in accordance with the manufacturer's protocol. The control group was prepared using DMSO (dimethyl sulfoxide, labeled as (−)SHAPE), following the same procedure as described above.

To capture the structure information around the cleavage site of miRNA target genes, we adopted the feature of 5PSeq (20). The whole CAP-STRUCTURE-seq procedure is illustrated in Figure 1. In our method, the (+)SHAPE and (−)SHAPE RNA samples were treated with Terminator™ 5′-Phosphate-Dependent Exonuclease (TER51020, EPICENTRE CO.), which processively digests RNA with 5′-monophosphate ends, thereby leaving mRNAs with 5′cap structures (Supplementary Figure S2B). Following the 5′cap enrichment, polyA+ selection was carried out using the PolyA purist Kit (Ambion™) leaving intact (precleaved) mRNAs with enriched 5′cap and 3′poly(A) tails. The resultant mRNAs were subjected to library construction following the STRUCTURE-seq procedure on Illumina HiSeq 4000 (BGI). The name of CAP-STRUCTURE-seq refers to 5′**CAP**-enriched and 3′ poly(A)-enriched RNA **structure** **seq**uencing.

CAP-STRUCTURE-seq analysis

We merged the biological replicates of the transcript-level reverse transcription (RT) stop counts to obtain a single (−)SHAPE library and a single (+)SHAPE library. We calculated the SHAPE reactivity using a slightly modified version of the formula in Ding *et al*. (8),

$$\text{SHAPE reactivity}_i = \frac{\log(1+P_i)}{\sum_i \log(1+P_i)} - \alpha \frac{\log(1+M_i)}{\sum_i \log(1+M_i)},$$

where P_i is the (+)SHAPE RT count and M_i is the (−)SHAPE RT count at nucleotide i. The factor, α (= $\min(1, \sum_i \log(1+P_i) / \sum_i \log(1+M_i))$ is a simple library size correction factor. Setting $\alpha = 1$ recovers the reactivity formula in Ding *et al*. (8). The reactivities were then normalised using the box-plot method (21). For the SHAPE

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

reactivity profiles, we extracted values in the 50 nt upstream and downstream of target sites and calculated a per nucleotide mean and standard error of the mean (SEM).

Degradome library construction

Five-day-old *A. thaliana* etiolated seedlings were grown as described above. They were collected and immediately frozen in liquid nitrogen and stored at −80°C. The seedlings were ground into powder. Total RNA was extracted using RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. On-column DNAaseI treatment was carried out according to RNase-Free DNase Set (Qiagen). To construct the Illumina library for degradome analysis, polyA+ selection was carried out using the Poly(A)Purist Kit (Ambion™). Selectively captured polyadenylated RNAs (1 μg) were ligated directly to an DNA/RNA hybrid adapter (5′-CTACAC GACGCTrC rUrUrCrCrGrArUrCrUrNrNrN-3′) using T4 RNA ligase (NEB) at 37°C for 30 min. The ligated RNAs were subjected to RT by SuperScript III First-Strand Synthesis System (Invitrogen) with random hexamers fused with Illumina TruSeq adapters (5′-CAGACGTGTGCTCTTCC GATCTNNNNNN-3′). Polymerase chain reaction (PCR) amplification was performed on the ligated cDNA using Illumina TruSeq Primers. Two different barcode indices were used for two degradome biological replicates. The final dsDNA degradome libraries were subjected to next-generation sequencing on Illumina HiSeq 4000 (BGI).

Degradome analysis

Raw reads were processed to remove 5′and 3′ adapter sequences. Degradome reads were mapped to the TAIR10 transcript reference and a degradome density file was generated. The degradation level of target genes was normalized by reads per kilobase per million mapped reads (RPKM).

miRNA library construction

The same seedling samples stored at −80°C, as described above, were ground into powder using liquid nitrogen. Total RNA was extracted using mirVana miRNA Isolation Kit (Ambion™, Austin, TX, USA) following the manufacturer's instructions. The integrity analysis was performed on a Bioanalyzer by the Beijing Genomics Institute (BGI), Shenzhen, China, which also performed the library construction according to standard protocols.

miRNA-seq analysis

The small RNA sequences were processed by BGI to filter out the 5′ adapter sequences, 3′ adapter sequences and low-quality reads. We mapped two biological replicates against 253 miRNA sequences confidently annotated as *A. thaliana* mature miRNAs (22). We used Bowtie (23) for the mapping using the command 'bowtie -f -a -S –best –strata -v 1'. pysam (23) was used to count the mapped reads.

Cleavage efficiency (CE) calculation

The cleavage efficiency (CE) can be estimated by,

$$\mathrm{CE} \propto \frac{\mathrm{mirDegradome\,[RPKM]}}{((-)\mathrm{SHAPE\ [RPKM]} + \mathrm{mirDegradome\,[RPKM]}) \times \mathrm{miRNAseq\,[RPKM]}}$$

Briefly, we first identified how much miRNAs were expressed in our samples using the RPKM values from miRNA-seq. We used TargetFinder (24) to predict the miRNA target sites on the expressed transcripts in our samples and removed any duplicated target sites from the same miRNA family. TargetFinder predicts target sites with high specificity in *A. thaliana* by assigning a sequence complementarity penalty score (25) (Supplementary Figure S3A). Then, we mapped degradome reads to the reference transcripts that had been identified as miRNA target genes in *Arabidopsis* (22). We counted the reads within the target sites as the degradation products causing miRNA-mediated cleavage. Then, we summed the RPKM values from the (−)SHAPE and Degradome library to yield an estimate of how many transcripts served as substrates of miRNAs. The benefit of combining the (−)SHAPE and the degradome libraries to calculate the CE lies in its focus on miRNA-mediated cleavage events. The CE pipeline is illustrated in Supplementary Figure S3B and the derivation of the CE formula is described in the Supplementary Methods.

Calculation of $\Delta G^{\ddagger}_{open}$ and $\Delta G^{\ddagger}_{cutting}$

$\Delta G^{\ddagger}_{open}$ measures the energy required to open the target sites during miRISC binding. $\Delta G^{\ddagger}_{open}$ was computed as the difference between the minimum free energy of the *in vivo* secondary structure and the minimum free energy of the 'hard constrained' transcript, in which the target nucleotides were required to be unpaired (6,26). By exploring a range of flanking region lengths upstream and downstream of the target site (Supplementary Figure S8), we chose the upstream and downstream flank lengths to be 50 nt for the majority of analyses. We used *RNAfold* from the Vienna RNA package (27) together with our SHAPE reactivity data to calculate the energy terms in $\Delta G^{\ddagger}_{open}$, the RNA structures and the base pairing probabilities (BPP).

$\Delta G^{\ddagger}_{cutting}$ measures the energy required to raise the initial substrate target RNA to the transition catalysis-compatible state (Figure 3A), and is given by:

$$\Delta G^{\ddagger}_{cutting} = \Delta G^{\ddagger}_{open} - \left|\Delta G_{duplex}\right| + \Delta G_{catalysis},$$

where ΔG_{duplex} is the binding free energy of the miRNA-target duplex, and $\Delta G_{catalysis}$ refers to the miRISC transition catalytic state energy. ΔG_{duplex} was calculated for the miRNA sequence and the target region sequence using *RNAduplex* from the Vienna RNA package (27).

The crystal structures of *Thermus thermophilus* Argonaute (TtAgo) (28,29), human AGO2 (30) and yeast *Kluyveromyces polysporus* Argonaute (KpAGO) (31) suggest that AGO proteins have a conserved catalytic mechanism. Furthermore, the transition cleavage model does not engage in any nucleotide interactions (28,29). Therefore, we assumed that the activation energy, $\Delta G_{catalysis}$, is a constant for the same type of AGO protein. Quantum mechanics simulations estimate the value to be ∼15 kcal mol^{-1} (32).

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

Therefore, $\Delta G^{\ddagger}_{cutting}$ is given by:

$$\Delta G^{\ddagger}_{cutting} = \Delta G^{\ddagger}_{open} - \left|\Delta G_{duplex}\right| + 15.$$

Plasmid construction

For CE validation, the miRNA156 target sites, followed by 0 or 2 Adenines (As) and ending with a G-quadruplex (GQS) or a stem-loop (SL) were synthesized and inserted into AflII and PacI of Firefly 3′UTR in vector inter2. We labeled the GQS constructs as 0A_GQS and 2A_GQS, and the stem-loop constructs as 0A_SL and 2A_SL, with the prefix indicating the number of Adenines. Antisense of miRNA156 target site constructs with the same flanking sequence were also synthesized as the control for each construct.

For the miRNA156 overexpression vector construction in AGO1 *in vivo* binding assay, the *MIR156B* genomic sequence was inserted into AscI and SacI of vector *pMDC32*. Primers are listed in Supplementary Table S1.

Arabidopsis protoplast transformation

Protoplasts from the stable *MIR156* over-expression line were prepared and transformed according to the Tape-Arabidopsis Sandwich method (33). 16 h after transformation, protoplasts were centrifuged at 100 *g* for 2 min. RNA was extracted with Qiagen RNeasy kit and qRT-PCR quantification was performed with Bio-Rad CFX. Primers are listed in Supplementary Table S1.

For the protoplast transformation analysis, we transformed the plasmids into a stable *MIR156* over-expression line, where the miRNA expression level should be formally the same in each cell. We assessed the TAM functionality by inserting the miR156 target site and the TAM sequence into the Firefly gene, then we used qRT-PCR to measure the non-cleaved transcript level. We calculated the protoplast transformation efficiency from the level of Renilla luciferase gene located on the same plasmid with Firefly gene but in different transcription units. The Firefly luciferase gene expression level was normalized by transformation efficiency in our results.

In vivo structure validation experiments

Four-week-old tobacco leaves were co-infiltrated with agrobacterium strains harboring plasmids of 0A_SL, 2A_SL, 0A_GQS or 2A_GQS. Two days after infiltration, the leaves were treated with 150 mM of the SHAPE reagent (NAI). The control group was treated with DMSO. Total RNA was extracted using RNeasy Plant Mini Kit (Qiagen) and then by on-column DNaseI treatment (following the respective manufacturer's protocol). Gene-specific reverse transcription was performed as previously described by Kwok *et al*. (34), with a few modifications. A total of 2 μg of *in vivo* total RNA was resuspended in 10 μl RNase-free water. Primer extension was performed with 2 pmol of DNA gene-specific primers (5′CATGCTTAACGTAATTCAACAGAAATTATATG) by Invitrogen SuperScript III reverse transcriptase. The resulting cDNA pellet was dissolved in RNase-free water and mixed with 1 μl 50 μm Mly1-HBLPCR-5′ssDNA linker modified by a 5′-phosphate and a 3′-3-Carbon spacer group (5′P-AGATCGACTCAGCGTCGTGTAGCTGAGTCGATCTNNNNNN-C3-3′), 10 μl Quick Ligase Reaction Buffer (2×), 1U Quick Ligase (New England Biolab) in a 20 μl system. The ligation was performed at 25°C for 1 h, followed by Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v, sigma) and Chloroform:Isoamyl (24:1, v/v, sigma) purification.

The ligated cDNA samples were dissolved in 10 μl of water and used for the PCR reaction. The PCR reaction contained final concentrations of 0.5 mM VIC-labeled DNA gene-specific primers (the same as that used in the reverse transcription primers except the 5′ end was labeled with Vic), 0.5 mM of linker reverse primer (AGATCGACTCAGCTACACGACGC), 200 mM dNTPs, 1× ThermoPol reaction buffer and 1.25U of NEB Taq DNA polymerase in 25 μl. The solution was then extracted with Phenol:Chloroform:Isoamyl (25:24:1, v/v, sigma) and incubated with Mly1 restriction enzyme, according to the manufacturer's protocol. Finally, the reaction pellets were dried and resuspended in Hi-Di formamide (Applied Biosystems/Life Technologies).

The Ned-labeled gene-specific primer (the same as that used in reverse transcription primer except the 5′ end was labeled with NED) was used to make sequencing ladders using linear DNA and 1 μl 5 mM ddTTP by Klenow DNA Polymerase I (New England Biolab) (35). Then, the reaction pellets were dried, resuspended in Hi-Di formamide (Applied Biosystems/Life Technologies) and run on an Applied Biosystems 3730xl Genetic Analyzer. The resulting data were analyzed using QuSHAPE (36).

AGO1 *in vitro* cleavage assay

HA-tagged $AGO1^{WT}$ was immuno-purified from *Arabidopsis* seedlings (37). The 0A_GQS and 2A_GQS designed RNAs were transcribed *in vitro* with T7 polymerase (NEB, 2040S) as substrates. To perform the slice assay, cleavage buffer (100 mM ATP, 10 mM GTP, 60 mM $MgCl_2$, 0.5 M CPO_4, 1 mg/ml CPK) was added to 20 μl beads in extraction buffer (1:1) bearing freshly purified HA-AGO1 from 3 g seedling on the beads' surface. A total of 50 cps of labeled substrate was added to the reaction and incubated at 25°C. A total of 10μl of the resultant liquid was added to 10 μl 2× RNA loading buffer (95% Formamide, 0.02% sodium dodecyl sulphate, 1 mM ethylenediaminetetraacetic acid (EDTA), 0.02%, Bromothymol Blue, 0.01% Xylene Cyanol), denatured for 5 min at 95°C and loaded into a 1 mm polyacrylamide gel electrophoresis gel (10% acrylamides:bis 19:1, 7M Urea, 1×Tris Borate EDTA). Then the gel was dried and exposed to a phosphor screen for image analysis.

AGO1 *in vivo* binding assay

Four-week-old tobacco leaves were co-infiltrated with agrobacterium strains harboring plasmids of 35S:MIR156B, $35S:HA\text{-}AGO1^{DAH}$ and 0A_GQS or 2A_GQS. Two days after infiltration, the leaves were collected and ground in liquid nitrogen. The protein/RNA

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

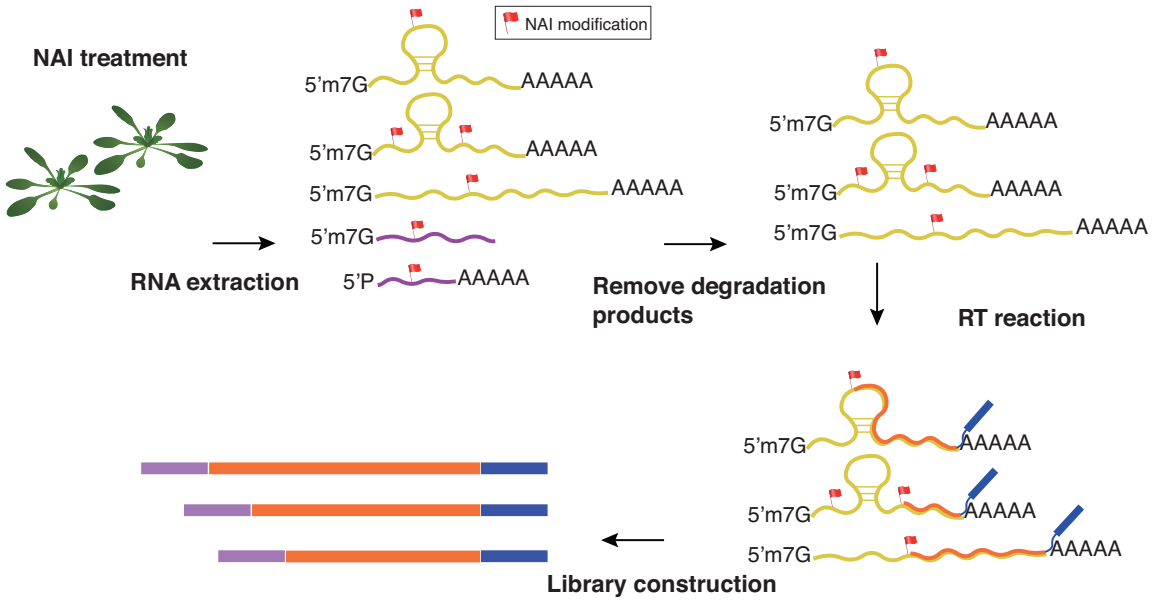

Figure 1. Schematic of CAP-STRUCTURE-seq. (+)SHAPE sequencing library generation showing NAI treatment, nucleotide modification and purification of intact mRNA steps. *Arabidopsis thaliana* etiolated seedlings were treated with NAI. After extraction of total RNA, degraded mRNAs (purple) were removed, leaving intact mRNAs characterized by 5′CAP and 3′ polyA+ (dark yellow). cDNAs (orange) were obtained by reverse transcription (RT) and subjected to an established library construction. RT, reverse transcription.

complexes were extracted using two volumes of IP buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 5% β-mercaptoethanol, 1 mM EDTA, 10% glycerol, 0.1% NP-40, 1 mM PMSF and 1× complete protease inhibitor cocktail). After removing insoluble debris by centrifugation, cell extracts were incubated with anti-HA antibody (Abcam ab9110) for 1h at 4°C with gentle mixing. The anti-HA-decorated extracts were then incubated with pre-washed protein G magnetic beads for 1 h. After incubation, the beads were washed 6 times with the IP buffer. The RNA produced after co-immunoprecipitation was precipitated with ethanol and glycogen, and analysed by RT-PCR. The miRNA156 expression levels were analysed by miRNA RT-PCR (38).

RESULTS

CAP-STRUCTURE-seq accurately probes intact mRNA structures *in vivo*

To investigate how mRNA structure affects miRNA-mediated cleavage, RNA structure models should be captured before cleavage occurs. We therefore developed a novel strategy to obtain the structure of intact mRNAs, named CAP-STRUCTURE-seq (Figure 1). To obtain the RNA structure of intact mRNAs, we performed *in vivo* SHAPE chemical probing (12) on *A. thaliana* with optimized conditions (Figure 1 and Supplementary Figure S2A). Next, we used terminator exonuclease treatment to enrich the intact mRNAs (Supplementary Figure S2B) (20) and then removed the degraded mRNAs by polyA+ purification. We generated two independent biological replicates of both the SHAPE-treated library and the control library (without SHAPE treatment) according to an established protocol (8,39). We labeled the SHAPE-treated libraries as (+)SHAPE and the control libraries as (−)SHAPE. Between 90 and 97% of 340–380 million reads were mapped onto mRNAs (Supplementary Figure S4A and B) with the reproducibility of the CAP-STRUCTURE-seq libraries confirmed by comparing the two biological replicates (Supplementary Figure S5A and B). We checked for a nucleotide bias in the (+)SHAPE and (−)SHAPE libraries and found the nucleotide occurrence to be consistent between the libraries (Supplementary Figure S4C). We then computed the SHAPE reactivity using the (+)SHAPE and (−)SHAPE libraries ('Materials and Methods' section). To validate CAP-STRUCTURE-seq, we compared the SHAPE reactivity of the 18S ribosomal RNA with the corresponding phylogenetic covariance structure (Supplementary Figure S6A) and the 3D structure (Supplementary Figure S6B). We found that the SHAPE reactivity from CAP-STRUCTURE-seq accurately reflects the RNA structure *in vivo*. Finally, we computed the predicted RNA structure using the SHAPE reactivity and found that it outperformed both the *in silico* prediction and the prediction from the DMS-based method, STRUCTURE-seq (8) (Supplementary Table S2).

To broaden out the validation of CAP-STRUCTURE-seq we performed meta-property analyses with over 16,576 transcripts of sufficient RNA structure information (Supplementary Figure S7A). Our CAP-STRUCTURE-seq SHAPE reactivity data for *A. thaliana* exhibits similar genome-wide *in vivo* RNA structural properties to previous results from a DMS-based method (8). For example, the region immediately upstream of the start codon showed particularly high SHAPE reactivity (Supplementary Figure S7B) further supporting the notion that less structured regions near the start codon may facilitate translation (40,41). We found a periodic trend in the reactivity along CDS but

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

this was absent along UTRs (Supplementary Figure S7C), which is consistent with previous studies (8,42). Similar to a RNase-based structure study in human (42), we also observed a unique asymmetric RNA structure signature at the exon–exon junction in *A. thaliana* (Supplementary Figure S7D). Taken together, these conserved RNA structure features suggest that CAP-STRUCTURE-seq successfully provides global RNA structure information in plants.

We then assessed whether CAP-STRUCTURE-seq can overcome the limitations of previous transcriptome-wide RNA structure probing methods. The miRNA-mediated cleavage in the mRNA target site occurs at the tenth nucleotide of miRNA complementary sites (22), which leaves endogenous degraded products. In the previous DMS Structure-seq data, the cleavage site led to reverse transcription stalling, and caused a skewed DMS reactivity profile due to false positive signals (Figure 2A). These degradation signals were excluded in our CAP-STRUCTURE-seq data (Figure 2B), thereby overcoming the limitations of previous methods that include degradation products (Supplementary Figure S1) (8,10,15,43–44). Overall, these data demonstrate that CAP-STRUCTURE-seq can accurately identify *in vivo* structures of intact mRNAs.

Cleavage efficiency robustly measures miRNA-mediated cleavage events

Deciphering the *in vivo* relationship between mRNA structure and miRNA cleavage requires an *in vivo* structure model of target genes before cleavage, and the outcome after miRNA-mediated cleavage. Having developed a method to measure the former we turned our attention to the latter. To estimate the *in vivo* miRNA-mediated CE, we drew inspiration from the definition of enzymatic activity (45). We quantified CE by measuring how many degradation products were generated from one unit of substrate mRNA by one unit of miRNA ('Materials and Methods' section; Supplementary Methods). Our CE calculation is based on two underlying facts (22,46): (i) miRNA-mediated cleavage is the major mRNA turnover pathway for target genes, (ii) the 5′ cleaved products are located within binding sites, which are temporally stable. Therefore, the degradation signal within target sites reflects the cleavage products from miRISC cleavage. We generated degradome libraries to estimate the degradation products (Supplementary Figure S3B and 'Materials and Methods' section) and miRNA-seq libraries to estimate miRNA abundance (Supplementary Figure S3B and 'Materials and Methods' section), with library reproducibility confirmed by comparing the biological replicates (Supplementary Figure S5C and D). We then combined the degradome, (−)SHAPE and miRNA-seq libraries to estimate CE (Supplementary Figure S3B, 'Materials and Methods' section and Supplementary Methods).

We verified the consistency of our CE against previously reported targets (Supplementary Table S3). We expect the CE of targets that have been shown to act via a non-cleavage mechanism or those that do not show evidence of miRNA cleavage to be zero. For example, the CE of *AP2* targeted by miRNA172, which has been shown to act through translational repression rather than mRNA cleavage (47–49), was zero as expected (Supplementary Table S3). *SNZ* (Supplementary Table S3) is another target of miRNA172, which also showed no evidence of miRNA cleavage, consistent with the previous result (50). In contrast, targets that have been shown to undergo miRNA cleavage are expected to have non-zero CE. Indeed, *TOE2*, which is cleaved by miRNA172, had relatively high CE (Supplementary Table S3). Additionally, *TAS1a* and *TAS2*, which must be cleaved by miRNA173 to then serve as templates for trans-acting siRNA (tasi-RNA) (51), had high CE (Supplementary Table S3). These observations were consistent with their previous reported biological functions (47–51). Since sequence complementarity was reported to affect miRNA target cleavage (3,4) we then systematically examined the relationship between sequence complementarity and CE. Globally, we found that sequence complementarity and CE were uncorrelated (Spearman correlation −0.015, Supplementary Figure S3E). In addition, targets with mismatches and/or GU wobble pairs were sometimes more effectively cleaved than targets with perfect complementarity (Supplementary Figure S3E). Our results indicate that other factors besides sequence complementarity between the miRNA and the mRNA may affect CE, with one possible example being mRNA structure. In summary, we can quantitatively measure both the RNA structure of the intact mRNAs and miRNA cleavage *in vivo*.

Target site structure unfolding plays a key role in AGO processing of target RNAs

With CAP-STRUCTURE-seq elucidating the RNA structure, we can begin to answer the elusive question about whether miRNA target sites are structurally accessible *in vivo*. Since our CAP-STRUCTURE-seq directly measured the *in vivo* structural accessibility via SHAPE reactivity (52), we assessed the SHAPE reactivity profiles across the miRNA target sites on the intact mRNAs. SHAPE reactivities of the target sites showed no significant difference from the upstream region (one-sided Mann–Whitney-U test, *P*-value is 0.34, Figure 2B), and were lower than the downstream region (one-sided Mann–Whitney-U test, *P*-value is 0.0035, Figure 2B). These features indicate that under physiological conditions the target sites are not highly accessible, which may provide a protective mechanism for target sites, mitigating against processing by other cellular ribonucleases.

These relatively inaccessible target sites prompted us to ask whether the target site structure affects miRNA cleavage *in vivo*. To address this question, we examined two alternative energetic landscapes associated with the miRISC cleavage process *in vivo*: an enzyme-limiting scenario and a structure-limiting scenario (Figure 3A). In the enzyme-limiting scenario, the energy barrier ($\Delta G^{\ddagger}_{open}$) between the inaccessible and accessible structural states (i.e. the unfolding of the target site) is lower than the barrier for catalytic cleavage (black line in Figure 3A). Thus, the target sites equilibrate quickly between inaccessible and accessible structural states during the binding step prior to the catalytic step of miRNA cleavage. In this scenario, the CE would vary with the free energy required to surmount the AGO catalytic barrier, $\Delta G^{\ddagger}_{cutting}$ ('Materials and Methods'

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

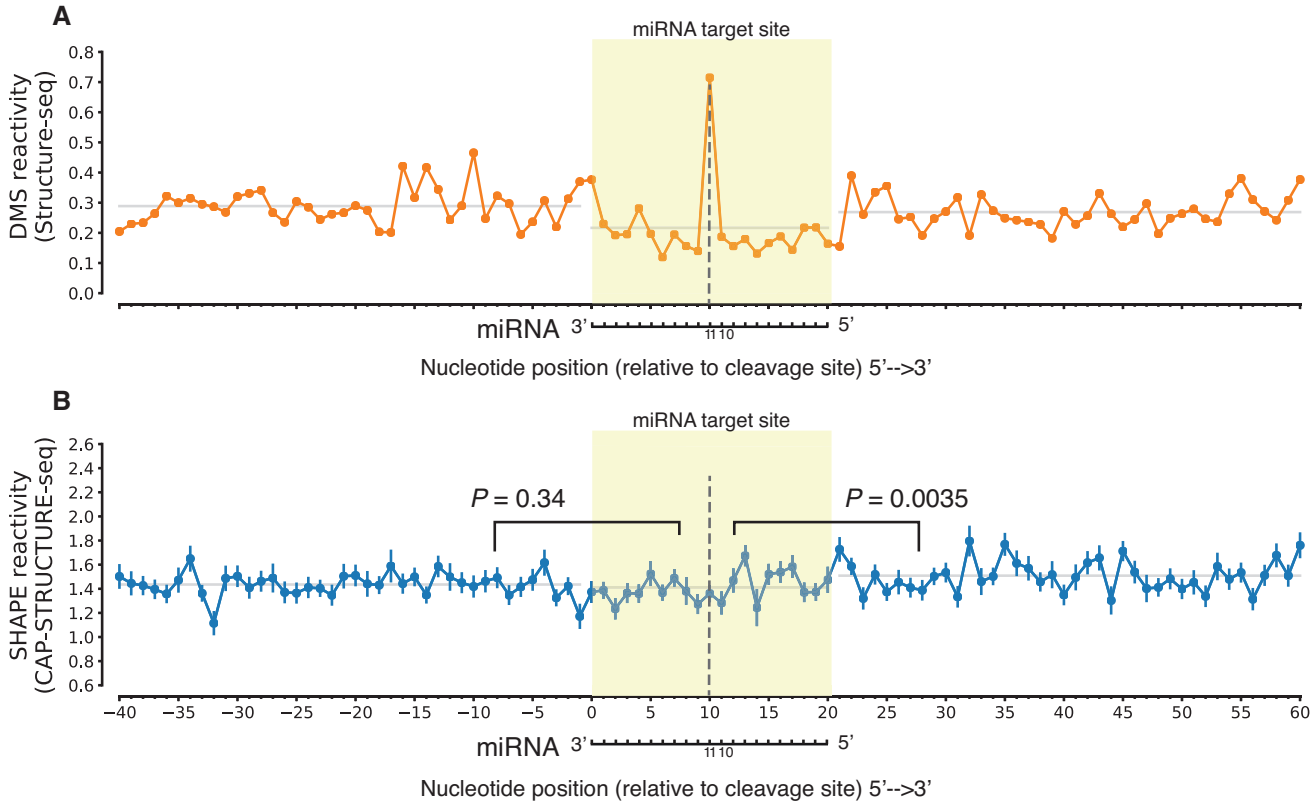

Figure 2. CAP-STRUCTURE-seq can accurately probe intact mRNA structures *in vivo*. (**A**) Reactivity profiles across the miRNA target sites for the DMS reactivity from Structure-seq (8). The miRNA-mediated cleavage in the mRNA target site occurs at the tenth nucleotide of miRNA complementary sites, which leads to a skewed DMS reactivity profile at the eleventh nucleotide of miRNA complementary sites. Mean reactivity is shown. (**B**) SHAPE reactivity profiles from CAP-STRUCTURE-seq. The miRNA cleaved products do not show a peak at the cleavage sites, indicating the elimination of false positive signals. Mean reactivity together with SEM is shown. In A and B the yellow shading indicates the target binding sites. The dotted lines refer to the 11th nucleotide opposite to the miRNA. The horizontal axis is labeled from the 5′ end of the target gene to the 3′ end. *P*-values from one sided Mann–Whitney-U tests.

section), and would be less affected by the RNA structure of the target site. In the structure-limiting scenario, the energy barrier ($\Delta G^{\ddagger}_{open}$) between the inaccessible and accessible structural states is higher than the barrier for cleavage (red line in Figure 3A). Therefore, the target sites cannot achieve equilibrium binding with miRISC before catalytic cleavage. In this scenario, CE would vary with the free energy of opening the target site structure, $\Delta G^{\ddagger}_{open}$, rather than $\Delta G^{\ddagger}_{cutting}$. We used our *in vivo* structures to computationally approximate these two scenarios and explored a range of flanking lengths upstream and downstream of the target site (Figure 3B and C; Supplementary Figure S8A and B). Analysis of our SHAPE reactivity-informed structures revealed that, for most flank sizes, CE anti-correlated with $\Delta G^{\ddagger}_{open}$ with a broad maximum centered around flanks of 50 nt upstream and downstream (Spearman correlation of −0.23, P = 6.3e-9) (Figure 3B and Supplementary Figure S8A). However, for most flank sizes, CE had no correlation with $\Delta G^{\ddagger}_{cutting}$ (Figure 3C and Supplementary Figure S8B), contrary to the reaction kinetics where the energy barrier is anti-correlated with reaction processivity. These results indicate that target site unfolding is likely to be the rate-limiting step that determines miRISC activity *in vivo*. Furthermore, this structure-limiting scenario reveals that the ribonuclease AGO undergoes 'sticky regime' activation (53), where substrate mRNAs associate and dissociate with AGO more slowly than they are being cleaved.

Our analysis favors the structure-limiting scenario rather than the enzyme-limiting scenario, implying that unwinding the miRNA target may be the limiting step in miRISC action, but once the miRNA is bound cleavage occurs quickly. This is consistent with AGO RIP-seq results, where few target transcripts have been captured in wild-type (WT) with less than 1-fold enrichment, while target mRNAs were enriched 7-fold in the catalytic mutant of ($AGO1^{DAH}$) (54). Further, we found that $\Delta G^{\ddagger}_{open}$ anti-correlated with the enrichment ratio of target RNAs from previous $AGO1^{DAH}$-RIP-seq results (54) (Spearman correlation −0.21, P = 0.05). In contrast, both the free energy of binding of the miRNA-target duplex (ΔG_{duplex}) and $\Delta G^{\ddagger}_{cutting}$ show no correlation with the enrichment (Spearman correlation 0.06 with $P = 0.32$ and −0.11 with $P = 0.16$, respectively). These observations suggest that the target sites are not structurally accessible *in vivo*, but rather the unfolding of the target site structure plays a key role in AGO processing of target RNAs.

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

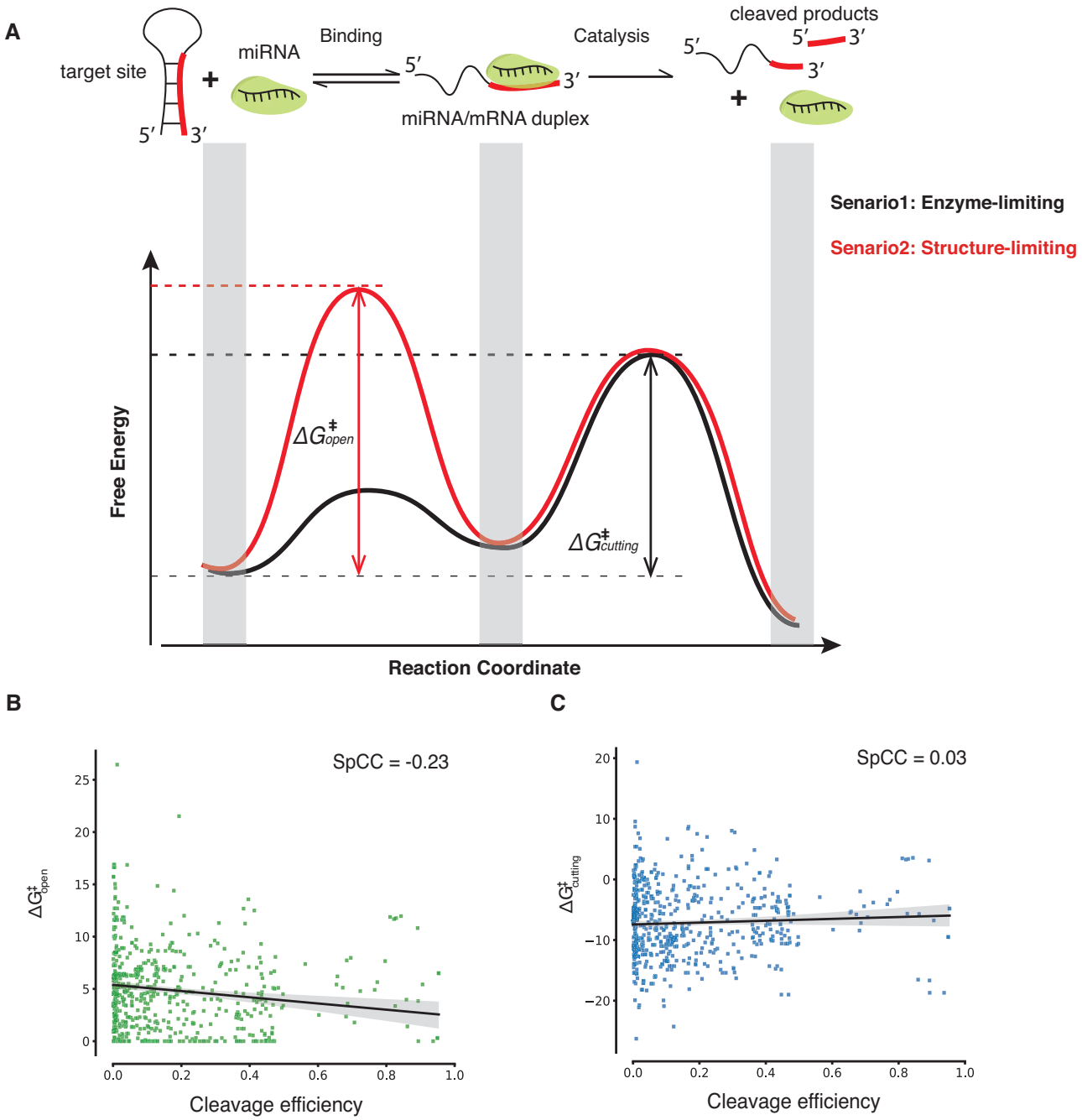

Figure 3. Energetic landscape of the miRISC cleavage process. (**A**) The miRISC cleavage reactions include target binding and cleavage catalysis. Two alternative scenarios demonstrate the energetic landscape of miRNA cleavage (black and red). In the enzyme-limiting scenario (black), target site structure equilibrates quickly between inaccessible (closed) and accessible (open) states in the binding step compared to the catalytic step of miRNA cleavage. In this scenario, the apparent activation energy is $\Delta G^{\ddagger}_{cutting}$, which measures the energy required to raise the initial substrate target RNA to the transition catalysis-compatible state. Alternatively, in the structure-limiting scenario (red), the target site cannot achieve equilibrium binding before cleavage. In this scenario, the energy barrier between the target site and the transient state is higher than the barrier for cleavage. And the apparent activation energy is equal to $\Delta G^{\ddagger}_{open}$, which measures the energy required to open the target site structure. (**B**) Spearman correlation between $\Delta G^{\ddagger}_{open}$ and CE (647 target sites with the upstream and downstream flank lengths of 50 nt, $P = 6.3e\text{-}9$ ***). (**C**) A similar analysis to B, but for $\Delta G^{\ddagger}_{cutting}$ and CE ($P = 0.46$).

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

Target Adjacent nucleotide Motif (TAM) contributes to miRNA-mediated cleavage *in vivo*

Having revealed that the target site structure affects cleavage *in vivo*, we then investigated whether the structure of the target site flanking regions is involved with miRNA cleavage. We assessed the RNA secondary structure by separating the RNA targets into non-cleaved (zero CE) and cleaved (positive CE) groups. We found higher SHAPE reactivity at the +1 and +2 nt immediately downstream of target sites in the cleaved group relative to the non-cleaved group (Figure 4A), suggesting that these two nucleotides are more single-stranded than their neighbors. To confirm this observation, we used the SHAPE reactivity with the ViennaRNA *RNAfold* utility (55) to calculate the BPPs for each nucleotide in the target site and the flank regions. We found that the BPPs of the +1 and +2 nt were significantly lower than their neighboring nucleotides (Figure 4B), indicating an increased likelihood of single-strandedness in the cleaved group compared to the non-cleaved group. Furthermore, the single-strandedness of the two nucleotides was unlikely to be due to sequence composition (Figure 5A) or AT content (Figure 5B) because there was no difference between the non-cleaved and cleaved groups. Our results reveal that a secondary structure feature, specifically single-strandedness of the two nucleotides adjacent to the 3′ end of the miRNA target site, generally exists *in vivo* in intact mRNAs that will undergo cleavage. We named this structure feature Target Adjacent nucleotide Motif (TAM).

TAM promotes miRNA cleavage but not miRNA binding

To explore the functional role of TAM in miRNA cleavage, we devised a structure assay ('Materials and Methods' section) involving the 20 nt miRNA156 target site and a designed stable structure module. We concatenated the target site with either a G-quadruplex structure or a stem-loop structure with the aim of mimicking the base-pairing state of the 2 nt immediately downstream of the target site (Figure 6A and Supplementary Figure S9A). To maintain the single-strandedness of the TAM we inserted two Adenines (AA) between the target site and the designed structure module (Figure 6A and Supplementary Figure S9A). We confirmed the formation of TAM *in vivo* by using capillary electrophoresis (34) to resolve the *in vivo* RNA structure ('Materials and Methods' section, Figure 6B and Supplementary Figure S9B) and then assessed miRNA cleavage *in vivo* by measuring the levels of non-cleaved substrate mRNA. We found that the mRNA level of non-cleaved target genes with TAM was significantly lower (Student's t-test P-value < 0.01) than those without TAM (Figure 6C and Supplementary Figure S9C). This suggests that the presence of TAM leads to higher cleavage than when it is absent.

To further confirm that the presence of TAM promotes target cleavage, we performed an *in vitro* AGO cleavage assay using HA immuno-affinity-purified WT AGO protein. We found that the target RNA was cleaved only when TAM was present (Figure 7A). Our results reveal that TAM is essential for miRISC nuclease activity. TAM in the target mRNA could facilitate AGO binding instead of directly triggering the nuclease activity of AGO proteins. To test the possibility that TAM affects target binding, we conducted an *in vivo* binding assay (Methods) by using the slicing-defective AGO1 mutant, $AGO1^{D762A}$. We found that AGO1 was able to bind the target RNAs with the same binding affinity irrespective of whether the TAM was present or absent (Figure 7B and C). Therefore, our data reveal that TAM promotes miRISC cleavage activity but does not affect target binding.

DISCUSSION

The SHAPE reagent, 2-methylnicotinic acid (NAI) is a well-characterized RNA-structure probing chemical which can detect accessibility of all four RNA nucleotides *in vivo* (19). We applied this chemical as part of our novel CAP-STRUCTURE-seq method that differentiates RNA structure information for intact RNAs from degradation fragments. The intact RNA structurome facilitates the discovery of causal relationships between RNA structure and miRNA-mediated cleavage. In addition, we generated the first *in vivo* RNA structure landscape of Arabidopsis with structure information covering all four nucleotides. The method outperforms the previous DMS-based Structure-seq *in vivo* RNA-structure probing method (8) (Supplementary Table S2) which only captured structural information for A and C nucleotides.

Armed with the RNA structurome we investigated target site accessibility, which has long been interpreted on the basis of spatial accessibility from a geometric viewpoint. Previous RNase-based in vitro RNA-structure studies on the nuclear RNAs (e.g. PIP-seq (56)) show that miRNA targets are less double-stranded in *A. thaliana*. However, in our study, we found that the target sites are not significantly spatially accessible *in vivo* (Figure 2B). Instead, we elucidated a structure-limiting scenario for miRNA cleavage (Figure 3A) from an energetic viewpoint. The differences between *in vitro* and *in vivo* studies may indicate that the cellular environment affects the RNA structure thus highlighting the possibility of a change in function. The spatially inaccessible target sites may provide a protective mechanism which prevents mRNAs from being targeted by other ribonucleases. Since miRISC has no helicase activity to unfold the RNA structure, miRISC has to take advantage of local structural variations, i.e. target site nucleotides becoming single-stranded ('breathing'), to find and bind its target site (Figure 3A). Thus, the equilibrium between a folded and an unfolded target site initially determines the binding rate (Figure 3A). This equilibrium is dependent on the energy required to open the target site ($\Delta G^{\ddagger}_{open}$). Thus, the lower the energy barrier the easier miRISC can bind to the target sites (Figure 8).

In living cells, many factors affect the final miRNA CE, including the miRNA precursor processing (reviewed by (57)), the miRNA methylation (reviewed by (58)), the miRNA exportation (59–61) and the miRNA localization and sequence complementarity (reviewed by (62)). Each factor contributes to the final miRNA CE. In our RNA structure study, we found that the target site unwinding ($\Delta G^{\ddagger}_{open}$) can contribute 23% to the final miRNA-mediated cleavage *in vivo* (Figure 3B). Considering the other contributors *in vivo*, our results suggest that under physiological condition the unfolding of the target site structure

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

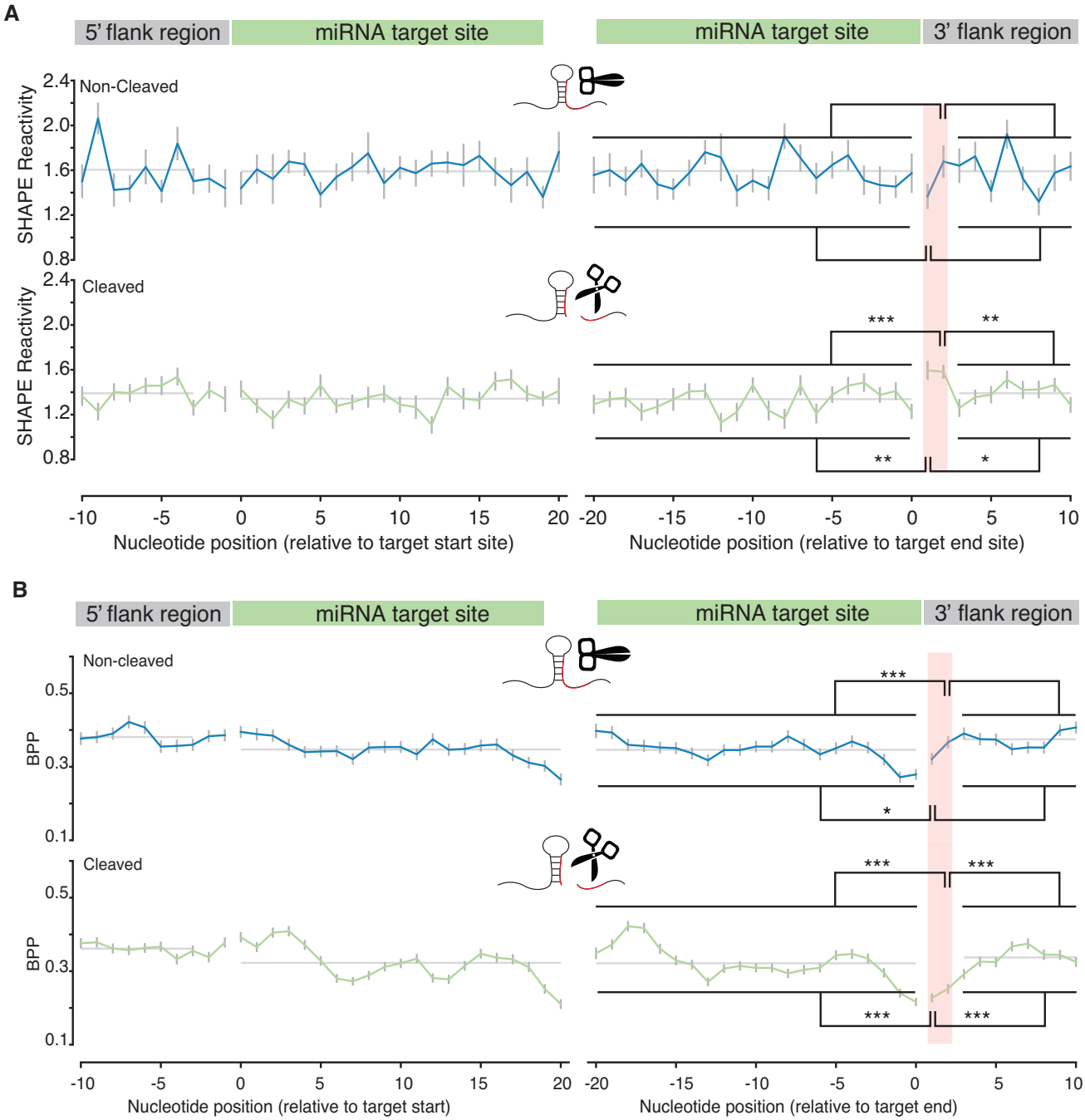

Figure 4. Discovery of the structure motif, TAM. (**A**) SHAPE reactivity profiles for miRNA target sites in the non-cleaved group (387 flanked target sites with reactivity values) and cleaved group (567 flanked target sites with reactivity values). The profiles show the per nucleotide mean ± SEM across transcripts, aligned by target site start (left panels) and end position (right panels). Two nucleotides in the cleaved group (positive CE group), immediately downstream of target sites (TAM region), show significantly higher SHAPE reactivities compared to their neighbors (by one-sided Mann-Whitney-U tests). Compared to the upstream region (target sites), $P = 0.002$** and 3.6e-05*** for first and second nucleotides, respectively; Compared to the downstream region, $P = 0.019$* and 0.0014** for first and second nucleotides, respectively. The two individual nucleotides of the TAM region in the non-cleaved group (zero CE group) are not significantly higher than their neighbors by one-sided Mann–Whitney-U tests. (**B**) BPP at TAM in the non-cleaved group and cleaved group. SHAPE reactivity-directed BPP ('Materials and Methods' section) for miRNA target sites in the non-cleaved group (387 target sites) and cleaved group (567 target sites). Corresponding to A, the two individual nucleotides within the TAM region (pink shading) show significantly lower BPP compared to the target site ($P = 4.2e\text{-}14$ *** for the first nucleotide and $P = 2.3e\text{-}12$ *** for the second nucleotide) and downstream region of TAM ($P = 5.1e\text{-}15$*** for first nucleotide and 1.3e-13*** for second nucleotide) in the cleaved group. Besides, the two individual nucleotides within the TAM region in the non-cleaved group are not significantly lower than downstream region by one-sided Mann–Whitney-U tests, but lower than upstream nucleotides ($P = 0.045$* for first nucleotide and 0.00054*** for second nucleotide). Since the SHAPE reactivity is the direct measurement of single-strandedness, here the subtle inconsistency between SHAPE reactivity and BPP may result from the uncertainty of the nearest neighbor parameter embedded in RNA structure prediction software (70).

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

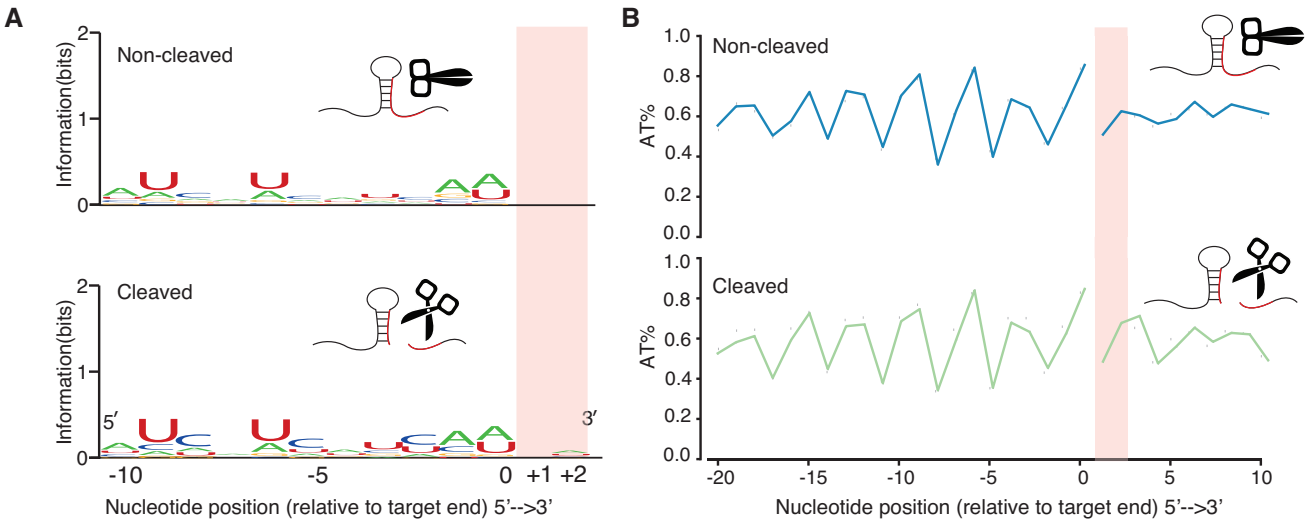

Figure 5. No sequence preference in the TAM region. (**A**) Sequence composition around the target sites for the total cleaved (567 target sites) and non-cleaved groups (387 target sites). (**B**) AT content around the target sites for the total cleaved (567 target sites) and non-cleaved (387 target sites) groups.

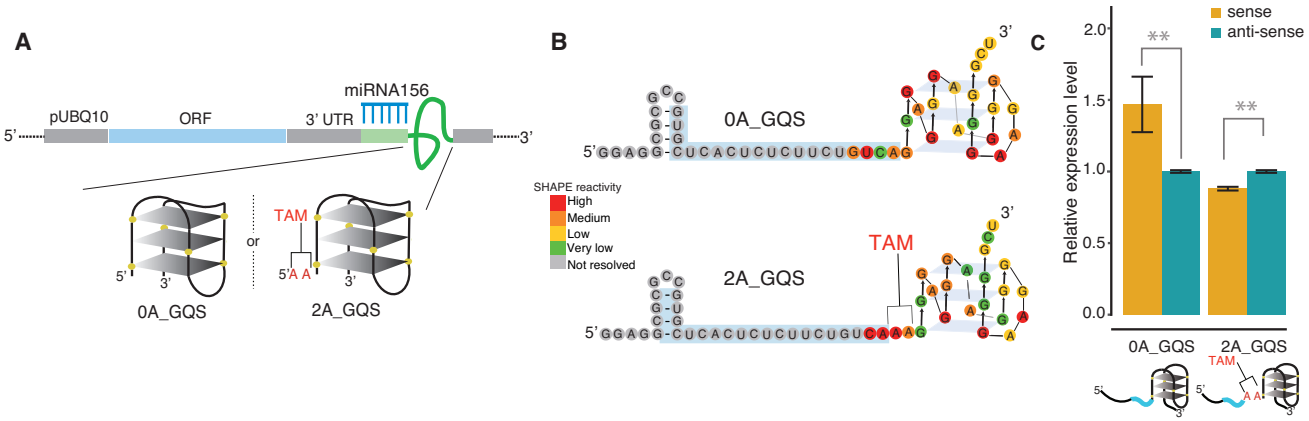

Figure 6. Validation of TAM functionality by a designed structure assay. (**A**) Cartoon representation of the protoplast transformation assay to validate the TAM functionality using a designed structure assay. GQS refers to a G-quadruplex. The miRNA156 target sites (blue comb) followed by 0 or 2 Adenines (As) and ending with a GQS. The prefixes, '0A' and '2A', indicate the number of Adenines. (**B**) *In vivo* RNA structures of 0A_GQS and 2A_GQS. (**C**) The non-cleaved mRNA abundance for the structures in (B) was measured by qRT-PCR (dark yellow bars). *P*-value < 0.01 by Student's *t*-test. The antisense target sites were used as controls (teal bars). Data are mean ± SEM from three independent biological replicates.

plays an important role in AGO processing target RNA. In contrast, another factor, sequence complementarity does not show a high correlation with CE globally (Spearman correlation −0.015, Supplementary Figure S3E), indicating a relatively smaller contribution.

Once miRISC binds to the target sites, it needs to adjust the target conformation to perform the catalytic cleavage activity. A set of high resolution (∼2.2 Å) ternary structures of *Thermus thermophilus* Argonaute (*TtAgo*) complexes has been solved (29), providing structural information about the transition between cleavage-incompatible and cleavage-compatible stages. AGO protein has been found to require conformation changes in three loops, L1, L2 and L3, to facilitate geometrical coordination of two magnesium ions (Mg^{2+}) with the AGO nuclease activation site, the phosphate oxygens and in-line water, in order to facilitate the attack on the cleavable phosphate. The single-stranded TAM may promote this conformation transition and trigger the nuclease activity of AGO. Since the TAM is located at the 3′ end of the target site on the RNA, which is in parallel with 5′ end position of the miRNA and the 5′ end of the miRNA interacts with the MID domain of AGO, we suspect that single-stranded TAM may engage or 'touch' amino acids in the MID domain (Figure 8), thereby reducing the energy of conformation transition and facilitating the nuclease reaction (Figure 8).

The distinct roles of the target site and the TAM region decouples the target binding from target cleavage of miRISC *in vivo* (Figures 6–8 and Supplementary Figure S9). These properties are reminiscent of the CRISPR-CAS system (CAS9 and CAS13a) where both CAS9 and CAS13 decouple their binding and cleavage activity (63–66). In addition, the endonucleolytic domains of CAS13 (HEPN domain), CAS9 (RuvC domain) (67) and RISC (PIWI do-

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

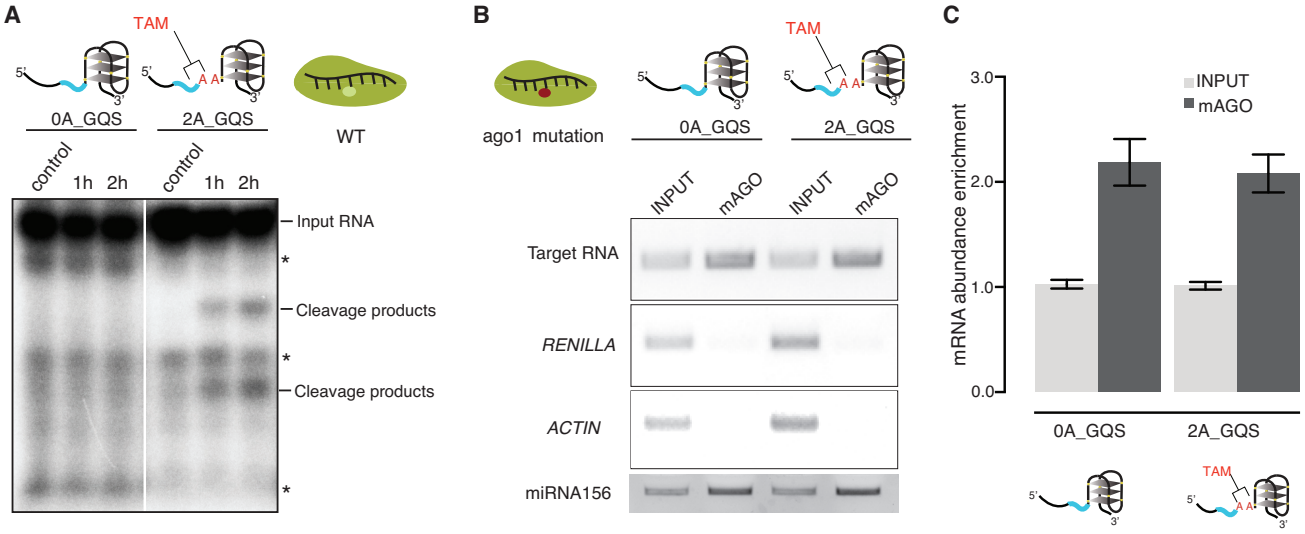

Figure 7. TAM promotes miRNA cleavage but not miRNA binding. (**A**) *In vitro* AGO1 cleavage assay shows that target RNA is cleaved when the TAM is present. The target RNAs were incubated for 1 or 2 h, where two cleavage products were present in the target with TAM on the X-ray film (as indicated). The asterisk indicates the background bands present in both control and experiment groups. The control represents the experiment at time 0. (**B**) *In vivo* AGO1 binding assay shows no difference between target RNAs with TAM and without TAM. *RENILLA* and *ACTIN* were used as the control. miRNA156 levels were measured in all the samples. (**C**) The RNA abundance enrichment in (B) was quantified by amplicon intensities and normalized by input. Data are mean ± SEM from three independent biological replicates.

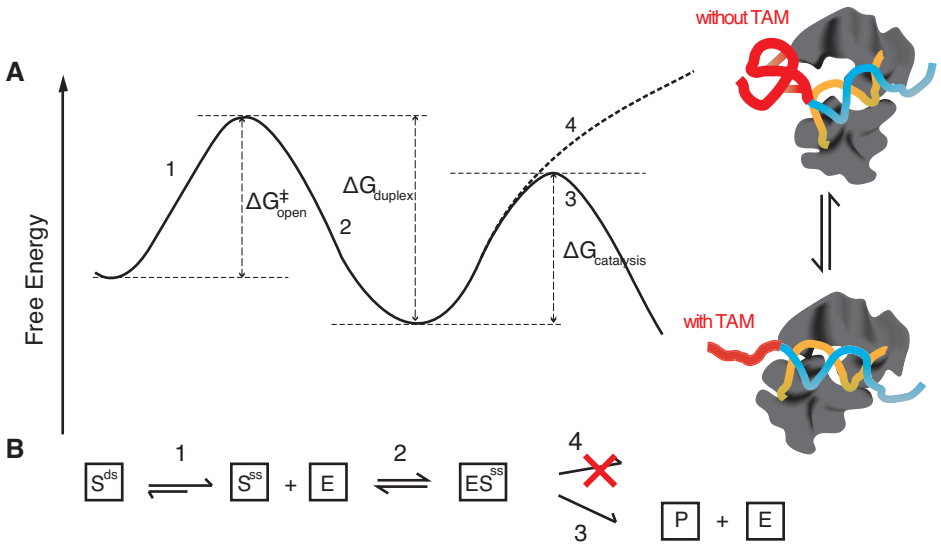

Figure 8. mRNA secondary structure-based model of miRNA-mediated cleavage. (**A**) Energetic landscape of the RISC cleavage process. (1) energy barrier for unwinding the target sites ($\Delta G^{\#}_{open}$); (2) energy released from miRNA-target duplex formation (ΔG_{duplex}); (3) energy barrier for cleavage-compatible conformation with TAM ($\Delta G_{catalysis}$); (4) energy barrier for cleavage-incompatible conformation without TAM. (**B**) Illustration of the biochemical process of miRNA cleavage. (1) miRISC takes advantage of local structural variations, i.e. target site nucleotides 'breathing' between a double-stranded (S^{ds}) and a single-stranded (S^{ss}) state, to find and bind its target site; (2) miRISC (E) binds its target through base pairing (ES^{ss}); (3) the TAM can trigger the AGO into a cleavage compatible conformation, which then cleaves and releases the cleavage products (P); (4) double-stranded di-nucleotide in the TAM region interacts with AGO protein and maintains the AGO in a cleavage-incompatible conformation. The TAM region of the miRNA target gene is colored red.

main) (68), contain an RnaseH-like fold and require Mg^{2+} as co-factor for catalytic activity. Furthermore, TAM can trigger the nuclease activity of miRISC. This mechanism, termed 'substrate-dependent enzyme activation', has also been found for CAS9 (64). This similarity indicates there may be a conserved mechanism between the CAS system and the miRISC system.

Our work indicates that the accessibility of the mRNA target site may be the primary determinant for RISC endonuclease efficacy. And we determined that a motif comprising two unpaired nucleotides immediately downstream of the pre-cleaved mRNA target site is required to direct RISC cleavage. Adaptation of this motif within the *Arabidopsis* genome appears to have selected mRNAs that are

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

readily cleavable, versus sites where miRNAs can bind, but miRISC cleavage does not occur. The presence of the TAM appears to promote RISC cleavage of its target mRNA and such knowledge has the potential to allow adjustment of the cleavability of RISC targets, potentially switching their mode of regulation. This supports the burgeoning hypothesis that RNAs may regulate RNA-binding protein (RBP) function rather than be regulated by RBPs (69). Furthermore, our results indicate that messenger RNA secondary structure may have important physiological functions in many biological processes.

In summary, by deciphering intact mRNA structures *in vivo* through CAP-STRUCTURE-seq, we found that miRNA target sites were not structurally accessible *in vivo* and we demonstrated that the unfolding of the miRNA target site structure predominantly affected miRISC activity *in vivo*. Furthermore, we discovered that the native RNA structure motif, TAM, was sufficient to regulate miRNA cleavage *in vivo*. The TAM mechanism provides evidence of mRNA structure-dependent regulation of biological processes *in vivo*. Our study reveals that *in vivo* mRNA structure serves as an additional regulator of miRISC activity, which could also facilitate the biotechnological engineering of gene silencing, and possibly provide an additional avenue toward crop improvement.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dame Prof. Caroline Dean (John Innes Centre), Prof. Giles Oldroyd (SLCU, Cambridge), Dr Desmond Bradley (John Innes Centre) and the group members in Ding lab for discussions with this work. We thank Dr Saima Shahid and Prof. Michael Axtell for the suggestions and discussions when we initiated our research project as well as providing us the validated miRNA target list. We thank Mr Mirko Ledda (UC Davis) for discussions on data analysis. We thank Jim Carrington's lab for providing us with the AGO construct for performing the binding assay.

Author contributions: M.L.Y. and Y.L.D. conceived the study. M.L.Y., Y.Y.Z. and Y.L.D. designed the experiments with assistance from Y.S., A.C., P.B. and J.W.W. M.L.Y. designed the analyses. M.L.Y, Y.Y.Z., Q.L., M.L.V., X.F.Y. and X.F.F. performed the experiments with assistance from P.B., J.W.W. and Y.L.D. H.C.W., J.C. and M.N. performed the analyses. M.L.Y. wrote the paper with input from all authors. M.L.Y., H.C.W., Y.Y.Z. and X.F.F. contributed equally to this work.

FUNDING

Biotechnology and Biological Sciences Research Council [BB/L025000/1]; the Norwich Research Park Science Links Seed Fund; and European Commission Horizon 2020 European Research Council, Starting Grant [680324]. Funding for open access charge: Biotechnology and Biological Sciences Research Council [BB/L025000/1]; the Norwich Research Park Science Links Seed Fund; and European Commission Horizon 2020 European Research Council, Starting Grant [680324].

Conflict of interest statement. None declared.

REFERENCES

1. Fang,W. and Bartel,D.P. (2015) The menu of features that define primary MicroRNAs and enable de novo design of MicroRNA genes. *Mol. Cell*, **60**, 131–145.
2. Yu,Y., Jia,T. and Chen,X. (2017) The 'how' and 'where' of plant microRNAs. *New Phytol.*, **216**, 1002–1017.
3. Zhang,C., Ng,D.W.K., Lu,J. and Chen,Z.J. (2012) Roles of target site location and sequence complementarity in trans-acting siRNA formation in Arabidopsis. *Plant J.*, **69**, 217–226.
4. Liu,Q., Wang,F. and Axtell,M.J. (2014) Analysis of complementarity requirements for plant MicroRNA targeting using a Nicotiana benthamiana quantitative transient assay. *Plant Cell*, **26**, 741–753.
5. Ameres,S.L., Martinez,J. and Schroeder,R. (2007) Molecular basis for target RNA recognition and cleavage by human RISC. *Cell*, **130**, 101–112.
6. Kertesz,M., Iovino,N., Unnerstall,U., Gaul,U. and Segal,E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
7. Long,D., Lee,R., Williams,P., Chan,C.Y., Ambros,V. and Ding,Y. (2007) Potent effect of target structure on microRNA function. *Nat. Struct. Mol. Biol.*, **14**, 287–294.
8. Ding,Y., Tang,Y., Kwok,C.K., Zhang,Y., Bevilacqua,P.C. and Assmann,S.M. (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, **505**, 696–700.
9. Rouskin,S., Zubradt,M., Washietl,S., Kellis,M. and Weissman,J.S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, **505**, 701–705.
10. Spitale,R.C., Flynn,R.A., Zhang,Q.C., Crisalli,P., Lee,B., Jung,J.W., Kuchelmeister,H.Y., Batista,P.J., Torre,E.A., Kool,E.T. *et al.* (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*, **519**, 486–490.
11. Wells,S.E., Hughes,J.M., Haller Igel,A. and Ares,M. Jr (2000) Use of Dimethyl Sulfate to Probe RNA Structure in vivo. *Methods Enzymol.*, **318**, 479–493.
12. Merino,E.J., Wilkinson,K.A., Coughlan,J.L. and Weeks,K.M. (2005) RNA structure analysis at single nucleotide resolution by Selective 2′-Hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.
13. Flynn,R.A., Zhang,Q.C., Spitale,R.C., Lee,B., Mumbach,M.R. and Chang,H.Y. (2016) Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nat. Protoc.*, **11**, 273–290.
14. Talkish,J., May,G., Lin,Y., Woolford,J.L. and McManus,C.J. (2014) Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA*, **20**, 713–720.
15. Zubradt,M., Gupta,P., Persad,S., Lambowitz,A.M., Weissman,J.S. and Rouskin,S. (2016) DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods*, **14**, 75–82.
16. Siegfried,N.A., Busan,S., Rice,G.M., Nelson,J.A.E. and Weeks,K.M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods*, **11**, 959–965.
17. Souret,F.F., Kastenmayer,J.P. and Green,P.J. (2004) AtXRN4 degrades mRNA in Arabidopsis and its substrates include selected miRNA targets. *Mol. Cell*, **15**, 173–183.
18. German,M.A., Pillay,M., Jeong,D.H., Hetawal,A., Luo,S., Janardhanan,P., Kannan,V., Rymarquis,L.A., Nobuta,K., German,R. *et al.* (2008) Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.*, **26**, 941–946.
19. Spitale,R.C., Crisalli,P., Flynn,R.A., Torre,E.A., Kool,E.T. and Chang,H.Y. (2013) RNA SHAPE analysis in living cells. *Nat. Chem. Biol.*, **9**, 18–20.
20. Pelechano,V., Wei,W. and Steinmetz,L.M. (2016) Genome-wide quantification of 5′-phosphorylated mRNA degradation

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

intermediates for analysis of ribosome dynamics. *Nat. Protoc.*, **11**, 359–376.
21. Deigan,K.E., Li,T.W., Mathews,D.H. and Weeks,K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 97–102.
22. Addo-Quaye,C., Eshoo,T.W., Bartel,D.P. and Axtell,M.J. (2008) Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis Degradome. *Curr. Biol.*, **18**, 758–762.
23. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, 25.
24. Fahlgren,N., Howell,M.D., Kasschau,K.D., Chapman,E.J., Sullivan,C.M., Cumbie,J.S., Givan,S.A., Law,T.F., Grant,S.R., Dangl,J.L. *et al.* (2007) High-throughput sequencing of Arabidopsis microRNAs: Evidence for frequent birth and death of MIRNA genes. *PLoS One*, **2**, e219.
25. Srivastava,P.K., Moturu,T.R., Pandey,P., Baldwin,I.T. and Pandey,S.P. (2014) A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. *BMC Genomics*, **15**, 348.
26. Dimitrov,R. (2014) MicroRNA gene finding and target prediction—basic principles and challenges. *MOJ Proteomics Bioinform.*, **1**, 105–110.
27. Tafer,H., Höner zu Siederdissen,C., Stadler,P.F., Bernhart,S.H., Hofacker,I.L., Lorenz,R. and Flamm,C. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 348–362.
28. Wang,Y., Juranek,S., Li,H., Sheng,G., Tuschl,T. and Patel,D.J. (2008) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature*, **456**, 921–926.
29. Sheng,G., Zhao,H., Wang,J., Rao,Y., Tian,W., Swarts,D.C., van der Oost,J., Patel,D.J. and Wang,Y. (2014) Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 652–657.
30. Schirle,N.T., MacRae,I.J., Liu,J., Meister,G., Behm-Ansmant,I., Braun,J.E., Huntzinger,E., Fauser,M., Izaurralde,E., Fabian,M.R. *et al.* (2012) The crystal structure of human Argonaute2. *Science*, **336**, 1037–1040.
31. Nakanishi,K., Weinberg,D.E., Bartel,D.P. and Patel,D.J. (2012) Structure of yeast Argonaute with guide RNA. *Nature*, **486**, 368–374.
32. Rosta,E., Nowotny,M., Yang,W. and Hummer,G. (2011) Catalytic mechanism of RNA backbone cleavage by ribonuclease H from quantum mechanics/molecular mechanics simulations. *J. Am. Chem. Soc.*, **133**, 8934–8941.
33. Wu,F.H., Shen,S.C., Lee,L.Y., Lee,S.H., Chan,M.T. and Lin,C.S. (2009) Tape-arabidopsis sandwich—a simpler arabidopsis protoplast isolation method. *Plant Methods*, **5**, 16.
34. Kwok,C.K., Ding,Y., Tang,Y., Assmann,S.M. and Bevilacqua,P.C. (2013) Determination of in vivo RNA structure in low-abundance transcripts. *Nat. Commun.*, **4**, 2971.
35. McGraw,R.A. (1984) Dideoxy DNA sequencing with end-labeled oligonucleotide primers. *Anal. Biochem.*, **143**, 298–303.
36. Karabiber,F., McGinnis,J.L., Favorov,O.V. and Weeks,K.M. (2013) QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA*, **19**, 63–73.
37. Qi,Y. and Mi,S. (2010) Purification of Arabidopsis Argonaute complexes and associated small RNAs. In: Walker,J.M. (ed). *Methods in Molecular Biology*. Humana Press, pp. 243–254.
38. Varkonyi-Gasic,E., Wu,R., Wood,M., Walton,E.F. and Hellens,R.P. (2007) Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods*, **3**, 12.
39. Ding,Y., Kwok,C.K., Tang,Y., Bevilacqua,P.C. and Assmann,S.M. (2015) Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. *Nat. Protoc.*, **10**, 1050–1066.
40. Studer,S.M. and Joseph,S. (2006) Unfolding of mRNA secondary structure by the bacterial translation initiation complex. *Mol. Cell*, **22**, 105–115.
41. Burkhardt,D.H., Rouskin,S., Zhang,Y., Li,G.W., Weissman,J.S. and Gross,C.A. (2017) Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *Elife*, **6**, e22037.
42. Wan,Y., Qu,K., Zhang,Q.C., Flynn,R.A., Manor,O., Ouyang,Z., Zhang,J., Spitale,R.C., Snyder,M.P., Segal,E. *et al.* (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature*, **505**, 706–709.
43. Talkish,J., May,G., Lin,Y., Woolford,J.L. and McManus,C.J. (2014) Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA*, **20**, 713–720.
44. Smola,M.J. and Weeks,K.M. (2018) In-cell RNA structure probing with SHAPE-MaP. *Nat. Protoc.*, **13**, 1181–1195.
45. NC-IUB (1979) Units of enzyme activity. *Eur. J. Biochem.*, **97**, 319–320.
46. Jackowiak,P., Nowacka,M., Strozycki,P.M. and Figlerowicz,M. (2011) RNA degradome-its biogenesis and functions. *Nucleic Acids Res.*, **39**, 7361–7370.
47. Aukerman,M.J. and Sakai,H. (2003) Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2 -Like target genes. *Plant Cell*, **15**, 2730–2741.
48. Li,S., Liu,L., Zhuang,X., Yu,Y., Liu,X., Cui,X., Ji,L., Pan,Z., Cao,X., Mo,B. *et al.* (2013) MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in arabidopsis. *Cell*, **153**, 562–574.
49. Chen,X. (2004) A MicroRNA as a translational repressor of APETALA2 in Arabidopsis flower development. *Science (80-.).*, **303**, 2022–2025.
50. Schwab,R., Palatnik,J.F., Riester,M., Schommer,C., Schmid,M. and Weigel,D. (2005) Specific effects of microRNAs on the plant transcriptome. *Dev. Cell*, **8**, 517–527.
51. Yoshikawa,M., Peragine,A., Park,M.Y. and Poethig,R.S. (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes Dev.*, **19**, 2164–2175.
52. McGinnis,J.L., Dunkle,J.A., Cate,J.H.D. and Weeks,K.M. (2012) The mechanisms of RNA SHAPE chemistry. *J. Am. Chem. Soc.*, **134**, 6617–6624.
53. Bisaria,N., Jarmoskaite,I. and Herschlag,D. (2017) Lessons from enzyme kinetics reveal specificity principles for RNA-guided nucleases in RNA interference and CRISPR-based genome editing. *Cell Syst.*, **4**, 21–29.
54. Carbonell,A., Fahlgren,N., Garcia-Ruiz,H., Gilbert,K.B., Montgomery,T.A., Nguyen,T., Cuperus,J.T. and Carrington,J.C. (2012) Functional analysis of three arabidopsis ARGONAUTES using slicer-defective mutants. *Plant Cell*, **24**, 3613–3629.
55. Lorenz,R., Hofacker,I.L. and Stadler,P.F. (2016) RNA folding with hard and soft constraints. *Algorithms Mol. Biol.*, **11**, 8.
56. Li,F., Zheng,Q., Vandivier,L.E., Willmann,M.R., Chen,Y. and Gregory,B.D. (2012) Regulatory impact of RNA secondary structure across the arabidopsis transcriptome. *Plant Cell*, **24**, 4346–4359.
57. Dolata,J., Taube,M., Bajczyk,M., Jarmolowski,A., Szweykowska-Kulinska,Z. and Bielewicz,D. (2018) Regulation of plant microprocessor function in shaping microRNA landscape. *Front. Plant Sci.*, **9**, 753.
58. Ji,L. and Chen,X. (2012) Regulation of small RNA stability: methylation and beyond. *Cell Res.*, **22**, 624–636.
59. Muqbil,I., Bao,B., Abou-Samra,A., Mohammad,R. and Azmi,A. (2013) Nuclear export mediated regulation of MicroRNAs: potential target for drug intervention. *Curr. Drug Targets*, **14**, 1094–1100.
60. Li,S., Le,B., Ma,X., Li,S., You,C., Yu,Y., Zhang,B., Liu,L., Gao,L., Shi,T. *et al.* (2016) Biogenesis of phased siRNAS on membrane-bound polysomes in arabidopsis. *Elife*, **5**, e22750.
61. Li,S., Liu,L., Zhuang,X., Yu,Y., Liu,X., Cui,X., Ji,L., Pan,Z., Cao,X., Mo,B. *et al.* (2013) MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in arabidopsis. *Cell*, **153**, 562–574.
62. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
63. Sternberg,S.H., Redding,S., Jinek,M., Greene,E.C. and Doudna,J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, **507**, 62–67.
64. O'Connell,M.R., Oakes,B.L., Sternberg,S.H., East-Seletsky,A., Kaplan,M. and Doudna,J.A. (2014) Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature*, **516**, 263–266.
65. Tambe,A., East-Seletsky,A., Knott,G.J., Doudna,J.A. and O'Connell,M.R. (2018) RNA binding and HEPN-Nuclease

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020

activation are decoupled in CRISPR-Cas13a. *Cell Rep.*, **24**, 1025–1036.
66. Dagdas,Y.S., Chen,J.S., Sternberg,S.H., Doudna,J.A. and Yildiz,A. (2017) A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.*, **3**, eaao0027.
67. Chen,J.S. and Doudna,J.A. (2017) The chemistry of Cas9 and its CRISPR colleagues. *Nat. Rev. Chem.*, **1**, 78.
68. Meister,G. (2013) Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.*, **14**, 447–459.
69. Hentze,M.W., Castello,A., Schwarzl,T. and Preiss,T. (2018) A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.*, **19**, 327–341.
70. Zuber,J., Joseph Cabral,B., McFadyen,I., Mauger,D.M. and Mathews,D.H. (2018) Analysis of RNA nearest neighbor parameters reveals interdependencies and quantifies the uncertainty in RNA secondary structure prediction. *RNA*, **24**, 1568–1582.

Downloaded from https://academic.oup.com/nar/article/48/15/8767/5870339 by 81263607 user on 07 September 2020