Supplementary Information for

Unique Functionality of 22 nt miRNAs in Triggering RDR6-Dependent siRNA Biogenesis from Target Transcripts

By Josh T. Cuperus, Alberto Carbonell, Noah Fahlgren, Hernan Garcia-Ruiz, Russell T. Burke, Atsushi Takeda, Christopher M. Sullivan, Sunny D. Gilbert, Taiowa A. Montgomery and James C. Carrington

SUPPLEMENTAL METHODS

Small RNA datasets. Small RNA data sets used in this paper are available from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo). Data for *Arabidopsis* wild type Col-0 and *dcl234* were previously described[1] and are available from GEO (GSE20197). Samples used in this work were GSM506656, GSM506657, GSM506658, GSM506662, GSM506663 and GSM506664 for Col-0; and GSM506680, GSM506681, GSM506686, GSM506687 and GSM506688 for *dcl234*. Data from rice are available from GEO series GSE11014 (GSM278532, GSM278533, GSM278534, GSM278535, GSM278571 and GSM278572)[2], GSE16350 (GSM409313, GSM409314, GSM409315, GSM409316, GSM409317, GSM409318, GSM409319, GSM409320, GSM409321, GSM409322, GSM409323 and GSM40924)[3], GSE16248 (GSM407071 and GSM407072)[3], GSE12317 (GSM309691, GSM309692 and GSM309693)[4], GSE19602 (GSM489087)[5], and GSE13152 (GSM329296, GSM329297, GSM329298 and GSM329299)[6].

***MIRNA* secondary and tertiary structure prediction**. *MIRNA* secondary structures were generated using MC-Fold, a knowledge-based folding algorithm that utilizes data from the Protein Data Bank (PDB) to take into account both canonical Watson-Crick and non-canonical RNA base pairs commonly found in solved RNA structures[7].

MIRNA tertiary structures were predicted using MC-Sym[7]. MC-Sym structures were constrained using the MC-Fold structure. Initially, 1000 tertiary models are returned. These initial models were sorted based on all-atoms force-fields and the ten models with lowest free energy were kept. Because MC-Sym constructs tertiary models using fragments from the PDB, the phosphate backbone has stretches of discontinuity. To correct for discontinuity, the models were first relieved until the atomic force's root-mean-square (RMS) was < 100 Kcal/mol/Å, then refined until the RMS was < 5 Kcal/mol/Å[7]. Tertiary structures were viewed using open-source PyMOL v1.2[8].

SUPPLEMENTARY REFERENCES

1. Garcia-Ruiz, H. et al. Arabidopsis RNA-dependent RNA polymerases and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during Turnip Mosaic Virus infection. *Plant Cell* 22, 481-96 (2010).

2. Zhu, Q.H. et al. A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. *Genome Res* 18, 1456-65 (2008).

3. Johnson, C. et al. Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. *Genome Res* 19, 1429-40 (2009).

4. Zhou, X., Sunkar, R., Jin, H., Zhu, J.K. & Zhang, W. Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in Oryza sativa. *Genome Res* 19, 70-8 (2009).

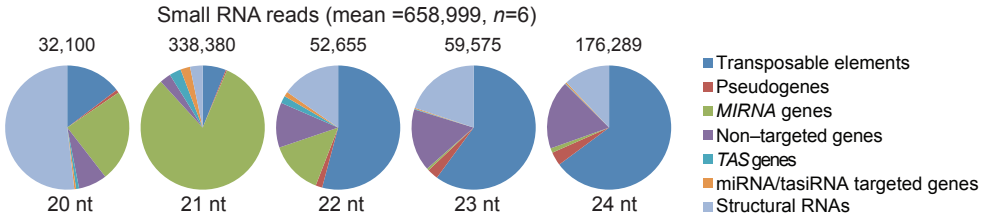
5. He, G. et al. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* 22, 17-33 (2010).

6. Heisel, S.E. et al. Characterization of unique small RNA populations from rice grain. *PLoS One* 3, e2871 (2008).

7. Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 452, 51-5 (2008).

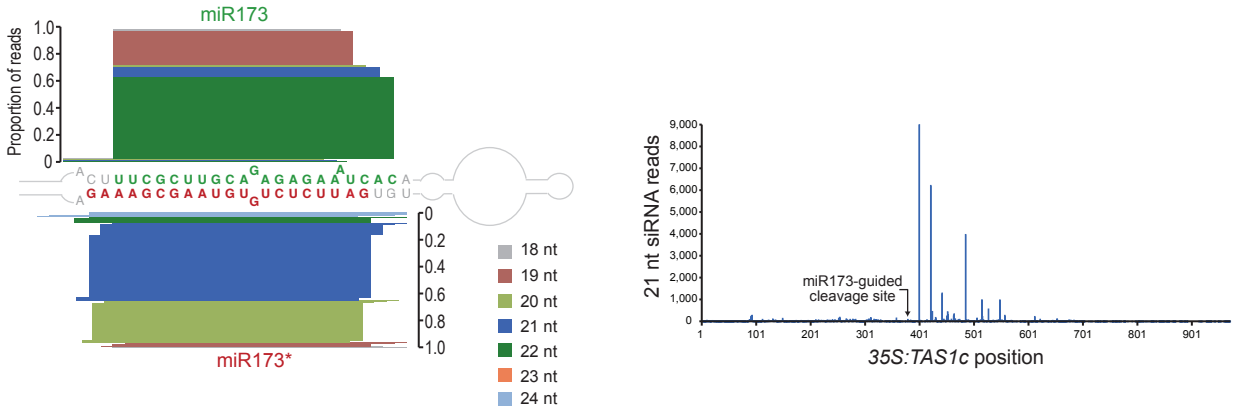
8. DeLano, W.L. The PyMOL molecular graphics system. 1.2 edn (DeLano Scientific LLC, Palo Alto, CA USA, 2008).

9. Rajagopalan, R. et al. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev.* 20, 3407-3425 (2006).

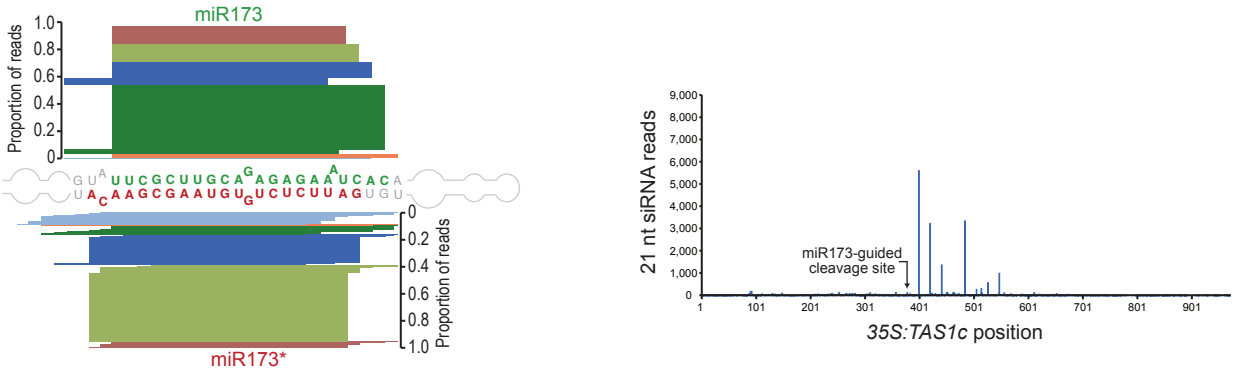


Supplementary Figure 1 Distribution of variously sized small RNA to the indicated features across the Arabidopsis genome. Data are from the means from six libraries.

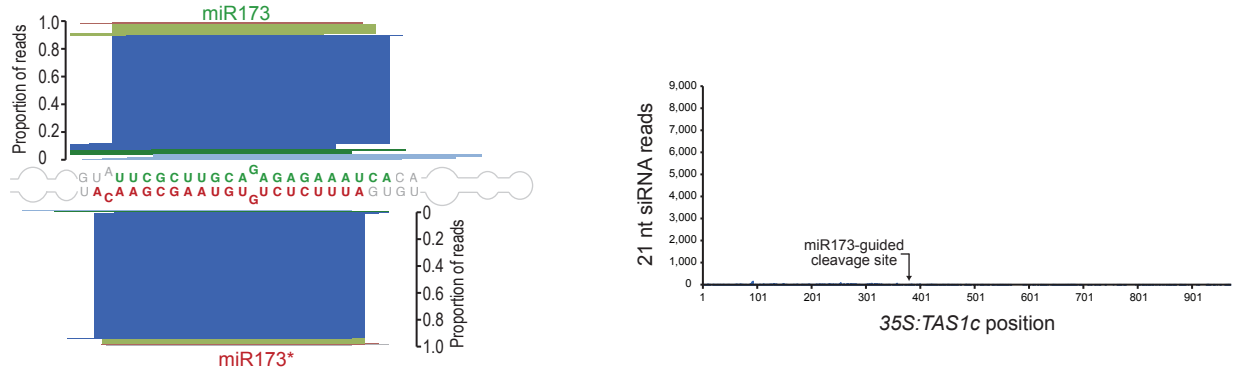
a *35S:MIR173 & 35S:TAS1c*-derived reads



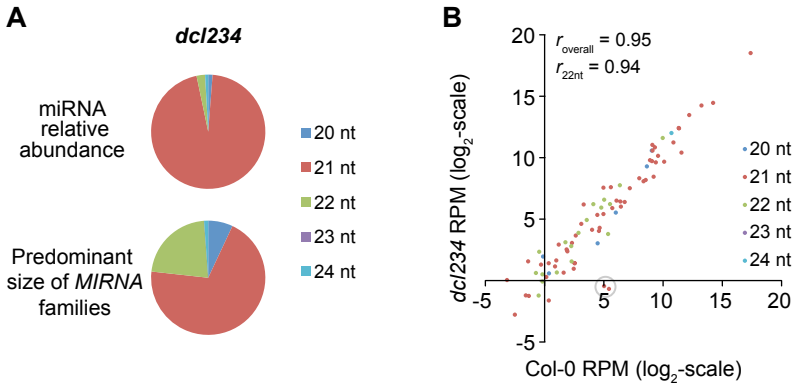
b *35S:amiR173 & 35S:TAS1c*-derived reads



c *35S:amiR173-21 & 35S:TAS1c*-derived reads

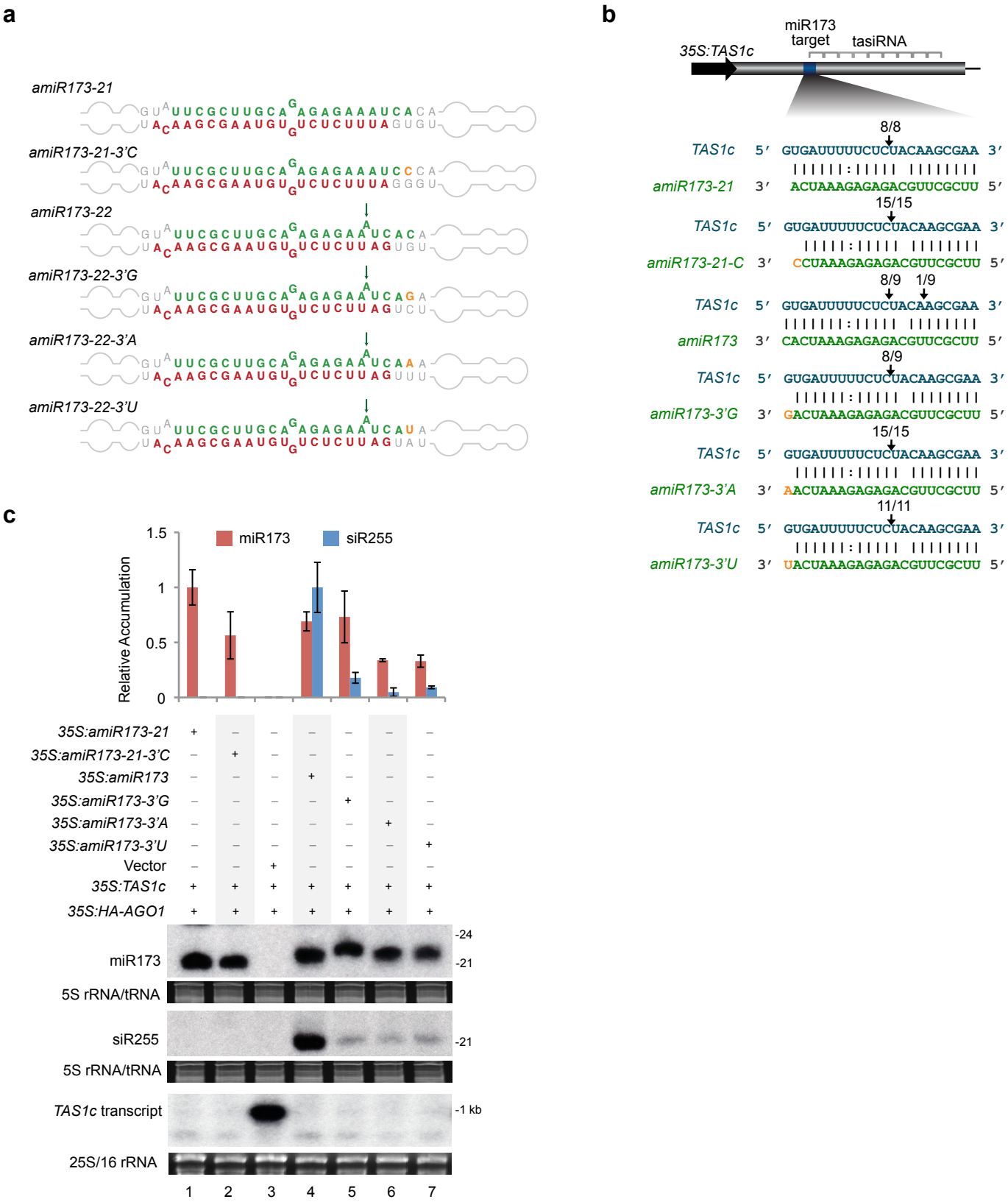


Supplementary Figure 2 Mapping of miR173 or miR173-related reads, and *TAS1c* transcript-derived reads, from the transient expression assay data shown in Fig. 3c. (**a,b,c**) Analysis of miR173, miR173* and *TAS1c* reads after transient co-expression of *35S:MIR173a*, *35S:amiR173* or *35S:amiR173-21* with *35S:TAS1c*. Proportions of reads are plotted as stacked bars based on size (color coded), 5' position and 3' position, with end positions aligned to the respective sequences shown in the foldbacks. miR173 and related sequences are plotted upwards, and miR173* sequences are plotted downwards. Twenty-one-nucleotide *TAS1c* siRNA reads are plotted in the bar graphs along the transcript length, with the miR173-guided cleavage site indicated. The bar graphs show consolidated "sense" and "antisense" reads, in which the siRNAs from the two strands were normalized by 2 nts to account for the known offset in duplexes resulting from DCL4-mediated cleavage.

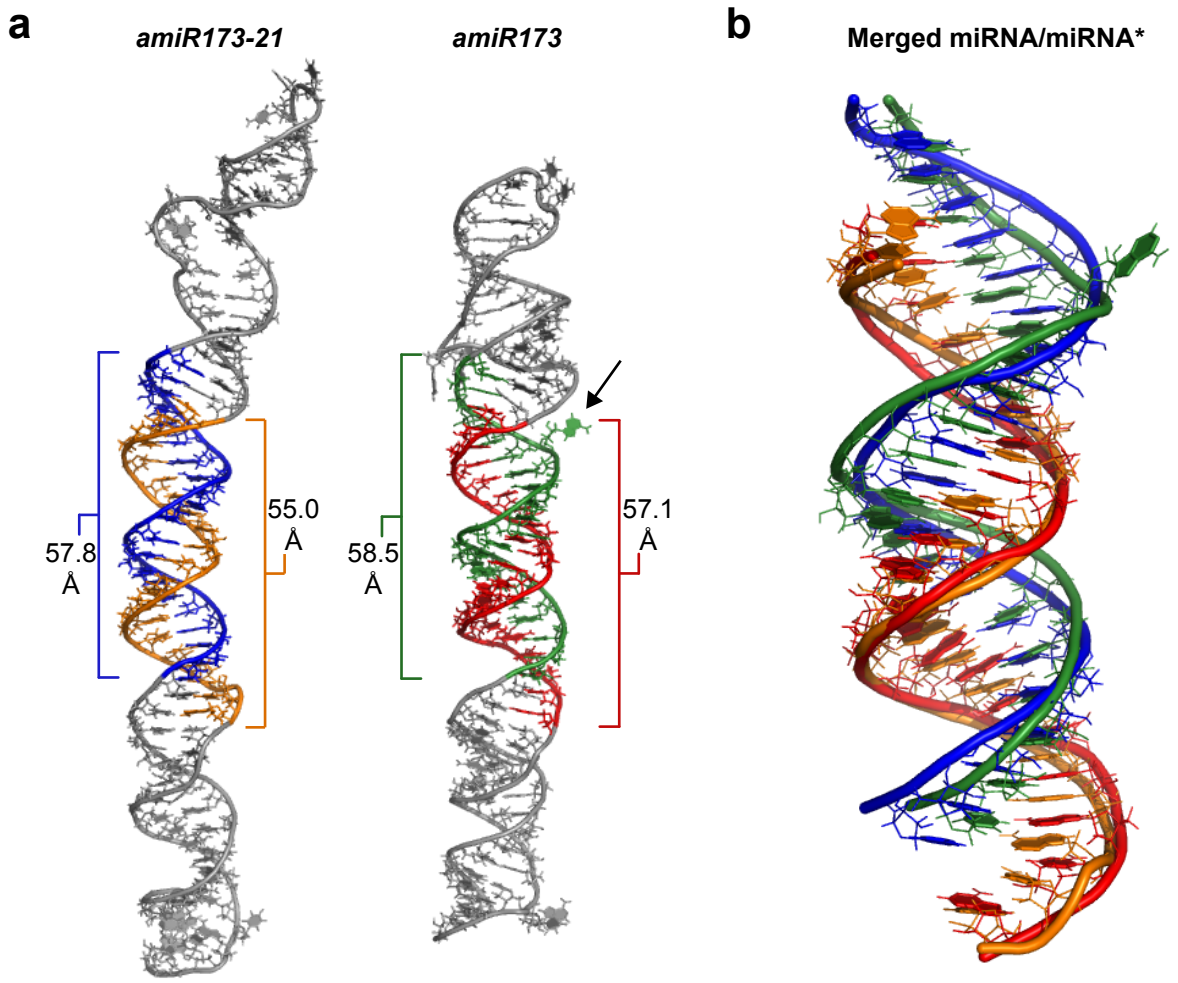


Supplementary Figure 3 miRNA accumulation in the Arabidopsis *dcl234* triple mutant. (**a**) Mean proportions of distinct miRNA size classes in read datasets (top), and of predominant size class for *MIRNA* families (bottom), from Arabidopsis *dcl234* triple mutants. (**b**) Scatterplot comparing average miRNA reads in wild-type Arabidopsis (Col-0) and *dcl234* triple mutants (six and five replicates, respectively). Reads were normalized per million (RPM). The Pearson product-moment correlation coefficients (r) for the overall and 22 nt sets of miRNA families are shown. DCL4-dependent miR822 and miR839 are indicated (grey circle)[9].

Supplementary Figure 4 Enrichment or depletion of miRNA in HA-AGO1 coimmunoprecipitates. Mature miRNA from total RNA (input fraction) and immunoprecipitated HA-AGO1 complexes (IP fraction) were identified by high-throughput sequencing (two replicates each). Enrichment or depletion of miRNA in the IP fraction was determined by plotting the average ratio of miRNA reads/million (RPM) in the IP samples divided by the miRNA reads in the input sample. Only miRNA with an average of at least 5 RPM in either the IP or input samples were plotted. The predominant size class for most groups was plotted. For miR397a and miR167a,b,d, two size classes are plotted, as 22 nt variants represented more than 30% of the 21 and 22 nt reads for these families (percentages of 21 and 22 nt reads are indicated as in **Fig. 2b**).



Supplementary Figure 5 Role of the 3' end nucleotide in miR173. (**a**) Foldbacks yielding 21 nt or 22 nt miR173 variants with distinct 3' end nucleotides. Mutagenized positions are shown in orange. (**b**) 5'RACE assays showing proportion of products mapping to the canonical cleavage site for each miR173 variant tested. (**c**) Accumulation of miR173 variants and *TAS1c* tasiRNA (siR255) in *N. benthamiana* co-expression assays with the constructs indicated above the blot panels. Mean (n=3) relative miR173 variant (red) and siR255 (blue) levels +/- SD (lane 1 and lane 4 = 1.0 for miR173 variant and siR255, respectively). Based on RNA migration, amiR173-22-3'G may be 23 nts in size rather than 22 nts.



Supplementary Figure 6 Predicted tertiary structures of *MIR390a*-derived *amiR173-21* and *amiR173* foldbacks using MC-Fold and MC-Sym. (**a**) Structures of foldbacks were predicted using MC-Fold and MC-Sym[7]. The *amiR173-21* guide and star strands are highlighted blue and orange, respectively. The *amiR173* guide and star strands are highlighted green and red, respectively. The position of the nucleotide of the asymmetric bulge is indicated with an arrow. The lengths (in angstroms) of the guide and star strands are shown. (**b**) Overlay of miRNA/miRNA* duplexes of *amiR173-21* and *amiR173* show similarity in calculated duplex lengths.