Chemistry & Biology, Vol. 11, 1619–1623, December, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.chembiol.2004.09.010

Architecture of a Validated MicroRNA::Target Interaction

Brief Communication

Monica C. Vella, Kristy Reinert, and Frank J. Slack*
Department of Molecular, Cellular, and Developmental Biology
Yale University
P.O. Box 208103
New Haven, Connecticut 06520

Summary

MicroRNAs are small ~22 nucleotide regulators of numerous biological processes and bind target gene messenger RNAs to control gene expression. The *C. elegans* microRNA *let-7* and its target *lin-41* were the first microRNA::target interaction to be validated in vivo. *let-7* molecules form imperfect duplexes with two required *let-7* complementary sites in the *lin-41* 3′ UTR. Here, we show that base pairing at both the 5′ and 3′ ends of the *let-7* binding site, as well as the presence of unpaired RNA residues in the predicted duplexes, are required for *lin-41* downregulation. In this study, our model for microRNA::target interactions also demonstrates that the context of a microRNA binding can be critical for function, revealing an unforeseen complexity in microRNA::target interactions.

Introduction

MicroRNAs (miRNAs) are genomically encoded, untranslated RNA molecules of approximately 20–25 nucleotides (nt) found in animals and plants [1–3]. MiRNAs control many biological processes, including development [1, 4], and act by binding to complementary sequences in the 3′ untranslated regions (UTRs) of the mRNA of their target genes. One of the founding members of this class of RNAs is *C. elegans let-7*, conserved across animal phylogeny [3, 5]. To identify animal miRNA targets and to understand the mechanism by which miRNAs interact with their target genes, an imperative first task entails describing features in the complementary sequences in the mRNA of the target genes needed for proper function of the miRNA. Few miRNA target genes have been validated, and thus little is currently known about the sequence requirements for miRNA::target binding. However, various attempts have been made to predict miRNA targets using bioinformatics approaches that rely heavily on base pairing and free energy (ΔG) calculations [6, 7–9]. The "miRNA seed" (positions 2–8; 5′-3′ of the miRNA sequence, complementary to the 3′ end of the target mRNA) has been defined by bioinformatics [7] and through analysis of artificial miRNA::mRNA target interactions [10] as the most important region for pairing in the miRNA::mRNA duplex, and these studies also suggest that the 3′ end of the miRNA has a minor role, if any, for function [10].

*Correspondence: frank.slack@yale.edu

The *let-7* microRNA and its target, *lin-41*, comprised the first miRNA::target pair to be verified in vivo [3, 11, 12] and remains the best understood natural miRNA::target interaction. In this report, we use *let-7* and *lin-41* as a model for understanding miRNA::target interactions. Previously, we have shown *let-7* to bind to complementary sites in the *lin-41* 3′ UTR and defined the minimal sequence required for proper regulation of *lin-41* in vivo [12]. The minimal sequence consists of three elements clustered in the 3′ UTR of *lin-41*: two *let-7* complementary sites (LCSs) and a 27 nucleotide intervening sequence (abbreviated LCS1-27-LCS2) (Figure 1A). Fusing LCS1-27-LCS2 downstream of a *lacZ* reporter gene caused downregulation of the reporter at the adult stage of *C. elegans* development ([12], see also Figure 1), coincident with *let-7* upregulation [3, 13].

Both LCS1 and LCS2 are 100% conserved in a related nematode species, *C. briggsae* [12] (which diverged from *C. elegans* 100 million years ago [14]), suggesting that LCS1 and LCS2 in their entirety are essential for function. Our previous data indicated that point mutations in the 3′ region of LCS1 and LCS2 compromised *lacZ* reporter gene downregulation [12]. In this work, we demonstrate that residues outside of the 3′ region of the complementary site are critical for LCS function in the *lin-41* 3′ UTR and that the context of miRNA complementary sites is also important.

Results and Discussion

The 3′ Region of an LCS Is Not All that Is Required for Function In Vivo

To test if regions in the LCSs other than just the 3′ region were important for function in the *lin-41* 3′ UTR, we mutated residues in the middle and 5′ regions of the LCSs. First, we disrupted a C:G base pair toward the 5′ region of LCS1 and LCS2 (pMV26) corresponding to the 3′ region of the *let-7* RNA. Disrupting this base pairing was predicted to reduce the stability of the RNA::RNA duplexes, and we found that it prevented downregulation of the reporter gene (Figure 1, compare pMV26 to pMV9 [wild-type LCS1-27-LCS2]). Second, certain predicted looped residues in the middle of LCS1 and LCS2 were forced to pair with the corresponding *let-7* sequence (one bulged U remained to prevent RNA interference [RNAi], which may result from a completely complementary RNA::RNA duplex) (pMV27). These unpaired, looped residues were also needed for proper LCS function (Figure 1), even though they reduced the relative predicted stability of the duplexes. Thus, unpaired RNA residues in the predicted *let-7::lin-41* duplexes, as well as base pairs with the 3′ region of the miRNA, are required for LCS function in vivo.

To examine the role of the very 3′ end of the *let-7* miRNA, the predicted 3′ UU *let-7* overhanging residues were forced to pair to LCS1 and LCS2 to eliminate the overhangs (pMV28). The G:U wobble base pair at position 20 of *let-7* was also changed to G:C in LCS1 and LCS2. Our results demonstrate that the 3′ UU *let-7*

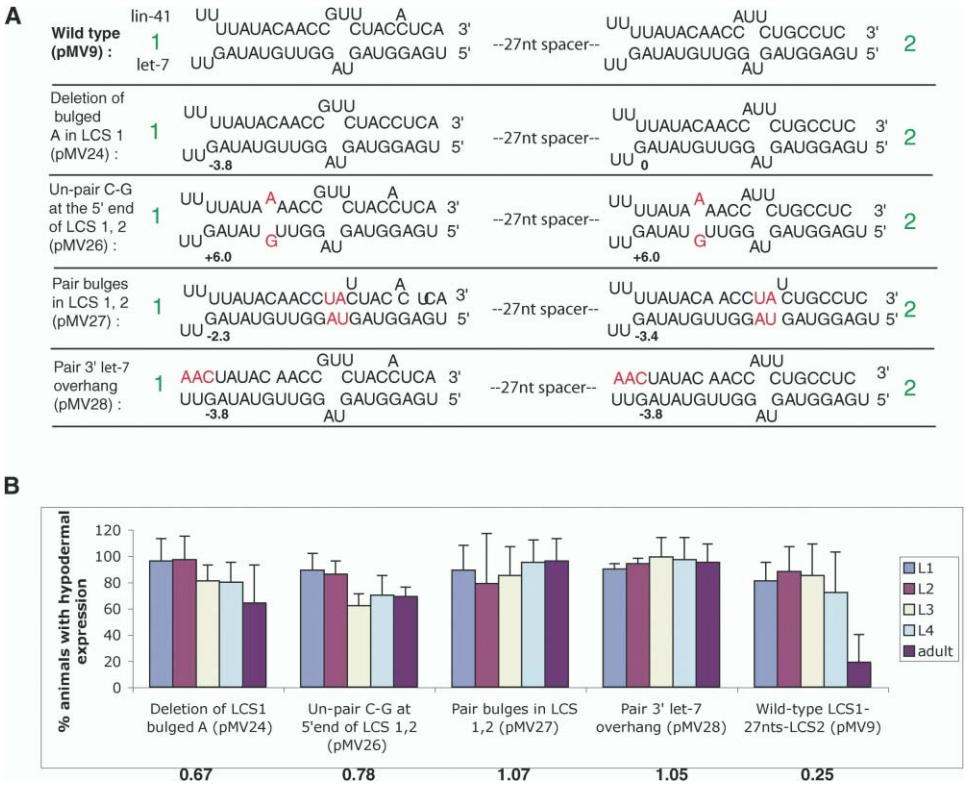

Figure 1. Base Pairs with the 3′ Region of an miRNA and Looped Residues in the Complementary Sites Are Required for Downregulation of Reporter Gene Expression

(A) Shown are the predicted duplexes between the *let-7* complementary sites (LCSs) (top) and the *let-7* miRNA (bottom). Below each duplex are predicted change in free energy (ΔG) values from wild-type (LCS1 = –22.0 kcal/mol; LCS2 = –23.5 kcal/mol) [18]. pMV9 contains the wild-type minimal sequence required for proper downregulation of *lin-41* in vivo. Mutations are shown in red and indicated on the left. Numbers 1 and 2 in green indicate LCS1 and LCS2, respectively.

(B) Transgenic animals carrying these constructs were assayed for β-galactosidase activity in vivo at indicated developmental stages. Shown are the averages of multiple lines for each construct. Error bars indicate the standard deviation between lines and/or experiments. Numbers below each construct refer to the ratio of adult animals with hypodermal reporter gene expression compared to the first larval stage (L1). See Experimental Procedures for number of lines generated and number of animals scored.

overhangs or the G:U wobble base pair in LCS1 and LCS2 are critical for proper downregulation of *lin-41*, even though they also reduce the relative predicted stability of the duplexes. It has been shown that the PAZ protein domain found in Argonaute and Dicer proteins (involved in RNAi and miRNA processing) can bind a small interfering RNA with a 3′ UU overhang [15], leading to the possibility that PAZ domain proteins may also bind the unpaired 3′ UU of *let-7* to regulate *lin-41* expression. The result presented here may assist in defining this interaction.

The Context of an LCS Is Important for Function In Vivo

We examined whether the context of the LCS was important for function. Constructs containing either two LCS1 (pMV20) or two LCS2 (pMV21) sites, respectively separated by the 27 nt spacer, did not downregulate *lacZ* at the adult stage (Figure 2), suggesting that function might be dependent on relative order of the sites or on slight structural differences between them, i.e., the predicted bulged A and GUU loop in LCS1 compared to the AUU loop in LCS2. When the order of the LCSs was reversed (pMV25), however, downregulation was still observed (Figure 2), reflecting the relative unimportance of LCS order. By contrast, deletion of the unpaired A in LCS1 (Figure 1) in LCS1-27-LCS2 failed to cause proper downregulation of the reporter, demonstrating that the unpaired A is important for *let-7* activity. Since LCS2 does not contain the bulged A, this explained why LCS2-27-LCS2 is likely not functional. While a predicted unpaired residue in the *lin-4* miRNA has been shown to be important for the *lin-4::lin-14* interaction [2, 16], our work shows that in our case an unpaired residue in a miRNA target site is essential for downregulation. Remarkably, deletion of the bulged A in LCS1 compromises function even though this deletion is predicted to increase the relative stability of the miRNA::mRNA duplex compared to the wild-type interaction. The bulged A may be a binding site for a particular protein specific to the *let-7::lin-41* interaction that is necessary for proper *lin-41* function at the correct developmental stage, or this less stable duplex may be required for function in some unknown way, e.g., for unwinding the duplex.

Since LCS1-27-LCS1 also fails to downregulate the

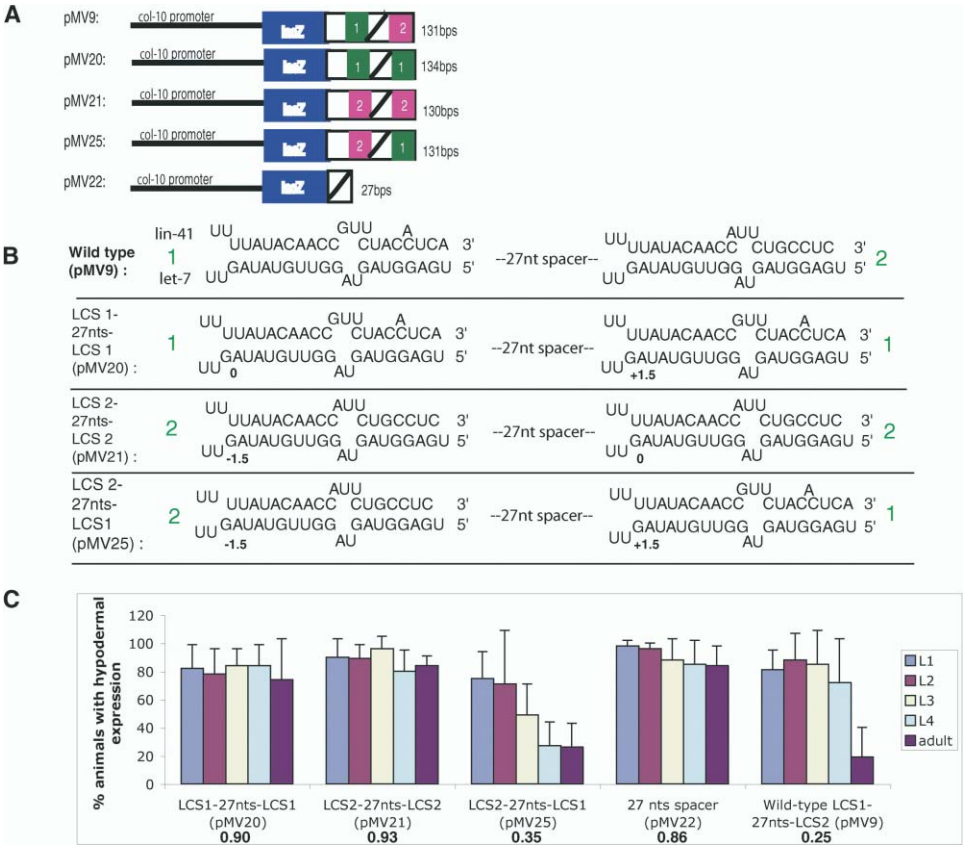

Figure 2. The Context of miRNA Binding Sites in the *lin-41* 3′ UTR Is Critical for Function

(A) Schematic of constructs. Box 1 (green) and box 2 (pink) refer to LCS1 and LCS2, respectively. The hatched white box between LCS1 and LCS2 indicates the 27 nt spacer.

(B) Shown are the predicted duplexes between the *let-7* complementary sites (LCSs) (top) and the *let-7* miRNA (bottom) for constructs in (A). Below each duplex are predicted change in free energy (ΔG) values from wild-type (LCS1 = −22.0 kcal/mol; LCS2 = −23.5 kcal/mol) [18]. pMV9 contains the wild-type minimal sequence required for proper downregulation of *lin-41* in vivo.

(C) Transgenic animals carrying these constructs were assayed for β-galactosidase activity in vivo at indicated developmental stages (refer to Figure 1B for analysis).

reporter gene, LCS1 and LCS2 are different enough to be distinguished in vivo. Our demonstration that LCS1 can function only in the context of LCS2 (and vice versa) reveals that the context of an LCS is also important, uncovering an unforeseen complexity in an miRNA::target interaction.

The 27 nt spacer between LCS1 and LCS2 reveals another layer of complexity in understanding the sequence requirements for *let-7::lin-41* interaction and function. While the 27 nt spacer itself does not downregulate the reporter at the adult stage (pMV22, Figure 2), it is specifically needed in the context of LCS1 and LCS2 to downregulate *lacZ* [12]. We have shown previously that both the length and specific sequence of the 27 nt spacer in the minimal sequence (LCS1-27-LCS2) was necessary for proper *lin-41* downregulation activity [12]. Shortening the length of the 27 nt spacer by replacing it with a 4 nt or 12 nt linker resulted in a failure to properly downregulate the reporter gene [12]. Altering the specific sequence of the 27 nt spacer without changing the length also resulted in a failure to downregulate a reporter gene in vivo [12]. Thus, the LCS1-27nt-LCS2 architecture, including 5′ and 3′ base pairs, the bulged and looped residues, 5′ overhangs, and the 27 nt spacer, is important for the *let-7::lin-41* interaction. This architecture may provide potential specificity that allows only the correct factors to bind to the *lin-41* mRNA. For example, this arrangement may be needed to prevent other members of the *let-7* miRNA family [17], which are similar in sequence, from binding to the LCSs in the *lin-41* 3′ UTR and downregulating *lin-41*. We speculate that the 27 nt spacer may bind additional factors, either RNA and/or protein, which are needed for *lin-41* downregulation in vivo. Future experiments to identify these factors will be critical to understanding the mechanism of *let-7* regulation of *lin-41*.

Our experimental results using a verified miRNA::target interaction demonstrate that specific features of an miRNA binding site as a whole as well as its context are important for function in vivo. It appears that sequences outside the “seed” area as well as unpaired residues are important for functional interactions. *let-7* is used extensively as a model miRNA, and we predict that our general findings will be relevant to other miRNA::target interactions. In particular, given the complexities we encountered in this one specific, well-studied case,

we imagine that other miRNA::target interactions will be equally complex. This work might thus dictate a revision of the general rules used to predict miRNA binding sites. miRNAs potentially regulate thousands of human genes, including disease genes, making these predictions very important [1, 7], especially if miRNAs prove to be effective small molecule therapeutics to specifically downregulate these disease genes.

Significance

MicroRNAs (miRNAs) are an abundant class of small, noncoding RNAs present in many different organisms, with roles in a variety of different biological processes. However, few RNA target genes are known (despite the fact that many have been predicted), and very little is known about the target sequence requirements that would allow accurate predictions to be made. A region toward the 3′ end of the complementary site predicted to base pair to the "miRNA seed" [7] was deemed important based on studies using artificial miRNA::target combinations, and is supported by bioinformatics analyses, which unfortunately have not provided structural information. In this report, we present evidence, using a specific validated miRNA:: target interaction as a model, that target sequences outside the miRNA seed are also required for function in vivo. In this study, we used *let-7* and *lin-41*, an miRNA::target interaction we had previously validated by demonstrating rescue of activity by compensatory mutations between *let-7* and *lin-41* in vivo [12]. Our data also demonstrate that the context of the LCSs in the *lin-41* 3′ UTR with respect to each other is critical and that unpaired residues in the duplex are important. These results demonstrate an unanticipated but striking complexity in an miRNA::target interaction that may dramatically affect the current bioinformatics approaches used to find targets. We anticipate that our findings will also be helpful in understanding other miRNA::target interactions and the miRNA mechanism.

Experimental Procedures

Prediction of *let-7::lin-41* 3′ UTR duplexes and free energy (ΔG) calculations were performed using *mfold*, an RNA folding prediction program [18] entering a single oligonucleotide containing the LCS of interest, a 4nt linker (UUUU), and the *let-7* miRNA sequence.

For all plasmids with designation "pMV" the following applies: Over-lapping oligonucleotide fragments with *Sac*II and *Nco*I overhangs, were annealed and ligated into the *unc-54* 3′UTR of the B29 vector cut with *Sac*II and *Nco*I [3]. B29 contains the *col-10* promoter (directing expression to the hypodermis) fused to *E. coli lacZ*. All plasmid sequences were verified by the Keck Facility, Yale University Medical School; oligonucleotide sequences are available upon request.

Cloning

pMV9, see [12]; for pMV20, the 134 bp region containing LCS1-27-LCS1 was created by annealing primers OMV68 and OMV69. OMV68 and 69 were PAGE purified by M.C.V. For constructs pMV21, pMV22, pMV25, pMV26, pMV27, and pMV28, primers were purchased 5′ phosphorylated and PAGE purified from IDT Technologies (www.idtdna.com). For pMV21, the 130 bp region containing LCS2-27-LCS2 was created by annealing primers OMV70b and OMV71b. pMV24, 130 bp region in which the bulged A in LCS 1 was deleted using pMV9 as template and primers OMV76 (forward) and OMV77 (reverse) with the GeneTailor Site-Directed Mutagenesis System, Invitrogen; pMV25, 131 bp region containing LCS2-27-LCS1 was created by annealing primers OMV78 and OMV79; pMV26, 131 bp region containing LCS1-27-LCS2 was created by annealing primers OMV80 and OMV81; pMV27, 131 bp region containing LCS1-27-LCS2 was created by annealing primers OMV82 and OMV83; pMV28, 131 bp region containing LCS1-27-LCS2 was created by annealing primers OMV84 and OMV85.

All resulting plasmid sequences were verified by sequencing using primer OMV3 (specific to the *unc-54* 3′ UTR) and injected at a final concentration of 5 ng/ul with 80 ng/ul pRF4(*rol-6*) injection marker into wild-type animals to generate extrachromosomal array lines. β-galactosidase activity was assayed as previously described [12]. We analyzed multiple lines for all constructs and averaged the data for each construct. For each stage (L1-adult) of each construct assayed, >50 animals were scored. pMV9, four experiments assayed (two lines, two times each line); pMV20, six lines assayed; pMV21, four lines; pMV22, six lines; pMV24, six lines; pMV25, six lines; pMV26, six lines; pMV27, nine lines; pMV28, ten lines.

Acknowledgments

We would like to thank the Slack Lab for their comments and suggestions and R. Breaker for critical reading of the manuscript. This work was supported by a grant from NIH (GM062594) to F.J.S.

Received: July 21, 2004
Revised: September 14, 2004
Accepted: September 24, 2004
Published: December 17, 2004

References

1. Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281–297.
2. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell *75*, 843–854.
3. Reinhart, B., Slack, F., Basson, M., Pasquinelli, A., Bettinger, J., Rougvie, A., Horvitz, H.R., and Ruvkun, G. (2000). The 21 nucleotide let-7 RNA regulates C. elegans developmental timing. Nature *403*, 901–906.
4. Lin, S.Y., Johnson, S.M., Abraham, M., Vella, M.C., Pasquinelli, A., Gamberi, C., Gottlieb, E., and Slack, F.J. (2003). The *C. elegans* hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. Dev. Cell *4*, 639–650.
5. Pasquinelli, A., Reinhart, B., Slack, F., Maller, B., and Ruvkun, G. (2000). Conservation across animal phylogeny of the sequence and temporal regulation of the 21 nucleotide *C. elegans let-7* heterochronic regulatory RNA. Nature *408*, 86–89.
6. Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in *Drosophila*. Genome Biol. *5*, R1.
7. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787–798.
8. Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003). Identification of Drosophila microRNA targets. PLoS Biol. *1*, e60. 10.1371/journal.pbio.0000060.
9. Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., and Hatzigeorgiou, A. (2004). A combined computational-experimental approach predicts human microRNA targets. Genes Dev. *18*, 1165–1178.
10. Doench, J.G., and Sharp, P.A. (2004). Specificity of microRNA target selection in translational repression. Genes Dev. *18*, 504–511.
11. Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., and Ruvkun, G. (2000). The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the *lin-29* transcription factor. Mol. Cell *5*, 659–669.
12. Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K., and Slack, F.J.

(2004). The *C. elegans* microRNA *let-7* binds to imperfect *let-7* complementary sites from the *lin-41* 3′UTR. Genes Dev. *18*, 132–137.

13. Johnson, S.M, Lin, S-Y., and Slack, F.J (2003). The time of appearance of the *C. elegans let-7* microRNA is transcriptionally controlled utilizing a temporal regulatory element in its promoter. Dev. Biol. *259*, 364–379.
14. Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al. (2003). The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol. *1*, e45. 10.1371/journal.pbio.0000045.
15. Ma, J.B., Ye, K., and Patel, D.J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature *429*, 318–322.
16. Ha, I., Wightman, B., and Ruvkun, G. (1996). A bulged *lin-4/lin-14* RNA duplex is sufficient for *Caenorhabditis elegans lin-14* temporal gradient formation. Genes Dev. *10*, 3041–3050.
17. Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003). The microRNAs of Caenorhabditis elegans. Genes Dev. *17*, 991–1008.
18. Zuker, M., Mathews, D.H., and Turner, D.H. (1999). Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In RNA Biochemistry and Biotechnology, J. Barciszewski and B.F.C. Clark, eds. (Emeryville, CA: Kluwer Academic Publishers), pp. 11–43.