ARTICLE

Published online: 1 April 2007; | doi:10.1038/nsmb1226

Potent effect of target structure on microRNA function

Dang Long, Rosalind Lee, Peter Williams, Chi Yu Chan, Victor Ambros & Ye Ding

Supplementary Fig. 1 (PDF 68K)
Analysis of alternative initiation energy values.

Supplementary Table 1 (PDF 84K)
Open blocks of nucleotides in *lin-41* UTR constructs.

Supplementary Table 2 (PDF 56K)
Analysis of published microRNA-target interactions.

Supplementary Table 3 (PDF 40K)
Comparison of folding programs.

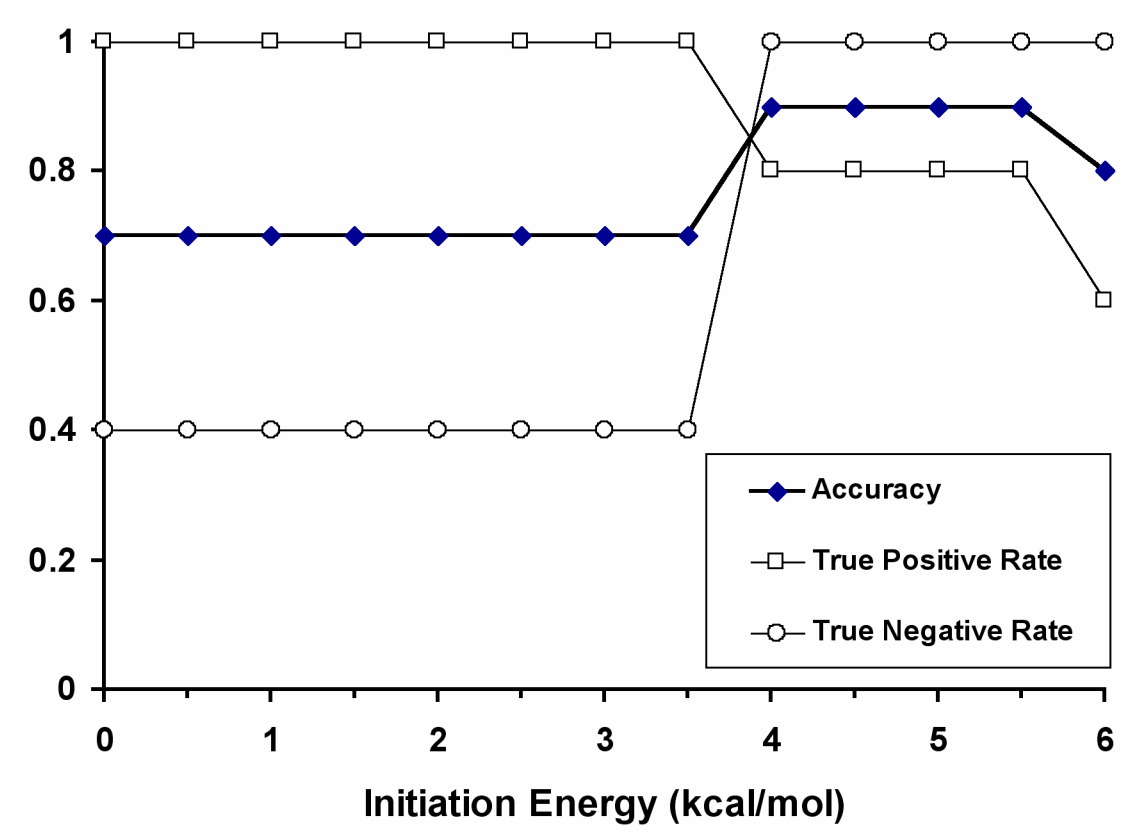
Supplementary Table 4 (PDF 32K)
UTR sequences of *lac-Z* reporter constructs.

Supplementary Table 5 (PDF 28K)
Spacer sequences.

Supplementary Data (PDF 116K)

Supplementary Methods (PDF 148K)

Top



Supplementary Figure 1: Performance of prediction for the 10 *lin-41* constructs in Vella et al (2004) by different values of the initiation energy (from 0.0 kcal/mol to 6.0 kcal/mol, in 0.5 kcal/mol increments). True positive rate (sensitivity) is the ratio of the number of true positive predictions to the total number of positive cases from *in–vivo* testing. True negative rate (specificity) is the ratio of the number of true negative predictions to the total number of negative cases by *in–vivo* testing. The overall accuracy of the prediction is the ratio of the total number of true predictions (positive or negative) to the total number of predictions.

References

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.

Supplementary Table 1. The number of open nucleotides or open blocks in the *let-7* complementary sites of experimentally-tested *lin-41* 3' UTR mutant constructs.

Mutant construct	Repression sensitivity [a]	Number of open nucleotides	Number of open 2-nt blocks	Number of open 3-nt blocks	Number of open 4-nt blocks
pMV1	+ +	20 (site 5); 8 (site 3); 12 (site 1); 14 (site 2)	18 (site 5); 4 (site 3); 9 (site 1); 10 (site 2)	16 (site 5); 2 (site 3); 7 (site 1); 9 (site 2)	14 (site 5); 0 (site 3) 6 (site 1); 7 (site 2)
pMV8	+ +	11 (site 3); 12 (site 1); 13 (site 2)	7 (site 3); 9 (site 1); 10 (site 2)	6 (site 3); 7 (site 1); 8 (site 2)	1 (site 3); 6 (site 1) 7 (site 2)
pMV9	+ + +	11 (site 1); 13 (site 2)	8 (site 1); 10 (site 2)	7 (site 1); 9 (site 2)	6 (site 1); 8 (site 2)
pMV5	+	8 (site 1); 7 (site 2)	5 (site 1); 4 (site 2)	4 (site 1); 3 (site 2)	3 (site 1); 2 (site 2)
pMV12	+	10 (site 1); 12 (site 2)	8 (site 1); 9 (site 2)	6 (site 1); 6 (site 2)	4 (site 1); 5 (site 2)
pMV19	−	8 (site 1); 10 (site 2)	3 (site 1); 4 (site 2)	1 (site 1); 3 (site 2)	0 (site 1); 0 (site 2)
pMV6	−	6 (site 1)	2 (site 1)	1 (site 1)	0 (site 1)
pMV16	−	10 (site 1); 10 (site 1); 11(site 1)	8 (site 1); 8 (site 1); 9 (site 1)	6 (site 1); 7 (site 1); 7 (site 1)	4 (site 1); 6 (site 1); 4 (site 1)
pMV7	−	10 (site 1)	6 (site 1)	4 (site 1)	1 (site 1)
pMV17	−	15 (site 2); 10 (site 2); 5 (site 2)	13 (site 2); 8 (site 2); 1 (site 2)	11 (site 2); 7 (site 2); 0 (site 2)	6 (site 2); 6 (site 2); 0 (site 2)

[a] Repression sensitivity as reported in Figure 1 of Vella et al. (2004)

References

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.

Supplementary Table 2. miRNA:target interaction energy computed by $\sum\Delta G_{total}$ and predicted functional interaction [a] for which experimental tests have been published [b]

Organism & references [c]	miRNA	Target	Experimental evidence	$\sum\Delta G_{total}$ (kcal/ mol)	Predicted functional interaction	Average $\sum\Delta G_{total}$ of randomers (kcal/mol)
Ce [1,2]	*let-7*	hbl-1	+	-191.69	+	-6.185
Ce [3]	*let-7*	lin-41	+	-83.65	+	-5.390
Ce [4]	*let-7*	daf-12	+	-157.42	+	-3.368
Ce [4]	*let-7*	pha-4	+	-24.72	+	-0.071
Ce [5,6]	*lin-4*	lin-14	+	-41.44	+	-1.069
Ce [7]	*lin-4*	lin-28	+	-12.72	+	-0.187
Ce [8,9]	*lsy-6*	cog-1	+	-38.89	+	-0.077
Ce [10]	*miR-84*	let-60	+	-84.02	+	-1.506
Ce [11]	*miR-273*	die-1	+	-11.23	+	-0.158
Dm [12]	*bantam*	Hid	+	-38.43	+	-0.212
Dm [13]	*miR-9a*	sens	+	-16.49	+	-1.184
Ce [14]	*let-7*	T14B1.1 [d]	+	-119.88	+	-3.535
Ce [14]	*let-7*	uba-1	+	-28.04	+	-0.014
Ce [14]	*let-7*	C35E7.4 [d]	+	-14.73	+	-1.167
Ce [14]	*let-7*	unc-129 [d]	+	-28.06	+	-1.348
Ce [14]	*let-7*	nhr-4 [d]	+	0.00	−	-1.010
Ce [14]	*let-7*	F29G9.4	+	-17.84	+	-1.811
Ce [14]	*let-7*	C27D6.4	+	-33.88	+	-1.040
Ce [14]	*let-7*	C48A7.2	+	0.00	−	-0.143
Ce [14]	*let-7*	K08F8.1	+	-38.85	+	-0.843
Ce [14]	*let-7*	K07A6.2	+	NA [e]		
Ce [14]	*let-7*	ceh-16 [d]	+	-39.16	+	-0.077
Ce [14]	*let-7*	oig-2 [d]	+	0.00	−	-0.214
Ce [15,16]	*miR-48*	hbl-1	+	-161.00	+	-6.184
Ce [15]	*miR-84*	hbl-1	+	-115.52	+	-5.541
Ce [15]	*miR-241*	hbl-1	+	-266.45	+	-4.177
Ce [17]	*let-7*	nhr-23	+	-38.84	+	-0.858
Ce [17]	*let-7*	nhr-25	+	-23.66	+	-0.472
Ce [17]	*miR-84*	nhr-23	+	-25.30	+	-0.013
Ce [17]	*miR-84*	nhr-25	+	-46.18	+	-0.754
Dm [18]	*miR-7*	Bearded	+	-23.29	+	-1.609
Dm [18]	*miR-7*	E(spl)m5	+	-69.92	+	-2.901
Dm [18]	*miR-4*	Bearded	+	-14.38	+	-1.609
Dm [18]	*miR-79*	Bearded	+	-17.47	+	-1.613
Dm [18]	*miR-7*	E(spl)mγ	+	-0.85	−	-0.062
Dm [18]	*miR-7*	Tom	+	-17.84	+	-0.912
Dm [18]	*miR-7*	Bob	+	NA [e]		
Dm [19]	*miR-7*	Hairy	+	-27.87	+	-2.277
Dm [18]	*miR-7*	Cut	+	-13.09	+	-1.493
Dm [18]	*miR-7*	Wingless	+	-1.25	−	-3.65
Dm [18]	*miR-4*	Tom	+	0.00	−	-0.912
Dm [18]	*miR-4*	E(spl)mδ	+	0.00	−	-2.258
Dm [18]	*miR-4*	E(spl)mγ	+	0.00	−	-0.062

Supplementary Table 2 (continued)

Organism & references [c]	miRNA	Target	Experimental evidence	$\sum\Delta G_{total}$ (kcal/ mol)	Predicted functional interaction	Average $\sum\Delta G_{total}$ of randomers (kcal/mol)
Dm [18]	*miR-4*	E(spl)mα	+	NA [e]		
Dm [18]	*miR-4*	E(spl)m4	+	NA [e]		
Dm [18]	*miR-4*	E(spl)m5	+	-13.00	+	-2.901
Dm [18]	*miR-79*	Tom	+	0.00	−	-0.912
Dm [18]	*miR-79*	E(spl)mδ	+	-10.68	+	-2.258
Dm [18]	*miR-79*	E(spl)mγ	+	0.00	−	-0.062
Dm [18]	*miR-79*	E(spl)mα	+	NA [e]		
Dm [18]	*miR-79*	E(spl)m4	+	NA [e]		
Dm [18]	*miR-79*	E(spl)m5	+	0.00	−	-2.877
Dm [19]	*miR-2b*	grim	+	-2.06	−	-2.373
Dm [19]	*miR-2a*	reaper	+	-16.81	+	-1.810
Dm [19]	*miR-2b*	sickle	+	-20.04	+	-1.671
Dm [18]	*miR-7*	E(spl)m3	+	-21.77	+	-1.611
Ce [9]	*lsy-6*	ZK637.13	–	-13.86	+	-2.124
Ce [9]	*lsy-6*	C02B8.4	–	-4.22	−	-0.081
Ce [9]	*lsy-6*	F55G1.1	–	0.00	−	-0.001
Ce [9]	*lsy-6*	C48D5.2a	–	-0.04	−	-1.833
Ce [9]	*lsy-6*	F59A6.1	–	-2.72	−	-4.541
Ce [9]	*lsy-6*	F40H3.4	–	0.00	−	-0.042
Ce [9]	*lsy-6*	T05C12.8	–	0.00	−	-0.001
Ce [9]	*lsy-6*	C27H6.3	–	-0.06	−	-0.530
Ce [9]	*lsy-6*	T23E1.1	–	0.00	−	-0.110
Ce [9]	*lsy-6*	T14G12.2	–	0.00	−	-0.001
Ce [9]	*lsy-6*	T20G5.9	–	0.00	−	-0.085
Ce [9]	*lsy-6*	R07E3.5	–	0.00	−	-0.027

[a] An interaction is predicted to be functional ("+") if for nucleation potential threshold of 4.09 kcal/mol, the sum of $\Delta G_{total} < -10$ kcal/mol; otherwise, the interaction is non-functional ("−");
[b] Positive interactions confirmed by conventional genetic epistasis are in shaded part of the table;
[c] Ce: *C. elegans*; Dm: *D. melanogaster*;
[d] Conflicting experimental evidence presented in the reference;
[e] 3′ UTR sequence is not available from the WormBase Release 1.44 (http://www.wormbase.org), or from the FlyBase Release 4.3 (http://www.flybase.org).

[1]Abrahante et al, 2003; [2]Lin et al, 2003; [3]Slack et al, 2000; [4]Grosshans et al, 2005; [5]Lee et al, 1993; [6]Wightman et al, 1993; [7]Moss et al, 1997; [8]Johnston and Hobert,2003; [9]Didiano and Hobert, 2006; [10]Johnson et al, 2005; [11]Chang et al, 2004; [12]Brennecke et al, 2003; [13]Li et al, 2006; [14]Lall et al, 2006; [15]Abbott et al, 2005 ; [16]Li et al, 2005; [17]Hayes et al, 2006; [18]Lai et al, 2005; [19]Stark et al, 2003.

References

1. Abrahante, J.E. et al. The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs. *Dev Cell* **4**, 625-37 (2003).
2. Lin, S.Y. et al. The C elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. *Dev Cell* **4**, 639-50 (2003).

3. Slack, F.J. et al. 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**, 659-69 (2000).
4. Grosshans, H., Johnson, T., Reinert, K.L., Gerstein, M. & Slack, F.J. The temporal patterning microRNA let-7 regulates several transcription factors at the larval to adult transition in C. elegans. *Dev Cell* **8**, 321-30 (2005).
5. Lee, R.C., Feinbaum, R.L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-54 (1993).
6. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**, 855-62 (1993).
7. Moss, E.G., Lee, R.C. & Ambros, V. The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell* **88**, 637-46 (1997).
8. Johnston, R.J. & Hobert, O. A microRNA controlling left/right neuronal asymmetry in Caenorhabditis elegans. *Nature* **426**, 845-9 (2003).
9. Didiano, D. & Hobert, O. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat Struct Mol Biol* **13**, 849-51 (2006).
10. Johnson, S.M. et al. RAS is regulated by the let-7 microRNA family. *Cell* **120**, 635-47 (2005).
11. Chang, S., Johnston, R.J., Jr., Frokjaer-Jensen, C., Lockery, S. & Hobert, O. MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature* **430**, 785-9 (2004).
12. Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B. & Cohen, S.M. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**, 25-36 (2003).
13. Li, Y., Wang, F., Lee, J.A. & Gao, F.B. MicroRNA-9a ensures the precise specification of sensory organ precursors in Drosophila. *Genes Dev* **20**, 2793-805 (2006).
14. Lall, S. et al. A genome-wide map of conserved microRNA targets in C. elegans. *Curr Biol* **16**, 460-71 (2006).
15. Abbott, A.L. et al. The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in Caenorhabditis elegans. *Dev Cell* **9**, 403-14 (2005).
16. Li, M., Jones-Rhoades, M.W., Lau, N.C., Bartel, D.P. & Rougvie, A.E. Regulatory mutations of mir-48, a C. elegans let-7 family MicroRNA, cause developmental timing defects. *Dev Cell* **9**, 415-22 (2005).
17. Hayes, G.D., Frand, A.R. & Ruvkun, G. The mir-84 and let-7 paralogous microRNA genes of Caenorhabditis elegans direct the cessation of molting via the conserved nuclear hormone receptors NHR-23 and NHR-25. *Development* **133**, 4631-4641 (2006).
18. Lai, E.C., Tam, B. & Rubin, G.M. Pervasive regulation of Drosophila Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. *Genes Dev* **19**, 1067-80 (2005).
19. Stark, A., Brennecke, J., Russell, R.B. & Cohen, S.M. Identification of Drosophila MicroRNA targets. *PLoS Biol* **1**, E60 (2003).

Supplementary Table 3. miRNA:target interaction energy computed by $\sum \Delta G_{total}$ and functional interaction [a] (in parentheses) predicted using various RNA folding programs for experimentally-tested *lin-41* 3′ UTR mutants.

Mutant construct	Repression Activity	$\sum \Delta G_{total}$ (kcal/mole) and predicted functional interaction					
		MFE structure			1000 lowest energy structures by RNASubOpt [d]	100 suboptimal structures by Mfold [e]	1000 structures sampled by Sfold
		by Mfold	by RNAfold [b]	by RNAstructure [c]			
pMV1	+ +	−60.5 (+)	−33.6 (+)	−35.1 (+)	−33.7 (+)	−59.8 (+)	−43.3 (+)
pMV8	+ +	−43.8 (+)	−33.3 (+)	−34.8 (+)	−35.0 (+)	−68.1 (+)	−43.1 (+)
pMV9	+ + +	−43.5 (+)	−33.4 (+)	−34.7 (+)	−35.1 (+)	−55.3 (+)	−43.4 (+)
pMV5	+	−0.2 (−)	−0.3 (−)	−16.8 (+)	−1.0 (−)	−27.4 (+)	− 5.0 (−)
pMV12	+	−36.6 (+)	−32.4 (+)	−33.0 (+)	−34.1 (+)	−28.2 (+)	−20.3 (+)
pMV19	−	−40.4 (+)	−15.9 (+)	−33.6 (+)	−32.4 (+)	−54.7 (+)	− 8.3 (−)
pMV6	−	−0.7 (−)	−0.0 (−)	−12.4 (+)	−0.1 (−)	−0.0 (−)	− 0.0 (−)
pMV16	−	−26.4 (+)	−42.9 (+)	−52.3 (+)	−45.8 (+)	−26.2 (+)	− 5.7 (−)
pMV7	−	−0.8 (−)	−15.3 (+)	−16.6 (+)	−16.2 (+)	−26.9 (+)	− 0.0 (−)
pMV17	−	−29.4 (+)	−47.7 (+)	−43.7 (+)	−47.9 (+)	−82.6 (+)	− 5.6 (−)

[a] An interaction is predicted to be functional ("+") if, for nucleation potential threshold of 4.09 kcal/mol, $\sum \Delta G_{total} < -10$ kcal/mol; otherwise, the interaction is non-functional ("−");
[b] Hofacker (2003)
[c] Mathews et al. (1999)
[d] Wuchty et al. (1999), Hofacker (2003),
[e] Generated with default parameter settings for mfold

References

Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Res* **31**, 3429-3431.

Mathews, D. H., Burkard, M. E., Freier, S. M., Wyatt, J. R., and Turner, D. H. (1999). Predicting oligonucleotide affinity to nucleic acid targets. *RNA* **5**, 1458-1469.

Wuchty, S., Fontana, W., Hofacker, I. L., and Schuster, P. (1999). Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* **49**, 145-165.

Supplementary Table 4. 3′ UTR sequences of *lac-Z* reporter constructs.

pMV9: Site 1 and Site 2, wild type linker.

TAATAGGCCTACTAGACCGCGGAACTCAAGTATACCTT***TTATACAACCGTTCTACACTCA***ACGCGATGTAAATATCGC
AATCCCTTT***TTATACAACCATTCTGCCTC***TGAACCATTGAAACCTTCTCCCGTACTCCCACCAACCATGGCCGCTGTC
ATCAGATCGCCATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCC
CTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAACTTCTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCA
CCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAA

pMV19: Site 1 and Site 2, mutant linker.

TAATAGGCCTACTAGACCGCGGAACTCAAGTATACCTT***TTATACAACCGTTCTACACTCA***AAGTGATGTAAATATAGG
AATGTATTT***TTATACAACCATTCTGCCTC***TGAACCATTGAAACCTTCTCCCGTACTCCCACCAACCATGGCCGCTGTC
ATCAGATCGCCATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCC
CTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAACTTCTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCA
CCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAA

pVT701: Sites 1 and 2; mutant linker designed to maintain accessibility.

TAATAGGCCTACTAGACCGCGGAACTCAAGTATACCTT***TTATACAACCGTTCTACACTCA***ACGGGATGTCCCTATCCC
AATCCCTTT***TTATACAACCATTCTGCCTC***TGAACCATTGAAACCTTCTCCCGTACTCCCACCAACCATGGCCGCTGTC
ATCAGATCGCCATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCC
CTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAACTTCTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCA
CCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAA

pVT702: Sites 1 and 2; mutant linker designed to maintain accessibility.

TAATAGGCCTACTAGACCGCGGAACTCAAGTATACCTT***TTATACAACCGTTCTACACTCA***ACCGCTAGCTTATTAGCG
AATTTTCCC***TTATACAACCATTCTGCCTC***TGAACCATTGAAACCTTCTCCCGTACTCCCACCAACCATGGCCGCTGTC
ATCAGATCGCCATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCC
CTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAACTTCTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCA
CCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAA

pVT704: Sites 1 and 2, mutant linker designed to impair accessibility.

TAATAGGCCTACTAGACCGCGGAACTCAAGTATACCTT***TTATACAACCGTTCTACACTCA***ACATGCGGCAGTGATACG
CTATTTCCC***TTATACAACCATTCTGCCTC***TGAACCATTGAAACCTTCTCCCGTACTCCCACCAACCATGGCCGCTGTC
ATCAGATCGCCATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCC
CTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAACTTCTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCA
CCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAA

pVT705: Sites 1 and 2; mutant linker aimed impair accessibility.

TAATAGGCCTACTAGACCGCGGAACTCAAGTATACCTT***TTATACAACCGTTCTACACTCA***AAGGGATGTAAATATAGG
AAACAATTT***TTATACAACCATTCTGCCTC***TGAACCATTGAAACCTTCTCCCGTACTCCCACCAACCATGGCCGCTGTC
ATCAGATCGCCATCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCC
CTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAACTTCTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCA
CCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATTAATTCGTAATAAA

pVT712: Site1- 27 nt Spacer-Site 1- 27 nt Spacer-Site1, designed for accessibility.

TAATAGGCCTACTAGACCGCGAACTCAAGTATACCTT***TTATACAACCGTTCTACACTC***AACGCGATGTAAATATCGCAA
TCCCTTT***TTATACAACCGTTCTACACTCA***ACGCGATGTAAATATCGCAATCCCTTT***TTATACAACCGTTCTACACTCA***T
GAACCATTGAAACCTTCTCCCGTACTCCCACCAACCATGGCCGCTGTCATCAGATCGCCATCTCGCGCCCGTGCCTCTG
ACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTTGTTATTATC
AAAAAACTTCTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAAT
AGAATTAATTCGTAATAAA

pVT713: Site 2- 27 nt Spacer-Site 2- 27 nt Spacer-Site 2, designed for accessibility.

TAATAGGCCTACTAGACCGCGAACTCAAGTATACCTT***TTATACAACCATTCTGCCTC***ACGCGATGTAAATATCGCAATC
CCTTT***TTATACAACCATTCTGCCTC***ACGCGATGTAAATATCGCAATCCCTTT***TTATACAACCATTCTGCCTC***TGAACCA
TTGAAACCTTCTCCCGTACTCCCACCAACCATGGCCGCTGTCATCAGATCGCCATCTCGCGCCCGTGCCTCTGACTTCT
AAGTCCAATTACTCTTCAACATCCCTACATGCTCTTTCTCCCTGTGCTCCCACCCCCTATTTTTGTTATTATCAAAAAA
CTTCTCTTAATTTCTTTGTTTTTTAGCTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATTCAAAAATAGAATT
AATTCGTAATAAA

3′ UTR sequences from the TAA stop codon through AATAAA polyadenylation signal. Modified *lin-41* 3′ UTR were inserted between the *sac*II (CCGCGG, underlined) and *nco*I (CCATGG, underlined) sites. (The *nco*I site in pVT712 and pVT713 was modified in the course of plasmid construction). Sequences flanking the *sac*II/*nco*I insert are from the *unc-54* 3′ UTR. *let-7* complementary site 1 and site 2 are indicated in bold italics; the 27 nt spacer sequence between site 1 and site 2 is underlined. Alignments of wild type (pMV9) and mutant (pMV19, pVT701-705) spacer sequences are shown in Supplementary Table 5.

Supplementary Table 5. Sequences of the 27 nt spacer between *let-7* site 1 and site 2 of the *lin-41* 3′ UTR in *lac-Z* reporter plasmids.

plasmid	Spacer sequence
pMV9	ACGCGATGTAAATATCGCAATCCCTTT
pMV19	A**A**G**T**GATGTAAATAT**A**G**G**AAT**GTA**TTT
pVT701	ACG**G**GATGT**CCC**TATC**C**CAATCCCTTT
pVT702	AC**CGCTA**G**CTT**AT**TAGCG**AAT**TTTCCC**
pVT704	AC**AT**G**CG**G**C**A**GTG**AT**ACGCTATTTCCC**
pVT705	A**A**G**G**GATGTAAATAT**A**G**G**AA**A**C**AA**TTT

Bold nucleotides indicate nucleotides that are mutant compared to the wild type spacer sequence (pMV9). See Supplementary Table 4 for complete 3′ UTR sequences.

Supplementary Data

Potent Role of Target Structure in MicroRNA Function

Dang Long, Rosalind Lee, Peter Williams, Chi Yu Chan, Victor Ambros, and Ye Ding

Supplementary Results

Assumptions for regression analysis

We first examined the assumption of independence of measurements for variables involved in linear regression analysis. Because β-gal ratios were measured for different constructs that grew independently, independence is not an issue for the β-gal ratios. It should be noted that slightly different 3′ UTRs can have substantially different $\Delta G_{\text{disruption}}$ values (e.g., pMV9 vs. pMV19) due to different 3′ UTR structure at and around the target sites. However, this does not guarantee independency between $\Delta G_{\text{disruption}}$ values, which were computed for highly similar UTRs, i.e., mutants of the same 3′ UTR. To statistically examine this issue, we used the 3′ UTR construct sequence for pMV9 as the baseline, and computed the pair-wise global alignment score for each of the other constructs in Table 3, using the EMBOSS:Align program (www.ebi.ac.uk/emboss/align). We found that, the alignment score, which takes into account of both similarity and gaps, is not at all correlated with the $\Delta G_{\text{disruption}}$ (Pearson's correlation of −0.3769, with a p-value of 0.3574). Thus, there was no evidence of dependency. We also note that the clustering of $\Delta G_{\text{disruption}}$ values for the data points in Figure 4 is due to the construct design which focused on highly accessible sites and highly inaccessible sites, and is not due to high similarity in construct sequences as it appears. The cluster patterns from multiple sequence alignment using ClustalW is quite different from that based on $\Delta G_{\text{disruption}}$ values.

We next performed diagnostic checks for the other underlying assumptions for linear regression model: linearity, independence of the errors (no serial correlation), normality and constant error variance. Specifically, a lack of unusual pattern on a residual plot (data not shown) verified linearity, a random pattern on an autocorrelation plot (data not shown) indicated a lack of serial correlation, and a usual linear pattern for a normal

quantile-quantile (Q-Q) plot validated the assumption of the normality and constant error variance. Thus, we did not detect a violation of any of the underlying assumptions.

Weighted regression analysis

When the standard deviations of the measured repression levels are available, an alternative to the usual linear regression analysis is the weighted regression analysis. In a weighted least-squares regression, the square term in the sum of squares for a data point is multiplied by a weight (Weisberg 2005). When the standard deviation of the repression level (as measured by β-gal ratios) from multiple measurements is available for every construct, $1/(\text{standard deviation})^2$ (i.e., 1/variance) can be used as the weight (Weisberg 2005). The weighted regression yielded a R^2 of 0.6211 with a significant *p*-value of 0.0202, which are highly similar to the results from the un-weighted regression analysis. The underlying assumptions were also examined for the weighted analysis, and no violation was detected. Thus, the conclusions from the un-weighted regression analysis are also valid for the weighted analysis.

Supplementary Discussion

The reliability of a computational RNA folding algorithm is known to vary from sequence to sequence. However, local structures of certain regions of an RNA sequence can be well-predicted with high probabilities. It is likely that the local structures for regulatory sites in 3′ UTRs are highly predictable, because a local conformation (e.g., a hairpin) favorable for a regulatory function may be "conserved" for the population of probable structures for the entire RNA molecule. In such cases, the better performance by Sfold is expected.

A two-step model for miRNA:target annealing has been considered previously (Rajewsky and Socci, 2004), although with important differences compared to our model. At the first step of miRNA target searching, Rajewsky and Socci (2004) employed a "binding nucleus", defined as a GC rich string which typically form 6-8 consecutive base pairs with the miRNA. After nucleus scoring, the second phase of the algorithm involves a thermodynamic calculation that models the completion miRNA:target hybridization. Although the Rajewsky and Socci (2004) algorithm involved a two-step model, it did not

incorporate target secondary structure and accessibility as a governing principle, and hence it is fundamentally different from our model. Specifically, in our model, a favorable nucleation potential requires an accessible target site. Thus, a "nucleus", as defined by Rajewsky and Socci (2004), may not be a good nucleation site as defined by our model, and vise versa. Moreover, the total hybridization energy for hybridization calculated through our model considers both the energy cost for disrupting the local target structure and the energy gain from miRNA:target hybridization. Only the energy gain component was considered in Rajewsky and Socci (2004).

References

Rajewsky, N., and Socci, N. D. (2004). Computational identification of microRNA targets. *Dev Biol* **267**, 529-535.

Vickers, T. A., Wyatt, J. R., and Freier, S. M. (2000). Effects of RNA secondary structure on cellular antisense activity. *Nucleic Acids Res* **28**, 1340-1347.

Weisberg, S. (2005) *Applied Linear Regression*, 3rd edition. John Wiley & Sons, New York.

Supplementary Methods

Potent Role of Target Structure in MicroRNA Function

Dang Long, Rosalind Lee, Peter Williams, Chi Yu Chan, Victor Ambros, and Ye Ding

Prediction of mRNA secondary structures

The secondary structure of an mRNA molecule influences the accessibility of that mRNA to numerous gene regulatory mechanisms that depend on base-pairing, including translational inhibition by antisense oligonucleotides (Vickers et al., 2000) and target cleavage by ribozymes (Zhao and Lemke, 1998) or siRNAs (Bohula et al., 2003; Kretschmer-Kazemi Far and Sczakiel, 2003; Overhoff et al., 2005; Schubert et al., 2005; Yoshinari et al., 2004). However, the determination of mRNA secondary structure presents theoretical and experimental challenges. One major impediment to the accurate prediction of mRNA structures stems from the likelihood that a particular mRNA may not exist as a single structure, but in a population of structures in thermodynamic equilibrium (Christoffersen et al., 1994; Altuvia et al., 1989; Betts and Spremulli, 1994). Thus, the computational prediction of secondary structure based on free energy minimization is not well suited to the task of providing a realistic representation of mRNA structures.

An alternative to free energy minimization for charactering the ensemble of probable structures for a given RNA molecule has been developed (Ding and Lawrence, 2003). In this approach, a statistically representative sample is drawn from the Boltzmann-weighted ensemble of RNA secondary structures for the RNA. Such samples can faithfully and reproducibly characterize structure ensembles of enormous sizes. In particular, this method has been shown to make better structural predictions (Ding et al., 2005) and to better represent the likely population of mRNA structures (Ding et al. 2006), and to yield a significant correlation between predictions and antisense inhibition data (Ding and Lawrence, 2001). A sample size of 1,000 structures is sufficient to guarantee statistical reproducibility in sampling statistics and clustering features (Ding and Lawrence, 2003; Ding et al. 2006). The structure sampling method has been

implemented in the Sfold software package (Ding et al., 2004) and is used here for mRNA folding.

To predict the secondary structure of the 3′ UTR for each of the tested targets for *C. elegans* or in *D. melanogaster*, we used Sfold to fold the 3′ UTR region retrieved either from the WormBase Release 1.44 (http://www.wormbase.org), or from the FlyBase Release 4.3 (http://www.flybase.org), together with 300 adjacent coding nucleotides. The addition of neighboring nucleotides in the coding region serves to accommodate potentially important secondary structure interactions between the 3′ UTR sequence and nearby nucleotides in the coding region. In other words, we do not assume that the 3′ UTR is always an independent folding domain, as RNA structures often involve long-distance base-pairing interactions. An alternative to the addition of 300 coding nucleotides would be to include the complete coding region and the 5′ UTR for folding, which is far more computationally intensive and less manageable for future genome-scale RNA folding. For a testing set of mRNAs, we did not observe a statistically appreciable difference in the folding results for the 3′ UTR region. We thus considered the addition of 300 coding nucleotides adequate.

Calculating nucleation potential between a small antisense nucleic acid and a structured mRNA target

In vitro hybridization studies using antisense oligonucleotides suggested that hybridization of a short oligonucleotide to a target RNA requires an accessible local target structure (Milner et al. 1997). Such a local structure includes a site of unpaired bases for nucleation, and duplex formation progresses from the nucleation site and stops when it meets an energy barrier. In a kinetic study, it was suggested that the nucleation step is *rate-limiting*, and that it involves formation of four or five base pairs between the interacting nucleic acids (Hargittai et al. 2004). When both of the two interacting RNAs have strong intramolecular structures, nucleation involves intermolecular base-pairing interactions between complementary loops (Hargittai et al. 2004; Kolb et al. 2001).

Nucleation potential for a miRNA hybridizing to a potential target site is calculated using the sample of 1000 structures predicted by Sfold for the target mRNA. Specifically, for sampled structure i ($1 \leq i \leq 1000$), the nucleation potential of a complementary site is

$\Delta G_{N,i} = \min(\sum_{1 \le j \le 3} \Delta G_{stack\,(j)})$, where $\sum_{1 \le j \le 3} \Delta G_{stack\,(j)}$ is the sum over the energies of three base-pair stacks for a single-stranded 4-bp block in this site (Fig. 1a), and the minimum is taken over all such blocks; in the absence of such a block, $\Delta G_{N,i} = 0$. The calculation is repeated for each of 1,000 sampled structures, and the average over the structure sample, i.e., $\Delta G_N = (\sum_{1 \le i \le 1000} \Delta G_{N,i})/1000$, is then calculated and is referred to as the nucleation potential for this putative binding site.

Construction of lin-41 3′ UTR reporter plasmids

Plasmids containing a *lacZ* reporter gene driven by the *col*-10 promoter and fused to various *lin-41* 3′ UTR sequences were generated as previously described (Vella et al. 2004) by insertion of PCR-generated, *ncoI/sacII*-digested DNA fragments between the *sacII* and *ncoI* sites of pFS1031. Details of the plasmid constructions are available on request. The sequences of the 3′ UTRs of *lacZ* reporters employed here are listed in Supplementary Tables 4 and 5.

Generation and analysis of C. elegans transgenics

Plasmids were transformed into *C. elegans* by microinjection (Mello et al. 1991) using an unc-119+ co-injection marker and recipient animals of genotype *unc-119* (*ed3*) (strain DP38). Transgenic lines were identified by heritable genetic rescue of *unc-119* (*ed3*). Two to five independent transgenic lines were produced for each plasmid construct. For each line, mixed-stage cultures were harvested, and animals were fixed and stained with X-gal for β-galactosidase activity as described (Vella et al. 2004). Several hundred animals were mounted per slide and examined in the dissecting microscope. Random, non-overlapping fields were chosen, and each animal in the field was scored visually for the presence of blue X-gal staining in hypodermal cells. Animals with one or more X-gal staining hypodermal cells were scored as β-gal+. In some cases, fields were recorded by digital imaging for scoring at a later date. Adults were distinguished from larvae by the presence of at least one developing embryo.

References

Altuvia, S., Kornitzer, D., Teff, D., and Oppenheim, A. B. (1989). Alternative mRNA structures of the cIII gene of bacteriophage lambda determine the rate of its translation initiation. *J Mol Biol* **210**, 265-280.

Betts, L., and Spremulli, L. L. (1994). Analysis of the role of the Shine-Dalgarno sequence and mRNA secondary structure on the efficiency of translational initiation in the Euglena gracilis chloroplast atpH mRNA. *J Biol Chem* **269**, 26456-26463.

Bohula, E. A., Salisbury, A. J., Sohail, M., Playford, M. P., Riedemann, J., Southern, E. M., and Macaulay, V. M. (2003). The efficacy of small interfering RNAs targeted to the type 1 insulin-like growth factor receptor (IGF1R) is influenced by secondary structure in the IGF1R transcript. *J Biol Chem* **278**, 15991-15997.

Christoffersen, R.E., McSwiggen, J.A. & Konings, D. (1994). Application of computational technologies to ribozyme biotechnology products. *J. Mol. Structure (Theochem)* **311**, 273-284 .

Ding, Y., Chan, C.Y. & Lawrence, C.E. (2006) Clustering of RNA secondary structures with application to messenger RNAs. *J Mol Biol* **359**, 554-71.

Ding, Y., Chan, C.Y. & Lawrence, C.E. (2005) RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA* **11**, 1157-66.

Ding, Y., Chan, C.Y. & Lawrence, C.E. (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* **32**, W135-41.

Ding, Y. & Lawrence, C.E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res* **31**, 7280-301.

Ding, Y. & Lawrence, C.E. (2001) Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucleic Acids Res* **29**, 1034-46.

Hargittai, M.R., Gorelick, R.J., Rouzina, I. & Musier-Forsyth, K. (2004) Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J Mol Biol* **337**, 951-68.

Kolb, F.A. et al. (2001) Bulged residues promote the progression of a loop-loop interaction to a stable and inhibitory antisense-target RNA complex. *Nucleic Acids Res* **29**, 3145-53.

Kretschmer-Kazemi Far, R., and Sczakiel, G. (2003). The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides. *Nucleic Acids Res* **31**, 4417-4424.

Mello, C.C., Kramer, J.M., Stinchcomb, D.,and Ambros, V. (1991). Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences *EMBO J* **10**:3959-3970.

Milner, N., Mir, K.U. & Southern, E.M. (1997) Selecting effective antisense reagents on combinatorial oligonucleotide arrays. *Nat Biotechnol* **15**, 537-41.

Overhoff, M. et al. (2005) Local RNA target structure influences siRNA efficacy: a systematic global analysis. *J Mol Biol* **348**, 871-81.

Schubert, S., Grunweller, A., Erdmann, V.A. & Kurreck, J. (2005) Local RNA target structure influences siRNA efficacy: systematic analysis of intentionally designed binding regions. *J Mol Biol* **348**, 883-93.

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.

Vickers, T. A., Wyatt, J. R., and Freier, S. M. (2000). Effects of RNA secondary structure on cellular antisense activity. *Nucleic Acids Res* **28**, 1340-1347.

Yoshinari, K., Miyagishi, M., and Taira, K. (2004). Effects on RNAi of the tight structure, sequence and position of the targeted region. *Nucleic Acids Res*. **32**, 691-699.

Zhao, J. J., and Lemke, G. (1998). Rules for ribozymes. *Mol Cell Neurosci*. **11**, 92-97.