# Potent effect of target structure on microRNA function

**Article** *in* Nature Structural & Molecular Biology · April 2007

DOI: 10.1038/nsmb1226 · Source: PubMed



**6 authors**, including:

Long Dang
University of Da nang
**17** PUBLICATIONS **825** CITATIONS
SEE PROFILE

Peter Williams
Stony Brook University
**12** PUBLICATIONS **1,267** CITATIONS
SEE PROFILE

nature structural & molecular biology

# Potent effect of target structure on microRNA function

Dang Long[1], Rosalind Lee[2], Peter Williams[2], Chi Yu Chan[1], Victor Ambros[2] & Ye Ding[1]

**MicroRNAs (miRNAs) are small noncoding RNAs that repress protein synthesis by binding to target messenger RNAs. We investigated the effect of target secondary structure on the efficacy of repression by miRNAs. Using structures predicted by the Sfold program, we model the interaction between an miRNA and a target as a two-step hybridization reaction: nucleation at an accessible target site followed by hybrid elongation to disrupt local target secondary structure and form the complete miRNA-target duplex. This model accurately accounts for the sensitivity to repression by *let-7* of various mutant forms of the *Caenorhabditis elegans lin-41* 3′ untranslated region and for other experimentally tested miRNA-target interactions in *C. elegans* and *Drosophila melanogaster*. These findings indicate a potent effect of target structure on target recognition by miRNAs and establish a structure-based framework for genome-wide identification of animal miRNA targets.**



miRNAs are a class of small noncoding RNAs found in plants and animals that regulate gene expression by base-pairing to mRNA targets, causing either target degradation or translational repression[1]. Experimental and computational studies have identified thousands of miRNAs encoded in animal genomes[1,2]. miRNAs have regulatory roles in developmental timing, patterning, embryogenesis, differentiation, organogenesis, stem cell and germline proliferation, growth control and apoptosis; they are also involved in physiological processes, including cancer, aging, hematopoiesis and endocrine function[1,3–5]. However, there is still much to be learned about the biological functions of miRNAs and the molecular mechanisms by which they regulate gene expression[1,6,7].

In plants, miRNAs bind their targets by complete or nearly complete complementarity, so target identification is straightforward[8]. By contrast, animal miRNAs are typically only partially complementary to their targets, which poses a challenge for the accurate computational identification of animal miRNA targets[6,9,10]. An improved understanding of the requirements for functional interactions between miRNAs and their targets is essential for precisely elucidating miRNA targets in animals.

To date, most studies of miRNA-target interactions have focused primarily on the characteristics of the sequence complementarity between the miRNA and putative target sites in the mRNA. Tissue-culture experiments[11,12] and computational analyses of base-pairing between miRNAs and mRNA targets[13,14] have suggested that perfect Watson-Crick complementarity of seven or eight consecutive bases (typically at positions 2–8) in the 5′ ends of miRNAs (the 'seed' region) is an important signal for target regulation. A study using an *in vivo* assay for miRNA repression in transgenic fruitflies has further suggested that strong complementarity in the 3′ end of a miRNA can contribute to miRNA efficacy and specificity by compensating for weak base-pairing in the 5′ end[6].

Although the base-pairing between a miRNA and target site is important, the sequence context surrounding miRNA-binding sites has been reported to influence sensitivity to repression by a miRNA[15,16]. Sequence context could influence miRNA efficacy by mediating the binding of hypothetical cofactor proteins or by affecting the secondary structure of a target site and hence its accessibility to binding by the miRNA[15]. Although it is reasonable to postulate that target structure could be a factor in miRNA efficacy, the conventional minimum free energy (MFE) approach for modeling RNA secondary structure has limitations when it comes to accurately representing the structure of an mRNA *in vivo*. An analysis of five known miRNA-target pairs in *C. elegans* and *D. melanogaster*, using mfold[17] to model mRNA structure, has suggested that the presence of three consecutive single-stranded nucleotides in the target facilitates pairing with nucleotides in the 5′ seed region of the miRNA[18]. A prediction method incorporating this feature results in perhaps fewer false positive predictions, but it also suffers from reduced accuracy. In particular, this method does not identify the two *let-7* binding sites in the *lin-41* 3′ untranslated region (UTR), because neither site meets the requirement for a fully Watson-Crick 5′ seed region. Although some existing target search methods such as RNAhybrid[19] can identify *lin-41* as a target of *let-7*, they do not account for the observed differences in sensitivity to repression by *let-7* of *lin-41* 3′ UTR mutants that differ only in sequences outside of the *let-7*–complementary sites[15].

Here, we use Sfold[20] to model mRNA structures and employ a novel two-step model for the hybridization between an miRNA and a structured target. In the two-step model, hybridization nucleates at an accessible target site, and then the hybrid elongates to form the complete miRNA-target duplex (**Fig. 1**). The reliable performance of the model strongly suggests a potent effect of mRNA secondary structure on target recognition by miRNAs. The two-step model

[1]Wadsworth Center, New York State Department of Health, 150 New Scotland Avenue, Albany, New York 12208, USA. [2]Department of Genetics, Dartmouth Medical School, Hanover, New Hampshire 03755, USA. Correspondence should be addressed to Y.D. (yding@wadsworth.org) or V.A. (vambros@dartmouth.edu).

Received 19 January; accepted 7 March; published online 1 April 2007; doi:10.1038/nsmb1226

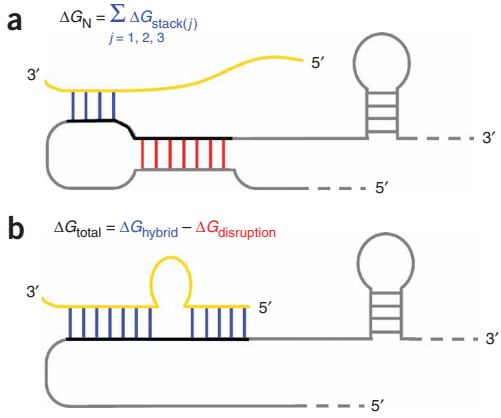


**Figure 1** A two-step model for hybridization between a structured mRNA and a partially complementary miRNA, illustrated for a single structural conformation of the target. (**a**) In the first step, the miRNA nucleates base-pairing at an accessible site of four unpaired nucleotides. The nucleation potential, $\Delta G_N$, is calculated by summing the stacking energies $\Delta G_{stack(j)}$ (where $(j)$ is the $j$th base pair stack, $1 \leq j \leq 3$) for the most stable 4-base-pair duplex at the complementary site; nucleation is considered favorable if $(\Delta G_N + \Delta G_{initiation}) < 0$ kcal mol$^{-1}$. (**b**) In the second step, the miRNA-target hybrid elongates, resulting in local disruption of the target secondary structure across the miRNA-complementary region. $\Delta G_{total}$, the total energy change for the hybridization, is the key energetic characteristic for this step (see Methods).

provides a structure-based, mechanistic framework for genome-wide identification of animal miRNA targets.

## RESULTS

### Previous data for *lin-41* 3′ UTR mutants

The 3′ UTR of the *C. elegans lin-41* mRNA contains six imperfectly complementary sites for *let-7*. Of these, site 1 and site 2 are well conserved in the orthologous gene in *Caenorhabditis briggsae*[15]. Vella and co-workers[15] have shown that a sequence from the *lin-41* 3′ UTR that contains only sites 1 and 2 (along with an intervening 27-nt spacer sequence) is sufficient, when placed in the context of the *unc-54* 3′ UTR, for repression of a *lacZ* reporter gene in transgenic worms[15]. Notably, this study[15] found that altering the 27-nt spacer abolishes repression by *let-7*. In particular, among ten tested constructs (**Table 1**), pMV9 and pMV19 both contain just sites 1 and 2; however, pMV9 contains the wild type 27-nt spacer, whereas pMV19 contains a mutated 27-nt spacer. pMV9 was repressed by *let-7*, but pMV19 was not repressed. Alteration of the relative configuration of sites 1 and 2 was also found to impair sensitivity to *let-7*. Construct pMV16 contains three copies of site 1 and no other complementary sites, and pMV17 contains only three copies of site 2. Neither of these latter constructs has normal sensitivity to *let-7 in vivo*. These results suggest that specific features of the configuration of sites 1 and 2 in the *lin-41* 3′ UTR are important for sensitivity to *let-7* repression. The authors considered altered local RNA secondary structure of the mutant 3′ UTR as a possible basis for the impaired functionality of pMV19, but they could draw no satisfactory interpretation of the results from structures predicted by mfold[15]. Below, we analyze the published *lin-41* 3′ UTR reporter data[15] using Sfold and the two-step model for miRNA-target hybridization (**Fig. 1**).

**Table 1 miRNA-target interaction energy predicts sensitivity of *lin-41* 3′ UTR reporters to repression by *let-7***

| | | $\sum\Delta G_{total}$ (kcal mol$^{-1}$)[a] | | |
|---|---|---|---|---|
| Mutant construct[b] | Reported repression sensitivity[b] | $\Delta G_{initiation}$ = 4.09 kcal mol$^{-1}$ [c] | $\Delta G_{initiation}$ = 0.0 kcal mol$^{-1}$ | $\sum\Delta G_{hybrid}$ (kcal mol$^{-1}$)[d] |
| pMV1 | + + | −43.3 (+)[e] | −69.5 (+) | −76.2 (+)[f] |
| pMV8 | + + | −43.1 (+) | −52.3 (+) | −57.0 (+) |
| pMV9 | + + + | −43.4 (+) | −43.4 (+) | −57.0 (+) |
| pMV5 | + | 0.0 (−) | −14.0 (+) | −51.7 (+) |
| pMV12 | + | −20.3 (+) | −36.6 (+) | −56.6 (+) |
| pMV19 | − | −8.3 (−) | −40.1 (+) | −57.0 (+) |
| pMV6 | − | 0.0 (−) | 0.0 (−) | −23.0 (+) |
| pMV16 | − | −5.7 (−) | −52.6 (+) | −83.3 (+) |
| pMV7 | − | 0.0 (−) | 0.0 (−) | −28.6 (+) |
| pMV17 | − | −5.6 (−) | −55.2 (+) | −87.0 (+) |

[a]See Methods. Because $\Delta G_{total}$ is computed for a site only if it has $(\Delta G_N + \Delta G_{initiation}) < 0.0$ kcal mol$^{-1}$, the value of $\Delta G_{total}$ for a site, and hence $\sum\Delta G_{total}$ for multiple sites on the same target, is dependent on the choice of the initiation energy $\Delta G_{initiation}$. [b]Ref. 15. [c]Ref. 24. [d]$\Delta G_{hybrid}$ is computed from RNAhybrid, and a site is counted for the sum if $\Delta G_{hybrid} \leq -14$ kcal mol$^{-1}$ (an energetic threshold previously considered for miRNA-target duplexes[35]; every *lin-41*–complementary site meets this threshold). This calculation ignores effects of target structure and nucleation. [e]For $\Delta G_{total}$, a *let-7*–target interaction is predicted to be functional (+) if, for *lin-41*–complementary sites[15], $\sum\Delta G_{total} < -10$ kcal mol$^{-1}$; otherwise, the interaction is predicted to be nonfunctional (−). [f]For $\Delta G_{hybrid}$, a *let-7*–target interaction is predicted to be functional (+) if, for *lin-41* complementary sites, $\sum\Delta G_{hybrid} \leq -14$ kcal mol$^{-1}$ (that is, at least one site meets the threshold); otherwise, the interaction is predicted to be nonfunctional (−).

### Frequencies of open nucleotides in target sites

To examine the potential importance of contiguous open nucleotides in a miRNA-complementary site, we used Sfold to generate a sample of secondary structures from the mRNA sequence formed by each of the mutant 3′ UTRs fused to the *lacZ* coding sequence[15]. We then used the sample of structures to identify blocks of open nucleotides of various lengths in the *let-7*–binding sites of the constructs. A nucleotide is considered open if its probability of being unpaired, as estimated from the Sfold structure sample, is greater than or equal to 0.5; a block of $n$ nucleotides (nt) is considered open if the probability that all $n$ nucleotides in the block are simultaneously unpaired is greater than or equal to 0.5. For putative binding sites in each construct, the numbers of open nucleotides and the numbers of open blocks are presented in **Supplementary Table 1** online.

For most of the constructs, a larger number of open single nucleotides in the complementary sites tends to correlate with greater repression. However, this trend does not hold for pMV16 and pMV17; moreover, for pMV9 and pMV19, the numbers of open single nucleotides do not contrast as sharply as do the repression sensitivities. Thus, the number of open single nucleotides is not sufficient to fully account for the data. In contrast, for pMV19 and pMV9, the repression sensitivity is better correlated with the frequency of open blocks of 4 nt. These findings suggest a potential effect of target structure in miRNA function; in particular, they suggest that effective miRNA-target interaction requires an open block of four or more consecutive complementary bases on the target.

### Sfold target accessibility profiling

To further examine the structural accessibility of *let-7*–complementary sequences in the 3′ UTRs of the *lacZ* reporter mRNAs corresponding to the pMV9 and pMV19 constructs, we used the Sfold samples of mRNA structures for these constructs to compute target accessibility profiles for sites 1 and 2 (**Fig. 2a**). The *let-7*–*lin-41* hybrid conformations predicted for the two sites in the wild-type 3′ UTR[15] are shown in **Figure 2b**. For a window length of 4 nt, the accessibility profile shows the probability that the four consecutive nucleotides starting at

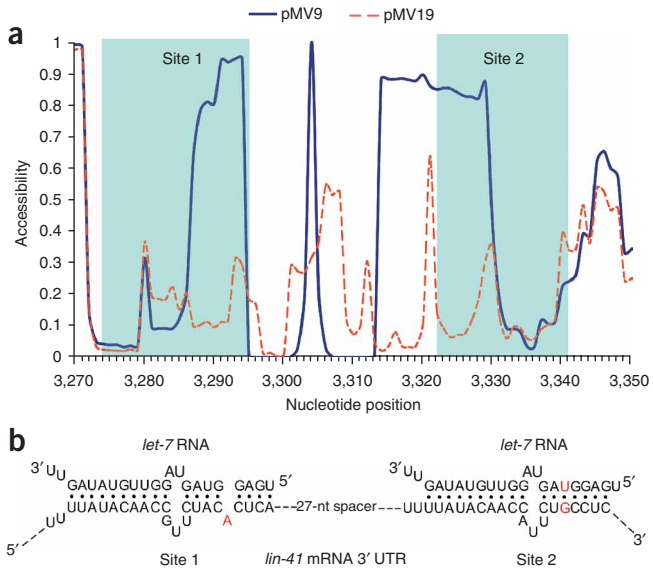


**Figure 2** Target accessibility profiling by S fold. (**a**) Sfold accessibility profiles for the region containing the two *let-7*–binding sites (shaded) in the *lin-41* 3′ UTR mutant construct pMV9 and in pMV19. For pMV9, which is sensitive to repression by *let-7* (ref. 15), the 3′ end of target site 1 is highly accessible for nucleation, as is the 5′ end of site 2. For the mutant 3′ UTR construct pMV19, which is insensitive to repression, neither site is accessible. (**b**) Neither site 1 nor site 2 satisfies the criteria for full Watson-Crick base-pairing across the 5′ seed region of the miRNA; a bulged A in site 1 and a wobble G•U base pair in site 2 are indicated in red.

the indicated nucleotide are all single-stranded, as predicted by Sfold[21,22]. A window length of 4 nt was chosen for this study, as use of this length led to a good correlation between our accessibility profile predictions and data from antisense experiments on rabbit β-globin[21]. Furthermore, a 4-nt accessible block can be sufficient for the nucleation step of the hybridization between a complementary nucleic acid molecule and its target[23]. Finally, we observed a better correlation with *lin-41* data for the block size of 4 nt than for block sizes of 1, 2 or 3 nt (**Supplementary Table 1**).

The accessibility profiles of pMV9 and pMV19 (**Fig. 2a**) are in substantial agreement with the sensitivities of the two constructs to repression by *let-7*. For pMV9, which is repressed by *let-7*, the 3′ end of target site 1 is highly accessible for nucleation, as is the 5′ end of site 2. In contrast, for pMV19, which is inactive for repression, neither *let-7* site is accessible. Notably, for pMV9, nucleation of hybridization is predicted to occur at the miRNA 5′ end (3′ end of the target site) for site 1, but at the miRNA 3′ end (5′ end of the target site) for site 2.

## Energetic analysis of miRNA-target hybridization

To model the potential function of each *let-7*–complementary site in the *lin-41* mutant 3′ UTR reporter constructs, we computed the nucleation potential ($\Delta G_N$) for each site (see Methods). For sites with favorable nucleation potential (where the nucleation potential overcomes the initiation threshold $\Delta G_{initiation}$, so that ($\Delta G_N + \Delta G_{initiation}$) $< 0$ kcal mol$^{-1}$, we calculated the total energy change ($\Delta G_{total}$) for the hybridization to each site, then $\sum\Delta G_{total}$ for multiple sites (see Methods). A miRNA–3′ UTR interaction is predicted to be functional if it has a favorable $\sum\Delta G_{total}$. To estimate a practical threshold for favorable $\sum\Delta G_{total}$, we considered the performance of various threshold values in accommodating the experimental results for *lin-41* 3′ UTR mutants. We found that a $\sum\Delta G_{total}$ of $-10$ kcal mol$^{-1}$ or less seems to separate efficient interactions from inefficient interactions (data not shown).

The energetic predictions for *lin-41* 3′ UTR reporter constructs are in good agreement with the reported repression sensitivities (**Table 1**). In general, repression by *let-7* was substantial for constructs with favorable nucleation sites and large negative hybridization energies. In contrast, constructs with weak or absent repression by *let-7* generally had poor nucleation sites, small hybridization energies or both. For example, despite a substantial number of open nucleotides in the pMV16 and pMV17 3′ UTRs (**Supplementary Table 1**), their *let-7*–complementary sites[15] are nevertheless predicted to be lacking in favorable nucleation potential, and overall the pMV16 and pMV17 3′ UTRs have small hybridization energies. This analysis predicts that pMV5 should not interact with *let-7*, yet this construct did show weak repression *in vivo*, and pMV5 and pMV12 were given the same ranking for repression sensitivity[15]. However, the measured level of repression for pMV5 is lower than that for pMV12 (ref. 15), which is consistent with the rankings of these two constructs by our energetic analysis (**Table 1**).

An accurate accounting for the *lin-41* 3′ UTR data requires application of both a nucleation-potential threshold (specified via an initiation energy) and a structure-based hybridization energy calculation. If we ignore the effects of target structure and nucleation by using only the energy gain from hybridization, $\Delta G_{hybrid}$, in the calculation, we find that all ten constructs have large negative $\Sigma\Delta G_{hybrid}$ values, even though five of them were observed to be insensitive to *let-7 in vivo* (**Table 1**). Even ignoring just the initiation energy, by setting it to 0.0 kcal mol$^{-1}$ instead of 4.09 kcal mol$^{-1}$ (the value empirically determined in ref. 24), results in rather poor predictions for pMV19, pMV16 and pMV17 (**Table 1**). Although inclusion of a nucleation-potential filter is crucial, the predictions are robust over a range of initiation energies from 4.0 kcal mol$^{-1}$ to 5.5 kcal mol$^{-1}$ (**Supplementary Fig. 1** online). For example, for 5.2 kcal mol$^{-1}$ (the other published value for initiation energy[23]), these predictions for the *lin-41* 3′ UTR were the same as for 4.09 kcal mol$^{-1}$. Although these two published values for the initiation energy generally performed similarly when applied as the nucleation-potential thresholds, we focus on results for the value of 4.09 kcal mol$^{-1}$ because of its somewhat superior predictive consistency for certain other miRNA-target interactions (see below).

## miRNA-target interactions in *C. elegans* and *Drosophila*

To examine the general applicability of the two-step hybridization model, we considered other miRNA-target interactions that had been predicted previously for *C. elegans* or *D. melanogaster*, and for which experimental validation had been published (**Supplementary Table 2** online). For some of these interactions, the validation experiments include tests of genetic epistasis (where targeting is validated by observing that loss of function of the target counteracts the effects of loss of function of the miRNA). For other interactions, validation is typically based on reporter gene expression. For each miRNA-target pair, we calculated $\sum\Delta G_{total}$ and compared it with the corresponding value computed for ten 'randomers', random control sequences generated by dinucleotide shuffling of the miRNA sequence with Dishuffle[25]. To statistically test whether the average $\sum\Delta G_{total}$ for the authentic miRNA sequences is lower than that for the randomers, we performed the one-sided, unequal-variance *t*-test and the nonparametric Wilcoxon rank-sum test.

We found that the degree of correspondence between our calculations and the published validation results depends somewhat on the

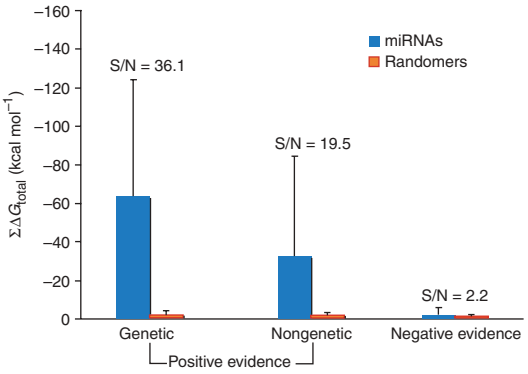


**Figure 3** The average $\sum\Delta G_{total}$ for miRNAs compared with that calculated for randomers, for positive miRNA-target interactions supported either by genetic epistasis evidence or by nongenetic evidence, and for the set of 12 putative *lsy-6*–target pairs predicted by conserved seed matching but having negative interactions *in vivo*[27] (**Table 2**). S/N, signal-to-noise ratio. Error bars represent s.d. for the three sets of interactions.

nature of the experimental supporting evidence. For the group of 11 predicted interactions for which the supporting evidence includes conventional genetic epistasis experiments, we calculated that all 11 should interact effectively, on the basis of $\sum\Delta G_{total}$ (**Supplementary Table 2** online). For 9 of these 11 miRNA-target pairs, a particularly strong interaction is predicted on the basis of hybridization energies. The mean $\sum\Delta G_{total}$ for the miRNAs (−63.70 kcal mol$^{-1}$) is statistically distinct from the mean for the randomers (−1.76 kcal mol$^{-1}$; **Fig. 3**): the *t*-test returned a highly significant $P = 3.55 \times 10^{-3}$, and the corresponding Wilcoxon rank-sum test yielded $P = 1.11 \times 10^{-12}$. The signal-to-noise ratio, defined here as the ratio of the average $\sum\Delta G_{total}$ for the miRNAs to that for the randomers, is 36.1.

There are two cases of relatively small hybridization energies: namely, the *lin-4*–*lin-28* pair and the *miR-273*–*die-1* pair. It has been reported that the repression of *lin-28* by *lin-4* may occur in conjunction with an additional '*lin-4*–independent circuit'[26], suggesting that effective repression of *lin-28* by *lin-4* requires additional factor(s). Similarly, the putative *miR-273*–*die-1* interaction may also require other unknown factors. For the initiation energy of 5.20 kcal mol$^{-1}$, *let-7* does not have a single favorable nucleation site on the *lin-28* 3′ UTR. Thus, we use 4.09 kcal mol$^{-1}$ in the subsequent analyses because of its better predictive consistency and higher inclusiveness of potential sites.

Our model predicts a functional interaction for 27 of 39 interactions whose *in vivo* efficacy is supported mainly by reporter gene tests, but for which genetic epistasis evidence is not available. Thus, the structure-based model performs less precisely in accounting for the experimental results for interactions tested by nongenetic criteria (69% true positive), compared with interactions validated genetically (100% true positive). Nevertheless, the average $\sum\Delta G_{total}$ for this set of miRNA-target interactions is statistically significant compared with that for randomers (**Fig. 3**): the average $\sum\Delta G_{total}$ for the miRNAs is −32.51 kcal mol$^{-1}$ and the average for the randomers is −1.67 kcal mol$^{-1}$, a difference of −30.84 kcal mol$^{-1}$, yielding $P = 3.89 \times 10^{-3}$ by the *t*-test and $P = 4.50 \times 10^{-7}$ by the Wilcoxon rank-sum test, and the signal-to-noise ratio is 19.5.

To examine the success of our model in accounting for negative results from experimental validation tests, we examined a set of 12 putative targets of the *C. elegans lsy-6* miRNA that had been predicted based on conserved seed matches, but for which *in vivo* tests did not validate the predicted interactions[27]. For only one of these 12 putative *lsy-6* targets (8%) was a functional interaction (albeit relatively weak) predicted by our structure-based model (**Table 2**). For 11 of the 12 putative targets (92%), we did not predict a functional interaction (**Table 2**), in complete agreement with the experimental results for these targets[27]. In contrast, if we ignore the effects of target structure and nucleation by using just $\sum\Delta G_{hybrid}$, functional interactions are predicted for 10 of the 12 negative cases (**Table 2**). This strongly suggests that mRNA secondary structure is a major factor behind the insensitivity of these putative targets to *lsy-6 in vivo*. Note that *cog-1*, which has previously been confirmed by various functional criteria to be regulated by *lsy-6* (refs. 27,28), is predicted by our model to have an effective interaction with *lsy-6* (**Table 2**). A comparison of $\Sigma\Delta G_{total}$ between the authentic miRNAs and randomers for these 12 targets yielded $P = 0.2364$ for the *t*-test and 0.9117 for the Wilcoxon rank-sum test, and the signal-to-noise ratio is merely 2.2 (**Fig. 3**); for the 11 false positive predictions by seed matches (see **Table 2**), there is no appreciable difference between the signal and the noise according to our model (for this subset, the $P = 0.559$ for the *t*-test and 0.964 for the Wilcoxon rank-sum test) (**Supplementary Table 2**).

### *In vivo* testing of newly designed *lin-41* 3′ UTR mutants

To further test for effects of target structure and accessibility on miRNA efficacy—in this case, on repression of *lin-41* by *let-7* in *C. elegans*—we carried out *in vivo* experiments using six new *lacZ* reporter transgenes containing rationally designed *lin-41* 3′ UTR mutations. For four of the new constructs, pVT701, pVT702, pVT704 and pVT705, the corresponding *lin-41* mutant 3′ UTR sequence contains one copy of site 1 and one copy of site 2 separated by a 27-nt spacer sequence. For pVT701 and pVT702, the wild-type 27-nt spacer was mutated to preserve the accessibility of both sites (a target site is considered highly accessible if it has both favorable nucleation potential and favorable $\Delta G_{total}$). For pVT704 and pVT705, the spacer was designed to reduce the accessibility of the sites. Constructs pVT712 and pVT713 contain, respectively, three copies of sites 1 or 2, in sequence contexts designed to permit good accessibility.

These new constructs, along with pMV9 and pMV19, were tested in transgenic worms (see Methods). There was substantial agreement

**Table 2 miRNA-target interaction energy predicts sensitivity of seed-match targets to repression by *lsy-6***

| Target[a] | Repression sensitivity[b] | $\sum\Delta G_{total}$[c] (kcal mol$^{-1}$) | $\sum\Delta G_{hybrid}$[c] (kcal mol$^{-1}$) |
|---|---|---|---|
| cog-1 | + | −39.89 (+)[d] | −75.9 (+)[b] |
| ZK637.13 | – | −13.86 (+) | −17.3 (+) |
| C02B8.4 | – | −4.22 (–) | −62.3 (+) |
| F55G1.1 | – | 0.00 (–) | 0.0 (–) |
| C48D5.2a | – | −0.04 (–) | −48.4 (+) |
| F59A6.1 | – | −2.72 (–) | −63.2 (+) |
| F40H3.4 | – | 0.00 (–) | −17.0 (+) |
| T05C12.8 | – | 0.00 (–) | −45.0 (+) |
| C27H6.3 | – | −0.06 (–) | −14.8 (+) |
| T23E1.1 | – | 0.00 (–) | −14.5 (+) |
| T14G12.2 | – | 0.00 (–) | −15.0 (+) |
| T20G5.9 | – | 0.00 (–) | −16.5 (+) |
| R07E3.5 | – | 0.00 (–) | 0.0 (–) |

[a]Ref. 27. [b]For $\Delta G_{hybrid}$, an interaction is predicted to be functional (+) if $\sum\Delta G_{hybrid} \leq -14$ kcal mol$^{-1}$ (see **Table 1**) and nonfunctional (–) otherwise. [c]Calculated as in **Table 1** for sites identified from *lsy-6*–3′ UTR alignment (see Methods). [d]For $\Delta G_{total}$, a *lsy-6*–target interaction is predicted to be functional (+) if $\sum\Delta G_{total} < -10$ kcal mol$^{-1}$ and nonfunctional (–) otherwise.

**Table 3 miRNA-target interaction energy predicts sensitivity of novel mutant *lin-41* 3′ UTR reporters to repression by *let-7***

| 3′ UTR construct[a] | Features of construct | $\sum\Delta G_{total}$[b] (kcal mol$^{-1}$) | $\sum\Delta G_{hybrid}$[b] (kcal mol$^{-1}$) | Observed repression[c] | Ratio % β-gal (adults/larvae) | Worm lines (*n*) |
|---|---|---|---|---|---|---|
| pMV9 | Wild-type sites and spacer | −43.44 (+)[d] | −57.0 (+)[d] | +++ | 0.39 ± 0.03 | 5 |
| pMV19 | Spacer mutation | −8.31 (−) | −57.0 (+) | − | 1.01 ± 0.25 | 3 |
| pVT701 | Spacer mutation | −43.02 (+) | −57.0 (+) | +++ | 0.38 ± 0.16 | 3 |
| pVT702 | Spacer mutation | −43.36 (+) | −57.0 (+) | ++ | 0.54 ± 0.06 | 3 |
| pVT704 | Spacer mutation | −6.35 (−) | −57.0 (+) | − | 0.88 ± 0.29 | 3 |
| pVT705 | Spacer mutation | −7.21 (−) | −57.0 (+) | + | 0.78 ± 0.15 | 2 |
| pVT712 | Three copies of site 1 | −54.79 (+) | −83.3 (+) | +++ | 0.42 ± 0.10 | 4 |
| pVT713 | Three copies of site 2 | −73.23 (+) | −87.0 (+) | ++ | 0.53 ± 0.26 | 2 |

[a]pMV9 and pMV19 were previously designed[15]; pVT701, pVT702, pVT703, pVT704, pVT705, pVT712 and pVT713 are new constructs designed for this study. [b]Calculated as in **Table 1**. [c]Empirical ranking of repression is based on the ratio (β-gal adults)/(β-gal larvae), adapting the published convention[15] (see **Table 1**), where the level of repression shown by pMV9 is scored as +++, the lack of repression for pMV19 as − and intermediate degrees of repression as + or ++. pVT704 was judged to be similar to pMV19; pVT701 and pVT712 are similar to pMV9. [d]Interactions are scored as functional (+) or nonfunctional (−) as in **Table 1**.

between predicted accessibility of the *let-7* binding sites and the observed *in vivo* temporal repression of the corresponding transgenes (**Table 3**). For pVT701, pVT702, pVT712 and pVT713, the expression of the *lacZ* reporter constructs is substantially reduced in adult worms, as compared with the expression of pMV19, which has two poorly accessible sites. For pVT704 and pVT705, the expression of the *lacZ* reporter constructs is substantially increased in adults, as compared with the expression of pMV9. Furthermore, the correlation between the *in vivo* repression and $\sum\Delta G_{total}$ for these constructs is 0.8083, with a significant *P*-value of 0.0152. In a linear regression analysis using $\sum\Delta G_{total}$ of the binding sites to predict the degree of the repression (**Fig. 4**), $R^2$ is 0.6534 and the *P*-value for the predictor is 0.0152. The underlying assumptions for a linear regression analysis were found to hold well, and an alternative weighted regression analysis yielded very similar results (see **Supplementary Data** online). The $R^2$ indicates that our model can account for about two-thirds of the variability in the repression sensitivity. Other factors are probably responsible for the remaining one-third of the variability, as well as for the appreciable difference in activity between pMV19 and pVT705, two constructs with comparable $\sum\Delta G_{total}$ values but different empirical rankings (**Table 3**).

Our results indicate that constructs with multiple copies of a single site (pVT712 and pVT713) can be repressed by *let-7*, provided that the sites are structurally accessible. In this regard, the results for pVT712 and pVT713 contrast with the activities of similar single-site constructs, pMV16 and pMV17, reported previously[15]. A major difference among these constructs is that adjacent sites in pVT712 and pVT713 are separated by the 27-nt spacer that is normally between site 1 and site 2 of the wild-type *lin-41* 3′ UTR, whereas pMV16 and pMV17 contain a 4-nt spacer. Notably, pVT712 and pVT713 are predicted to contain accessible sites, whereas pMV16 and pMV17 are predicted to contain inaccessible sites. We interpret these results to indicate that what is most essential for *let-7* activity is the presence of structurally accessible target sites, not the specific tandem configuration of site 1 and site 2 or the sequence content of the spacer between adjacent sites. Finally, $\sum\Delta G_{hybrid}$ cannot account for the weak or absent repression of pMV19, pVT704 and pVT705, as $\sum\Delta G_{hybrid}$ does not take into account the effects of target structure and nucleation (**Table 3**).

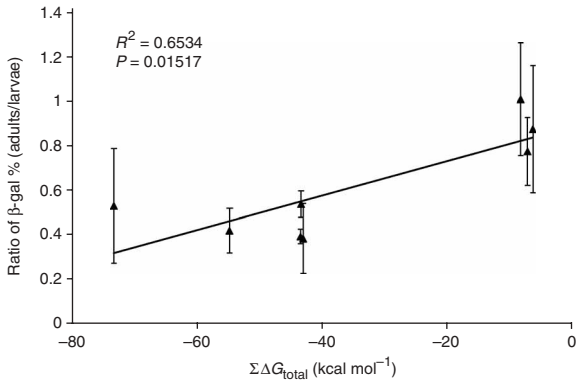


**Figure 4** Linear regression prediction of *in vivo* repression sensitivity (measured by β-galactosidase (β-gal) expression ratios in adult and larval stages) by the $\sum\Delta G_{total}$ for the *lin-41* 3′ UTR mutant constructs (see **Table 3**). The correlation between repression sensitivity and $\sum\Delta G_{total}$ is 0.8083. Error bars represent s.d. for multiple worm lines (see **Table 3**).

## DISCUSSION

In this study, we examined the potential effects of the secondary structure of a target mRNA on the mRNA's sensitivity to regulation by miRNAs. We analyzed published data on the *in vivo* activity of *C. elegans* reporter genes containing modified *lin-41* 3′ UTR sequences, as well as other experimentally studied miRNA-target interactions from *C. elegans* and *D. melanogaster*. We found that a two-step model for miRNA-target hybridization that incorporates target structure as a governing principle reliably accounts for these published data. We also tested *in vivo* a set of new *C. elegans lin-41* 3′ UTR mutants that were specifically designed to change the predicted accessibility of *let-7*–complementary sites without changing the sequences of the sites themselves. We found that our model accounts for the relative sensitivities of these mutants to *let-7* repression, further supporting the idea that target structure exerts a potent effect on miRNA activity.

We model the binding of an miRNA to a complementary target site as a two-step process. In the first step, the miRNA nucleates base-pairing with a block of four contiguous, single-stranded nucleotides of the target. In the second step, the miRNA completes base-pairing with the mRNA, accompanied by the disruption of target secondary structure in the region of hybridization. Computational implementation of each of these steps requires an accurate representation of the secondary structure of the mRNA target, followed by the identification of open blocks for nucleation, and then accurate accounting for both the free-energy gain and energy loss in the miRNA-target base-pairing transaction. We represent the secondary structure of a particular mRNA by a statistical sample of probable secondary structures

generated by Sfold. Compared with the conventional MFE approach, Sfold has been shown to make improved structure predictions and to better represent likely structure populations for mRNAs[29,30]. We also compared the performance of our model implemented with Sfold to the performance of models based on (i) the MFE structure, (ii) a heuristic set of suboptimal folds or (iii) a set of the lowest-energy structures. We found that the Sfold implementation performed best in accounting for the experimentally observed *in vivo* sensitivities of *C. elegans lin-41* 3′ UTR mutants (**Supplementary Table 3** online).

We found that accurate modeling of miRNA-target interactions requires incorporation of an initiation-energy penalty (or nucleation-potential threshold) for overcoming the energy threshold required for bimolecular initiation of RNA-RNA base-pairing. When we omitted the nucleation-potential threshold, we found that differences in the hybridization energies alone were not sufficient to account for the *lin-41* 3′ UTR reporter data. In contrast, when the nucleation-potential threshold was introduced, the model was able to more consistently account for the experimentally tested miRNA-target interactions. Our model seems to generally accommodate values for the nucleation-potential threshold in the range encompassed by the two published empirical values for the initiation energy, 4.09 kcal $mol^{-1}$ (ref. 24) and 5.20 kcal $mol^{-1}$ (ref. 23). However, the value of 4.09 kcal $mol^{-1}$ seemed to perform somewhat better than 5.2 kcal $mol^{-1}$ in our analyses, and so, for the purposes of this study, we considered 4.09 kcal $mol^{-1}$ as the best currently available approximation for the miRNA-target nucleation-potential threshold *in vivo*.

Our approach is limited to the prediction of RNA secondary structures; it does not incorporate the potential effects of RNA-binding proteins and other associated factors, including components of the RNA-induced silencing complex (RISC), that could affect the secondary structure and accessibility of an mRNA target site or the energetics of its interaction with miRNAs *in vivo*. However, although mRNA secondary structure is unlikely to be the only factor influencing target accessibility, our results suggest that it is an important factor in most cases that we examined and therefore that it probably contributes to accessibility in most miRNA-target interactions. The degree to which accessibility is also governed by associated proteins may depend on the particular miRNA-target interaction, the cellular context or both. Consistent with our findings that target structure can affect miRNA activity, previous studies of factors governing the efficiency of short interfering RNAs have suggested the importance of target structure[31,32].

Some previous studies of miRNA-target interactions have considered the role of mRNA secondary structure. It has been proposed that in vertebrates, miRNAs preferentially target 3′ UTRs with less stable structures[33]; however, in another study of mammalian miRNAs, an effect of target structure was not detected[34]. We found that the frequency of open blocks of 4 nt is better correlated with *let-7*–*lin-41* data than is the frequency of 3-nt blocks, which was previously considered[18]. The accuracy of those previous analyses of target structure was probably compromised by a reliance on predictions of the MFE structures.

A number of existing algorithms for miRNA target searching are based primarily on the identification of conserved continuous Watson-Crick base-pairing in the 5′ seed region (nucleotides 2–8) of the miRNA[6,10,35,36]. We believe that a strength of our structure-based approach is that it does not specifically stipulate the positions of hybridization-nucleating bases in the miRNA or the final configuration of base-pairing between the miRNA and the target. Thus, our approach should encompass interactions that would be identified by algorithms based on the 5′ seed model and also identify interactions that do not fit the strict 5′ seed pairing criteria. The 5′ seed approach has been successful in identifying many bona fide functional miRNA-target interactions, probably reflecting an important role for seed pairing in general; however, the range of permissible seed-target base-pairing configurations is unknown, and evidence indicates that a continuous Watson-Crick helix for the seed region is not always necessary or sufficient for effective repression[6,15,27,37]. Relaxation of the seed base-pairing requirement would encompass more authentic targets, but this leads to a lower signal-to-noise ratio and thus a greater chance of false positive predictions[13,36]. Functional interactions may also include cases where nucleation occurs outside of the 5′ seed region. The results of our accessibility profiling analysis suggest that nucleation can occur in the 3′ region of the miRNA, at least for the *let-7* miRNA and site 2 of the *lin-41* 3′ UTR (**Fig. 2a**).

For a set of *C. elegans* and *D. melanogaster* miRNA-target interactions predicted by seed pairing models, our structure-based model performed well, particularly for those interactions that had negative validation tests (**Table 2**). This suggests that sites meeting 5′ seed criteria can be functional if the sites are also structurally accessible, whereas other 5′ seed sites can be nonfunctional owing to structural inaccessibility. Thus, our findings provide one explanation for the observation that some targets predicted purely by seed pairing models test positively in validation tests, and others test negatively. Indeed, the lack of repression of certain *lin-41* mutant constructs by *let-7* is explained by our model, but not by other current approaches (**Tables 1** and **3**). Thus, we anticipate that the structure-based model can substantially improve the specificity of target predictions based on 5′ seed pairing or other algorithms such as RNAhybrid, by reducing the rate of false positive predictions.

It is currently not possible to accurately assess the relative rates of false negative predictions by our method compared with previous target-prediction approaches. This would require systematic testing of targets not predicted by each method, and such data are not yet available. However, we did find that our method confirmed 100% of the set of interactions supported by genetic epistasis tests (**Fig. 3** and **Supplementary Table 2**). The agreement with experimental data was less (69%) for predictions that had been validated only by nongenetic criteria (**Fig. 3** and **Supplementary Table 2**), suggesting a false negative rate of 31%. Nevertheless, we observed a highly significant difference (see Results) between miRNAs and randomers for both the genetic set and the nongenetic set, and the difference is much greater for the genetic set (**Fig. 3** and **Supplementary Table 2**). There are a number of possible explanations for our false negative predictions. First, the interactions that are currently supported only by nongenetic tests are generally less extensively studied than the other set, and some of these could become reclassified as negative after further testing. Second, some of the negative predictions could be erroneous on account of misannotated mRNA sequences, which could lead to inaccurate structural modeling of the target mRNA in such cases. Perhaps the most intriguing explanation would be the possible existence of a subclass of mRNA targets that are conditionally accessible to miRNA binding—where target accessibility could be modulated *in vivo* by RNA-binding proteins and/or RISC components to counteract otherwise intrinsically closed local target secondary structure.

Nearly all of the miRNA-target interactions that we examined here had been predicted based on phylogenetic conservation of the predicted base-pairing. Conservation of miRNA target sites in orthologous genes is a powerful criterion for identifying potentially functional targets. However, as we do not fully understand what

factors exert selective pressure on conserved 3′ UTR sequences, it is possible that some conserved sites are authentic, whereas others are not functional *in vivo*. Our findings suggest that inclusion of target-accessibility criteria could reduce the frequency of such false positive predictions from conserved sequences. Conservation-based algorithms can also be prone to false negative predictions in cases where sites that are not conserved in sequence alignments are nevertheless demonstrably functional *in vivo*[34] or in cases of poorly conserved human or virally encoded miRNAs[2,38]. The structure-based model described here, combined other known features of functional miRNA target sites, including aspects of the 5′ seed pairing model, could provide the basis for the reliable identification of both conserved and non-conserved miRNA regulatory sites in genome-wide searches.

## METHODS

**Prediction of mRNA secondary structures.** Accurate modeling of mRNA structure must take into account the principle that a specific mRNA molecule probably exists *in vivo* in dynamic equilibrium among a population of structures. Sfold is used for this purpose, because its algorithm generates a statistically representative sample from the Boltzmann-weighted ensemble of RNA secondary structures for the RNA[22]. Samples of 1,000 structures can faithfully and reproducibly characterize structure ensembles of enormous sizes[22,30]. Therefore, sampling estimates of structural features for a particular RNA are statistically reproducible from one sample to another. Additional information on RNA folding can be found in the **Supplementary Methods** online.

**Energetic modeling of two-step hybridization.** On the basis of studies on nucleic acid hybridization[23,39–42], we modeled the miRNA-target hybridization as a two-step process: (i) nucleation (the rate-limiting step), involving four consecutive complementary nucleotides in the two RNAs, and (ii) elongation of the hybrid to form a stable intermolecular duplex (**Fig. 1**). An miRNA itself is unlikely to have stable secondary structure owing to its short length. Therefore, for the identification of potential nucleation sites for an miRNA on a structured target, the miRNA itself was assumed to be unstructured.

**Computation of nucleation potential.** Nucleation of base-pairing between two nucleic acid molecules requires a gain in free energy from base-pairing at the nucleation site that is greater than the energy cost for the translational and rotational entropy loss when two RNA strands are fixed in a conformation by intermolecular base-pairing[24]. This latter energy cost—that is, the penalty of 'initiation energy'—presents an energy threshold for nucleation. Therefore, we determined the nucleation competence of a putative nucleation site by computing the free energy gained from base-pairing at that nucleation site (the 'nucleation potential', $\Delta G_N$) and comparing it to a chosen value of the initiation-energy threshold ($\Delta G_{initiation}$). There are two published values for the initiation-energy threshold, 4.09 kcal mol$^{-1}$ (ref. 24) and 5.2 kcal mol$^{-1}$ (ref. 23). We considered both values in our analyses. Sites with nucleation potentials that are favorable compared with the initiation-energy threshold are considered to be competent for nucleation.

In our model, we assume that four unpaired bases in the target can initiate nucleation[23], and we therefore identified possible miRNA binding sites defined by miRNA-mRNA complementary matches of at least four contiguous nucleotide matches anywhere in the miRNA. (For *lin-41* mutants, we considered the previously described complementary sites[15] as possible *let-7*–binding sites). We then computed the nucleation potential of a possible miRNA-binding site by identifying the particular open (single-stranded) 4-nt block within the site that would form the most stable 4-base-pair duplex with the miRNA. To identify open blocks of nucleotides in the target site, we used a sample of 1,000 structures for the target RNA predicted by Sfold. We considered each of the 1,000 structures in the sample separately and calculated the average nucleation potential over the whole set of structures in the sample (see **Supplementary Methods**). Nucleation is considered favorable if the calculated average nucleation potential can overcome the initiation energy threshold—that is, $\Delta G_N + \Delta G_{initiation} < 0$ kcal mol$^{-1}$.

**Computation of total hybridization energy.** After nucleation of miRNA-target binding, a hybrid elongation process follows, involving breakage of intramolecular base pairs in the target to accommodate the formation of an optimal intermolecular miRNA-target duplex. Because nucleation is the rate-limiting step for nucleic acid hybridization, we calculated the total energy change for the hybridization ($\Delta G_{total}$) only for those sites on the target with favorable nucleation potential ($\Delta G_N + \Delta G_{initiation} < 0$ kcal mol$^{-1}$). $\Delta G_{total} = \Delta G_{hybrid} - \Delta G_{disruption}$, where $\Delta G_{disruption}$ is the energy for the disruption of the intramolecular base pairs that involve binding-site nucleotides, and $\Delta G_{hybrid}$ is the energy gain owing to the complete intermolecular hybridization of the miRNA with the binding site (**Fig. 1b**). The complete hybrid may contain one or more loops (**Fig. 1b**) and is assumed to be in the conformation of the lowest free energy. Therefore, we calculated $\Delta G_{hybrid}$ using RNAhybrid, which finds the most stable miRNA-target hybrid conformation.

To calculate $\Delta G_{disruption}$, we adopted the simplifying assumption that the binding of an miRNA to a relatively much longer mRNA should cause a local structural alteration at the target site but not longer-range effects on overall target secondary structure. Specifically, we defined local structural alteration as the breakage of only those target intramolecular base pairs that must be broken to permit formation of the miRNA-target duplex predicted by RNAhybrid. $\Delta G_{disruption}$ was calculated as the energy difference between $\Delta G_{before}$, the free energy of the original mRNA structure, and $\Delta G_{after}$, the free energy of the new locally altered structure ($\Delta G_{disruption} = \Delta G_{before} - \Delta G_{after}$). We calculated $\Delta G_{before}$ as the average energy of the original 1,000 structures predicted by Sfold and $\Delta G_{after}$ as the average energy of all of the 1,000 locally altered structures. Therefore, under the assumption of local disruption, the calculations do not require refolding of the rest of the target sequence.

To model the cooperative effects of multiple binding sites on a single 3′ UTR, we computed $\sum \Delta G_{total}$ for sites with favorable nucleation potentials. This energetic additivity implicitly assumes independent contributions from multiple sites. However, it is possible that occupancy at one site could alter the accessibility for another site. Modifications of the above energetic calculations will be required for modeling potential structure-mediated interactions among multiple sites.

***In vivo* testing of *lin-41* 3′ UTR mutants.** Mutant constructs were generated and tested as described[15] (see also **Supplementary Methods** and **Supplementary Tables 4** and **5** online).

**Software availability.** We have developed STarMir, a new application module for the Sfold software. STarMir implements the approach described here to perform energy calculations for miRNA-target hybridization and is available through the Sfold web server at http://sfold.wadsworth.org.

*Note: Supplementary information is available on the Nature Structural & Molecular Biology website.*

ACKNOWLEDGMENTS

We acknowledge the Computational Molecular Biology and Statistics Core at the Wadsworth Center for providing computing resources. This work was supported in part by US National Science Foundation grants DMS-0200970 and DBI-0650991 and US National Institutes of Health grant GM068726 to Y.D., and by US National Institutes of Health grants GM34028 and GM066826 to V.A. We thank F. Slack of Yale University for gifts of plasmids, and A. Lee, G. Ambros and members of the Ambros lab for technical help and advice.

AUTHOR CONTRIBUTIONS

D.L. and Y.D. designed the algorithm and performed computational modeling of RNA structure and thermodynamics, R.L. and V.A. analyzed *lin-41* reporter genes in *C. elegans*, P.W. performed computational modeling and C.Y.C. developed the web interface.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
2. Bentwich, I. *et al.* Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.* **37**, 766–770 (2005).

3. Boehm, M. & Slack, F. A developmental timing microRNA and its target regulate life span in *C. elegans*. *Science* **310**, 1954–1957 (2005).
4. Calin, G.A. & Croce, C.M. MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res.* **66**, 7390–7394 (2006).
5. Cuellar, T.L. & McManus, M.T. MicroRNAs and endocrine biology. *J. Endocrinol.* **187**, 327–332 (2005).
6. Brennecke, J., Stark, A., Russell, R.B. & Cohen, S.M. Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85 (2005).
7. Sethupathy, P., Corda, B. & Hatzigeorgiou, A.G. TarBase: a comprehensive database of experimentally supported animal microRNA targets. *RNA* **12**, 192–197 (2006).
8. Jones-Rhoades, M.W. & Bartel, D.P. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* **14**, 787–799 (2004).
9. Lai, E.C. Predicting and validating microRNA targets. *Genome Biol.* **5**, 115 (2004).
10. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
11. Doench, J.G. & Sharp, P.A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
12. Kiriakidou, M. *et al.* A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.* **18**, 1165–1178 (2004).
13. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. & Burge, C.B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
14. Rajewsky, N. microRNA target predictions in animals. *Nat. Genet.* **38** Suppl, S8–S13 (2006).
15. Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K. & Slack, F.J. The *C. elegans* microRNA *let-7* binds to imperfect *let-7* complementary sites from the *lin-41* 3′UTR. *Genes Dev.* **18**, 132–137 (2004).
16. Vella, M.C., Reinert, K. & Slack, F.J. Architecture of a validated microRNA:target interaction. *Chem. Biol.* **11**, 1619–1623 (2004).
17. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
18. Robins, H., Li, Y. & Padgett, R.W. Incorporating structure to predict microRNA targets. *Proc. Natl. Acad. Sci. USA* **102**, 4006–4009 (2005).
19. Rehmsmeier, M., Steffen, P., Hochsmann, M. & Giegerich, R. Fast and effective prediction of microRNA-target duplexes. *RNA* **10**, 1507–1517 (2004).
20. Ding, Y., Chan, C.Y. & Lawrence, C.E. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.* **32**, W135–W141 (2004).
21. Ding, Y. & Lawrence, C.E. Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucleic Acids Res.* **29**, 1034–1046 (2001).
22. Ding, Y. & Lawrence, C.E. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.* **31**, 7280–7301 (2003).
23. Hargittai, M.R., Gorelick, R.J., Rouzina, I. & Musier-Forsyth, K. Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J. Mol. Biol.* **337**, 951–968 (2004).
24. Xia, T. *et al.* Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **37**, 14719–14735 (1998).
25. Clote, P., Ferre, F., Kranakis, E. & Krizanc, D. Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. *RNA* **11**, 578–591 (2005).
26. Seggerson, K., Tang, L. & Moss, E.G. Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene lin-28 after translation initiation. *Dev. Biol.* **243**, 215–225 (2002).
27. Didiano, D. & Hobert, O. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat. Struct. Mol. Biol.* **13**, 849–851 (2006).
28. Johnston, R.J. & Hobert, O. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* **426**, 845–849 (2003).
29. Ding, Y., Chan, C.Y. & Lawrence, C.E. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA* **11**, 1157–1166 (2005).
30. Ding, Y., Chan, C.Y. & Lawrence, C.E. Clustering of RNA secondary structures with application to messenger RNAs. *J. Mol. Biol.* **359**, 554–571 (2006).
31. Overhoff, M. *et al.* Local RNA target structure influences siRNA efficacy: a systematic global analysis. *J. Mol. Biol.* **348**, 871–881 (2005).
32. Schubert, S., Grunweller, A., Erdmann, V.A. & Kurreck, J. Local RNA target structure influences siRNA efficacy: systematic analysis of intentionally designed binding regions. *J. Mol. Biol.* **348**, 883–893 (2005).
33. Zhao, Y., Samal, E. & Srivastava, D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* **436**, 214–220 (2005).
34. Farh, K.K. *et al.* The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science* **310**, 1817–1821 (2005).
35. Enright, A.J. *et al.* MicroRNA targets in *Drosophila*. *Genome Biol.* **5**, R1 (2003).
36. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat. Genet.* **37**, 495–500 (2005).
37. Miranda, K.C. *et al.* A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell* **126**, 1203–1217 (2006).
38. Cullen, B.R. Viruses and microRNAs. *Nat. Genet.* **38** Suppl, S25–S30 (2006).
39. Paillart, J.C., Skripkin, E., Ehresmann, B., Ehresmann, C. & Marquet, R. A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl. Acad. Sci. USA* **93**, 5572–5577 (1996).
40. Reynaldo, L.P., Vologodskii, A.V., Neri, B.P. & Lyamichev, V.I. The kinetics of oligonucleotide replacements. *J. Mol. Biol.* **297**, 511–520 (2000).
41. Milner, N., Mir, K.U. & Southern, E.M. Selecting effective antisense reagents on combinatorial oligonucleotide arrays. *Nat. Biotechnol.* **15**, 537–541 (1997).
42. Kolb, F.A. *et al.* Bulged residues promote the progression of a loop-loop interaction to a stable and inhibitory antisense-target RNA complex. *Nucleic Acids Res.* **29**, 3145–3153 (2001).