nature genetics

The role of site accessibility in microRNA target recognition

Michael Kertesz[1,3], Nicola Iovino[2,3], Ulrich Unnerstall[2], Ulrike Gaul[2,3] & Eran Segal[1,3]

MicroRNAs are key regulators of gene expression[1–4], but the precise mechanisms underlying their interaction with their mRNA targets are still poorly understood. Here, we systematically investigate the role of target-site accessibility, as determined by base-pairing interactions within the mRNA, in microRNA target recognition. We experimentally show that mutations diminishing target accessibility substantially reduce microRNA-mediated translational repression, with effects comparable to those of mutations that disrupt sequence complementarity. We devise a parameter-free model for microRNA-target interaction that computes the difference between the free energy gained from the formation of the microRNA-target duplex and the energetic cost of unpairing the target to make it accessible to the microRNA. This model explains the variability in our experiments, predicts validated targets more accurately than existing algorithms, and shows that genomes accommodate site accessibility by preferentially positioning targets in highly accessible regions. Our study thus demonstrates that target accessibility is a critical factor in microRNA function.

Previous experimental and computational studies have addressed the mechanisms by which microRNAs recognize their targets[5–14]. These efforts have focused primarily on the quality of the sequence match between microRNA and target rather than on the role of the mRNA secondary structure in which the target is embedded. It is likely that secondary structures contribute to target recognition, because there is an energetic cost to freeing base-pairing interactions within mRNA in order to make the target accessible for microRNA binding. Several studies have indeed considered the secondary structure of the target[15–20], but the effect of site accessibility on microRNA-mRNA interactions had not been quantified by systematic experimentation, and the utility of incorporating site accessibility in the genome-wide identification of microRNA targets had not been tested by a principled computational model (**Supplementary Note** online).

To examine the importance of target accessibility in microRNA-mRNA interactions, we developed a quantitative luciferase assay in *Drosophila melanogaster* tissue culture (S2) cells for measuring translational repression by microRNAs[21]. Notably, our assay uses endogenously expressed microRNAs combined with mild overexpression of the target reporter; it has a sensitive and linear readout that permits the study of individual target sites. Moreover, we validated that measured reductions in luciferase activity are attributable to translational repression and not to transcript degradation[22] (**Supplementary Table 1** online). The microRNAs we examined are all robustly expressed in S2 cells, with between 500 and 5,000 copies per cell (**Supplementary Fig. 1** online). We used this assay to measure over 60 microRNA-target interactions, spanning a wide range of target types and mutation designs.

We first focused on validated sites with near-perfect sequence complementarity to the microRNA in the 3′ UTRs of the three proapoptotic genes *hid* (targeted by *bantam*), *rpr* (*miR-2*) and *grim* (*miR-2*)[12,21]. All three targets mediated significant repression in our assay, both when we used the full-length UTRs and when we used the ~200-base-pair (bp) fragments centered on the site (**Fig. 1a**). To test the effect of site accessibility on the strength of microRNA repression, we forced these targets into highly paired stem structures by engineering mutations that introduced their reverse complement in proximity to the target. These closed structures markedly reduced microRNA-mediated repression for the *hid* and *rpr* sites. However, only a small differential was seen for the *grim* site. Notably, this difference in behavior can be explained by target accessibility: the original *grim* site is predicted[23] to be part of a closed stem structure, in contrast to the original *rpr* and *hid* sites, which lie in open loop structures (**Fig. 1a**). Next, we tested whether the importance of site accessibility is asymmetric relative to the 5′ or 3′ end of the microRNA, as is the case for sequence complementarity[11,12]. We found that stem structures covering the 5′ seed or 3′ complementary region of *rpr* result in comparable intermediate repression, suggesting that for a site with significant 3′ pairing, accessibility of both target ends is equally important (**Fig. 1a**). Finally, we explored the effects of the wider sequence context on the efficacy of microRNA repression by swapping each site into the other ~200-bp UTR fragments. Both the *hid* and *rpr* native sites functioned similarly well across all sequence contexts, consistent with their predicted open structure in all UTRs. In contrast, the native *grim* site functioned poorly in the *grim* and *hid* UTRs,

[1]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel. [2]Laboratory of Developmental Neurogenetics, Rockefeller University, New York, New York 10021, USA. [3]These authors contributed equally to this work. Correspondence should be addressed to U.G. (gaul@mail.rockefeller.edu) or E.S. (eran@weizmann.ac.il).

Received 29 May; accepted 17 August; published online 23 September 2007; doi:10.1038/ng2135

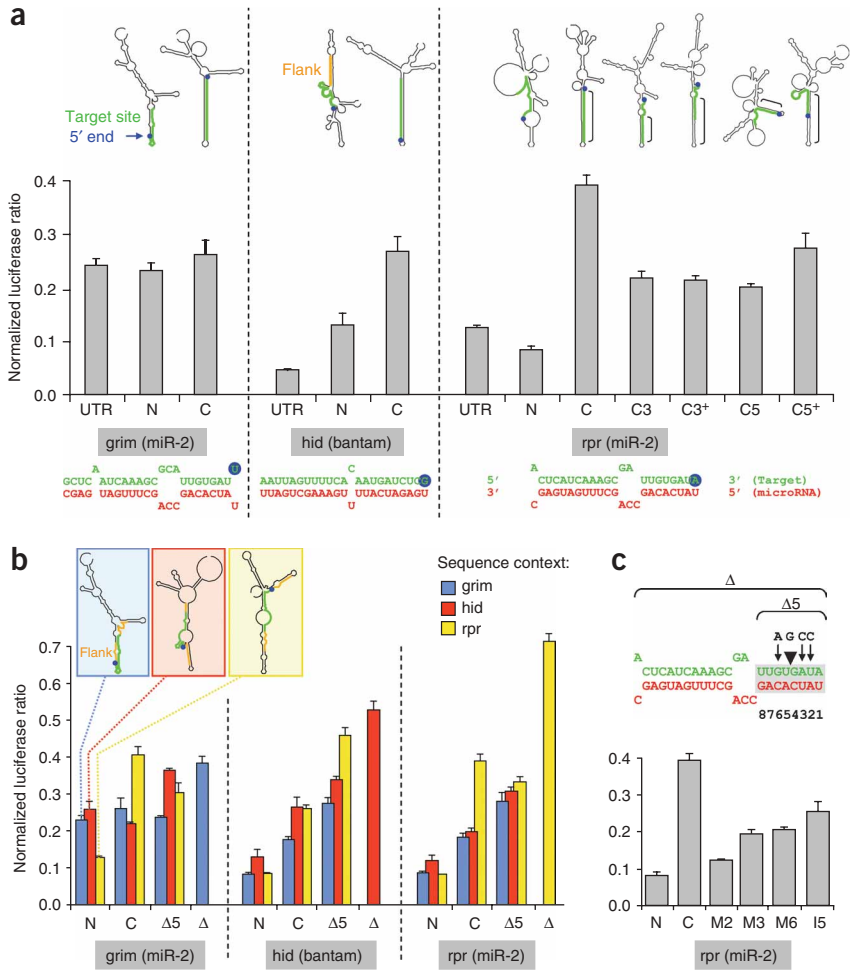

Figure 1 Role of microRNA target-site accessibility in microRNA-mediated repression. (**a**) Expression level mediated by various 3′ UTR constructs for three different microRNA targets (sequence pairing of microRNA to target site is indicated below the graph). Shown are results for full-length 3′ UTRs (UTR), ~200-bp native UTR fragments centered on the microRNA binding site (*N*), and ~200-bp constructs in which the sequence flanking the target site has been changed to force it into highly paired structures (*C*). For *rpr*, we also tested constructs in which only the 3′ half (*rpr-C3*) or 5′ half (*rpr-C5*) of the site are forced into a stem structure; in *rpr-C3*$^{+}$ and *rpr-C5*$^{+}$, this stem is extended to a total length of 22 nucleotides, such that its length is equal to the *rpr-C* stem. RNA structure icons above columns show predicted minimum free energy secondary structure around the target site for each construct (green, 5′ end marked blue, closed flank marked light brown). (**b**) Effect of 3′ UTR sequence context on site function. Shown are results of swapping each of the three target sites into the other two 3′ UTR fragments, as well as the effects of deleting the 5′ seed (Δ5) and the entire target (Δ); see schema in **c**. RNA structure icons are shown for *grim-N* in all three sequence contexts, with marking of the site itself (green) and the flanking region (light brown). (**c**) Effect of mutations in the 5′ seed, compared to mutations that close the secondary structure. Seed of the *rpr miR-2* target site is highlighted in gray; arrows point to the mutated bases (2, 3 and 6); the triangle points to the insertion point (before base 5). All results shown are average values of normalized *Renilla*/firefly luciferase ratios obtained from 4–8 replicates. Error bars, s.e.m.

where its structure is closed, but markedly better in the *rpr* UTR, where its structure is more open (**Fig. 1b**).

To gauge the relative importance of site accessibility, we experimentally compared these effects to those resulting from disruptions of sequence complementarity to the microRNA. As expected, deleting either the 5′ seed or the entire binding site was in most cases more detrimental than placing the site in a closed stem structure, suggesting that even closed structures retain a low level of functionality; these results also persisted when the truncated sites were swapped into the other sequence contexts (**Fig. 1b**). Notably, single-base mutations, insertions or deletions in the 5′ seed had an effect similar to that of closing the structure (**Fig. 1c** and **Supplementary Fig. 2** online). Taken together, our results indicate that site accessibility is critical for efficient repression, and is no less important than individual nucleotide matches in the seed.

Existing methods for microRNA target prediction score microRNA-target interactions based only on the sequence of the target site. These methods assign the same score to all targets with the same sequence and cannot explain the variability that we observed in our experiments, which arises from differences in accessibility imposed by the sequence surrounding the target. For example, a model based only on the binding energy of the microRNA-target duplex, ΔG_{duplex}, achieves poor correlation ($r = 0.36$, $P < 0.11$) to the observed degree of repression (**Fig. 2b**). To account for the effect of accessibility on the strength of microRNA repression, we devised an energy-based score for microRNA-target interactions, $\Delta\Delta G$, equal to the difference between the free energy gained by the binding of the microRNA to the target, ΔG_{duplex}, and the free energy lost by unpairing the target-site nucleotides, ΔG_{open} (**Fig. 2a**). To compute these quantities, we used energy-based secondary structure prediction algorithms; such algorithms have good accuracy[24] and should be particularly reliable for our purposes, because we considered the ensemble of all possible secondary structures and extracted only the free energy of the ensemble. Notably, although our model used a purely thermodynamic score with no parameters and arbitrary cutoffs, its predicted interaction energy, $\Delta\Delta G$, is strongly correlated ($r = 0.7$, $P < 4 \times 10^{-4}$) with the measured degree of repression (**Fig. 2c**).

Despite this success, in some cases our predictions disagreed with the experimental measurement. In most of these instances, the sequences immediately surrounding the target show strong base-pairing interactions within the mRNA. Because microRNAs are held within a large protein complex (the RNA-induced silencing complex, RISC)[25,26], it seemed plausible that tight secondary structures surrounding the target site may restrict the access of the RISC-bound microRNA to the target and would have to be unpaired for a microRNA-mRNA interaction to occur. To test this idea, we modified the component of our model that computes the energetic cost of making the target accessible, ΔG_{open}, to include the cost of unpairing additional bases flanking the target and explored a range of flank sizes upstream and downstream of the target (**Fig. 2d**). Indeed, for most moderate flank sizes, the predictions of this revised model show markedly improved correlation to the repression measurements,

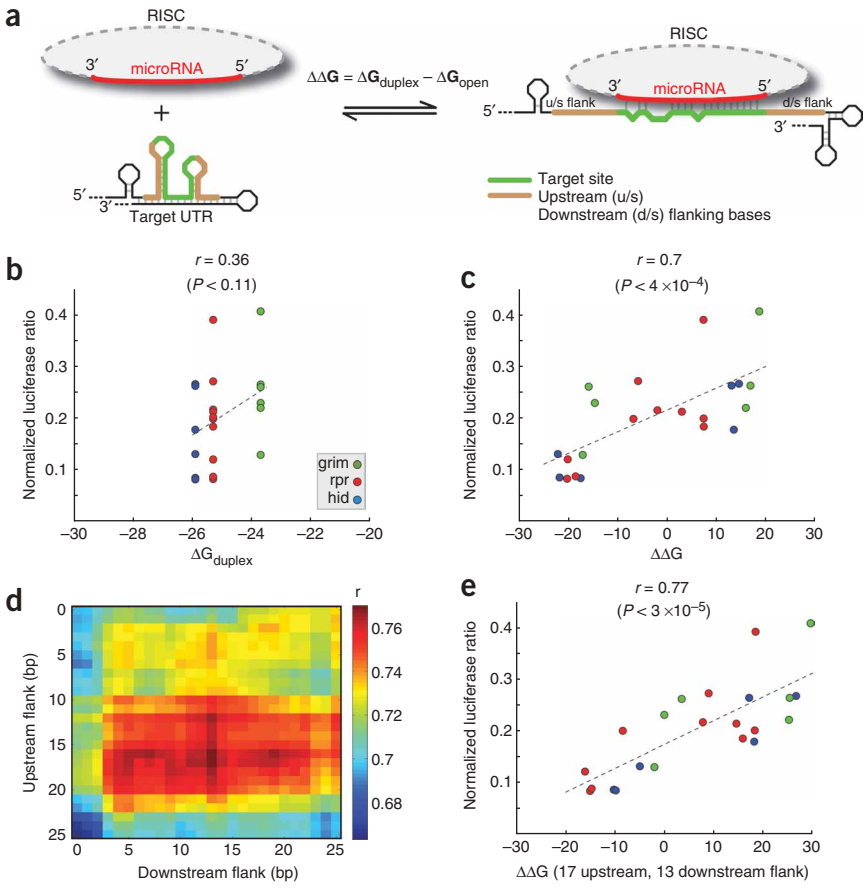

Figure 2 Our microRNA-target interaction model explains variability in target strength due to differences in accessibility. (**a**) Illustration of interaction energy $\Delta\Delta G$ for microRNA-target interactions, computed as the free energy gained by transitioning from the state in which microRNA and target are unbound (left) to the state in which the microRNA binds its target (right). The region of the target site that needs to be unpaired for a microRNA-target interaction to occur includes the microRNA bound region (green) and likely additional flanking nucleotides (brown). (**b**) Methods that consider only the target-site sequence do not explain the variability in our experiments. Shown is a scatter plot of the free energy of microRNA-target binding, ΔG_{duplex} (*x* axis), and the observed expression level of the microRNA target constructs shown in **Figure 1** (*y* axis). Constructs with identical target site sequence but mutated proximal sequence have the same ΔG_{duplex} and are thus located at the same position along the *x* axis. (**c**) As in **b**, but for our model score, $\Delta\Delta G$. Constructs with identical target sites but different proximal sequences have different unpairing energies, ΔG_{open}, and thus differ in their *x*-axis position. The correlation between our model score and the measured expression level is shown along with its associated *P* value (top). See **Supplementary Figure 7a** (online) for the correlation between ΔG_{open} and the measured expression. (**d**) Heatmap showing the correlation between our model score, $\Delta\Delta G$, and the measured expression level of the microRNA target constructs from *hid*, *grim* and *rpr*, when requiring unpairing of different numbers of flanking nucleotides upstream and downstream of the target site. The correlation shown in each entry of the heatmap is computed as in **Figure 2c**. (**e**) As in **c**, when requiring unpairing of the target-site nucleotides plus 17 and 13 nucleotides upstream and downstream of the target, respectively.

with a broad maximum centered around a flank of 17 upstream and 13 downstream nucleotides ($r = 0.77$) (**Fig. 2e**).

Binding sites with near-perfect complementarity to the entirety of microRNA, such as those tested above, are thought to represent only a small portion of functional targets[2,12]. We therefore extended our tests to a broader range of naturally occurring sites. We chose *miR-184*, which is expressed at intermediate levels in S2 cells, and 12 of its predicted targets, and we analyzed them in their native environment (full-length 3′ UTR and ~200-bp fragments centered on the site) as well as when swapped into the *rpr* ~200-bp UTR fragment (**Fig. 3a**). These targets all showed good to excellent 5′ seed complementarity to the microRNA, with varying degrees of 3′ complementarity. While a model using only the binding energy of the microRNA-target duplex, ΔG_{duplex}, achieved almost no correlation with the degree of measured repression ($r = 0.04$, $P < 0.9$), the correlation improved when site accessibility was incorporated ($r = 0.2$, $P < 0.4$) and was strong when a flank requirement was added ($r = 0.5$, $P < 0.03$) (**Fig. 3b,c**). Notably, the differentials in measured site repression between the native and the *rpr* sequence context were strongly correlated to the differences in interaction energy $\Delta\Delta G$ predicted by our model ($r = 0.87$, **Supplementary Fig. 3** online). These results indicate that target accessibility has a critical role in microRNA-mRNA interactions for a wide range of target types, and that our model accurately captures these effects.

To further test the predictive power of our model and compare it to existing methods, we applied it to all 190 microRNA-mRNA interaction pairs experimentally tested in *Drosophila* to date (**Supplementary Table 2** online). Because these experiments typically used full-length UTRs, which can include multiple sites, we followed an approach similar to that of most other prediction methods and first scanned the 3′ UTR of each target for potential microRNA sites, using standard seed parameters that require near-perfect matches to the 5′ end of the microRNA. We then used our model to compute the $\Delta\Delta G$ score of each putative site and appropriately summed these $\Delta\Delta G$ scores to derive a total interaction energy for each microRNA-target pair. Because microRNA-target pairs are primarily reported in binary format, as being either functional or non-functional, we used the standard area under the curve (AUC) measure to evaluate the sensitivity and specificity of our method. Our model achieved an AUC of 0.76, higher than the 0.71 and 0.74 scores achieved by two other state-of-the-art prediction methods, PicTar[10] and the algorithm developed by Stark *et al.*[8], respectively, and a substantial improvement over the 0.64 score of miRanda[6] (**Fig. 4a**). This is particularly notable because these other algorithms employ various filters, such as conservation of sites in related species and additional statistical criteria; our model employs no such parameters or thresholds. Our integration of site accessibility in microRNA target prediction thus represents both a measurable improvement and a simplification over existing methods.

As was the case with our own experimental data, including the requirement to unpair bases flanking the target improved the performance of our model on the literature-derived targets for most flank sizes, with a maximum around a flank of 3 upstream and 15

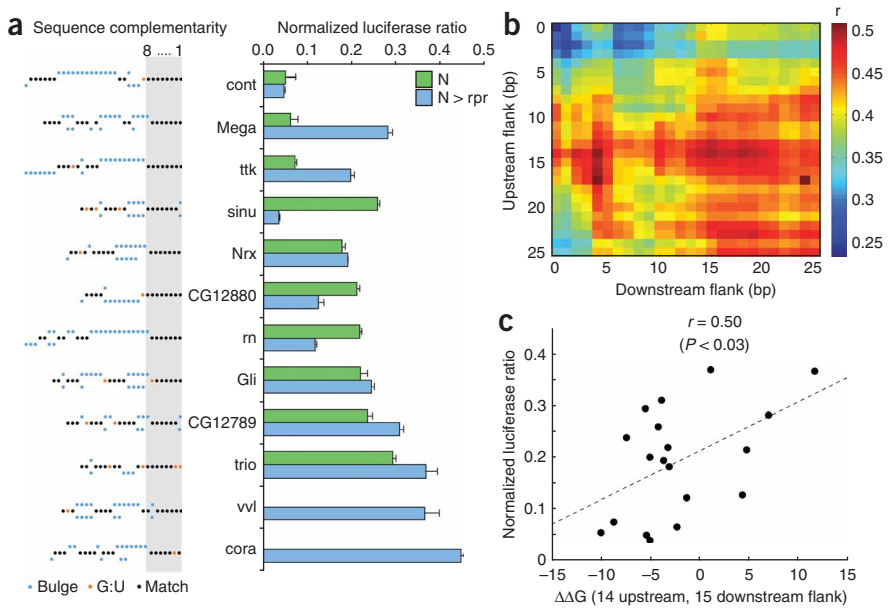

Figure 3 Our model predicts microRNA-target interactions across a wide range of target types. (**a**) Expression mediated by 3′ UTR fragments containing *miR-184* complementary sites (*N*) and by constructs containing the *miR-184* target site in the *rpr* sequence context (*N* > *rpr*). Sequence complementarity is depicted schematically at left (black dots indicate sequence match, orange dots indicate G:U pairing, and offset blue dots indicate unpaired nucleotides on either the target site (top) or the microRNA (bottom); gray underlay indicates seed region). Results shown are average values of normalized *Renilla*/firefly luciferase ratios obtained from 4–8 replicates. Error bars, s.e.m. (**b**) As in **Figure 2d**, heatmap showing the correlation between our model score, $\Delta\Delta G$, and the measured expression level of *miR-184* targets, when requiring unpairing of different numbers of flanking nucleotides upstream and downstream of the target site. The correlation shown in each entry of the heatmap is computed as in **Figure 2e**. (**c**) Correlation between our predicted interaction energy, $\Delta\Delta G$, and the measured expression level of each construct for the *miR-184* targets from **a**, when also requiring unpairing of target-site nucleotides and 14 and 15 nucleotides upstream and downstream of the target, respectively. See **Supplementary Figure 7b** for the correlation between ΔG_{open} and the measured expression.

downstream nucleotides, where the score is 0.79 (**Fig. 4b**). The fact that most moderate flank sizes result in better predictions for all three datasets strongly suggests that efficient microRNA-target interaction requires unpairing of bases flanking the target in addition to those in the target site itself. However, further experiments will be needed to determine the size of the required flank.

The only other available method that incorporates secondary structure into microRNA-target prediction is STarMir, which is based on a two-step nucleation-expansion model[19]. STarMir's predicted scores showed poor correlation to measured expression levels for our two sets of experimental constructs (**Supplementary Fig. 4** online), and its performance on the literature-derived targets was only on par with that expected by chance (**Fig. 4a**). In contrast, our parameter-less model produced scores with high correlation to measured expression values for the constructs used to validate STarMir ($r = 0.8$, **Supplementary Fig. 4**).

Because our results indicate that site accessibility is critical in microRNA-mRNA interaction, we factored it into a genome-wide target prediction algorithm, Probability of Interaction by Target Accessibility (PITA). PITA uses standard settings to identify initial seeds for each microRNA in 3′ UTRs, applies our model to each such putative site, and then combines sites for the same microRNA to obtain a total interaction score for the microRNA and UTR. We

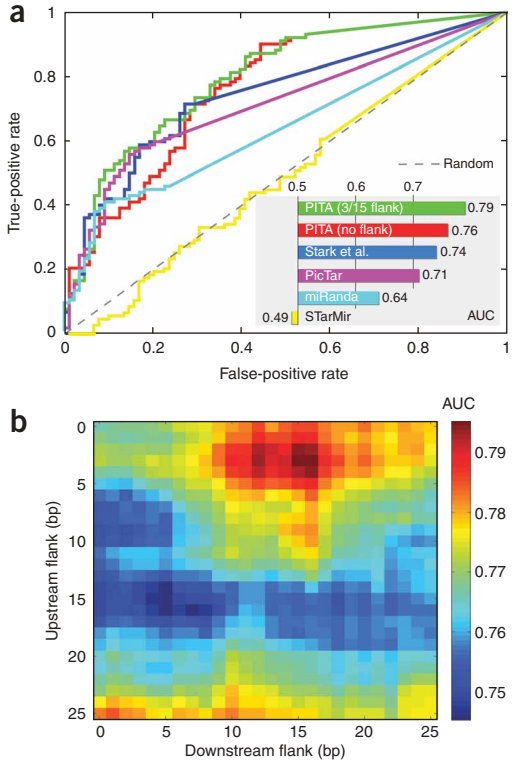

Figure 4 Our model accurately predicts previously published microRNA-target interactions. (**a**) Graphs showing the sensitivity and specificity of different microRNA-target prediction methods on all 190 microRNA-target interaction pairs that have been experimentally tested in *Drosophila* to date (**Supplementary Table 2**). Results are shown for four existing microRNA target prediction methods, including the recent structure-based method of ref. 19, and for two versions of our model, one requiring unpairing of only target-site nucleotides and another that also requires unpairing of 3 and 15 flanking nucleotides upstream and downstream of the target site, respectively. For each prediction method, the targets were sorted by score, and the rate of false predictions (*x* axis) and true predictions (*y* axis) was plotted for each possible score prediction threshold. The area under the curve (AUC) for each method is shown, computed by extending each plot to the upper right corner as is customary. The results obtained by a random sorting of the targets are shown by a dotted line. See **Supplementary Figure 7c** for the performance of a model based only on ΔG_{open}. (**b**) Heatmap for the performance of the model on the set of validated microRNA-target pairs from **a**, when unpairing of different numbers of flanking nucleotides upstream and downstream of the target site was required, and where each entry is reported by the AUC measure, computed as in **a**.

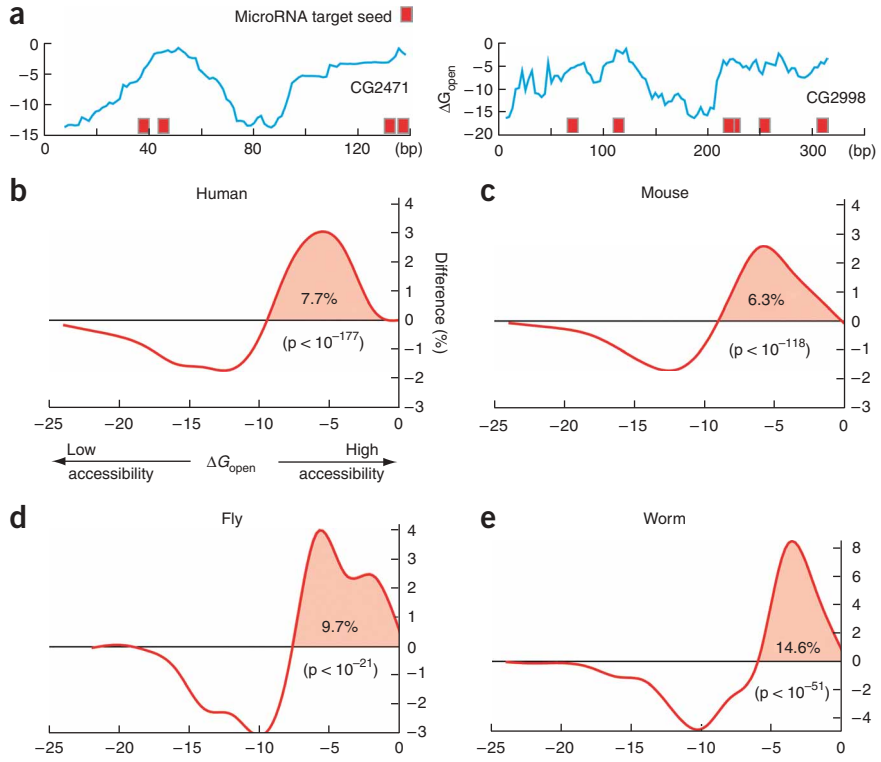

Figure 5 MicroRNA targets in animal genomes are preferentially located in regions of high accessibility. (**a**) Demonstration of the concept of preferential positioning of microRNA targets in regions of high accessibility on two 3′ UTRs from the fly. The target locations and accessibility are shown along each UTR. (**b**–**e**) Shown is the difference between the distribution of target accessibility scores, ΔG_{open}, for all microRNA seeds that have a conservation of at least 0.9 in the 3′ UTRs of human (**a**), mouse (**b**), fly (**c**) and worm (**d**), and the distribution of an equal-sized set of seeds whose locations were chosen at random from each of the respective organisms. For each accessibility value x (x axis), we plotted the difference between the fraction of real seeds and the fraction of randomly placed seeds whose accessibility is x. Because the ΔG_{open} measure depends on the number of G and C nucleotides, random seed locations were selected such that they had the same GC content distribution as that of the real seeds. The fraction of real seeds that are more accessible than the random seeds is indicated (shaded area), along with the significance level of the difference between the distributions, as measured by the Kolmogorov-Smirnov test. See **Supplementary Figure 6** for the same comparison without the conservation requirement.

applied PITA to all 3′ UTRs of fly, worm, mouse and human, resulting in prediction catalogs for these organisms, which are available through our laboratory website (see Methods for URL) (**Supplementary Fig. 5** online).

In comparing our predictions to those of two other prediction methods[8,10], we found an overlap ranging from 19% to 31%, depending on the organism and method being compared (**Supplementary Table 3** online). Although these overlaps are statistically significant, they also highlight the major influence that accessibility has on predictions: many sites predicted by other algorithms have low accessibility and thus are not predicted by PITA; by contrast, many highly accessible sites predicted by our approach do not pass the prediction threshold of other methods. We reasoned that if site accessibility has such a substantial effect on microRNA interactions, genomes may have evolved to accommodate this constraint, perhaps by preferentially positioning targets in regions that have open structures and are thus more accessible (**Fig. 5a**). Indeed, we found that microRNA seeds in all four organisms showed a notable preference for highly accessible regions, as compared to an equal-sized collection of seeds whose genomic locations were chosen at random (**Supplementary Fig. 6** online). Moreover, when the analysis was limited to conserved seeds, the fraction of seeds in more accessible regions increased, suggesting that the preferential positioning of targets at accessible regions is evolutionarily selected for and thus likely to be functionally important (**Fig. 5b–e**).

In conclusion, our results show that site accessibility is as important as sequence match in the seed for determining the efficacy of microRNA-mediated translational repression, and they suggest that effective microRNA binding also requires unpairing of the local region flanking the target. We introduce a parameter-free thermodynamic model that explains these effects and shows that preferential positioning of microRNA target sites in regions of high accessibility is a conserved feature in genomes. Our findings thus indicate that the thermodynamics of intra- and intermolecular base pairing account for a significant portion of the microRNA-target interaction, consistent with observations for siRNA-target interactions[27]. However, our model does not explain the entire variance in our experiments. This may be in part due to limitations of RNA structure prediction algorithms and their inability to account for the effects that RNA binding proteins have on secondary structures. Moreover, microRNAs bind as part of the RISC complex, and although there is no evidence for ATP-dependent unwinding of the mRNA[27,28], RISC proteins are likely to constrain base-pairing interactions between microRNA and target site and otherwise influence the mechanics of the duplex formation. Thus, further experiments will be needed to understand how different aspects of the interaction contribute to the strength of microRNA-mediated repression. Nevertheless, our results provide an important cornerstone for deciphering the rules that govern microRNA-mRNA interactions.

METHODS

Reporter constructs and luciferase assay. Our assay used a dual luciferase system in which two luciferase enzymes, one (from *Renilla reniformis*) containing the experimental target sequence and another (from firefly) containing the control[21], are expressed from a single plasmid. Both luciferase genes were controlled by heterologous promoters (SV40, HSVTK; Promega) that in S2 cells increased transcript levels tenfold over endogenous levels (tested for *Gli*, *ttk*; data not shown). Notably, even when strong microRNA target sequences were introduced into the *Renilla* luciferase, their transcript abundance showed no significant change, indicating that any observed reduction in luciferase activity was attributable to translational repression and not to transcript degradation (**Supplementary Table 1**).

To facilitate larger scale analysis of 3′ UTR sequences, we created a gateway (Invitrogen) version of the dual luciferase vector psicheck-2 (Promega) by ligating a blunt-ended cassette containing *attR* sites flanking the *ccdB* gene and the chloramphenicol resistance gene into the *Pme*I site within the psicheck-2 polylinker. Full-length and truncated UTR sequences were amplified by PCR

from genomic or plasmid DNA, cloned into pEntr vector and then recombined into the psicheck-2 destination vector using the LR recombinase kit (Invitrogen). For the *hid*, *grim* and *rpr* 3′ UTRs, we engineered ~200 bp truncated versions using 50-bp-long synthetic oligonucleotides such that the central microRNA target site was flanked by restriction sites to facilitate mutagenic replacements (5′–*Not*I–target site–*Xho*I–3′). The oligos were purified by PAGE, annealed, ligated and cloned into psicheck2. We confirmed all clones by DNA sequencing (Genewiz). The sequences of all constructs used in this study are compiled in **Supplementary Table 4** (online). We transfected 10^6 S2 cells with reporter plasmid (1 μg) using Cellfectin (Invitrogen), and after 20 h, we lysed the cells and tested for luciferase activity using the dual luciferase assay kit (Promega). The *Renilla*/firefly luciferase ratios were normalized against the empty psicheck-2 vector and averaged over 4–8 replicates. We assessed statistical significance using one-factor and two-factor ANOVA with the Student-Newman-Keul *post hoc* test. Note that none of our test constructs, even those in which the entire site was deleted, completely reverted to the expression level of the empty vector, suggesting that any significant addition of 3′ UTR sequence (>150 bases) reduces translation efficacy.

Thermodynamic model for microRNA-target interactions. Our model scores microRNA-target interactions by an energy score, $\Delta\Delta G$, equal to the difference between the energy gained by binding of the microRNA to the target, ΔG_{duplex}, and the energy required to make the target region accessible for microRNA binding, ΔG_{open}. ΔG_{duplex} is the binding free energy of the microRNA-target duplex structure in which the microRNA and target are paired according to pairing constraints imposed by the seed. To compute this value, we modified RNAduplex[23] such that in addition to microRNA and target sequences, it is given explicit seed pairings upon which the target was chosen and considers only structures that comply with these constraints. The code then computes the binding free energy of each of the complying structures (a computation that can be done efficiently using dynamic programming) and selects the minimum free energy structure as ΔG_{duplex}. ΔG_{open} is computed as the difference between the free energy of the ensemble of all secondary structures of the target region (see below) and the free energy of all target-region structures in which the target nucleotides (and additional nucleotides upstream and downstream in the case of flanking) are required to be unpaired. The free energies of these two ensembles are computed using RNAFold[23], by iterating over all possible structures (for the second term, all structures subject to the above unpairing constraints) and appropriately summing their respective free energies. These computations can be done efficiently using dynamic programming. The area of the target that is given to RNAFold for folding consists of the target (and flanking region when applicable) and 70 additional nucleotides upstream and downstream. The value of 70 nucleotides was chosen to reduce the time complexity of the above computations and is based on the fact that there is a low probability of secondary structure base-pairing interactions between nucleotides that are separated by more than 70 nucleotides (data not shown). Finally, the total interaction score, $\Delta\Delta G$, is equal to the difference between ΔG_{duplex} and ΔG_{open}. To integrate multiple sites with $\Delta\Delta G$ scores $s_1,\ldots, s_n$, for one microRNA on the same UTR into an overall microRNA-UTR interaction score, T, we compute the statistical weight of all configurations in which exactly one of the sites is bound by the microRNA according to $T = \log \sum_{i=1}^{n} e^{s_i}$. We chose this simple method of integrating multiple sites over computing the actual probability of microRNA binding, and over computations that include configurations in which two sites can be bound simultaneously, because such computations would require knowledge of an additional (unknown) microRNA free concentration parameter.

Identifying seeds for microRNAs. We follow standard seed parameter settings and consider seeds of length 6–8 bases, beginning at position 2 of the microRNA. No mismatches or loops are allowed, but a single G:U wobble is allowed in 7- or 8-mers.

MicroRNA-target interaction pairs experimentally tested in the literature. The compilation of experimentally tested *Drosophila* microRNA-mRNA pairs (**Supplementary Table 2**) was based on a previously published list[8], to which we added targets reported in TarBase[29] and ref. 30, and 25 targets of *miR-2* and *miR-184* from our own experiments. The resulting non-redundant list consists of 190 total pairs, of which 102 were reported as functional microRNA targets and 88 were reported as non-functional.

UTRs for genome-wide microRNA-target predictions. The fly (dm2), mouse (mm8), human (hg17) and worm (ce4) genomic sequences were downloaded from the University of California, Santa Cruz (UCSC) Genome Bioinformatics Site (http://genome.ucsc.edu), along with gene annotations and conservation tracks. For genes lacking a 3′ UTR annotation, we used predicted 3′ UTRs, with a stringent length cutoff of 500 bp (fly), 800 bp (human and mouse) or 300 bp (worm) downstream of the annotated end of the coding sequence. For the genome-wide PITA catalogs, which can be downloaded from our website, we used a looser cutoff of 1 s.d. above the mean length of the annotated 3′ UTR. Predicted UTRs account for 21% (3,914 genes) of the UTRs in fly, 5% (843 genes) in human, 10% (1,858 genes) in mouse and 54% (10,821 genes) in worm.

Testing preferential positioning of microRNA seeds for regions of open secondary structure. For each organism, we scanned the 3′ UTRs for perfect microRNA target seeds at least seven bases long, disallowing G:U wobbles, mismatches or loops. Because ΔG_{open} is affected by GC content, an equal-sized collection of locations with the same GC content distribution as that of the true microRNA seeds was randomly chosen as a control set. We computed the accessibility (ΔG_{open}) score of each true and control seed (without flank) and compared their distributions using the Kolmogorov-Smirnov test. For the conservation-restricted test, conservation tracks based on a phylogenetic hidden Markov model (phastCons) were downloaded from the UCSC site (comparing 12 *Drosophila* species, mosquito, honeybee and red flour beetle for the fly track and 17 vertebrates, including mammalian, amphibian, bird and fish species for the human and mouse tracks). Only microRNA seeds having an average (along the seed) conservation score higher than 0.9 were considered. For these comparisons, the control set of random seeds was also chosen from areas of the 3′ UTR with a conservation score higher than 0.9.

URLs. For our data, results, an online tool for predicting microRNA-target interactions on any microRNAs and target sequences of interest, and the downloadable PITA executable, see http://genie.weizmann.ac.il/pubs/mir07.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank O. Manor for assistance with genome-wide predictions, Y. Lubling for creating the supplementary website, T. Tuschl and B. Darnell for the use of equipment, and K. O'Donovan and J. Fak for technical assistance. This work was supported by the Israel Science Foundation (M.K., E.S.), a PhD fellowship from the University of Rome "La Sapienza" (N.I.) and the Rockefeller University (U.G.). E.S. is the incumbent of the Soretta and Henry Shapiro career development chair.

Published online at http://www.nature.com/naturegenetics
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. Carthew, R.W. Gene regulation by microRNAs. *Curr. Opin. Genet. Dev.* **16**, 203–208 (2006).
3. Esquela-Kerscher, A. & Slack, F.J. Oncomirs—microRNAs with a role in cancer. *Nat. Rev. Cancer* **6**, 259–269 (2006).
4. Bartel, D.P. & Chen, C.Z. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat. Rev. Genet.* **5**, 396–400 (2004).
5. 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).
6. Enright, A.J. *et al.* MicroRNA targets in *Drosophila*. *Genome Biol.* **5**, R1 (2003).
7. 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).
8. Stark, A., Brennecke, J., Bushati, N., Russell, R.B. & Cohen, S.M. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. *Cell* **123**, 1133–1146 (2005).
9. Stark, A., Brennecke, J., Russell, R.B. & Cohen, S.M. Identification of *Drosophila* MicroRNA targets. *PLoS Biol.* **1**, e60 (2003).
10. Grun, D., Wang, Y.L., Langenberger, D., Gunsalus, K.C. & Rajewsky, N. microRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. *PLoS Comput. Biol.* **1**, e13 (2005).

11. Doench, J.G. & Sharp, P.A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
12. Brennecke, J., Stark, A., Russell, R.B. & Cohen, S.M. Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85 (2005).
13. 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**, 1067–1080 (2005).
14. Rehmsmeier, M., Steffen, P., Hochsmann, M. & Giegerich, R. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**, 1507–1517 (2004).
15. Zhao, Y., Samal, E. & Srivastava, D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* **436**, 214–220 (2005).
16. Zhao, Y. *et al.* Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1–2. *Cell* **129**, 303–317 (2007).
17. Vella, M.C., Reinert, K. & Slack, F.J. Architecture of a validated microRNA:target interaction. *Chem. Biol.* **11**, 1619–1623 (2004).
18. Robins, H., Li, Y. & Padgett, R.W. Incorporating structure to predict microRNA targets. *Proc. Natl. Acad. Sci. USA* **102**, 4006–4009 (2005).
19. Long, D. *et al.* Potent effect of target structure on microRNA function. *Nat. Struct. Mol. Biol.* **14**, 287–294 (2007).
20. Muckstein, U. *et al.* Thermodynamics of RNA-RNA binding. *Bioinformatics* **22**, 1177–1182 (2006).
21. Leaman, D. *et al.* Antisense-mediated depletion reveals essential and specific functions of microRNAs in *Drosophila* development. *Cell* **121**, 1097–1108 (2005).
22. Valencia-Sanchez, M.A., Liu, J., Hannon, G.J. & Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* **20**, 515–524 (2006).
23. Hofacker, I.L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431 (2003).
24. Wiese, K.C. & Hendriks, A. Comparison of P-RnaPredict and mfold–algorithms for RNA secondary structure prediction. *Bioinformatics* **22**, 934–942 (2006).
25. Sontheimer, E.J. Assembly and function of RNA silencing complexes. *Nat. Rev. Mol. Cell Biol.* **6**, 127–138 (2005).
26. Parker, J.S., Roe, S.M. & Barford, D. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**, 663–666 (2005).
27. Ameres, S.L., Martinez, J. & Schroeder, R. Molecular basis for target RNA recognition and cleavage by human RISC. *Cell* **130**, 101–112 (2007).
28. Haley, B. & Zamore, P.D. Kinetic analysis of the RNAi enzyme complex. *Nat. Struct. Mol. Biol.* **11**, 599–606 (2004).
29. Sethupathy, P., Corda, B. & Hatzigeorgiou, A.G. TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA* **12**, 192–197 (2006).
30. Rehwinkel, J. *et al.* Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in *Drosophila melanogaster*. *Mol. Cell. Biol.* **26**, 2965–2975 (2006).