Published in final edited form as:

Nat Struct Mol Biol. ; 19(3): 321–327. doi:10.1038/nsmb.2230.

An alternative mode of microRNA target recognition

Sung Wook Chi^{1,2,3}, Gregory J. Hannon², and Robert B. Darnell¹

¹Laboratory of Neuro-Oncology, The Rockefeller University, Howard Hughes Medical Institute, New York, New York 10065, USA

²Cold Spring Harbor Laboratory, Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor, New York 11724, USA

³Graduate School, Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Samsung Medical Center, 50 Irwon-dong, Kangnam-ku, Seoul 135-710, Korea

Abstract

MicroRNAs (miRNAs) regulate mRNA targets through perfect pairing with their seed region (position 2-7). Recently, a precise genome-wide map of miRNA interaction sites in mouse brain was generated by high-throughput sequencing of clusters of ~50 nucleotide RNA tags associated with Argonaute (Ago HITS-CLIP). By analyzing Ago HITS-CLIP “orphan clusters” – Ago binding regions from HITS-CLIP that cannot be explained by canonical seed matches – we have identified an alternative binding mode used by miRNAs. Specifically, G-bulge sites (position 5-6) are often bound and regulated by miR-124 in brain. More generally, bulged sites comprise 15% (1441 sites) of all Ago-miRNA interactions in mouse brain and are evolutionally conserved. We have termed position 6 the “pivot” nucleotide and suggest a model in which a transitional “nucleation-bulge” leads to functional bulge mRNA-miRNA interactions, expanding the number of potential miRNA regulatory sites.

Diversity generated through the regulation of RNA transcripts has been proposed as an explanation for the discrepancy between organismal complexity and the relatively limited number of primary coding transcripts^{1,2}. Regulation by miRNAs is one such mechanism, since each miRNA directly binds hundreds of mRNAs to induce post-transcriptional repression, leading to diverse cellular phenotypes^{3,4}. The best characterized features determining miRNA-target recognition are short 6 nucleotide seed sites, which perfectly complement the 5' end of the miRNA (positions 2-7)⁵. So-called “seed pairing rules” are widely used to predict functional miRNA target sites, often in combination with evolutionary conservation⁶, secondary structure⁷, or neighboring context information⁸.

Seed rules have been informative to identify targets through miRNA overexpression or knockdown studies, especially in combination with microarray⁹ or proteomic approaches^{10,11}. However, such strategies suffer from both false positive (~40-66%)^{10,12}

and false negative predictions (~50-70%)^{10,11,13} and cannot identify non-canonical target sites. Several biological studies have functionally validated that perfectly matched miRNA seeds are neither necessary nor sufficient for all functional miRNA-target interactions¹⁴⁻¹⁷. For instance, genetically verified *lin-4*¹⁴, *let-7*¹⁵ or *lys-6*¹⁷ targets in *C. elegans* as well as some mouse miRNA targets¹⁶ contain only imperfect binding sites, with bulges or G-U wobble pairs in the seed region. Although these studies strongly suggested the existence of non-canonical target sites, they have not been recognized as general features of miRNA-mRNA interactions, in part due to the difficulty in determining how frequently such atypical sites are used *in vivo* and what general rules could be used to predict them. False negative predictions from seed rules could be interpreted as target transcripts with non-canonical target sites, especially for false negative predictions from co-immunoprecipitation of RNA-Ago complex (~21-50%)^{13,18,19}, although uncertainty regarding the specificity and resolution of these experiments has made it difficult to interpret and identify such non-canonical sites.

Recently, a precise, genome-wide map of miRNA-binding sites in mouse brain was decoded by applying a biochemical method which involves direct recovery of crosslinked RNA-protein complexes (CLIP)²⁰ containing Ago, followed by high-throughput sequencing of the isolated RNA interaction sites (HITS-CLIP)^{21,22} and bioinformatic analysis of the sequences based on seed matches²³. Ago HITS-CLIP has been used to map Ago-mRNA binding footprints (~45-62 nt) with high specificity (~93% specificity, ~13-27% false-positives and ~15-25% false-negatives)²³ and also has been applied to *C. elegans*²⁴ and cultured cells^{25,26}, allowing the identification of *in vivo* miRNA binding sites on a genome wide scale. However, not all identified Ago binding sites followed classical seed rules. In fact, 27% of Ago-mRNA clusters (normalized overlapping tags from Ago crosslinked mRNAs) in mouse brain were ‘orphans’ with no predicted seed matches among the top Ago-bound miRNAs (~88%, from the top 20 miRNA families). These results suggest that a significant number of Ago-miRNA binding sites *in vivo* might fail to follow the rule that mRNA recognition is dictated by canonical seed matches²³.

To search for such non-canonical seed pairing, we analyzed Ago HITS-CLIP “orphan clusters”. mRNAs harboring G-bulge sites were often bound by miR-124 in mouse brain, and conformed to a rule in which formation of a transitional “nucleation-bulge” is determined by the annealing of 5 consecutive nucleotides in positions 2-6 (a “pivot pairing” rule). Applying this rule globally we found that bulged sites are common, comprising > 15% (> 1441 sites) of all Ago-miRNA interactions in mouse brain, thereby expanding the number of potential regulatory sites for miRNAs and providing insight into the biochemical mechanisms by which miRNA-Ago complexes bind their targets.

RESULTS

Identification of G-bulge sites pairing to miR-124 in brain orphan clusters

To uncover new rules for miRNA binding, we first performed an unbiased search for all 6–8-nucleotide sequence motifs enriched within orphan clusters detected in mouse brain Ago HITS-CLIP by using the MEME program²⁷ (Fig. 1a and Supplementary Fig. 1B). Six motifs were identified (E-value < 0.01) and their locations relative to cluster peaks²³ were further

examined. UGGCCUU was identified as the most significant motif enriched (Fig. 1a). This motif corresponds to a match for miR-124, but only if a G-bulge in the mRNA (a G corresponding to position 5-6 of the miRNA) is allowed (Fig. 1b). miR-124 is a brain-specific miRNA previously shown to be the eighth most frequent miRNA associated with Ago, and its canonical seed sites are the most enriched in Ago mRNA clusters identified in mouse brain²³. Of note, among all 11,463 Ago-mRNA clusters²³, 684 clusters are miR-124 G-bulge clusters (~6% of total clusters, enriched relatively to the fraction of miR-124 (~3%, among all brain miRNAs bound to Ago), equal to ~46% of the miR-124 seed clusters (1480 clusters)). Further analysis revealed that the same G-bulge motif was present in *de novo* Ago miR-124 clusters, those clusters that appear after miR-124 transfection of HeLa cells (Fig. 1c), consistent with this motif acting as a *bona fide* miR-124-dependent Ago binding site. G-bulge sites were enriched in *de novo* Ago miR-124 footprints comparable to seed sites (418 G-bulge and 691 seed clusters were identified, 6-mers from position 1-8 in 62 nt). Among all possible bulge types, only G-bulge sites were significantly enriched in *de novo* Ago miR-124 clusters (Fig. 1d, kurtosis (k) = 4.02 in G-bulges, versus k = 1.82 in C-bulges, k = 2.23 in A-bulges and k = 2.25 in U-bulges; the same distribution as seed sites (k = 4.38)) and in several additional analyses evaluating all possible bulge and wobble sites in Ago clusters (Supplementary Table 1 and Supplementary Fig. 1). Considered together, these data provide direct evidence that miR-124 may act on brain transcripts harboring non-canonical bulge sites and that such interactions are specific for G-bulges relative to other nucleotides.

Validation of G-bulge sites in Ago-miR-124 clusters

To measure the extent of G-bulge binding to miR-124 *in vivo*, we analyzed a set of Ago-miR-124 bulge clusters detected from both brain and miR-124 transfected HeLa cells (Supplementary Fig. 2A); such sites are identified by a remarkably stringent test in that they denote native mouse brain Ago-mRNA footprints that are conserved in HeLa cells, referred to as conserved *de novo* Ago clusters. From this dataset we conservatively estimate (Fig. 2a) that one fourth of all Ago-miR-124-mRNA interactions are mediated through binding to bulge sites, with three-quarters of these attributable to G-bulge sites (~18% of total). We used luciferase reporter assays to validate G-bulge sites in *Mink1* and *Epb41* (Fig. 2b and Supplementary Fig. 2B,C). Generation of point mutants demonstrated that in these contexts G-bulge sites were able to mediate miR-124-dependent repression (Fig. 2c,d) as efficiently as canonical seed sites. We then extended these validation analyses to a large number of mRNAs (Fig. 2e and Supplementary Fig. 3A) which showed miR-124 dependent changes in transcripts levels after miR-124 transfection ($n = 2694$, $P < 0.05$)⁹. In analysis of the cumulative distribution of miR-124-dependent changes, transcripts with G-bulge sites identified in *de novo* Ago-miR-124 HeLa clusters were downregulated ($P < 0.05$, relative to the distribution of total miR-124 dependent transcripts, Kolmogorov-Smirnov (KS) test), although less so than the transcripts with seed sites in the clusters ($P < 0.01$, KS test). Notably, transcripts with G-bulge sites were downregulated also relative to total Ago-miR-124 bound transcripts (Supplementary Fig. 3C **AU: correct to change from Supp Fig 4C?**), showing that repression observed is G-bulge specific effect. To compare the extent of repression mediated by perfect and bulged seed matches for miR-124, the downregulated transcripts were examined further, and this analysis indicated that G-bulge sites were present in abundance comparable to seed sites, whereas possible binding sites with other types of

bulges were not (Fig. 2f). The degree of repression mediated by G-bulges versus seed sites overlaps those seen with canonical sites significantly (Supplementary Fig. 3B). In addition, transcripts with G-bulge sites were also downregulated at the protein level¹⁰, although this was only evident in the smaller number of transcripts that had larger changes (Supplementary Fig. 3C). We also observed evidence that nucleotide changes in G-bulge sequences could abrogate interactions with Ago *in vivo* (Supplementary Fig. 3D) by examining species specific nucleotide changes between two Ago HITS-CLIP datasets (HeLa vs. mouse brain, $P < 0.01$, Fisher's exact test, Supplementary Fig. 3E). Overall, these data lead us to conclude that G-bulge sites bound by miR-124 are functional.

An alternative mode of miRNA target recognition

We explored the mechanisms that might account for the preference for G-bulges in miR-124 orphan clusters. The free energy calculation for G, C, U or A bulges in position 5-6 was the same (Fig. 3a and Supplementary Fig. 4A) and hence did not explain the specificity for G. However, when we considered an intermediate in the miR-124-mRNA duplex in which the G-bulge nucleotide was temporarily used to bind to miR-124, 5 consecutive nucleotides were available for annealing (position 2-6) compared to only 4 consecutive nucleotides (position 2-5) for all other possible bulges, and this intermediate yielded a free energy that was significantly lower than that seen with only 4 consecutive annealing nucleotides (Fig. 3a and Supplementary Fig. 4B). This intermediate (termed the “transitional nucleation”) may occur *in vivo*, such that stability of the transition state would be largely determined by pairing in miRNA position 6 (a C nucleotide, termed the “pivot”).

To test the validity of such a “pivot pairing” rule, we initially examined whether bulge sites with this pattern (resulting in a “nucleation bulge”, in which the bulge sequence was competent to pair to the miRNA pivot nucleotide in a transitional nucleation intermediate, Fig. 3a) were enriched in Ago footprint regions. In an analysis of the top 30 Ago-bound miRNAs (~92% of all miRNAs bound to Ago), nucleation bulges were slightly but significantly enriched near Ago-mRNA cluster peaks ($k = 1.92$, within -50 and +50 nucleotides, $P < 0.05$, relative to background uniform distribution ($k = 1.80$), KS test. Fig. 3b), as also seen with canonical seed sites ($k = 2.33$, $P < 0.01$, KS test)²³. Such an enrichment was not seen with bulge sites harboring with only 4 nucleotide transitional interactions ($k = 1.81$, $P = 0.95$, KS test, the same as background uniform distribution ($k = 1.80$)), termed “non-nucleation bulges” (in which the position 5-6 bulge sequence was identical to the pivot nucleotide and therefore unable to pair to pivot, Fig. 3a and Supplementary Fig. 4B).

Nucleation bulge sites are evolutionally conserved, functional and widespread

Evolutionary conservation of miRNA binding sites can provide strong evidence for their biological significance. Comparison of the conservation of canonical seed sites and nucleation bulges revealed that within the context of otherwise poorly conserved 3' untranslated regions (3'UTR, median conservation rate = ~5.4%, 7-mers), nucleation bulges for all known miRNAs were significantly conserved over the background (~6.3 vs. 5.4%, $P < 0.01$, KS test) as well as seed sites (~6.6 vs. 5.4%, $P < 0.01$) and were distributed to the same degree of conservation rates observed in seed sites (~6.3 vs. 6.6%, respectively, Fig.

3c), whereas non-nucleation bulges were not conserved and as background distribution (~5.1 vs. 5.4%, respectively). Among all Ago footprints in mouse brain (average conservation rate = ~25%, 7-mers), nucleation bulges for the top 20 Ago-bound miRNAs (~88% of all miRNAs bound to Ago) were also evolutionarily conserved over background rates (~32% vs. 25%, $P = 0.06$, t-test, Fig. 3d), but less so than canonical seed sites (~51% vs. 25%, $P = 0.02$, t-test); non-nucleation bulges were even less conserved than the background rate (~22% vs. 25%, $P = 4.0 \times 10^{-5}$, t-test). When non-nucleation bulges were used as control, nucleation bulges were also significantly conserved (~32% vs. 22% $P = 0.01$, t-test) but less so than canonical seed sites (~51% vs. 22%, $P = 1.9 \times 10^{-7}$, t-test). In addition, the 3'UTR and coding sequence were analyzed separately with similar results (Supplementary Fig. 4C).

We further investigated the functional significance of the nucleation bulge by compiling brain expressed transcripts from a meta-analysis of microarray experiments performed with seven different miRNAs transfected into cell lines²⁸ (excluding the previous analysis of miR-124). Transcripts harboring nucleation bulges in Ago clusters showed significant miRNA-dependent repression ($P = 0.03$, KS test, Fig. 3e); while this was to a lower degree than miRNA-dependent repression of transcripts with canonical seed sites ($P = 1.05 \times 10^{-24}$, KS test), transcripts with non-nucleation bulges showed no repression ($P = 0.95$, KS test). Repression mediated by sites harboring a nucleation bulge showed a broad range of fold changes, overlapping with and similar to the degree of repression through canonical seed-match sites (-0.79 vs. -0.97 median log₂ fold changes; Fig. 3f and Supplementary Fig. 4D).

We also expanded our analysis to a second organism by using recently available Ago HITS-CLIP and microarray data from the wild type versus *alg-1(gk214)* mutant *C.elegans*²⁴. In this dataset, the degree of transcript repression from nucleation bulge and canonical seed sites also overlapped (Supplementary Fig. 4E-G). The same enrichment of functional nucleation bulges (~20% number of seed sites, respectively to non-nucleation bulges; < 10% number of seed sites, Supplementary Fig. 5A-C) was also seen in *de novo* Ago-miR-124 or Ago-miR-7 clusters identified by PAR-CLIP²⁵. However, nucleation G-bulge sites were more specifically enriched in *de novo* Ago-miR-124 clusters identified by HITS-CLIP than by PAR-CLIP (Supplementary Fig. 5D).

We tested whether the pivot pairing rule could explain “orphan transcripts” evident in previously published datasets in which miRNAs had been exogenously added to Dicer-null cells. These orphan transcripts were defined as those showing miRNA-dependent changes but in none of these could 6-mer seed matches for the added miRNA be found (the same as false positive predictions from seed rule). In a cumulative distribution analysis for miR-430 injected Dicer-null zebrafish (MZdicer, Fig. 3g)²⁹ or a meta-analysis of data generated for eight different miRNAs in Dicer-null cell (HCT116 Dicer -/-, Fig. 3h)³⁰, orphan transcripts with predicted nucleation bulges were significantly downregulated ($P < 0.05$ and $P = 0.005$, KS test, relatively to total orphan transcripts. Of note, total orphan transcripts (Fig. 3g,h) showed an increase in fold change, probably caused by subtraction of transcripts with seed matches, which show higher fold decreases.

We examined the correlation between the number of conserved bulge sites predicted by the pivot rule to be present in 11,463 Ago footprints and the number of miRNAs associated with

such bulge sites, using linear regression analysis. Nucleation bulge sites for the top 100 Ago-bound miRNAs showed a similar positive correlation (Fig. 3i) as did canonical seed sites (Fig. 3j), although the correlation was not as high as for canonical seed sites. However, non-nucleation bulge sites showed no correlation (Fig. 3k). Notably, we also confirmed that 5mer seed matches (position 2-6), which could confound the result from bulge sites, are not functional, since we observed that those cannot interact with Ago-miRNA (Supplementary Fig. 5E) nor mediate miRNA-dependent repression in vivo (Supplementary Fig. 5F). Considered as a whole, our analyses demonstrate that nucleation bulges are biologically significant, as they are evolutionarily conserved, functional and widespread

Application and validation of pivot pairing rule for Ago HITS-CLIP analysis

To define and apply the pivot pairing rule with more precision, we searched for nucleation bulges in orphan clusters predicted to have let-7-bulge mRNA interactions, since let-7 is a well-studied miRNA and is highly expressed (ranked #5 in Ago-miRNA binding) with canonical seed sites enriched in brain (second only to miR-124 in Ago-mRNA clusters)²³. Interestingly, let-7 has a pivot nucleotide (U) capable of wobble pairing, and this in theory might contribute transitional nucleation energy to either A or G residues. We identified a let-7 nucleation A-bulge in a *Kif5b* orphan cluster (Supplementary Fig. 6A) and found that it mediated a level of repression comparable to a canonical seed site in luciferase reporter assays (Fig. 4a). We noted that the extent of repression by let-7 was roughly proportional to the absolute value of free energy from transitional nucleation ($|G|$); a mutated G-bulge site with similar $|G|$ showed let-7 dependent repression comparable to that seen with the native A-bulge likely due to G:U wobble pivot pairing, and even a mutated C-bulge site showed some repression, possibly through non-canonical Watson-Crick pivot pairing with extended extra G:U pairing at position 8, while an U-bulge site (a “non-nucleation” bulge) was not functional.

To extend this observation, we also examined miR-708, which has a G pivot nucleotide, and also has the potential to have extended transitional nucleation pairing from the pivot onward (position 6-9), including the possibility of G:U pairing in three of four nucleotides and potentially more promiscuous bulge partners (upper panel of Fig. 4b). Therefore, stable nucleation states of miR-708 are predicted by 5 consecutive base pairs at the 5' end of miR-708 (with a G pivot nucleotide at position 6 as indicated) as well as potential additional G:U and/or A:U pairing in position 8-9, resulting in the prediction that in addition to C-bulges, substitution of other bulge nucleotides might have similar transitional nucleation stability (inset, lower panel of Fig. 4b). We addressed miR-708 bulge interactions by analyzing Ago HITS-CLIP and microarray data following miR-708 transfection of HeLa cells (Supplementary Fig. 6B,C). From among 3433 *de novo* Ago-miR-708 clusters with predicted target sites, 63% matched the canonical miR-708 seed sequence while 37% had bulge sites. All four possible bulge sites were detected, in rough proportion to the predicted free energy of binding to the G-pivot nucleotide in the nucleation state (C: 10%, U: 14%, A: 11% and G: 2%, Fig. 4b), although we recognize that factors other than free energy predictions are likely to contribute to efficiency of nucleation (see discussion below). The 3433 *de novo* Ago-miR-708 clusters with predicted target sites were used for this analysis, since a relatively small number of transcripts with Ago-miR-708 clusters in the brain were

also expressed in HeLa cells (140, versus 423 observed for miR-124). Although the numbers were small, we also observed a similar frequency of seed (72%) and bulge predictions (C: 6%, U: 15%, A: 6%, and G: 0%) in the 47 conserved *de novo* Ago miR-708 clusters (overlapping set of *de novo* miR-708 HeLa cluster and p13 mouse clusters) as in the *de novo* cluster analysis alone. Importantly, transcripts harboring these *de novo* clusters were significantly downregulated after miR-708 transfection (all $P < 0.01$; relative to total transcripts, KS test), and again G-bulge sites were the least effective in miR-708 dependent repression ($P = 0.01$, KS test), correlating with the lowest $|G|$ for G-G interactions. These experiments demonstrate that miR-708 transfection induces Ago-mRNA binding clusters that include bulge sites enabling pivot nucleotide interactions.

Finally, we applied the pivot pairing rule to the previous Ago ternary map established for the 20 most abundant Ago-miRNAs in the brain. This analysis identified 1441 clusters decoded as nucleation bulge sites (2162 clusters when G:U wobble pivot pairing was considered, ~15-22% of Ago-miRNA-mRNA interactions). Using gene ontology (GO) analysis, we examined the functions encoded by these additional targets, and found that they were also enriched in brain-function GO categories, less so than Ago-mRNA targets with canonical seed sites but more than control, non-nucleation bulge targets (Fig. 4c and Supplementary Fig. 7A), suggesting that neuronal functions are also regulated through interactions between nucleation bulges and Ago-miRNAs. These results demonstrate that the application of the pivot pairing rule and identification of miRNA bulge interactions can expand the Ago-ternary map and our understanding of the biological actions mediated by miRNAs.

DISCUSSION

Previous work demonstrated that most miRNA targets are mediated by seed pairing^{6,9-11}. Although early biological studies demonstrated several instances in which bulged sequences are functional¹⁴⁻¹⁷, they have not been recognized as general features of miRNA-mRNA interactions. Recently, Ago HITS-CLIP has been applied to the mouse brain²³, *C. elegans*²⁴ and cultured cells^{25,26}, allowing the identification of *in vivo* miRNA binding sites on a genome wide scale. The ability to map Ago-mRNA binding footprints with high resolution (~45 nt footprints, with ~93% specificity)²³ and the identification of Ago-mRNA orphan clusters motivated us to analyze the genome-wide use of atypical miRNA sites. We observed that Ago binds to a large number of bulged sites *in vivo*, expanding on the observations of individual bulged seed sites¹⁴⁻¹⁷. Importantly, from these *in vivo* miRNA-mRNA interactions, we demonstrate that nucleation bulges occur at a specific position in the mRNA seed site (position 5-6) when that nucleotide is able to bind sufficiently and robustly with the miRNA pivot nucleotide (position 6). On average, the degree of Ago-interaction and repression seen with nucleation bulges are somewhat less than seen with canonical seed sites, but are statistically significant in all experiments and analyses at the same order with a number of studies of suppression reported with canonical seed sites (with significance of $P < 0.05$)³¹⁻³³. The pivot-pairing rule provides a qualitative change in our understanding and assessment of miRNA-mRNA regulation, and is a mechanism conserved across a range of species, from *C. elegans* to mammals.

We propose a “transitional nucleation model” in which the pivot-bulge interaction serves as a general means of enabling a transitional nucleation state by stabilizing nucleation base pairing (position 2-6), allowing subsequent bulge formation and propagation of the seed interaction (Fig. 5). In support, this model is consistent with structural studies of Ago³⁴⁻³⁷. In the structure of the Ago-miRNA binary complex, the same residues (position 2-6) identified here are exposed, enabling to initiate nucleation for mRNA pairing (other nucleotides in seed regions were embedded (position 1) or partially buried (position 7-8))³⁶. In the Ago ternary complex, bulges at position 4-5 and 5-6 in the mRNA but not miRNA can function *in vitro*³⁵ and nucleation and subsequent propagation is believed to be important for cleavage of perfectly matched miRNA-target duplexes³⁷. Interestingly, our nucleation model is presaged by a similar hypothesis in which 4 nucleotides within the seed (position 1-6) were used to calculate the initiation potential of miRNA-mRNA interactions in efforts to improve the prediction of target sites considering secondary structure⁷. Although the nucleation potential calculated from 4 nucleotides was not able to explain the rules of miRNA binding in Ago orphan clusters, these prior studies provide conceptual support for our findings^{7,18}. Here we define the rules for such pairing, and find their frequent occurrence as detected by Ago HITS-CLIP in brain (~15% of Ago-miRNA-mRNA interactions among the top 20 Ago-bound miRNAs) or miRNA transfected HeLa cells (~7-18% or one third of seed sites estimated from brain-conserved *de novo* Ago-miR-708 or Ago-miR-124 clusters, respectively), demonstrating that they are widely used for miRNA-mRNA interactions *in vivo*.

Our results also provide evidence that non Watson-Crick pivot pairings can support bulge formation if they can contribute sufficient ΔG to stabilize the transitional nucleation state. However, non Watson-Crick base pairings are incompletely understood³⁸, such that accurate calculation of ΔG to precisely predict nucleation bulges for all miRNAs is challenging. Moreover, such calculations are likely to serve as only approximate guides to more complex biology. For example, formation of a tethered complex between a Piwi protein from *A. fulgidus*, which contains a MID/PIWI domain like Ago, and a short RNA had a higher seed-to-target binding affinity than did the short RNA in isolation (up to ~300 fold enhancement)³⁹. Thus the contribution of Ago in stabilizing miRNA pivot pairing might contribute more substantially to nucleation than estimated on the basis of free energy calculations alone. For this reason, we limited our analysis to Watson-Crick pivot-pairings for target prediction in this study; while this enabled us to decode new functional sites comprising ~15-22% of Ago-miRNA-mRNA interactions, it may also be of interest to examine non Watson-Crick pivot pairing.

We noted that bulge sites include a potential 5 nucleotide seed sequence (position 2-6). Therefore, in our all analyses we used 7-mer motifs to search for nucleation bulges, which differentiate longer bulge sites, from 5 nucleotide seeds; we also confirmed that sites harboring only 5mer seed matches (position 2-6) are not functional (Supplementary Fig. 5E,F). It is possible that some 6-mer seed matches to position 1-6 might function as 5-mer sites, as there is some indirect⁶ and *in vitro* evidence³⁹ that the first position of a miRNA cannot base pair. However, in our previous Ago HITS-CLIP study²³, we were able to address this question *in vivo* and observed that all 6-mers matching to miRNA position 1-8

are enriched in Ago-mRNA footprints. Further study would be needed to clarify if and when the first position participates in functional miRNA-mRNA interactions.

Although bulge sites account for many of the orphan clusters, there are still remaining unexplained orphans (~25% of the conserved *de novo* Ago miR-124 clusters, Fig. 2a), and these have some evidence for being functional (Fig. 2e). Although we have not found a consensus motif or free energy change associated with miR-124 binding in these remaining orphans (Supplementary Fig. 7B), they may be useful to uncover additional rules or expand the study of non-canonical miRNA binding sites, such as functional “seedless” elements⁴⁰ or centered pairing sites⁴¹. Such remaining orphans could include Ago-mRNA only interactions, such as those containing G-rich motifs recently reported in Dicer null mouse embryo stem cells by performing Ago HITS-CLIP²⁶. Interestingly, our initial analysis of orphan cluster also identified some as harboring G-rich motifs (Supplementary Fig. 1B).

In summary, Ago HITS-CLIP analysis enabled us to identify a new class of miRNA target sites “nucleation bulges” and an alternative mode of miRNA target recognition by a “pivot pairing rule”. From these findings, we propose a transitional nucleation model in which a transitional nucleation state determines the binding of miRNAs to nucleation bulge mRNAs. The identification of functional non-canonical miRNA-mRNA interactions have great importance in understanding mechanisms of miRNA target recognition, discovering new miRNA targets, and applying RNA interference for experimental and clinical purposes, where understanding the specificity of target recognition is especially important.

METHODS

Bioinformatics

In general, bioinformatics analysis used UCSC genome browser (<http://genome.ucsc.edu/>), Galaxy (<http://galaxy.psu.edu>) and Python scripts as described previously²³. The minimum free energy (ΔG) of miRNA-mRNA duplexes were calculated by RNAhybrid⁴². To calculate ΔG of intermediate duplex (transitional nucleation), helix constraint (-f, 2-6) was used. All statistical tests were performed by Scipy (<http://www.scipy.org/>) and other analyses were performed by the Python scripts utilizing Biopython (<http://biopython.org/>). Motif analysis used MEME (<http://meme.sdsc.edu>) and WebLogo (<http://weblogo.threeplusone.com>). Detailed methods for bioinformatics analysis were described in Supplementary Method.

Transfection of miRNA mimics and inhibitors

Transfections of miRNA mimics and inhibitors were performed as described previously²³. Briefly, HeLa cells were transfected using Lipofectamine 2000 (Invitrogen) in 100 mm² plates with 75 nM RNA duplexes, miR-124, miR-708, and negative control miRNA from miRIDIAN miRNA Mimic (Dharmacon) or 75 nM miRNA inhibitors against let-7 and negative control inhibitor from miRIDIAN miRNA Hairpin Inhibitors (Dharmacon).

Luciferase reporter assay

Luciferase reporter assays were performed according to manufacturer's protocol for Dual-luciferase assays (Promega). Psi-check2 plasmid (Promega) was used for luciferase reporter and co-transfected with miR-124, miR-708, control miRNA mimic, let-7 inhibitor or control miRNA inhibitor (Dharmacon) using Lipofectamine 2000 (Invitrogen). Primers for Mink1 (Forward: CCGCTCGAGAGCAGCAAGTAACCCTTCTCCTCCCTCCCCCTCCCC CTCCTCAATGTAG, Seed-reverse: ATTTGCGGCCGCTTAACAAACAGGATATCCAAGGCACTACATTGAGGAGGGGG GAGGGG, G-bulge-reverse: ATTTGCGGCCGCTTAACAAACAGGATATCCAAGGCCACTACATTGAGGAGGGG GGAGGGG, No seed-reverse: ATTTGCGGCCGCTTAACAAACAGGATATCCCTACATTGAGGAGGGGGGAGGGG), for Epb41 (Forward: CCGCTCGAGACATGGAAGTTGCTTCAGATATCTGATACTGTGAATGTTTGAACA TATCCG, G-reverse: ATTTGCGGCCGCAAGGGTAGCTGGAGAGGTGAAGGCCACGGATATGTTCAAAC ATTCACA, A-reverse: ATTTGCGGCCGCAAGGGTAGCTGGAGAGGTGAAGGTCACGGATATGTTCAAAC ATTCACA, U-reverse: ATTTGCGGCCGCAAGGGTAGCTGGAGAGGTGAAGGACACGGATATGTTCAAAC ATTCACA, G-reverse: ATTTGCGGCCGCA AGGGTAGCTGGAGAGGTGAAGGGCACGGATATGTTCAAACATTCACA) and for Kif5b (Forward: CCGCTCGAGTTGAAAAGTAATTGAAGTTTGAAGAGGACATAAAATCAGTCTTT CACTAAC, A-reverse: ATTTGCGGCCGCCTACAATCCCAAGGAATAGAGGTTAGTTAGTG AAAGACTGATTTTATG, U-reverse: ATTTGCGGCCGCCTACAATCCCAAGGAATAGAGGATAGTTAGTGAAAGACTGAT TTTATG, G-reverse: ATTTGCGGCCGCCTACAATCCCAAGGAATAGAGGCTAGTTAGTGAAAGACTGAT TTTATG, C-reverse: ATTTGCGGCCGCCTACAATCCCAAGGAATAGAGGGTAGTTAGTGAAAGACTGAT TTTATG, Seed-reverse: ATTTGCGGCCGCCTACAATCCCAAGGAATAGAGGTAGTTAGTGAAAGACTGATT TTATG) were used to generate fragments (104 nucleotides) containing the bulge and various mutation sites and cloned into downstream of Renilla luciferase in Psi-check2 plasmid. 24hr after transfection, relative activity (renilla luciferase activity normalized to firefly luciferase) was measured by Dual-luciferase assays (Promega) according to the manufacturer's protocol.

Microarrays

RNA from miR-708 or control miRNA transfected HeLa was extracted using Trizol and RNAeasy kit and mRNA was amplified and labeled by the method provided by Affymetrix. Human HuEx 1.0ST Arrays were used and the data were analyzed by using Affymetrix Power Tools as described previously²³.

Ago HITS-CLIP

Ago HITS-CLIP was performed as described²³ using monoclonal Ago antibodies, 2A8⁴³ and 7G1-1* (an old batch of 7G1-1 from Iowa Developmental Hybridoma (IDH) bank, mixture of clones of mouse specific anti-FMRP antibody and anti-Ago antibody, as confirmed previously). Because of its specificity to only mouse FMRP⁴⁴, 7G1-1* could be used as anti-Ago antibody in any human cell line, such as HeLa without peptide blocking FMRP epitope. In brief, miR-708 or control miRNA transfected HeLa were UV-irradiated to covalently crosslink RNA-protein complexes. After lysing the cells, extracts were partially digested with RNase A to reduce the modal size of cross-linked RNA bound to Ago to ~50nt. After immunoprecipitating Ago complex with 2A8 or 7G1-1*, Ago-miRNAs (~110 kD) and Ago-mRNAs (~130 kD) were purified by SDS-PAGE followed by nitrocellulose transfer and further purified RNAs by proteinase K treatment. After generating cDNA libraries by PCR, sequences were analyzed with an Illumina Genome Analyzer. Degenerate barcodes (4 nucleotides tag followed by G) were introduced in the 5' fusion linker to increase the complexity in unique tags and to avoid artifacts from PCR contamination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the members of the Darnell and Hannon laboratories for helpful discussions. This work was supported in part by grants from the National Institutes of Health (R.B.D., G.J.H.) and a grant of the Korean Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (A111989; S.W.C) R.B.D. and G.J.H. are Investigators of the Howard Hughes Medical Institute.

References

1. Sharp PA. The centrality of RNA. *Cell*. 2009; 136:577–80. [PubMed: 19239877]
2. Licatalosi DD, Darnell RB. RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet*. 2010; 11:75–87. [PubMed: 20019688]
3. Ambros V. The functions of animal microRNAs. *Nature*. 2004; 431:350–5. [PubMed: 15372042]
4. He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet*. 2004; 5:522–31. [PubMed: 15211354]
5. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009; 136:215–33. [PubMed: 19167326]
6. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005; 120:15–20. [PubMed: 15652477]
7. Long D, et al. Potent effect of target structure on microRNA function. *Nat Struct Mol Biol*. 2007; 14:287–94. [PubMed: 17401373]
8. Grimson A, et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell*. 2007; 27:91–105. [PubMed: 17612493]
9. Lim LP, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*. 2005; 433:769–73. [PubMed: 15685193]
10. Baek D, et al. The impact of microRNAs on protein output. *Nature*. 2008; 455:64–71. [PubMed: 18668037]
11. Selbach M, et al. Widespread changes in protein synthesis induced by microRNAs. *Nature*. 2008; 455:58–63. [PubMed: 18668040]
12. Mourelatos Z. Small RNAs: The seeds of silence. *Nature*. 2008; 455:44–5. [PubMed: 18769430]

13. Easow G, Teleman AA, Cohen SM. Isolation of microRNA targets by miRNP immunopurification. *RNA*. 2007; 13:1198–204. [PubMed: 17592038]
14. Ha I, Wightman B, Ruvkun G. A bulged lin-4/lin-14 RNA duplex is sufficient for *Caenorhabditis elegans* lin-14 temporal gradient formation. *Genes Dev*. 1996; 10:3041–50. [PubMed: 8957004]
15. Vella MC, Choi EY, Lin SY, Reinert K, Slack FJ. The *C. elegans* microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. *Genes Dev*. 2004; 18:132–7. [PubMed: 14729570]
16. Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*. 2008; 455:1124–8. [PubMed: 18806776]
17. Didiano D, Hobert O. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat Struct Mol Biol*. 2006; 13:849–51. [PubMed: 16921378]
18. Hammell M, et al. mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat Methods*. 2008; 5:813–9. [PubMed: 19160516]
19. Karginov FV, et al. A biochemical approach to identifying microRNA targets. *Proc Natl Acad Sci U S A*. 2007; 104:19291–6. [PubMed: 18042700]
20. Ule J, et al. CLIP identifies Nova-regulated RNA networks in the brain. *Science*. 2003; 302:1212–5. [PubMed: 14615540]
21. Licatalosi DD, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*. 2008; 456:464–9. [PubMed: 18978773]
22. Darnell RB. HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdisciplinary Reviews: RNA*. 2010; 1:266–286. [PubMed: 21935890]
23. Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*. 2009; 460:479–86. [PubMed: 19536157]
24. Zisoulis DG, et al. Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nat Struct Mol Biol*. 2010; 17:173–9. [PubMed: 20062054]
25. Hafner M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 2010; 141:129–41. [PubMed: 20371350]
26. Leung AK, et al. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat Struct Mol Biol*. 2011; 18:237–44. [PubMed: 21258322]
27. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*. 1994; 2:28–36. [PubMed: 7584402]
28. Khan AA, et al. Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat Biotechnol*. 2009; 27:549–55. [PubMed: 19465925]
29. Giraldez AJ, et al. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science*. 2006; 312:75–9. [PubMed: 16484454]
30. Linsley PS, et al. Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol Cell Biol*. 2007; 27:2240–52. [PubMed: 17242205]
31. Kawahara Y, et al. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science*. 2007; 315:1137–40. [PubMed: 17322061]
32. Zhao Y, et al. Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell*. 2007; 129:303–17. [PubMed: 17397913]
33. Gehrke S, Imai Y, Sokol N, Lu B. Pathogenic LRRK2 negatively regulates microRNA-mediated translational repression. *Nature*. 466:637–41. [PubMed: 20671708]
34. Song JJ, Smith SK, Hannon GJ, Joshua-Tor L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*. 2004; 305:1434–7. [PubMed: 15284453]
35. Wang Y, et al. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature*. 2008; 456:921–6. [PubMed: 19092929]
36. Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ. Structure of the guide-strand-containing argonaute silencing complex. *Nature*. 2008; 456:209–13. [PubMed: 18754009]
37. Wang Y, et al. Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature*. 2009; 461:754–61. [PubMed: 19812667]

38. Tinoco I Jr, Bustamante C. How RNA folds. *J Mol Biol.* 1999; 293:271–81. [PubMed: 10550208]
39. Parker JS, Parizotto EA, Wang M, Roe SM, Barford D. Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol Cell.* 2009; 33:204–14. [PubMed: 19187762]
40. Lal A, et al. miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to “seedless” 3'UTR microRNA recognition elements. *Mol Cell.* 2009; 35:610–25. [PubMed: 19748357]
41. Shin C, et al. Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell.* 2010; 38:789–802. [PubMed: 20620952]
42. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. *RNA.* 2004; 10:1507–17. [PubMed: 15383676]
43. Nelson PT, et al. A novel monoclonal antibody against human Argonaute proteins reveals unexpected characteristics of miRNAs in human blood cells. *RNA.* 2007; 13:1787–92. [PubMed: 17720879]
44. Brown V, et al. Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell.* 2001; 107:477–87. [PubMed: 11719188]
45. He L, et al. A microRNA component of the p53 tumour suppressor network. *Nature.* 2007; 447:1130–4. [PubMed: 17554337]

Figure 1. Identification of G-bulge sites pairing to miR-124 by Ago HITS-CLIP analysis

(a) Over-represented motifs in orphan clusters, in which Ago footprint regions have no predicted seed matches among the top 20 Ago-miRNAs families (left panel, Supplementary Fig. 1A). 6 significantly enriched motifs (E-value < 0.01) were identified in the orphan Ago footprint regions by MEME analysis (Supplementary Fig. 1B) and their E-values (MEME expectation) are indicated in middle panel. Right, distribution of the motifs relative to peaks of Ago mRNA clusters with the same colors represented in middle panel. UGGCCUU (orange line) is the most significant enriched motif near peaks ($k = 2.2$ vs. $k = 1.8$ in uniform distribution as control). (b) UGGCCUU (blue) is a G-bulge (in position 5-6) match to miR-124 (position 2-8, red). The bulge nucleotide (between position 5 and 6) is highlighted. (c) MEME analysis of the G-bulge motif in 2392 orphan *de novo* Ago miR-124 clusters after miR-124 transfection into HeLa cells²³. Relative height of the individual bases represents the frequency; error bars indicate sampling errors caused by small number of sample at each position. (d) The position of seed (red) and 4 possible bulge matches (G; orange, C; green, A; purple, U; blue) to miR-124 are plotted relative to the peak of 3083 *de novo* Ago miR-124 clusters (normalized, BC = 2). Of note, *de novo* miR-124 clusters could be generated as the consequence of miR-124 overexpression which reduced or precluded Ago binding to sites occupied in untransfected cells, as previously observed.^{23,28}

Figure 2. Validation of G-bulge sites in Ago miR-124 clusters

(a) Pie chart showed seed (red, 105 clusters), G-bulge (orange, 36 clusters), C-bulge (green, 6 clusters), A-bulge (cyan, 5 clusters), U-bulge (purple, 2 clusters) and others (blue, 52 clusters) in 206 conserved de novo Ago miR-124 clusters. (b) A G-bulge site (middle panel: black in dotted box, seed sites; red) of *Mink1* 3'UTR (Hs: human, Mm: mouse) in conserved de novo miR-124 clusters (top panel: red) and Ago mRNA clusters from brain (bottom panel: grey). (c) Luciferase reporter assay for the wild type G-bulge site ("G-bulge": GUGGCCUU) or other mutant sites ("Seed": GUGGCCUU; "No site": a deletion of GUGGCCUU; "Luc": luciferase vector with no insert) of *Mink1* in the presence of a miR-124 (red bar) or a control miRNA mimic (blue bar). Relative activity, average renilla luciferase activity normalized to firefly luciferase in three replicates; error bars, s.d. Asterisks denote instances of $P < 0.01$ (t-test). (d) The same luciferase reporter assay as (c) but performed for a G-bulge site in *Epb41* with more single-point mutations in bulged position. Asterisks denote instances of $P < 0.05$ (t-test). (e) Transcripts with G-bulge sites in de novo Ago miR-124 clusters (orange line) showed miR-124 dependent suppression relative to previous analysis of all regulated transcripts in miR-124 transfected HeLa cells⁹. (f) Numbers of miR-124 dependent transcripts containing seed (150 transcripts), G-bulge (98 transcripts) and other bulge sequences (U;19, A;13, C;17 transcripts) in de novo Ago miR-124 clusters are plotted in all ranges of repression.

Figure 3. Nucleation bulges are widely used and evolutionarily conserved as functional miRNA target sites

(a) Minimum free energy (ΔG) calculation for miR-124 pairing to a G-bulge site (left) and a transition nucleation model, showing a nucleation bulge (middle) with pivot pairing (blue shade) and a non-nucleation bulge (right) with no pivot pairing (grey shade). (b) The position of seed (“Seed”;), nucleation bulge (“Nuc”; and non-nucleation bulge sites (“Non-nuc”; from the top 30 Ago-bound brain miRNAs plotted relative to the peak of 11,463 Ago footprint regions (BC 2). (c) Distribution of conservation rates for three different sites (seed, nuc and non-nuc) of mouse miRNAs and all heptamers (dotted line) in 3' UTRs. (d) Average conservation rates of the three different sites (as in (c)) for the top 20 Ago-bound miRNAs were calculated in Ago footprint regions; error bars indicate standard error. (e) Meta analysis of brain-expressed transcripts harboring the three different sites (as in (c)) in compiled microarray data²⁸ using 7 different miRNA transfections. (c-e) are colored as in (b). (f) Numbers of the 7 different miRNA dependent transcripts containing the three different sites analyzed in (e) plotted in all ranges of repression. (g) Cumulative distribution of miR-430 orphan transcripts. (h) Meta-analysis of orphan transcripts for 8 miRNA families. (i) Linear regression analysis to comparing the number of conserved miRNA seed matches (7-mers to position 2-8) or (j) nucleation bulges (7-mers) with the frequency of the top 100 miRNAs experimentally determined by mouse brain Ago HITS-CLIP. (k) Non-nucleation bulges analyzed by linear regression (7-mers).

Figure 4. Functional nucleation bulges in let-7, miR-708, and GO analysis

(a) Transitional nucleation model of let-7 (top) and luciferase reporter assay (bottom) for the A-bulge site in *Kif5b* (Supplementary Fig.6). Luciferase assays were performed as described in Fig. 2d except reporter activities were compared between let-7 inhibitor transfected (let-7-in) and control transfected HeLa cells (Control), to inhibit high endogenous levels of HeLa let-7. Asterisks denote instances of $P < 0.01$ (t-test). Bracket without asterisk indicates $P < 0.05$ (t-test). (b) A transition nucleation state model of miR-708 (top) and free energies of transitional nucleation (ΔG, bottom). The frequencies of seed and different bulge sites identified in *de novo* clusters from miR-708 transfected HeLa are also indicated in lower panel as percentage; composition of seed and bulge motifs in the Ago footprint region (64nt): seed (red, 864 clusters), C-bulge (brown, 143 clusters), U-bulge (orange, 185 clusters), A-bulge (cyan, 153 clusters) and G-bulge (purple, 28 clusters). Cumulative distribution of transcripts with canonical seeds or each type of bulge (colored as indicated) in conserved *de novo* Ago miR-708 clusters are shown with all transcripts (black) in lower panel. (c) Heat maps derived from gene ontology (GO) analysis of target transcripts for each of the top 20 miRNAs based on Watson-Crick pivot pairing versus seed rules (seed sites; Seed, nucleation bulge; Nuc and non-nucleation bulge; Non-nuc) show false discovery rate (FDR). The tree indicates the hierarchical clustering of GO sub-categories for brain function.

Figure 5. Pivot pairing and transitional nucleation models

Nucleation bulges enabling pivot pairing (upper panel) induce transitional nucleation (5 consecutive pairing in position 2-6) and initiate miRNA-mRNA duplex by stabilizing thermodynamics (e.g., $-11.6 \text{ kcal mol}^{-1}$; miR-124). This transition state is followed by formation of a bulge (position 5-6) and propagation of base pairing distally, leading ultimately to post-transcriptional repression by Ago-miRNA. Non-nucleation bulges without pivot pairing only make 4 consecutive matches, probably not stable enough for nucleation to initiate miRNA-mRNA duplex (e.g., $-7.2 \text{ kcal mol}^{-1}$; miR-124, lower panel).