**The contribution of both the miRNA seed- and 3′- region to target RNA binding affinity are not independent, and vary for different miRNAs.**

**Target features that determine binding thermodynamics to the 3′-end of the miRNA**

**Thermodynamics of miRNA 3′-pairing relies on an interplay of miRNA-independent and dependent factors.**

Namita Bisaria1,2,3,6, Sean E. McGeary1,2,3,6, Kathy S. Lin1,2,3,4 and David P. Bartel1,2,3,4,5

1Howard Hughes Medical Institute, Cambridge, MA, 02142, USA

2Whitehead Institute for Biomedical Research, Cambridge, MA, 02142, USA

3Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

4Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

5Correspondence to: dbartel@wi.mit.edu.

6These authors contributed equally to this work.

**Abstract:**

The current understanding of animal microRNAs (miRNAs) largely pertains to the nucleotides at positions 2–8 of the miRNA guide sequence, referred to as the seed region. The function of ~14 positions 3′ of the seed, which are highly conserved for many miRNAs, remains enigmatic, with discordant findings across a handful of studies. We therefore performed RNA bind-n-seq (RBNS) with purified AGO2–miRNA complexes and libraries with partially randomized RNA sequence to measure the relative binding affinity of over ~12,000 distinct 5′- and 3′-bipartite pairing configurations for each of three naturally occurring miRNAs and five synthetic derivatives. These data demonstrate upwards of 100-fold increase in binding affinity contributed by extensive pairing to the miRNA 3′ region in comparson to seed pairing alone. In addition, we find that the extent of this increase is highly variable between miRNAs, and is influenced by 1.) the predicted stability of the pairing between the miRNA 3′ region and the target RNA sequence, 2.) the length and sequence content of the unpaired segment of the target RNA between the miRNA seed– and 3′–paired regions, and 3) the specific miRNA–target pairing configuration within the seed region. Our results suggest that pairing to the 3′ end may be utilized by miRNAs to enhance affinity and modulate specificity.

**Introduction:**

miRNAs are ~22-nt regulatory RNAs that are processed from hairpin precursors and, upon association with an Argonaute (AGO) protein, pair to sites within mRNAs to direct their destabilization and translational repression {Guo:2010kh}{Eichhorn:2014db}{Bartel:2018gz}. For most sites that confer repression in mammalian cells, pairing to miRNA nucleotides 2–7, referred to as the miRNA seed, is critical for target recognition, with an additional pair to miRNA position 8 or an A across from miRNA position 1 often enhancing targeting efficacy {Lewis:2005cb}{Bartel:2009fh}{Schirle:2015ct}. Such sites with a perfect 6–8-nt matches to the miRNA seed region (Figure 1A) are heuristically predictive of repression, with longer sites being more effective than shorter ones and more sites being more effective than fewer sites {Grimson:2007cy}{Agarwal:2015bw}. Features of site context can also influence targeting efficacy. For example, sites are more effective if they reside near to other sites that can act cooperatively, within a region that does not form occlusive secondary structure, and in the 3′-untranslated region (UTR), i.e., out of the path of the scanning initiation complex and the ribosome {Grimson:2007cy}{Wan:2014ih}.

Pairing to the miRNA 3′ region, particularly nucleotides 13–16, can supplement perfect seed pairing to enhance targeting efficacy, and more extensive pairing to this region can compensate for imperfect seed pairing to enable targeting {Grimson:2007cy}. These two bipartite site types are referred to as 3′-supplementary and 3′-compensatory sites, respectively (Figure 1A). Although much less common than sites with only a seed match, thousands of sites with preferentially conserved 3′-supplementary pairing are present in human 3′-UTRs {Friedman:2009km}. Indeed, while conserved 3′-compensatory sites are even less common, such sites can have crucial biological functions. For example, let-7-mediated repression of *lin-41* is required for viability in *C. elegans* through two 3′-compensatory {Reinhart:2000dh}. Moreover, the use of 3′-compensatory sites rather than canonical sites provides a mechanism by which to confer differential target specificity between miRNAs that share common seed sequences but differ at their 3′ regions, which is proposed to prevent precocious repression of *lin-41* by earlier-expressed let-7 paralogs (REFs).

Although analyses of site conservation can detect the preferential conservation of 3′-supplementary and 3′-compensatory pairing {Grimson:2007cy}, and a global contribution to targeting efficacy can be detected for 3′-supplementary pairing after introducing a miRNA to cells (REF), the biochemical contribution of pairing to the miRNA 3′ region is poorly understood. Studies of a few target sites of a few miRNAs indicate that pairing to the 3′ region of the miRNA confers either negligible-to-small (2–7-fold) {Wee:2012df}{Salomon:2015fb} or moderate (>20-fold) increases in binding stability. With a limited amount of biochemical information, it is difficult to understand the role of pairing to the miRNA 3′ region (hereafter referred to as “3′ pairing”) for different miRNAs and targeting contexts. The variation across miRNA 3′-compensatory sites can decomposed into four distinct characteristics, being 1.) the length of contiguous pairing of the site to the miRNA 3′ end, 2.) the 5′-most nucleotide of paired 3′ region of the miRNA (hereafter referred to as the “register” of pairing, 3.) the difference between the number of unpaired target nucleotides and that of the miRNA (hereafter referred to as the “3′-pairing offset”, and 4.) the nature of the imperfect pairing to the the seed. (Fig 1B). Without a more general model of miRNA 3′ pairing and the consistency of its binding capacity between different miRNAs, it is difficult to predict the occurrence and efficacy of these types of sites *in vivo*.

RNA bind-n-seq (RBNS) allows for unbiased, high-throughput assessment the relative binding activities of RNA *k*-mers of variable length embedded within a larger random-sequence context {Lambert:2014jm}{Dominguez:2017fk}. We recently adapted RBNS for the study of miRNA targeting, and we built a computational analysis pipeline enabling calculation of relative equilibrium dissociation constants (*K*D values) of many thousand distinct RNA sequences. These *K*D values allowed for quantitative comparisons of putative site types and sequence features that would not have been possible using *k*-mer enrichment due to their variation as a function of the concentration of the AGO–miRNA complex in the binding reaction {McGeary:2018jl}. Applying this AGO-RBNS platform to AGO–miRNA complexes with six different miRNAs revealed unanticipated site preferences for these miRNAs and enabled a principled biochemical model of miRNA-targeting efficacy {McGeary:2018jl}.

Given the current absence of a systematic characterization of the features of 3′ pairing that contribute to affinity for even a single miRNA sequence, AGO-RBNS provides an ideal platform to construct such *a prior* profiles, and to compare such profiles across miRNAs to identify any consistent features of this pairing. However, AGO-RBNS performed with fully randomized libraries does not provide information on sites with more than ~5 supplementary/compensatory pairs because such sites, which involve >12 nt of total pairing (Fig 1A), are too rare in input random library sequencing for accurate calculation *K*D values. Here, we construct RNA libraries suitable for characterizing 3′-compensatory sites and use these libraries to generate AGO-RBNS data that enable the systematic, comprehensive analysis of pairing to the miRNA 3′ region for individual miRNA sequences.

**Results**

**Binding-affinity profiles for 3′-compensatory sites of let-7a**

As previously implemented, AGO-RBNS begins with the incubation of purified AGO–miRNA complex with an RNA library that contains a central region of 37 random-sequence nucleotides flanked by constant sequences enabling construction of RNA-seq libraries (Fig 2A, top). Five binding reactions were performed, each with a constant library concentration of 100 nM and a different miRNA-AGO concentration spanning a 100-fold range. After reaching binding equilibrium, each reaction is applied to a nitrocellulose filter membrane under vacuum, which retains the AGO–miRNA complex and any bound RNA library molecules. The bound library molecules are isolated and subjected to high-throughput sequencing, along with the input RNA library. This allows for the fractional abundance of *k*-mers of various lengths in the bound libraries to be compared to that of the input library (Fig 2B). For any *k*-mers of interest that are ≤12 nt, relative *K*D values can be fit to its observed abundance profile constructed from multiple samples spanning a range of AGO–miRNA concentrations, using a biochemical model of the binding equilibrium between the *k*-mers within library sequences and the AGO–miRNA complex. This procedure simultaneously determines relative *K*D values for tens of thousands of *k*-mers, which allows the binding affinity of the same 5–8 nt site of a particular miRNA to be evaluated in the context of many longer *k*-mers, thereby report on the effects of different flanking nucleotide possibilities {McGeary:2018jl}.

3′-compensatory sites typically require >12 nt of pairing (Fig. 1A), and so would be expected to be sequences less than twice for every million reads from a 37-nt random library. Such extreme rarity of relevant sites renders these libraries fundamentally unsuited for comprehensive identification and dissection of the complex features within the seed and 3′ region that together modulate overall binding affinity. We therefore constructed an RNA library to investigate the 3′-compensatory sites of let-7a; this library included a programed region of imperfect seed pairing with 25- and 4-nt of random-sequence RNA (Fig 2A, bottom) separating the programed region from the 5′ and 3′ constant sequences, respectively. The programed region contained imperfect seed pairing with a match to let-7a at position 8, an A nucleotide across from position 1, and a match at all but one position to the seed nucleotides 2–7, such that each library molecule contained one of 18 possible single-nucleotide seed mismatches (including wobbles) in approximately equal proportion. With this programed region of imperfect seed pairing, each library contained 3′-compensatory sites at a ~16,000-fold greater frequency than expected with a fully randomized RNA library.

AGO-RBNS performed with let-7a and this programmed RNA library yielded binding-affinity profiles for a massive number of bipartite site architectures combinatorially varying in pairing to both the seed region and the miRNA 3′ region—two regions of each bipartite site hereafter termed the “seed site” and “3′ site,” respectively. Within the randomized region 5′ of the programed seed site (Fig 2A), most library molecules varied with respect to 1) the length of the 3′ site, 2) the register of the 3′ site, as defined by the 5′-most miRNA position paired to the 3′ site, 3) the position and nucleotide identity of the seed mismatch, and 4) the length and sequence composition of the loop in the target that bridges the seed site and 3′ site (Fig 1B).

AGO-RBNS was performed using this programed library and purified AGO2–let-7a. Anticipating that 3′ sites would have preferred registers and preferred loop lengths, we searched for the top enriched *k*-mer, considering enrichments at each position across the read of the bound library molecules. Among *k*-mers ≤8 nt, the most enriched one was AUACAACC—the perfect Watson–Crick match to positions 11–18 of the let-7a miRNA, which was enriched up to 16-fold in the sample with 840 pM AGO2–let-7a (Fig. 2C). This 8-nt 3′ site was most strongly enriched when occurring at positions 14–21 of the library, thereby creating an asymmetric internal loop between seed pairing and 3′ pairing, which included miRNA positions 8–9 and library positions 13–9 (Fig. 2C, right). This 8-mer 3′ site was also ≥5-fold enriched at eight other contiguous positions of the library, indicating that other lengths of the loop in the target were functional, albeit to a lower degree (Fig. 2C).

The second-most enriched *k*-mer was UACAACCU—the perfect Watson–Crick match to let-7a positions 10–17 (Fig 2C). Although this 3′ site also had an optimal length of 5 nt separating the seed site and the 3′ site, its distribution of suboptimal loop lengths was shifted to be about 1 nt shorter than that of the AUACAACC site, consistent with the single-nucleotide shift in its pairing register. The next 18 most enriched *k*-mers revealed additional 3′-site possibilities, together with their preferred and tolerated pairing registers and loop lengths (Fig 2C). Each had a contiguous stretch of 6–8 perfect Watson–Crick pairs to the let-7a 3′ region, with no bulges, wobbles, or mismatches, and usually including the ACAACC 6-mer, which matched miRNA positions 11–16.

**3′ pairing can substantially increase affinity and compensate for a seed mismatch**

For more comprehensive examination of 3′ sites of varied lengths and within different contexts (Fig. 1B), we enumerated, for each of the 18 seed-mismatch possibilities, 3′ sites of lengths 5–9 nt that perfectly paired to the miRNA starting at or 3′ of miRNA nucleotide 9, allowing loops of 0–17 nt to separate the seed site from the 3′ sites. Of these 12,960 site possibilities, 12,638 has sufficient read coverage for analysis, and for each of these, the enrichment profile across AGO2–let-7a concentrations was used to calculate its relative *K*D value, using a previously described computational pipeline (McGeary, Lin, et al., 2019). These relative *K*D values spanned a >100-fold range, with strong agreement observed between the results of replicate experiments performed independently with different preparations of both the AGO2–let-7a complex and the let-7a 3′-compensatory RNA library (*r*2 = 0.86, Fig 3A). Furthermore, for 3′-compensatory sites that had 5-nt 3′ sites, which by virtue of their small number of specified residues could be evaluated using results from a conventional random-sequence library, the *K*D values determined from the programmed library correlated well with those determined from a random-sequence library (Fig S3A), indicating that the fitting procedure was robust to library composition.

Indeed, the majority of the 12,638 potential 3′-compensatory sites examined had *K*D values resembling those of their seed-mismatched sites alone, as might be expected when performing an unbiased analysis of putative sites. However, the detection of \_\_\_\_\_\_ 3′‑compensatory sites with affinities as highas that of canonical 7mer sites (Fig 3A) motivated us to perform a systematic investigation of which features of the seed- and 3′-portion of a 3′-compensatory site can together impart such affinity.

We first examined the relative *K*D values of 3′-compensatory sites as stratified by both 3′-site pairing length and target-RNA loop length (Fig 3B). Although all possibilities examined were at least 1.5-fold better than the seed-mismatched site alone, only 3′ sites ≥6 nt in length and with loop lengths ranging from ~3–8 nt conferred affinities exceeding those of the canonical 6-nt site. As also indicated in our analysis of enriched 8-nt k-mers across from the miRNA 3′ end (Fig 2C), the optimal loop length for 3′ sites beginning at register 11 was 5 nt (Fig 3B). At this loop length, an 8-nt 3′ site conferred a ~60-fold increase in affinity over the seed-mismatched site alone, enabling the affinity of the 3′-compensatory site to surpass those of the canonical 7-nt seed sites and approach that of the canonical 8-nt seed site (Fig 3B).

This 60-fold increase in affinity conferred by optimally positioned site with 7-nt of 3′ pairing was substantially greater than the 2-fold increase observed for 3′ pairing in the context of a perfectly paired target (Wee et al., 2012), perhaps because of the shorter suboptimal loop length. It was also somewhat greater than the 20-fold increase measured for miR-122 pairing to the GCGG site at register 13 and loop-length of 4 nt (MacRae REF). Nonetheless, it was substantially less than expected based on the predicted free energy of RNA pairing in solution. For example, the predicted free energy (∆*G*) of a 7-nt segment paring to let-7a positions 11–17 is 7.1 kcal/mol, which corresponds to a 164,000-fold difference in binding affinity (from ∆*K*eq = *e*–∆∆*G*/*RT*). This large discrepancy is presumably due to the energetic costs of 1) displacing favorable contacts between the miRNA 3′ region and AGO2, and 2) undergoing any conformational changes needed to accommodate 3′ pairing (REF haley). Indeed, the benefit of extending the 3′ site plateaued at about 7 or 8 nt (Fig 3B), implying that these costs were even more severe when pairing to the last five nucleotides of let-7a.

**The type of seed-mismatch affects the affinity of 3′ pairing**

To examine the influence of position and nucleotide identity of seed mismatches, we examined another slice through the multidimensional feature space of 3′ pairing, holding 3′-site length, register, and loop length at a near-optimal values (pairing to let-7a positions 11–17 with a loop length of 4 nt) and examined the overall binding affinity of the 18 different seed mismatch types (Fig 3C). We observed a >100-fold range in relative *K*D for this seed- and 3′-site architecture depending on position and identity of the seed mismatch; by comparison, we observed only a XX-fold range in relative *K*D for these 18 seed mismatch sites in the absence of any 3′ pairing. This demonstrates that the energetics of the seed- and 3′-sites are not independent, such that the potential benefit of 3′ pairing is differentially available when paired with different seed mismatch types. To directly analyze the dependence of 3′-pairing on seed mismatch type, we divided the relative *K*D value of each 3′-compensatory site by that of its seed site to generate a *K*D fold-change value representing the contribution of the 3′ site to overall affinity (Fig 3C). Even after this normalization for the differential effects of the 18 mismatches on seed-site affinity, a 11.7-fold range in affinities was observed at the optimal loop lengths (Fig 3D). This variation was reproducible, with only 1.X-fold attributable to experimental variability (Fig. S3), and it did not correlate noticeably with either the type of seed mismatch or the measured affinity of the seed site (Fig 3D). Notably, a let-7a compensatory site that occurs naturally within the 3′ UTR of *C. elegans* *lin-41* mRNA contains a seed mismatch, register, and loop length with one of the higher 3′-pairing affinity contributions (Fig 3D, bottom).

The strong effect of the seed-mismatch identity suggested a long-range conformational coupling allowing the pairing state at the seed region to communicate with the pairing potential at the 3′ region. This inferred coupling is reminiscent of mismatches between the CRISPR-Cas13a guide RNA and RNA target having differential effects on HEPN nuclease activation compared to their effects on binding affinity (Tambe et al., 2018). Structures of loaded human AGO2 with targets that pair to the miRNA seed (g2-g7, g2-g8, and g2-g9) display a change in the 3′ residues of the miRNA to be more available for pairing than without a seed-matched target (REF), suggesting there exists coordination to relay the seed-pairing state to protein residues that contact the 3′ end of the miRNA. Moreover, in the recent structure of loaded AGO2 bound to a fully complementary target, the seed conformation is slightly different, in particular at guide nucleotides 7 and 8 (REF). It is possible that the criteria for the conformational change is simply having a target bound such that any seed-match or mismatch will have an equal likelihood of promoting the conformational change of the 3′ end. In this scenario, the contribution of the 3′ pairing to overall binding affinity would be the same regardless of different amounts of seed pairing, or mismatches. On the other hand, various mismatches in the seed region may promote or inhibit this conformational change, and thus change the fraction of time the complex is competent to make 3′ base pairs. There is precedent for analogous behavior; it has been shown previously that a G-G mismatch at position 6 in the miRNA guide promote zebrafish’s Argonaute2 target cleavage reaction (Chen et al., 2017). This second model predicts that the contribution of 3′ pairing to overall binding affinity would be different for different seed-mismatches.

**The preferred pairing register differs for different miRNAs**

Repeating the analysis of Fig 3D for each of the seven possible pairing registers for a 7-nt 3′ site for let-7a revealed that the register of pairing had a dramatic effect on the affinity of compensatory pairing. This affinity sharply diminished at register 12 and did not exceed background levels at registers 13–16 (Fig 4A, top row). At registers 9 and 10, affinity of compensatory pairing was retained, although somewhat diminished. At register 9, two loop-length optima were observed, one of which was at a loop length of 0 nt, consistent with pairing of seed and 3′ sites forming a single contiguous helix. However, in the context of most seed mismatches, the optimum that occurred at a loop length of 2 nt was somewhat more favorable, implying preference for non-contiguous pairing. Indeed, even at a loop length of 0 nt the potential pairing at positions 9 and 10 might not form, as indicated by the similar affinities observed at this loop length when substituting a U for the A at target position 9 (Fig SX). This apparent lack of pairing to miRNA position 9 indicates that, as with the miRNA 3′ terminal nucleotides, the energetic costs of displacing favorable contacts and undergoing conformational changes prevent stable pairing to miRNA nucleotide 9.

The optimal register for 3′ pairing to let-7a, occurring at miRNA nucleotides 11–17 (Fig 4B), was somewhat offset from the optimal register previously determined from the repression and conservation data, which centered on miRNA nucleotides 13–18 (Fig. 1A) (REF Grimson). Because the previous analyses pooled data from multiple miRNAs, and because different miRNAs can have diverse site-type preferences (McGearyREF), we acquired the 3′-pairing profiles of two other miRNAs, miR-1 and miR-155, for comparison to the let-7a profiles. As with let-7a, we synthesized programed libraries enriched for all possible single-nucleotide seed mismatches at positions 2–7, performed AGO-RBNS, and calculated *K*D, Rel values for 7-nt 3′ sites in all possible registers, over a 0–16-nt range of loop lengths, in the context of each of the 18 seed mismatches. Normalizing for the contribution of the mismatched seed sites yielded 3′-pairing profiles at each register (Fig 4A).

Clear evidence for stabilizing 3′ pairing was observed for miR-1 at registers 11–13 and for miR-155 at registers 10–13 and 15–17 (Fig 4C). As observed for let-7a, 3′-site affinity for miR-1 and miR-155 varied with the position and identity of the seed mismatch. The optimal register for 3′ pairing to miR-1 occurred at miRNA nucleotides 12–18, whereas the optimal register of 3′ pairing to miR-155 occurred at miRNA nucleotides 15–21 (Fig 4C and D). Thus, on the whole, the different optimal pairing registers observed for the three miRNAs (11–17, 12–18 and 15–21) concurred with the conclusion derived from pooling repression and conservation data from multiple miRNAs, which stated that optimal 3′ pairing centers on miRNA nucleotides 13–18 (Grimson). However, the current results revealed unanticipated differences between miRNAs in optimal register of pairing. We also observed differences between miRNAs in the strength of 3′ pairing; compared to 3′-site affinities observed for let-7a, affinities were somewhat lower for miR-1 and miR-155 (maximal *K*D fold-change values for 7nt of 3′ pairing, 43, 15, and 14 for let-7a, miR-1 and miR-155, respectively) (Fig 4D). Thus, our results indicate that the sequence of the miRNA, not common interactions with the protein, influence the contribution and the preferred register for 3′ pairing.

These trends were recapitulated using AGO-RBNS data from random-sequence libraries (McGeary et al., 2018), albeit with more error due to lower counts of site occurrences, even when confining our analysis to shorter, 5-nt 3′ sites (Fig S4 ). Of the six miRNAs examined, lsy-6 had the highest-affinity 3′ pairing—up to 100-fold for a 5-nt 3′ site pairing to miRNA nucleotides 13–17. In contrast, miR-7 had the lowest-affinity 3′ pairing, with little signal for a contribution of 5-nt 3′ sites detected above background (Fig S4).

Our finding that the identity and position of the seed mismatch can exert such a striking influence on the added affinity conferred by 3′ pairing raised the question of how this added affinity might differ for canonical sites, which have no seed mismatches. To answer this question, we analyzed the AGO-RBNS results from random-sequence libraries, focusing on 5-nt 3′ site sites in the optimal register and aggregating results for the 5 most optimal loop lengths (Fig S5). Overall, the additional affinities conferred by 3′ pairing associated with canonical sites fell within the range of affinities conferred by 3′ pairing associated with seed-mismatched sites (Fig S5). Thus for most miRNAs, even the highest-affinity seed-matched sites can be improved with optimally positioned 3′-supplementary pairing. The exception was miR-7, for which a contribution of 3′ pairing was difficult to detect for both seed-mismtached and seed-matched sites.

**The optimal 3′-pairing register depends on the identity of the 3′ sequence.**

To begin to examine the determinants of the optimal 3′-pairing register, we repeated the AGO-RBNS procedure with let-7a variants that had single-nucleotide insertions and deletions that shifted the let-7a 3′ sequence by a single nucleotide in either direction while maintaining the length of the miRNA (Fig 5A, let-7a(−1) and let-7a-(+1)). Analysis of these new datasets revealed that the optimal 7-nt 3′ site (UACAACC) remained the same, regardless of the shift in the register of the corresponding segment within the let-7a variant. Thus, the preferred register of 7-nt 3′ sites, which was 11–17 for let-7a, shifted to 10–16 for let-7a(−1) and to 12–18 for let-7a(+1) (Fig 5B and 5C). These results showed that particularly favorable 3′ pairing, such as that involving nucleotides 11–17 of let-7a, can dictate the preferred register of pairing and that this phenomenon can help to confer different optimal registers to different miRNAs.

This observation can be further quantified by comparing the correlation between 3′ pairing matrices of the register shifted let-7a variants to the wildtype let-7a matrices (Fig S4); specifically, register 10 and 11 are most correlated between the minus1/WT 3′ pairing matrices, and 12 and 11 for the plus1/WT 3′ pairing matrices (R2 = 0.9x and 0.9y, respectively). The high correspondence of 3′ pairing contribution for offset registers to offset guide miRNAs strongly suggests that the sequence of the 3′ end of the miRNA dictates its preferred register for pairing.

**AGO2 inhibits pairing to the end of most miRNAs but can be overcome by basepairing energetics**

Structures of human AGO2 show that the PAZ domain holds the last 3-4 3′-nucleotides of the miRNA in a conformation that might preclude pairing. Moreover, in structures that contain the a guide-target pairing that forms supplemental pairing, the 3′ end is still bound by the PAZ domain (Sheu-G). Since the contribution of 3′ pairing to overall affinity is much smaller than expected for RNA-RNA interactions, this, along with previous biochemical and aforementioned structural data suggests that the protein inhibits access to the 3′-most residues of the miRNA (Model 1 Fig S5A; Haley, Wee,Soloman). On the other hand, there is evidence for miRNA target sites that include extensive and exclusive (REF McGeary) pairing to the end of the miRNA and machinery that recognizes the 3′ of the guide to tail and trim the guide, indicating there are cases where the 3′ of the guide is free from the protein. It could be that a certain amount of pairing energy to a target may be sufficient to favor a state where the 3′ end of the miRNA is dislodged from the protein and bound to a target (Model 2 Fig S5A).

Fig 4B show that for let-7a and miR-1, the 3′ pairing contribution matrices do not have signal for registers that correspond to pairing to the end of the miRNA, while miR-155 shows a relatively stronger signal for pairing at the end of the miRNA compared to registers around guide nucleotides 11-13. To determine which guide nucleotide to which pairing iis inhibited, we compared the 3′ pairing contributions of a 7nt target to the contribution 6nt region that spans the 7nt target except for the final (5′) residue. This comparison will determine whether the addition of that extra nucleotide confers and increase in affinity. For let-7a, comparing the 3′ pairing contribution matrices to g11-g17 to g11-g16 showed an increase in contribution due to the addition of g17 (Fig S5B). A similar comparison for g18 shows a slight increase, while g19 showed no increase in affinity. miR-1 shows an increase in affinity due to the addition of g18 and g19, but not g20 (Fig S5C). These data quantitatively support the model that let-7a and miR-1 don’t use pairing past g18 and g19 respectively, which encompasses the last 3-4 nt of the guide. Similar comparisons for miR-155 show that guide nucleotides 21-23 all contribute to a 7nt site of 3′ pairing affinity (Fig S5D), strongly suggesting that miR-155 pairs through the end of the miRNA. An alternative model is that the binding signal for miR-155 at these 3′-most registers is due to previously described “3′-only” sites, whereby the 3′ end of the miRNA pairs to the target without seed-sites pairing (McGeary, Lin, *et al*., 2019). While we are unable to deconvolute the enrichments due to binding through these two modes, there are variations in the binding energy across different seed-mismatches for miR-155 at these 3′-most registers, supporting the model that these energetics are, in some part, due to bipartite binding at both the seed and 3′ sites (Fig 4B).

These data suggest that the miR-155 pairing stability at these registers is sufficient to stabilize the target bound state compared to being held in the protein, unlike let-7a and miR-1. Notably, miR-155 includes four consecutive guanosine residues in its far 3′ end (g19-g22), which represents as significant amount of potential binding energy. Furthermore, miR-155 is 23 nucleotides long, which others (Sheu-G) have shown slight increases in the dissociation rate of target sequences compared to shorter guides (22 and 21 nt). These data suggest that strong “enough” pairing to the 3′ end likely can dislodge the miRNA 3′ end, and expose it. We speculate that these types of miRNAs, specifically longer miRNAs with strong pairing at their 3′end, may be more prone to target-mediated degradation (REF).

**The seed-mismatched and 3′-sequence effects act independently**

The large range of 3′ pairing stabilities observed across different seed mismatches and positions (Fig 2C) raised the question of whether these differences depended on the seed-mismatch type and position alone, or whether this variability also depended on the sequence of the 3′ region. To answer this question, we performed AGO-RBNS with two synthetic guide RNAs: one being a chimeric guide sequence bearing the seed miRNAs, one fusing the seed of miR-155 to the 3′ region of let-7a and the other fusing seed of let-7a to the 3′ region of miR-155 (Fig 6A), and then we performed RBNS using the respective seed-mismatched libraries. Natural and chimeric miRNAs containing the same seed sequences had remarkably similar relative affinities of mismatched seed sites (Fig S6A). Moreover, miRNAs containing the same 3′ ends had very similar preferences for 3′-pairing registers, with similar rise and fall across the registers and similar overall magnitude of *K*D fold-change (Fig 6B–D), which further supported the conclusion that the identity of the miRNA 3′ region dictates the preferred pairing register and the potential contribution of 3′ pairing. Indeed, the largest deviation in the average profile, which was the one observed at register 14 when comparing results for let-7a and the let-7a–miR-155 chimera (Fig 6B), could be explained by potential partial pairing of the let-7a seed to that 3′ site in the let-7a 3′ region, which would augment its enrichment (Fig S6B).

Importantly, when comparing miRNAs that contained the same seed sequence but different 3′ sequences, the effects of the 18 seed mismatches on the efficacy of 3′ pairing were strikingly recapitulated (Fig 6D). Thus, within each matrix of 3′-pairing fold-change values, the magnitude differences between rows were similar for let-7a and the let-7a–miR-155 chimera, and the same was true for miR-155 and the miR-155–let-7a chimera, as shown for matrices depicting results for 7-nt 3′ sites at the optimal register of each miRNA (Fig 6D). For example, the pyrimidine–pyrimidine mismatches at position 6 of let-7a inhibit 3′ pairing contributions for both let-7a and the let-7a–miR-155 chimera (cells outlined in red). Similarly, a A-G mismatch a position 4 and AA mismatch in position 3 in miR-155 confer a large decrease in 3′ pairing contribution and a U-G mismatch at position 5 has a larger increase, both of which are also seen in the miR-155-let-7a chimera (see blue outlines). Indeed, the matrix of 3′-pairing fold-change values observed for the let-7a–miR-155 chimera could be accurately predicted by multiplying the magnitudes across the rows of the let-7a matrix (which predicted the effects of the seed mismatch on 3′-pairing affinity) by the magnitudes across the columns of the miR-155 matrix (which predicted the effects of the 3′ sequence on 3′-pairing affinity), and an analogous procedure predicted the fold-change values observed for the miR-155–let-7a chimera (Fig 6E). These results showed that the influence of the seed mismatch on the magnitude of 3′-pairing affinity depended primarily on the seed-mismatch type and position, with relatively little dependence on the sequence of the 3′ region.

To quantify this result, we compared the correlation values of the 3′ pairing contributions across seed mismatches across a single maximal loop value (Fig 6E), which shows that the miRNAs harboring the same seeds sequences have similar effects of mismatches on their 3′ pairing contributions (R2 = 0.53 and 0.62) (Fig 6E). As a positive control we compared the effects of mismatches on the 3′ pairing contribution of the let-7a replicates, which had an R2 value of 0.96 respectively. To rule out the effects of seed position on the mismatch effects, the average effect across position was compared for miRNAs containing the same seed (Fig 6F; R2 = 0.54 and 0.74) versus the same 3′ regions but different seed sequences, which showed significantly lower correlations (Fig 6G; R2 = 0.02 and 0.41), except in the case of let-7a versus the let-7a/miR-155 chimera, which had an R2 of 0.68 .

**3′ pairing contributions can largely be predicted by nearest neighbor rules**

The results from the previous sections suggest that the RNA sequence of the guide in the major determinant of the preferred pairing register. We wondered if the thermodynamic preferences, such as register preferences, or even over 3′ pairing contribution, could be predicted by nearest neighbor (NN) rules for RNA duplex hybridization in solution (i.e. the RNAfold algorithm). If so, pairing to the 3′ end may be more generally predicted by NN rules than previous work had concluded (Ref Grimson). We first compared the predicted NN energetics for a sequence that corresponds to the sequence of different registers across miRNAs to the mean of each 3′ pairing matrix, as a proxy for the average increase in affinity due to that particular sequence pairing to the miRNA (Fig 7A, left). The values of the mean observed 3′ pairing energy correlated well with predicted NN affinity (R2 = 0.57-0.80 for each miRNA, slope ~0.1). We also find this relationship is not just due increasing the length of pairing, as there exists strong correlation among all 6,7, and 8mer sequences of 3′ pairing across miRNAs (Fig 7A, right).

**An linear model of 3′ site features predicts affinities and improves target site prediction**

While we are able to identify features that contribute to the affinity of 3′ paired sites, we wanted to determine the relative contributions of each feature to aid in design of future high affinity target sites and to improve prediction of potential endogenous targets. We built a linear model for the contribution of 3′ target sites parameterized by previously denoted features: register (reg), loop length (loop len), seed-mismatch type (seedMMtype), seed-mismatch position (seedMMpos), affinity of the seed (dGseed), length of the 3′-paired region (len3p), and the NN predicted affinity for the sequence of the 3′ paired region (dG\_3pred). xAfter normalization of the continuous variables to range between 0 and 1 (as in Agarwal et al, ), the model was trained on 70% of all the affinity data for each natural miRNA (let-7a, miR-1, and miR-155) and then predicted the affinity for the 30% of the data that was left out of the training.

A linear model including just the NN predicted affinity can predict 47% of the variation in the affinity data (R2 = 0.47, Fig 7B). A linear model including all previously denoted features above (Fig 1B) resulted in an increase in predictive power of the combined affinity data (R2 = 0.55 +/-, Fig 7C). Moreover, the affinity data was reparameterized to include two additional features that we postulated might be important to target pairing: 1) the offset of target site, defined as the difference in the size of the internal loop in the miRNA versus target between seed and 3′ paired sites 2) the center of target pairing with respect to the guide (Fig 7D). These two additional features lead to a modest increase in prediction power of the linear model (R2 = 0.61; Fig S7A) .

We found that this model trained on each miRNA separately did better than all the miRNAs combined (Fig. S8B-D, R2 = 0.50-0.79). There were notable similarities in parameter coefficients between each miRNA, such as the loop length dependencies. Moreover, data trained on let-7a and miR-155 and can predict the both chimeric miRNA affinities with an R2 of 0.55, suggesting that training on miR-155 and let-7a together has enough information to predict a chimera of the two miRNAs, and that there is not new behavior along these parameters that arises from a chimeric sequence (Fig S7D).

We observe differences between miRNAs, such as in the register coefficients; these differences were exemplified by the observation that training on two miRNAs was unable to predict another as well as itself (Fig S7E-G, R2 = 0.35-0.44). A correlation of each parameter’s coefficients of the natural miRNAs to the that of the chimeric miRNAs resolves two classes of parameters (Fig 7F)--ones that relate to the seed region (seed mismatch position and type) and ones that report on the miRNA 3′ region (register, and register center). For example, there is a high correlation of the let-7a-miR-155 chimera seed mismatch position and seed mismatch type coefficients with let-7a, let-7a +1, and let-7a -1’s, while a low correlation with their register coefficients. On the other hand, let-7a-miR-155 and miR-155 have a highly correlated register coefficients. Furthermore, there are features that seem similar across miRNAs regardless of sequence composition, such as offset, length, and loop length, as seen as high correlations across all miRNAs.

Since there exists miRNA-specific differences, we included interaction terms between the miRNA and various other parameters to determine whether they increased the prediction power of the linear model, and found that an interaction between miRNA and register had the largest increase in prediction power for left-out data (R2=0.61; Fig 7D). The observation that there are miRNA-specific differences in model for predicting the effects of 3′-pairing is not surprising, given that models for seed-pairing required miRNA-specific terms as well. While this makes a general model for 3′-pairing for the effects on *in vivo* repression more distant, we find that this model does significantly better than the current 3′ pairing score (Fig 7E) in predicting affinity, and as shown below, for repression of 3′ paired sites for the miRNAs assayed herein.

**Displaced scraps worth considering:**

, and it seems that certain seed-mismatch positions have higher 3′ pairing contributions; however, with data for only a single miRNA that cannot span all possible seed mismatch locations and types it is difficult to draw generalizable conclusions.

In the simplest model, AGO2 is interpreting different seed-mismatches through interactions with the major and minor groove of the seed-target duplex (REF), and these interactions could change the likelihood of undergoing a conformational change to promote formation of the 3′ paired sequences. Alternatively, the sequence of the 3′ end of the miRNA could affect the conformation of AGO, such that it is more or less likely poised to allow formation of 3′ pairing interactions--i.e. some synergistic interaction between the seed and 3′ pairing regions of the guide RNA.

**Figure legends**

**Figure 1. miRNA target sites and 3′-compensatory site features.**

(A) Canonical (top), 3′ supplemental (middle), and 3′ compensatory (bottom) site architectures (Grimson *et al*., 2007, and McGeary, Lin, *et al*., 2019). Canonical sites (top) have at least six nucleotides of contiguous Watson–Crick pairing (blue) to the miRNA seed region (red), sometimes with an A opposite miRNA nucleotide 1, or an additional match to miRNA nucleotide 8. 3′-supplementary sites (middle) contain canonical seed pairing with additional pairing (green) to miRNA 3′ end (yellow), typically centered around miRNA positions 13–16. 3′-compensatory site (bottom) contains fewer than six nucleotides of contiguous Watson–Crick pairing to the seed region (due to a mismatch, wobble, or single-nucleotide bulge) with additional pairing in 3′ region, typically centered around positions 13–16. (miRNA sequence N, any nucleotide; mRNA sequence N, a Watson–Crick match to the miRNA; M, a mismatch to the miRNA; B, not A.

(B) Four independent features contributing to the diversity of possible 3′-compensatory sites. These are 1.) the length of Watson–Crick complementarity to the 3′ end of the miRNA (top left), 2.) the 5′-most miRNA nucleotide at which the 3′ pairing begins (top right), 3.) the length of the loop between the seed and 3′ pairing regions of the target RNA (bottom left), and 4.) the particular seed pairing configuration of that target mRNA (bottom right). The two 3′ pairings lengths depicted are 5 and 9 nt (top left) and the two different imperfect seed pairing configurations contain a mismatch to miRNA nucleotide 2 and 7, respectively (bottom right).

**Figure 2. AGO-RBNS with miRNA-specfic libraries enables comprehensive interrogation of 3′ site affinity.**

(A) AGO-RBNS RNA libraries. The fully randomized RNA libraries used in prior studies (top) contain 37 positions of randomized sequence, enabling identification and quantitative analysis of seed- and 3′-only sites. Programmed, miRNA-specific RNA libraries (bottom) enable detection of bipartite sites that contain both seed- and 3′-regions of pairing. These libraries contain 25 nucleotides of randomized sequence followed by an 8-nt region containing all single-nucleotide mismatches to the miRNA nucleotides 2–7 (shown here for let-7a), followed by four additional nucleotides of randomized sequence before the 3′ constant sequence used for reverse transcription and library generation. The library is generated by combining the transcription products of six synthetic DNA oligos that each contain a single mismatch position (pink). (D, not C; V, not U; and B, not A).

(B) AGO-RBNS experimental design. The RNA library (right) is incubated with increasing concentrations of purified AGO2–miRNA complexes until the reaction reaches equilibrium. Each binding reaction is subjected to a filter-binding step to recover the AGO-bound RNA library molecules. The bound library molecules from each reaction are purified and, in addition along with the input RNA library are submitted for high-throughput sequencing. Enrichment of particular sequences are determined through comparison of the sequence frequencies in the bound versus input libraries. The sequence enrichments are fit using a biochemical model of the binding reaction to determine relative *K*D values of different sequences.

(C) Enrichment of top 20 positional 8-nt *k*-mers. *k*-mers are ranked by the sum of their top five enrichments at each position across the sequencing read. Left, *k-*mers are aligned by the CAAC motif shared by all 20 *k*-mers, with nucleotides that are not a Watson–Crick match to the miRNA sequence shown in red. The x-axis is numbered to indicate the position of the 3′-most nucleotide of the *k-*mer relative to the pairing position of the miRNA nucleotide 1 in the programmed region of the target sequence. Right shows the target-site architecture for the *k-*mer–position pair with the greatest enrichment.

**Figure 3.** **Let-7a relative affinity measurements of compensatory target sites.**

(A) Comparison of replicate compensatory library measurements. Independently prepared AGO2-let-7a and compensatory site RNA libraries were used to perform RBNS and generate relative affinity measurements for 12,XXX enumerated 3′ paired site architectures. Relative affinities for the canonical seed sites are also determined by occurrences in the random region of the library and are denoted.

(B) Relative affinity of let-7a sites across different site lengths and loop lengths. The architecture of the sites examined is shown, bottom right: sites begin pairing at register 11 and contain an AC seed mismatch at position 7. Relative affinities for the seed mismatch site alone (black), the 6mer seed site (green) and the 8mer seed site (red) are denoted for comparison.

(C) Comparison of the relative affinities of a let-7a 7nt 3′ paired site beginning at register 11 across all possible seed mismatched positions and identities (N = 18) and loop lengths. Pairing architecture is shown, bottom. The rows are ranked by the affinity of the seed-mismatched site alone, which is shown in the first column of the heatmap. Color intensity designates higher affinity sites.

(D) Comparison of the change in *K*D due to 3′ pairing across seed mismatch position, identity, and loop length. Values are determined by computed the fold change in affinity of the seed mismatched site with a 3′ paired site to the seed mismatched site with no 3′ pairing. Color intensity designates a larger fold increase in affinity due to 3′ pairing. Schematic to the right shows a single example of a site with all the major features, register, loop length, site lengths, and seed mismatch, shown. Bottom schematic shows the endogenous C. elegans let-7a miRNA and lin-41 compensatory target site, with that site architecture denoted in the heatmap with a white box.

**Figure 4. 3′-site pairing energetics between different miRNAs**

(A) The change in *K*D due to a 7-nt 3′ site across different seed mismatches, loop-lengths, register, and miRNAs. Each heatmap is as in Figure 3D, where the intensity denotes the increase in *K*D due to a 3′ site, each row is a different seed mismatches, and each columns is increasing loop size (nt). Rows are ranked by the seed-mismatch affinity alone without 3′ paired sites. Horizontally, heatmaps are displayed by pairing to different registers of the miRNA, from register 9 to the last possible pairing register given the length of the miRNA. Vertically, data for different miRNAs, let-7a, miR-1, and miR-155 are shown. For all miRNAs the labels for the rows of the heatmaps are shown in Figure 3B for let-7a, and Figure 4B for miR-1 and miR-155. Registers containing the most increase in affinity due to 3′ sites are denoted by a blue box. Bottom shows schematic of miR-155 pairing at register 9 and register 17.

(B) Heatmaps for registers containing the most increase in affinity due to 3′ sites. Left) miR-1 at register 12 and right) miR-155 at register 15. Rows are seed site mismatch and position and are ranked by seed site affinity. Columns are displayed by increasing loop length (nt).

(C) Average increase in affinity due to 3′ sites across register for different miRNAs. Values are the mean of all seed mismatched and position sites across the top 5 loop lengths, which are different for each miRNA and register, as seen in Figure 4A.

**Figure 5. Let-7a sequence mutants show that register preferences are determined by sequence**

(A) Sequences of let-7a mutants with top: let-7a -1 where a single adenosine residue is removed at guide position 10 and placed at guide position 21 to maintain the total length of the miRNA. Middle: the let-7a natural sequence. Bottom: let-7a +1 where the an additional adenosine reside is added between guide positions 10 and 11 and a uridine reside is removed at guide position 21 to preserve the 21nt length of the guide RNA.

(B) Heatmaps of the increase in *K*D due to a 7nt length 3′ site, whereby each row is a different seed mismatch and each column is increasing loop length (nt). The rows are ranked by seed site affinity and the labels for the seed mismatch and affinity are found in Figure 3B. Horizontally, the heatmaps are arranged by the pairing to different registers of the three let-7a variants, starting at register 9, which includes pairing to guide residues 9-15, to register 14, which includes pairing to guide residues 14-20. Vertically, heatmaps are arranged by miRNA, with let-7a -1, let-7a, and let-7a +1 from top to bottom.

(C) Average increase in affinity due to 3′ sites across register for the different let-7a variants. Values are the mean of all seed mismatches and positions across the top 5 loop lengths.

**Figure 6. Chimeric miRNA demonstrate that the seed mismatch and 3′ sequence pairing energetics are separable**

(A) Sequences of natural and chimeric let-7a and miR-155 miRNA variants.

(B) Average increase in affinity due to 3′ paired sites across register for top: let-7a and the miR-155-let-7a chimera, bottom: miR-155 and the let-7a-miR-155 chimera. Values are the mean of all seed mismatched and position sites across the top 5 loop lengths.

(C) Scatter plot and correlation of the average increase in affinity due to 3′ paired sites between chimeric and natural miRNAs harboring the same sequence from guide nucleotides 9 to the length of the miRNA: let-7a and miR-155- let-7a variant (green) and miR-155 and the let-7a-miR-155 chimera (pink).

(D) Comparison of the effects of mismatches 3′ pairing energetics between the let-7a and let-7a-miR-155 chimera. All heatmaps are shown at the maximal register and rows are ranked by seed-mismatch affinity and columns are show by increasing loop length (nt). Seed mismatch position and sequence is denoted. From left the right, heatmpas of : 1) let-7a miRNA at register 11. 2) let-7a-miR155 chimeric miRNA at register 15. 3) Prediction of increase in affinity due to 3′ paired sites for the let-7a-miR-155 chimera assuming average loop length dependence of miR-155 at register 15 and normalized seed mismatch effects of let-7a (see Methods). 3) Absolute difference between actual let-7a-miR-155 data at register 15 and predicted increase in affinity assuming average loop length dependence at seed mismatch effects shown in 2). 4) Absolute difference between actual let-7a-miR155 data at register 15 and prediction assuming average loop length dependence of let-7a at register 11 and normalized seed mismatch effects of miR-155.

(E) Comparison of the effects of mismatches 3′ pairing energetics between the miR-155 and miR-155-let-7a chimera. All heatmaps are shown at the maximal register and rows are ranked by seed-mismatch affinity and columns are show by increasing loop length (nt). Seed mismatch position and sequence is denoted. From left the right, heatmpas of : 1) miR-155 miRNA at register 15. 2) miR155-let-7a chimeric miRNA at register 11. 3) Prediction of increase in affinity due to 3′ paired sites for the miR-155-let-7a chimera assuming average loop length dependence of let-7a at register 11 and normalized seed mismatch effects of let-7a. 3) Absolute difference between actual miR-155-let-7a data at register 11 and predicted increase in affinity assuming average loop length dependence at seed mismatch effects shown in 2). 4) Absolute difference between actual miR155-let-7a data at register 11 and prediction assuming average loop length dependence of miR-155 at register 15 and normalized seed mismatch effects of let-7a.

McGeary, S.E., Lin, K.S., Shi, C.Y., Bisaria, N., and Bartel, D.P. (2018). The biochemical basis of microRNA targeting efficacy. bioRxiv 414763.

McGeary, S.E., Lin, K.S., Shi, C.Y., Bisaria, N., and Bartel, D.P. (2018). The biochemical basis of microRNA targeting efficacy. bioRxiv 414763.

**Materials and Methods**

**Analysis of *k*-mer enrichments (Figure 2C)**

Positional enrichments of all 8-nt *k*-mers were calculated by comparison of the binding sample containing NNNNN pM AGO2–let-7a complex and its corresponding programmed library to that of the input sequencing of the programmed library. For each of the two libraries, reads that both contained one of the 18 possible 8mer mismatch sites in the correct position (such that the C nucleotide that pairs with nucleotide 8 of let-7a is at position 26 within the read) and also did not contain either a 6mer, a 6mer-A1, or a 6mer-m8 site against let-7, were used to calculate the total number of each *k*-mer at each position across the library. A pseudocount of 0.001 × the average number of counts per *k*-mer was added to the positional count of each *k*-mer in the library, and then these tables were normalized to 1. The normalized count table corresponding to the NNNNN pM sample was divided by that of the input to arrive at the enrichment of each *k*-mer at each position within the library. These *k*-mers were ranked according to the sum of the top five positional enrichments of each, considering positions 1–19 of the library (where the kmer at position 19 overlaps the programmed 8mer mismatch region by one nucleotide).