Nucleic Acids Research, 2018 **1**
doi: 10.1093/nar/gky546

The sequence features that define efficient and specific hAGO2-dependent miRNA silencing guides

Yifei Yan[1,2], Mariana Acevedo[1], Lian Mignacca[1], Philippe Desjardins[2], Nicolas Scott[2,3], Roqaya Imane[2], Jordan Quenneville[2], Julie Robitaille[1,2], Albert Feghaly[2], Etienne Gagnon[2,4], Gerardo Ferbeyre[1] and François Major[2,3,*]

[1]Département de biochimie et médecine moléculaire, Université de Montréal, Montréal, Québec H3C 3J7, Canada, [2]Institut de recherche en immunologie et en cancérologie (IRIC), Université de Montréal, Montréal, Québec H3C 3J7, Canada, [3]Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Québec H3C 3J7, Canada and [4]Département de microbiologie, infectiologie et immunologie, Université de Montréal, Montréal, Québec H3C 3J7, Canada

Received April 07, 2018; Revised May 10, 2018; Editorial Decision June 04, 2018; Accepted June 05, 2018

ABSTRACT

MicroRNAs (miRNAs) are ribonucleic acids (RNAs) of ∼21 nucleotides that interfere with the translation of messenger RNAs (mRNAs) and play significant roles in development and diseases. In bilaterian animals, the specificity of miRNA targeting is determined by sequence complementarity involving the seed. However, the role of the remaining nucleotides (non-seed) is only vaguely defined, impacting negatively on our ability to efficiently use miRNAs exogenously to control gene expression. Here, using reporter assays, we deciphered the role of the base pairs formed between the non-seed region and target mRNA. We used molecular modeling to reveal that this mechanism corresponds to the formation of base pairs mediated by ordered motions of the miRNA-induced silencing complex. Subsequently, we developed an algorithm based on this distinctive recognition to predict from sequence the levels of mRNA downregulation with high accuracy ($r^2 > 0.5$, *P*-value $< 10^{-12}$). Overall, our discovery improves the design of miRNA-guide sequences used to simultaneously downregulate the expression of multiple predetermined target genes.

INTRODUCTION

A microRNA (miRNA) and an Argonaute (AGO) protein associate to form an essential component of the miRNA-induced silencing complex (miRISC). The miRISC-targeting specificity is mainly determined by sequence complementarity between the miRNA seed (nucleotides 2–8) and target RNA. The miRNA complementary region in the target RNA is called the miRNA regulatory element (MRE) (1,2). Base complementarity between seeds and MREs is the predominant feature of most miRNA target prediction algorithms (3,4). When the base complementarity between a miRNA and its target mRNA is complete, they form a perfect duplex and the miRISC cleaves the mRNA (1,5). However, beyond the seed, bilaterian miRNAs are rarely fully complementary to their targets (6). Therefore, in most cases, the miRISC downregulates the expression of a gene by either removing its poly-A tail or 5′ cap structure (5,7–9), or repressing its translation (1,5).

A loop in the central region of the miRNA–mRNA duplex is tolerated (10). Studies revealed that the loop size and sequence influence the silencing efficiency, and hence loop scores have been assigned to improve miRNA target prediction (11,12). Mismatches at the 3′ end of the miRNA–mRNA duplex (miRNA 3′ end positions) were found to hamper the release of the miRNA from the miRISC and to enhance gene silencing (13). Base pairs (bps) involving miRNA positions 13–16 were shown to rescue silencing when the base complementarity in the seed region is low (14,15). To delineate the precise contribution of all bps in the duplex, researchers performed mutagenesis studies introducing mismatches along the entire duplex (16–23). Beyond confirming that non-seed bps contribute to silencing efficiency, *in vitro* experiments further revealed that many of them along the duplex play different roles in bilaterian animals AGO2-mediated cleavage and silencing (24,25). While the seed bps were shown to affect K_m, which is a measure of affinity of the targeting process, those in the central region were found to contribute mostly to k_{cat}, which is a measure of its endonuclease activity.

These findings were sought to resolve the controversy surrounding mismatches and their effects on the efficiency of silencing. Several attempts were made to incorporate them

*To whom correspondence should be addressed. Tel: +1 514 343 6752; Fax: +1 514 343 7383; Email: Francois.Major@UMontreal.CA

into prediction programs. This involved calibrating each position and trying to fit a free-energy model accordingly (1,12,14,15,26,27). Despite providing some additional predictive power, considerable discrepancies still exist between experimental silencing measurements and predictions. As such, the intrinsic requirements of the miRNA-induced silencing mechanism remained elusive.

Structural studies of AGO became essential to provide mechanistic details that may help reveal miRNA-induced inhibitory action. AGO is found in all three domains of life (28). Even though AGO was initially discovered in eukaryotes (29), the first structural insights originated from their prokaryotic orthologs (30). The AGO structure is well conserved and display a bi-lobed conformation that consists of the N, PAZ, MID and PIWI domains. The PAZ domain is connected to the N- and MID domain by two loops, L1 and L2, respectively. This renders the PAZ domain flexible enough to move as a rigid body relative to all other domains (30–34). The 5′ end of the guide RNA is bound to the MID domain in a pre-shaped A-form (34). A kink occurs at nucleotide (nt) 6, and the guide assumes an extended form all the way to the PAZ domain. The PAZ domain through hydrogen bonds and ring-stacking interactions holds the 3′ extremity of the guide (31,34,35).

Target recognition by the miRNA seed has been well characterized in crystal structures (31,34,35). The PIWI-PAZ channel accommodates the seed region of the guide-target duplex, strongly favoring perfect complementarity at nts 2–6. In all crystal structures with the bound RNA guide (31,34–37), the nts 2–5 in the seed are exposed to the solvent and hence available for recognition.

As the formation of the RNA duplex proceeds in the 5′ to 3′ direction along the guide, the 3′ end of the guide strand is released from the PAZ domain, relieving the steric hindrance of the two intertwining strands. This model is supported by a crystal structure in which a bacterial AGO (*T. thermophilus*) accommodates a guide-target duplex of 15 bp in its MID-PIWI cradle (34). The duplex formation allows for the correct positioning of the scissile phosphate (between nts 10–11) in the nuclease active site so that the cleavage of the target strand takes place precisely. This model, called the nucleation and propagation model, was proposed to be also true for human AGO2 (hAGO2), which is the only RNase-capable isoform among the four human AGO proteins. It entails that hAGO2 exists in two states in action: the 3′ end bound and the 3′ end released states (31,34,35). Transition between these two states is achieved through the base pairing propagation.

However, the bacterial model is incompatible with experimental data collected from bilaterian animals as their AGO protein tolerates a central loop in the guide::target duplex (6). Moreover, it would also imply that base pairing downstream of the seed should increase the guide-target affinity, which was not observed at significant levels (24). The advent of human AGO2 protein structures offered an explanation (35). When hAGO2 is only loaded with a guide strand, the seed region of the guide is bound to highly conserved residues at the 5′ end, while nts 9–11 are occluded by the α-7 helix and a ten-residue loop (residues 600–609, PDB4W5N), and the 3′ end is bound to the PAZ domain by ring stacking and hydrogen bonds (35).

Using miB, a perfectly complementary small hairpin RNA (shRNA) against the *tat* gene of HIV (16,38), we demonstrate that base pairing beyond the seed exerts a spectrum of effects on reporter gene downregulation. A distinctive pattern linked to AGO-binding events allows us to predict induced silencing efficiency from mutated miB sequences. We used this model to develop a rule-based algorithm to compute the silencing efficiency of guide RNA sequences, validated against mRNAs in a pooled dataset from published data. We depicted this pattern at the molecular level and deduced the motions in the AGO2 upon RNA binding using molecular modeling.

MATERIALS AND METHODS

Plasmid construction

The Renilla luciferase control vector, SVR, was obtained by replacing the CMV promoter in the pcDNA-RlucII plasmid (a gift from the Mader lab) with an SV40 promoter. Briefly, the CMV promoter was removed by restriction enzymes SpeI and HindIII (New England Biolabs). The resulting linearized vector was gel-purified with QIAEX II® Gel Extraction Kit. The SV40 promoter from the pGL3-control luciferase vector fragment was obtained by digesting the vector with NheI and HindIII. Gel purified SV40 promoter fragment was inserted upstream of the RlucII gene in pcDNA-RLucII vector by ligation using T4 DNA ligase (NEB).

The firefly-Renilla opposite-sense target site reporter is referred to as the FR(-)TS construct, which contains both firefly and Renilla luciferase reporter genes oriented in the opposite directions. In addition, a 76 bp region of the HIV genome containing the miB shRNA target site in the center (pNL4-3 vector, Accession number: AF324493, nts 5968–6044) was inserted into the 3′UTR of the firefly gene. Cloning of the target site was carried out by inserting the annealed oligonucleotides into the *XbaI* site upstream of the poly-A signal in the pGL3-Ctl reporter. For FR(-)TS vector, the annealed oligos are the following: the forward oligo sequence is CTAGAATGGCAGGAAGAAGCGG AGACAGC**GACGAAGAGCTCATCAGAACA**GTCA GACTCATCAAGCTTCTCTATCAAAGCAT; and, the reverse oligo sequence is CTAGATGCTTTGATAGAG AAGCTTGATGAGTCTGAC**TGTTCTGATGAGCTCT TCGTC**GCTGTCTCCGCTTCTTCCTGCCATT. Bold letters represent the miB binding site. The reporter that contains six times of the target site does not include the flanking regions; rather, the 3′UTR insert is a tandem repeat of the target site only. Renilla luciferase gene was removed from pcDNA-RlucII plasmid by digesting the vector with SpeI and XbaI restriction enzymes. The gel purified (QIAEX II® Gel Extraction Kit) Renilla luciferase fragment was then inserted in the NheI site in Promega pGL3-control luciferase vector.

The FR(-)*tat* dual luciferase vector was constructed as follows. The FR(-)TS vector without insertion of the miB shRNA binding site from the previous step was used as a starting material. The vector was digested with restriction enzymes XbaI and HindIII from NEB, which creates a linearized vector for upstream insertion of the Renilla gene. Subsequent gel purification was performed using QIAEX

II® Gel Extraction Kit. The first exon of the *tat* gene was amplified from pNL4.3-luc vector (a gift from the Cohen lab) with forward primer (5′ to 3′): ATCCAAGCTTCC CGCCACCATGGCAGGAAGAAGCGGA, and reverse primer (5′ to 3′): CGACTCTAGATGCTTTGATAGAGA AGCT. The PCR was carried out using 55 ˚C as annealing temperature. The amplified fragment was ethanol precipitated and digested with restriction enzymes XbaI and HindIII. Upon gel purification, the fragment was ligated with the digested vector at 16˚C overnight. The ligation mix was transformed into DH10B.

The vector pPRIME (a gift from the Pelletier lab) has been previously optimized for shRNA cloning (39–41). Designed guide-RNAs were cloned into the vector following miR-30-based shRNA cloning protocols (42). Briefly, complementary oligonucleotides that contain the shRNA sequences (ordered from Biocorp) were diluted to 100 μM in deionized water. Annealing reaction was carried out at 95˚C in annealing buffer for 5 min followed by slow cooling to room temperature. The annealed double-stranded oligonucleotides were then phosphorylated by T4 PNK (NEB). Ligation reaction was performed by combining doubly digested pPRIME by XhoI and EcoRI with the phosphorylation product of annealed oligos in T4 DNA ligase (NEB) reaction mix at 16˚C overnight.

Cell culture and monitoring shRNA efficiencies

HEK 293T (c17) cells (from ATCC) were maintained according to established conditions (43). Cells were grown in DMEM (+L-glutamine) (Life Technologies) supplemented with 10% FBS, 100 U/ml penicillin/streptomycin at 37˚C and 5% CO_2. Cells were grown to confluence before plating. For testing the efficiencies of mismatched guides, cells were plated in 96-well plates at ∼20 000 cells per well 24 h prior to the transfection. For assays that required growth in 24-well plates, cells were plated at ∼100 000 cells per well.

The reporter plasmids and the shRNA plasmids were co-transfected into the cells using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Along with 20 ng of shRNA plasmid, 5 ng of pNL-luc and 2 ng of SVR control vector were co-transfected into each 96-well; alternatively, 50 ng of the shRNA construct, 20 ng of the pNL-luc, and 10 ng of the SVR control vector were co-transfected into each 24-well. When an AGO2 expression construct is used, 25 ng of the AGO2D597A vector (44) (A gift from the Diederichs lab) was combined with the DNA mix described above and subsequently co-transfected into the cells

Luciferase assays were performed accordingly to established protocols adapted from the Duo-Glo Luciferase System (Promega). Forty-eight hours post-transfection, cells were lysed with 1× Passive lysis buffer (Promega) and luciferase activity was assayed using the Dual-Glo Luciferase System (Promega). Luminescent light was measured on Veritas Microplate Luminometer (Turner Biosystems) (a gift from the Bouvier Lab). The ratio between the reporter and the control luciferase bioluminescence light was taken and then normalized to that of the negative control shRNA or empty vector, resulting in the percentage residual expression of the reporter gene.

Measuring reporter transcript and mature RNA guide abundance using qRT-PCR

RNA extraction was performed using TRIzol® reagent following manufacturer's protocol. RNA was extracted from the same cells used in the luciferase assay. Either oligo-dT primer or random primer were used for the synthesis of cDNA from total RNA extracted according to previously established protocols (45). Eight hundred nanograms of total RNA was used for each synthesis reaction in 20 μl of total volume using Invitrogen reagents (M-MLV Reverse Transcriptase, Cat. No. 28025-021, Invtitrogen™). RNA was extracted from the same cells that were used in the luciferase assay and M-MLV was used to perform the cDNA synthesis.

PCR of the mRNA was performed according to previously established protocol (46). Briefly, newly synthesized cDNA was diluted by a factor of 100 prior to real-time PCR. Each real-time PCR reaction mixture contained the diluted cDNA (1 μl), forward and reverse primers (250 nM), $MgCl_2$ (2.5 mM), dNTPs (0.2 mM), SYBR green (0.33×), buffer for Jumpstart *Taq* DNA polymerase and Jumpstart *Taq* DNA polymerase (0.25 U; Sigma) in a final volume of 10 μl. After denaturation at 95˚C for 6 min, samples went through 50 cycles of amplification (20 s at 95˚C, 20 s at 58˚C and 30 s at 72˚C). Melt curves were determined for each reaction and qPCR was performed using a LightCycler 480 (Roche Applied Science, Canada). Data was normalized using Renilla and HPRT as controls.

The detection of mature RNA guide molecules was performed following the polyA-based RT-qPCR protocol established previously (47,48). Briefly, 20 μl of reaction contained 1 μl of reverse transcription products diluted 10-fold, 10 μM of forward primer, and 10 μM of universal reverse primer, 2 μl of Taq polymerase buffer (10×), 4 μl of 2.5 mM each dNTP, 0.6 U Taq and 10 μM of universal TaqMan probe. The mix is heated to 95˚C for 2 min prior to entering 45 cycles of 95˚C for 15 s followed by 60 ˚C for 1 min. The reactions were carried out and measurements were taken on a StepOnePlus™ Real-Time System from Applied Biosciences. The forward primer sequences are as follows:

miB: GTGCTGTTCTGATGAGCTCTTCGTC;
miB-A: GTGCTGTTCTGAACTGCTCTTCGTC;
miB-B: GTGCTGTTCTGATGACGACTTCGTC;
miB-C: GTGCTGTTCTGATGAGCTGAACGTC;
miB-D: GTGCTGTTCTGATGAGCTCTTGCAG;
U6: ACGCAAATTCGTGAAGCGTTCCAT;
Puromycin: TGACCGAGTACAAGCCCAC.

Cells and Retroviral-Mediated gene transfer

PC3 were obtained from American Type Culture Collection (ATCC) and cultured in RPM1 (Wisent) supplemented with 10% FBS (Wisent), 1% penicillin/streptomycin sulfate (Wisent) and 2 mmol/l L-glutamine (Wisent) at 37˚C and 5% CO_2. Gene transfer was performed using retroviral particles produced in Phoenix packaging cells. Phoenix cells were transfected by calcium-phosphate precipitation with 20 μg of a retroviral plasmid (15 h at 37˚C). The plasmids used were: shNTC (non-targeting control), MiR20, MT E2F(1), E2F Afa, E2F Afb, E2F Afc, E2F Afd and

E2F Afe. After 48 h, the virus-containing medium was filtered (0.45 μm filter, Millipore) and supplemented with 4 μg/ml polybrene (Sigma) (first supernatant). Viruses were collected for an additional 8 h as before (second supernatant). For infections, the culture medium was replaced by the appropriate first and second supernatant on PC3 cells. Sixteen hours later, infected cell populations were purified by selection with 2 μg/ml puromycin for 48 h.

Growth curve

Twenty thousand cells per well were plated into 6-well plates. At the indicated times, cells were washed with PBS, fixed in 4% formaldehyde, and rinsed with distilled water. Cells were stained with 0.1% crystal violet (Sigma) for 30 min, rinsed extensively, and dried. Cell-associated dye was extracted with 2.0 ml 10% acetic acid. Aliquots were diluted 1:4 with H_2O, transferred to 96-well microtiter plates, and the optical density at 590 nm was determined. Values were normalized to the optical density at day 0 for the appropriate condition. Within an experiment, each point was determined in triplicate.

Western blot

PC3 cells were washed with cold PBS and then scraped on ice into 500 μl of PBS buffer containing 1× Complete-EDTA free Protease Inhibitor Cocktail (Roche Applied Science) and 1× PhosSTOP Phosphatase Inhibitor Cocktail (Roche Applied Science). Cells were spun at maximum speed for 5 min. The pellet was re-suspended in 100 μl of Laemmli-β-Mercaptoethanol buffer, sonicated 5 s at a low intensity, heated 5 min at 95°C and then cleared by centrifugation at 13 000 RPM for 10 min. The proteins were quantified with the Bradford reagent and 30 μg were loaded on a 10% SDS-PAGE and transferred to Immobilon-P PVDF membranes (Millipore). Membranes were blocked 1 hour at room temperature in PBS containing 0.1% Tween 20 (PBS-T) and 5% dry milk and then washed for 5 min 3 times with PBS-T. The membranes were incubated with the primary antibodies diluted in PBS-T + 3% BSA + 0.05% Na-azide overnight at 4°C. The following primary antibodies were used: anti E2F1 (1:1000, clone H-137; rabbit polyclonal; #SC22820, Santa Cruz); anti E2F2 (1:1000, clone L-20; rabbit polyclonal; #SC632, Santa Cruz); anti E2F3 (1:1000; clone PG-37, mouse monoclonal, #5551, Millipore); anti-α-tubulin (1:20000, mouse monoclonal clone B-5-1-2, T6074, Sigma-Aldrich). Membranes were washed three times 5 min with PBS-T and then incubated with the secondary antibodies diluted in PBS-T + 5% dry milk 1 h at room temperature. The following secondary antibodies were used: goat anti-rabbit IgG conjugated to HRP (1:3000, #170-6515, Bio-Rad) or goat anti-mouse IgG conjugated to HRP (1:3000, #170-6516, Bio-Rad). Finally, the membranes were washed three times 5 min with PBS-T. Immunoblots were visualized using enhanced chemiluminescence (ECL) detection systems and Super RX X-Ray films (Fujifilm) or a ChemiDocTM MP system (Bio-Rad). Band quantification was done using ImageJ or Image Lab 4.0 (Bio-Rad).

Molecular modeling of AGO protein structures

Protein structure files were downloaded from the Protein Data Bank (PDB) website. Modeling was performed in UCSF Chimera version 1.10.2 following the software documentation. Molecular surface calculation, sequence alignments, and distance measurements were performed according to established procedures (49). Briefly, the default settings were used for all the calculations: molecular surface computation and distance measurement was performed using the built-in function of Chimera. The coulomb potential surface was calculated using the built-in function under the 'surface/binding analysis' option among the 'Tools'. Default parameters were used: dielectric constant = 4.0, distance from surface = 1.4 and histidine protonation was assumed for structures without explicit hydrogens. Hydrogen bond predictions were performed with the 'FindHBond' function in 'Structural Analysis' with default parameter settings to relax H-bond constraints by 0.4 Å and 20°. The MatchMaker function of Chimera performed the structural alignment between the two protein structures. The function's default settings include using Needelman-Wunsch algorithm and BLOSUM-62 matrix for sequence alignment, where gap-opening penalties for intra-helix and intra-strand are both 18, and 6 for any others, and the program iterates by pruning long atom pairs until non-pair exceeds 2.0 Å.

Implementation and validation of *MicroAlign*

Fold inhibition of miR-21, miR-122 and miR-22 were taken from (23). For each miRNA, the dataset chosen represented what the authors defined as the inhibitor concentration whose efficacy most accurately captured the effects of the dinucleotide mismatches. The concentrations were: 20 nM for miR-21, 2 nM for miR-122 and 0.3 nM for miR-22. As a pre-filtering step, mismatched inhibitors in the first position and seed region were excluded. The data were transformed into residual target proportions (1/fold inhibition), and because all three miRNAs do not share the same concentration, the residual target proportions had to be linearly scaled to give relative target expression levels. The linear scaling was performed by fixing the lowest residual target proportion to 0 and the positive control value, represented by the fully matched inhibitor, to 100.

The catalytic efficiency (k_{cat}/Km) measured for AGO2 was used as a proxy to infer relative target expression levels (24). The less efficient the catalysis, the higher is the expression of the siRNA target. As for the Fold inhibition dataset, mismatched guide siRNAs in the first position and seed region were excluded from this dataset. For the sake of uniformity, the catalytic efficiency values were normalized to the most efficient siRNA guide to get k_{cat} /K_m percentage values comparable to the other datasets. Relative target expression was defined as 100 – the percentage catalytic efficiency of the siRNA guide.

In a previous study, miRNAs were ectopically expressed in cell lines followed by expression profiling mRNA and protein levels using microarray and SILAC (50) (data obtained from www.nature.com/articles/nature07242). The corresponding 3′UTR sequences were downloaded from

the NCBI Reference Sequence Database (release 49) (51) using a method previously established in our laboratory (52), and the *MicroAlign* program was applied on them. We compared the hits and the corresponding protein and mRNA expression levels.

The evaluation program MicroAlgin was implemented in MicroSoft Visual Studio Express 2012 C++ as a standalone windows application. Experimental measurements were plotted against the predicted miScores (see code below) to calculate Pearson correlations.

The pseudocode of *MicroAlign* evaluation algorithm implements the DFA described in Figure 3D. The set of transition functions (δ) of the DFA is described in the figure, where state set $Q = \{q_0, q_1, q_2, q_3, q_4\}$, alphabet set $\Sigma = \{\text{seed, A, B, C, D}\}$, and transition function set $\delta: Q \times \Sigma \rightarrow Q$. The start state is q_0 and the accepted state is q_4. The configuration of bps between the miRNA and the target now can be regarded as a regular expression that is recognized by this DFA, simulating the AGO2 mechanism.

The bps are predicted by Needleman-Wunch algorithm and evaluated by regions following the discovered order. We described this DFA using a recursive algorithm. The 'bottom' of the recursion is the evaluation of region-D, where the contribution of bps is little for accessible 3′UTR sites. The algorithm is implemented as a Windows application and a copy of it is available online: http://major.iric.ca/MajorLabEn/MiR-Tools.html.

```
list_of_regions = ( B, C, A, D )
Evaluate_score( list_of_regions ) {
     current_region = car( list_of_regions )
     %pair = paired bases in current_region / number of
      bases in current_region
      if( current_region == D )
            score = 0
      else if( %pair > 0 )
            score = %pair * ( pairing_score
            + Evaluate_score( cdr( list_of_regions )))
      return score}
```

RESULTS

Mismatched modules cause disturbance in silencing efficiency

To study the role of base pairing as a determinant of the efficiency of AGO2-mediated silencing, we chose an shRNA, miB, which was reported to target a structurally open region of the HIV genome and inhibit viral gene expression (16,53,54). Its MRE is located in exon 1 of the HIV-1 *tat* gene (nts 5993-6013). We mutated miB's nts in the non-seed region in short stretches of 3 or 4 nts at a time (modules) such that they mismatch the corresponding nts in the target sequence: from 5′ to 3′ module **A** (nts 9–11), **B** (nts 12–14), **C** (nts 15–17) and **D** (nts 18–21) (FSig. 1A). The mismatched positions were engineered by copying the nt from the target strand (A:A, G:G, C:C, and U:U). We named the guide strands containing these mismatches miB-**A**, -**B**, -**C** and -**D**, respectively, and cloned them into pPRIME (42), an shRNA expression vector based on the miR-30 backbone (Figure 1B).

To test the mismatched shRNAs, we used the pNL4.3-luc reporter construct, which contains the complete HIV genome with a disabled *env* gene (Figure 1C). Since effective endogenous MREs are often located in the 3′UTR of their mRNA (55), we constructed a dual luciferase reporter, FR(-)TS, which embeds the miB MRE in the 3′UTR of the firefly luciferase (Figure 1D). The MRE is located 29 nts downstream of the firefly luciferase stop codon, which is within a region (15–300 nts from the stop codon) associated with a high density of mRNA-bound AGO2 protein in the HITS-CLIP assays conducted by the Darnell group (56). To test whether this reporter construct functions properly, we mutated individual nts to their complementary nts in the seed of miB between position 1 and 6. As a result, we observed a significant abolishing effect of the repression compared to miB (Supplementary Figure S1A), confirming the reporter system is capable of measuring one-nt mismatch effects.

Consistent with the previous report, miB effectively repressed pNL-4.3 reporter gene expression, with a 75–80% knockdown efficiency relative to vector-only transfected cells (16). However, reporter gene silencing by mismatched small RNA guide was greatly abolished except for miB-D, which retained more than 50% of the silencing capability (Figure 1E). ShRNAs (or miRNAs) that partially base pair with the HIV target sequences in the non-seed regions were strikingly ineffective in repressing the viral target (17,38). As previously reported, we observed at least 80% loss of repression due to a mismatch of three nts in module A, B or C. When FR(-)TS was used as the target construct, all guide strands showed improved silencing efficiency compared to the pNL-luc reporter construct (Figure 1F). Also, a profile of repression efficiency emerges: miB and miB-D were the most efficient, followed by miB-A, then miB-C and miB-B, with more than 60% remaining expression. These results corroborate the findings that some non-seed nts are important for silencing (14,15), as well as the results of mono- and di-nucleotide mismatching guide RNAs (6,13,24).

Variation in target concentration is not a dominant factor that perturbs the silencing efficiencies

Previous studies have shown that concentration of the target or the miRNA affects the repression efficiency due to threshold effects and competition from ceRNAs (17,57). We optimized our assay conditions so that target concentration will not affect repression significantly in this study. Both the pNL-luc and FR(-)TS reporters were titrated at a concentration range of 25-fold difference (4, 20 and 100 ng) with no significant alteration of the repression pattern. At higher concentrations of the target, the downregulation is less efficient in general (Figure 1G). However, the efficiency is maintained with the FR(-)TS reporter in cells transfected with miB, miB-A or -D (Figure 1H), even at the highest level. For these three guide sequences, the enhancement of repression did not exceed 20% for any guide even when the target concentration decreased 25-fold.

Confirmation of the effects of MRE location, accessibility, and repeats

Local structures in the target RNA may hinder the action of miRNA (15), and the RNA genome of the HIV is known to contain rich secondary structure (58). To make sure that the improvement of silencing efficiency when moving the MRE from the pNL-4.3 to FR(-)TS construct is not due

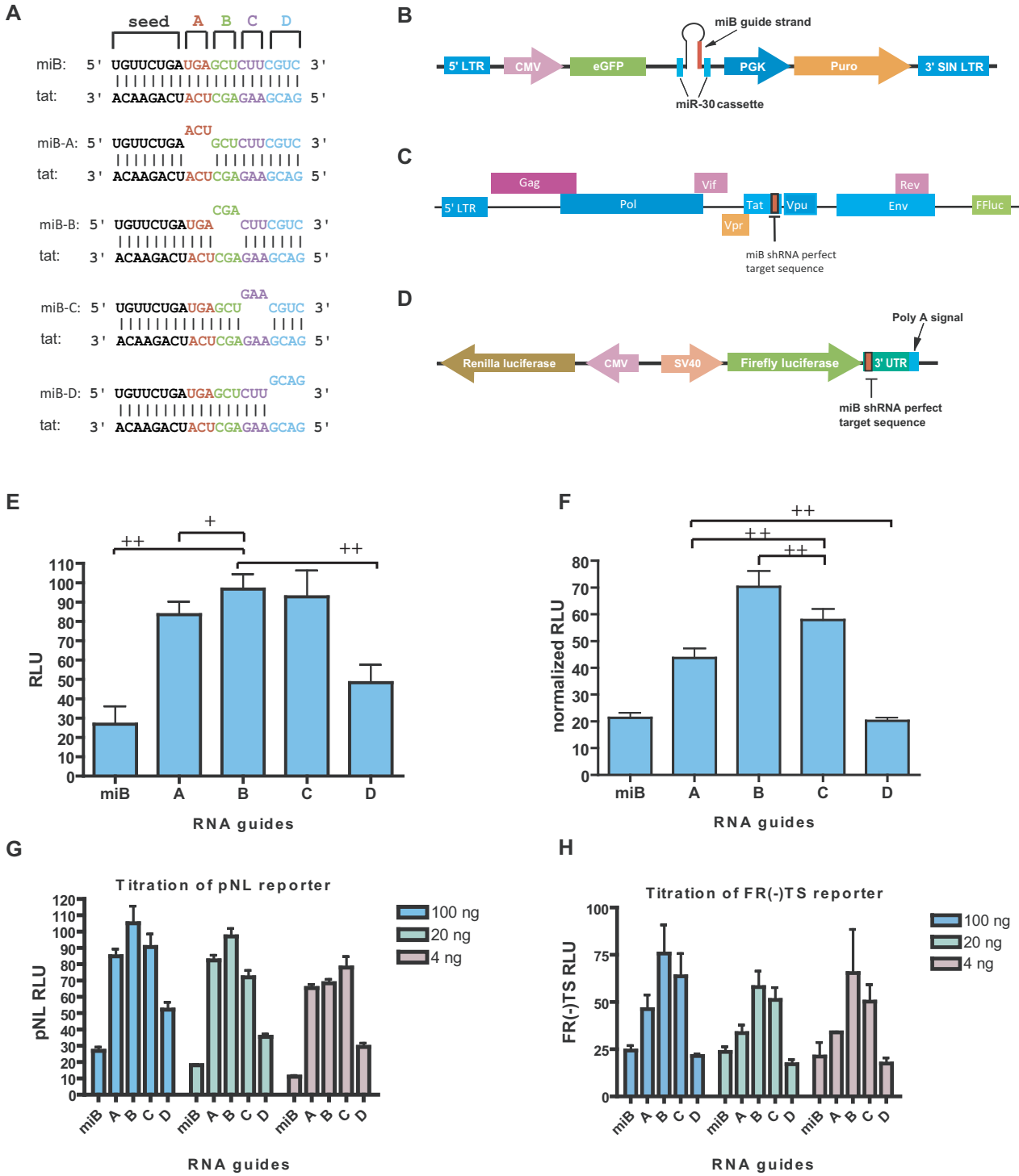


Figure 1. Silencing profile of the coding region and 3′UTR sites in the reporter plasmid. (**A**) MiB and designed single-module guides. The perfect complementary shRNA miB; then, from top to bottom, the mismatched guide RNAs miB-A (nts 9–11); miB-B (nts 12–14); miB-C (nts 15–17); and, miB-D (nts 18–21). (**B**) The pPRIME vector used to clone all shRNAs. The guide strand is located at the 3′ half of the stem loop structure (red). (**C**) The pNL-luc plasmid is a luciferase reporter that contains the HIV-1 genome. The luciferase reporter gene (green) is located near the 3′ LTR of the viral genome. (**D**) The dual luciferase reporter plasmid FR(-)TS in which the cloning site of the target sequence is located in the 3′ UTR of the firefly reporter gene. The firefly and Renilla luciferases are transcribed in opposite directions. (**E**) The repression profile of mismatched shRNA on the pNL-luc reporter. The plus sign (+) indicates the student t-test for the comparing columns yields $P < 0.05$; double plus signs (++) $P < 0.01$. The same convention is followed for panels FGH. (**G**) Titration of the 3′UTR FR(-)TS reporter has limited effects on silencing. (**H**) Titration of the HIV pNL-luc target reporter has limited effects on silencing.

to the removal of global structure of the viral mRNA, we cloned exon1 of the *tat* gene into the dual luciferase construct after removing the *miB* MRE. Exon 1 of *tat* is inserted in-frame with and upstream of the Renilla luciferase (Figure 2A). We name this vector FR-*tat*. As the result, a fusion protein of *tat* and Renilla luciferase is synthesized upon translation. Despite reduced light intensity, the Renilla luciferase remains active and its expression is still sensitive to the downregulation of miB shRNA (Figure 2B).

Mismatches in module A, B or C greatly abolish silencing. This resembles the repression pattern displayed when pNL4.3-luc viral genome construct was used as a target. Energy calculation following the approach developed by the Ding group using *sFold* (59) indicated the absence of stable local RNA structure (Figure 2C, $P > 0.5$), rendering the MRE accessible (seed position 41–47). This corroborates with the high throughput screen results from the Elledge Lab, where an shRNAs library tiling the entire genome of the HIV was screened to probe for the accessible regions of the viral RNA genome (53). Therefore, the enhancement of repression reflects the fact that the MRE has been moved from the coding to a non-coding region of the mRNA, rather than the removal of either global or local secondary structure.

Multiple MREs in close proximity were shown to have enhancing effect on miRNA-mediated repression (6,10,60). To see whether the number of MREs on each target RNA could alter the repression profile significantly, we inserted the miB target site into the FR(-)TS vector six times in tandem, and tested the 1- and 6-MRE target constructs side by side with the pNL-luc reporter (Figure 2D). The 6-MRE in the 3′UTR has enhancing effects on silencing for miB, miB-A and miB-D. However, no significant changes were observed for miB-B and miB-C. We concluded that the number of MREs in the 3′UTR influences the silencing efficiency, but to a much less extent than their location.

The pattern of repression levels is not associated with the levels of mature guide RNAs

We transfected shRNA constructs with 8-fold differences in quantity. Downregulation levels appeared to be resistant to such perturbations, indicating that the guide-AGO2 biogenesis pathway was already saturated at half of the amount of guide RNA constructs used (i.e. 20 ng) (Supplementary Figure S1B). To further confirm that the pattern is not due to differences in mature guide RNA levels, we measured them using TaqMan RT-qPCR (47) (Supplementary Figure S1C). We observed no significant differences for miB, miB-C, or miB-D. However, the levels of miB-A and miB-B are significantly different, respectively 1.5 and 0.4 times that of miB.

To address the concern of whether the profile of repression efficiency truthfully reflects the positional effects of the mismatched nts during the targeting process rather than the efficiency of processing and AGO-loading, we altered the sequence in the target, rather than the guide, to create the same mismatches in the four modules when using miB as a guide. Using the same design rationale for mismatches in the guide, four mutated target sequences, tat-A, -B, -C and D, were cloned into the 3′UTR of the same dual luciferase reporter (Figure 2E), and we observed a similar profile (Figure 2F). To confirm that this profile is stable with different amounts of mature miB guide RNA, we titrated the miB construct at eight different fold concentrations. Again, the same profile emerged (Supplementary Figure S1D). Then, using TaqMan RT-qPCR, we quantified the mature miB at each transfected concentration, and found that variations in mature miB abundance is not related to the observed pattern (Supplementary Figure S1E). We used 20 ng of each shRNA construct for transfection, where the mature levels can vary linearly with that of transfected DNA. However, within the variation range, no significant difference in repression levels could be detected. This confirmed that although the mature levels of the guide RNAs may differ by up to 1.5 times, such as in the cases of miB-A and miB-B in the previous experiments, the repression profile is not affected and is solely due to the positional effects of the mismatches in the targeting process.

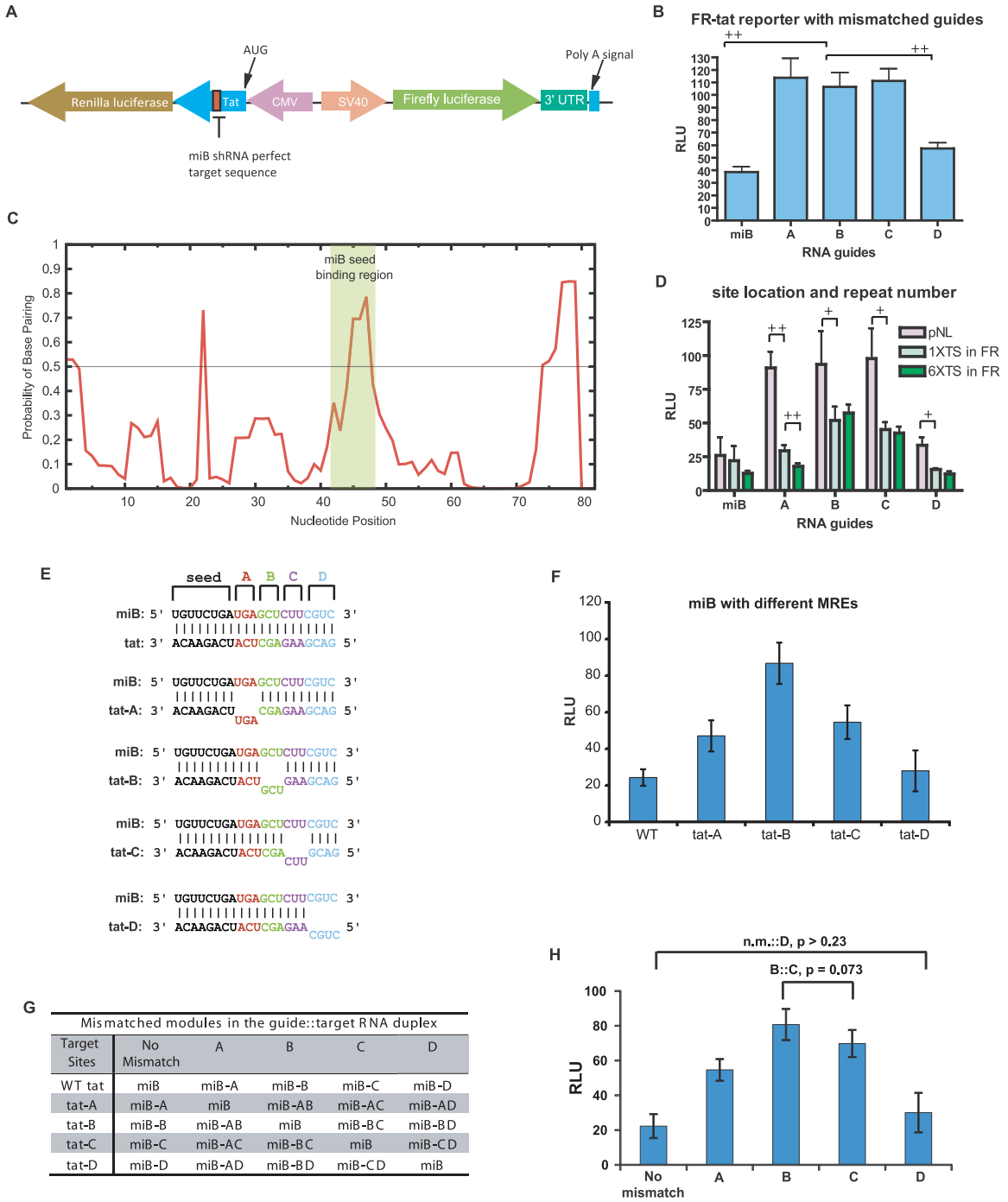
To make sure that these observations were not biased by a particular guide RNA or a particular MRE, we reconstituted the mismatches of the four modules in different target sites and shRNA-target combinations. Along with the wild-type, four additional sites were tested in combination with five sets of guides. Along with the fully complementary guide for each site, 25 different combinations were tested in total (Figure 2G). The nts at the mismatches as well as the surrounding sequences of the modules differ in each combination (Supplementary Figure S2A). For each mismatched module, we averaged the repression values obtained from all sites to produce a synthesized repression profile (Figure 2H). Again, the four modules are pair-wise distinguishable (Supplementary Figure S2B).

Sequence alteration in the non-seed region display a decidable pattern on repression levels

To grade the relative importance of each module in their ability to influence gene silencing, we combined the wild-type site with all six possible double-module variants of miB: miB-AC, -BD, -AD, -AB, -BC and -CD (Figure 3A). This produced a spectrum of silencing effects when luciferase expression was monitored (Figure 3B). We grouped the reporter level of each single-module with those of the double-module variants that contain it (Supplementary Figure S3A). A pattern of indistinguishable reporter levels emerged from these expression levels (Figure 3B; compare the columns of the same color), as well as from their associated *P*-values (Figure 3C; Student's *t*-test; *P*-values in Supplementary Figure S3B). We deciphered the following information: when the seed is perfectly matched, the B-module has the most decisive effect on silencing because it determines how base pairing in the rest of the non-seed nts contribute to silencing. Such decisive power of the modules decreases following the order of module C, A and D.

Establishing a computation model using the pattern

To consistently apply this rule to evaluate the targeting efficiency of miRNA-mediated repression, we built a com-



miB: 5' UGUUCUGAUGAGCUCUUCGUC 3'
tat: 3' ACAAGACUACUCGAGAAGCAG 5'

miB: 5' UGUUCUGAUGAGCUCUUCGUC 3'
tat-A: 3' ACAAGACU UGA CGAGAAGCAG 5'

miB: 5' UGUUCUGAUGAGCUCUUCGUC 3'
tat-B: 3' ACAAGACUACU GCU GAAGCAG 5'

miB: 5' UGUUCUGAUGAGCUCUUCGUC 3'
tat-C: 3' ACAAGACUACUCGA CUU GCAG 5'

miB: 5' UGUUCUGAUGAGCUCUUCGUC 3'
tat-D: 3' ACAAGACUACUCGAGAA CGUC 5'

G

Mismatched modules in the guide::target RNA duplex					
Target Sites	No Mismatch	A	B	C	D
WT tat	miB	miB-A	miB-B	miB-C	miB-D
tat-A	miB-A	miB	miB-AB	miB-AC	miB-AD
tat-B	miB-B	miB-AB	miB	miB-BC	miB-BD
tat-C	miB-C	miB-AC	miB-BC	miB	miB-CD
tat-D	miB-D	miB-AD	miB-BD	miB-CD	miB

Figure 2. Silencing profile in FR(-)*tat* and pNL-luc reporters resemble. (**A**) Dual luciferase construct FR(-)*tat* contains the first exon of the *tat* gene of HIV upstream of the Renilla luciferase, creating a fusion protein of *tat* and Renilla luciferase; the miB MRE is in the *tat* gene. (**B**) Repression profile on the FR(-)*tat* reporter. The plus sign (+) indicates the student t-test for the comparing columns yields $P < 0.05$; double plus signs (++) $P < 0.01$. The same convention is followed for panel D. (**C**) Secondary structure calculation of the miB target sequence within the *tat* gene by *SFold*. The vertical axis indicates the probability of being single-stranded. A horizontal line indicates the threshold of $P = 0.5$. (**D**) The silencing profile is more sensitive to MRE location than MRE repeat numbers. Reporter expression levels of pNL-luc vector (blue bars), or FR(-)TS vector that contains the target sequence either one time (red bars) or six times in tandem (green bars) in the presence of mismatched miB variants. (**E**) Base pairing between engineered sites that mismatch the miB shRNA at modules A–D. Each site is cloned into the same FR-reporter with the flanking regions from the *tat* gene. (**F**) Dual luciferase assay when miB shRNA construct is used in combination with all four site reporters. Firefly luciferase levels were first normalized to Renilla light, then normalized to the non-repressed level of each particular reporter construct ($N = 4$). (**G**) A table of shRNAs used in combination with each target site reporter to reconstitute the same mismatched positions in modules A, B, C and D. The first row of the table indicates which module is mismatched in the guide::target duplex. The first column on the left is a list of mismatched module site reporters. Each entry in the table is a miB-modified shRNA with mutated modules used in combination with the target site of that row. The sequences of the guide::target duplexes are listed in Supplementary Figure S2. (**H**) Synthesized repression profile from reporter assays results by testing the 25 guide-target combinations in the table. Indistinguishable columns heights are indicated by bars on top of the figure.

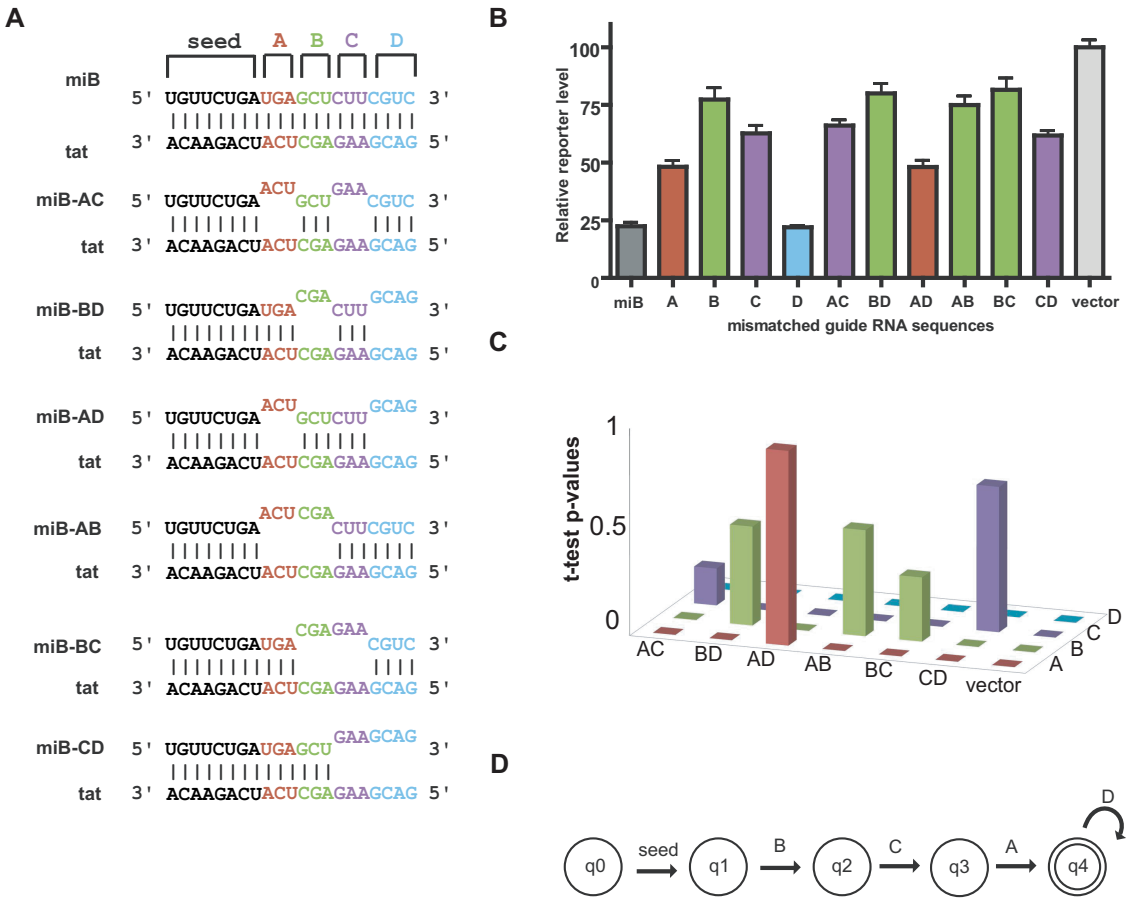
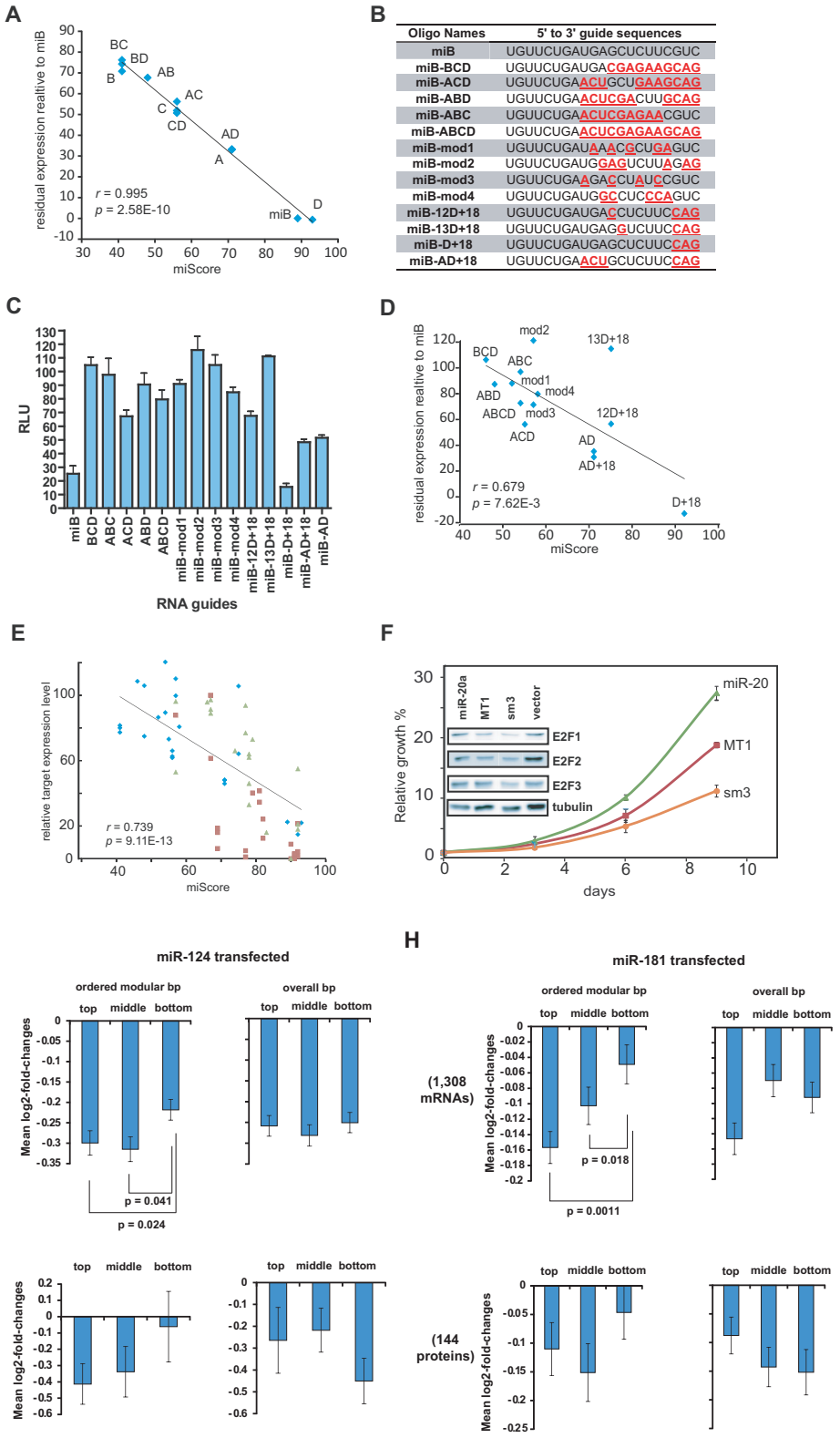


Figure 3. Combined effects of mismatches reveal the interdependency between the modules. (**A**) MiB and double-module guides. All possible combinations of two mismatched modules are listed. (**B**) The repression profile of all miB variants ($N = 4$). The columns where the expression levels cannot be distinguished are of the same color. (**C**) A 3D representation of the *P*-values of the Student's *t*-test results of comparing the efficiencies of all miB variants. When the pairwise comparison is not able to distinguish the two guide RNAs by their residual reporter expression levels, a large *P*-value shows up as a tall column on the graph. (**D**) State diagram of the proposed sequential recognition model for AGO2 slicing. Double circles represent the accepted state, which is defined as the most efficient slicing state.

putational tool that emulates the decision-making process of AGO2. AGO2 can be modeled as a multi-state machine, depicted in a Deterministic Finite Automaton (DFA) (Figure 3D). The guide-loaded AGO2 first recognizes bps in the seed. Seed pairing is followed by base pairing of the nts in module-B. When the bps in the module-B are recognized, AGO2 transitions to the next state, allowing base pairing of the nts in the C-module to be recognized, followed by the A-module. Since a mismatched module-D is indistinguishable from miB, the slicer activity is likely to be fully functional once modules A, B and C are all base paired. For this reason, we defined the accepted state of the DFA, q_4, i.e. where slicing can occur. This is consistent with the fact that the miRISC tolerates a loop in module-A, and that mismatches in module-D enhances the release of the miRNA from the miRISC (13), which is an independent step of the mechanism of the slicer activity of AGO2. The DFA describes a recursive algorithm that asserts the rule of evaluating the efficiency of a guide RNA. We implemented this model in a program called *MicroAlign* as a stand-alone Windows application. The first step of the program is to align the guide and the target strands to make sure that a reasonable conformation of the duplex is scored. Then, the miScore, which quantitatively reflects the silencing efficiency, is calculated. The miScores capture accurately the silencing efficiency. We observed a very strong correlation between miScores and the expression levels of our reporters (Figure 4A; $r^2 > 0.98$, $P < 2.6 \times 10^{-10}$).

Correlation with larger mismatched regions

We engineered RNA guides that contain at least three of the four mismatched modules (Figure 4B, rows 2–6), as well as combinations of random mismatches (Figure 4B, rows 7–14). From the reporter assay results (Figure 4C), we observed again a high accuracy of the miScores (Figure 4D, $r^2 \sim 0.50$, $P < 0.01$). This holds even when no alignment is performed (Supplementary Figure S4A). Inaccuracies of the free energy model mostly occur when the mismatches occupy more than two modules. Our alignment algorithm identified alternative bps (Supplementary Figure S4B) that improves the ranking of predicted activities of such guides (Supplementary Figure S4C).



Oligo Names	5' to 3' guide sequences
miB	UGUUCUGAUGAGCUCUUCGUC
miB-BCD	UGUUCUGAUGACGAGAAGCAG
miB-ACD	UGUUCUGAACUGCUGAAGCAG
miB-ABD	UGUUCUGAACUCGACUUGCAG
miB-ABC	UGUUCUGAACUCGAGAACGUC
miB-ABCD	UGUUCUGAACUCGAGAAGCAG
miB-mod1	UGUUCUGAUAAACGCUGAGUC
miB-mod2	UGUUCUGAUGGAGUCUUAGAG
miB-mod3	UGUUCUGAAGACCUAUCCGUC
miB-mod4	UGUUCUGAUGGCCUCCCAGUC
miB-12D+18	UGUUCUGAUGACCUCUUCCAG
miB-13D+18	UGUUCUGAUGAGGUCUUCCAG
miB-D+18	UGUUCUGAUGAGCUCUUCCAG
miB-AD+18	UGUUCUGAACUGCUCUUCCAG

Figure 4. Validation of the non-seed model. (**A**) Pearson correlation between reporter assay results and miScores (single- and double-module guides). (**B**) miB and additional 13 guide sequences. Mismatched nts are in red and underlined. (**C**) RLU values of the guides in (C) as measured in the FR(-)TS reporter assay. (**D**) Pearson correlation between the reporter assay results and the miScores without alignment (13 additional guides). (**E**) Pearson correlation between published silencing efficiencies and miScores. miB-based guides (blue diamonds); Wee *et al.* dataset (green triangles); and, Robertson *et al.* dataset (red squares). (**F**) PC3 cell growth curves of miR-20a, MT1, and sm3. Juxtaposed are the Western blots of E2F factors when miR-20a, MT1 and sm3 were present. (**G**) Mean $\log_2$ fold changes of the targeted mRNAs (top two panels) and proteins (bottom two panels) in miR-124 transfected cells binned by miScores (left) and scores not considering the modular order of the base pairs (right). (**H**) Same as (G), but for miR-181 transfected cells.

Correlation with other siRNA studies

The analysis of third-party published data further confirmed the strong correlation between miScores and silencing. We used: (i) the catalytic efficiency (k_{cat}/K_m) measured for AGO2 *in vitro*, where mismatches were systematically generated in the guide RNA (24) and (ii) miRNA sponges engineered with dinucleotide mismatches tiling the entire non-seed region (23) (see Supplementary Figure S4DE; and with alignment Supplementary Figure S4FG). We pooled these data and computed the Pearson correlation between miScores and experimental expression levels (Figure 4E; $r^2 > 0.5$; $P < 10^{-13}$).

Enrichment in designing effective artificial miRNAs

We further validated the model by showing how it enriches the design of efficient smartRNAs. As previously established, when the synthesis of multiple isoforms of the E2F protein is inhibited using smartRNAs, PC3 cell growth and proliferation are compromised (43). In the previous study, we used the program *MultiTar* developed in our laboratory to obtain a list of possible guide sequences against three E2F isoforms (E2F1-3). Here, with the same design principles of *MultiTar*, we used *MicroAlign* to score the efficiency of the designed anti-E2F smartRNAs. We then tested the top five scored designed smartRNAs, sm1-5 (Supplementary Figure S5A), alongside with the previous best smartRNA we tested, MT1. We compared the protein levels of the E2Fs (Supplementary Figure S5B) and found that three smartRNAs, sm3-5, significantly knockdown (>30%) all three isoforms (Supplementary Figure S5C). Plotting relative protein levels against the predicted miScores, we found that a cut-off score of 55 selects efficient guide strands (Supplementary Figure S5E). Comparing to the positive control, MT1, three of the five new smartRNAs knocked down E2F1 to a similar degree or more, while four of the five new ones knocked down E2F2 or E2F3 more effectively. Following a nine-day growth assay of PC3 cells, sm3 inhibited cell growth more efficiently than MT1 (Figure 4F), while sm4 and sm5 inhibit cell growth comparably to MT1 (Supplementary Figure S5D).

Enrichment effect in public data from genome-wide studies

To further validate the suggested modular base pairing mechanism beyond the seed is playing a significant role in the targeting process of cellular miRNAs, we used standard public data used to benchmark miRNA target prediction programs (50). These data were generated by transfecting cells with three miRNAs, miR-124, miR-181 and miR-1, followed by mRNA and protein quantification, using, respectively, expression profiling and Stable Isotope Labeling with Amino acids in Cell culture (SILAC) and LC–MS/MS (liquid chromatography-mass spectrometry/mass-spectrometry). If the modular order of base pairing beyond the seed is a significant factor in target repression efficiency, then the targets that are top-ranked by the *MicroAlign* program should be enriched by effectively repressed mRNAs and proteins.

We pooled the mRNA and protein levels of the three miRNA target genes, and calculated the mean differential repression levels as $\log_2$ fold changes. We first established the mean of the 293 protein targets, which is –0.15 (Supplementary Figure S4H). Then, we sorted the target protein levels by their *miScores* and split them into three equal sized bins, which we labeled 'top', 'mid' and 'bottom'. We calculated the mean of each bin (Supplementary Figure S4H) and observed enriched repression efficiencies in the top and mid bins (near –0.2). The mid bin significantly differs from the bottom bin ($P < 0.05$). This shows that the *miScores* significantly enrich for more effectively repressed targets in the top two bins. Then, we took the mean of the top-30 proteins from each transfected sample, and we consistently observed the enrichment (mean < –0.22, $P < 0.05$, Mann–Whitney U test). Previously, similar mean repression at the protein level was achieved by the top-scored target predictions from PicTar and PITA; and an even better mean was observed from those of *TargetScan* (near –0.28). As for the programs that do not consider evolutionary conservation, they mostly yielded less significant means (> -0.1) (50).

To confirm that the enrichment is due to the base pairing order beyond the seed, we modified the *MicroAlign* program so it calculates scores according to the total number of base pairs, without considering their order. We considered the enrichment for the miR-124 and miR-181 mRNA targets in the three bins. Using the non-modified *MicroAlign* to analyze 1334 miR-124 targets, we consistently observed top and middle bins enriched in more efficiently repressed targets (Figure 4G, top left panel), as well as a significant difference between the bottom and the top two bins ($P < 0.05$ in both cases, Mann–Whitney test). When we removed the base pairing order, the enrichment of efficiently repressed targets weakened in the top and middle bins, while the bottom bin got more efficiently repressed targets (Figure 4G, top right panel). The same pattern was observed for the 98 protein levels measured by SILAC; however, the statistical significance could not be established due to the low number of data points (Figure 4G, two bottom panels).

For the 1308 miR-181 mRNA targets, the same gradual enrichment from the bottom to the top bin was observed (Figure 4H, top left), with a significant difference between the bottom and the two top bins ($P < 0.02$ and $P < 0.002$, respectively). Once again, more efficiently repressed targets are found in the bottom bin when the base pairing order was not considered (Figure 4H, top right). The same pattern was observed at the protein expression levels (Fig 4H, bottom panels).

Taken together, when miRNAs were ectopically expressed, *MicroAlign* resolves the difference in repression efficiency of the targets solely based on the hierarchical order of base pairing beyond the seed. Hence, this modular base pairing mechanism beyond the seed is playing a significant role in the targeting process of cellular miRNAs and can be used to determine the repression efficiency of AGO2-dependent miRNA silencing guides.

Structural analysis supports the modular functioning of AGO2

Published data provided us with underlying AGO2 structural information to further substantiate our hierarchical

model. We found that, in addition to the seed (Supplementary Figure S6A, left; PDB 4W5N), nts in positions 13–15 (B-module) are also exposed to the solvent when the seed of the RNA guide is annealed to an mRNA target (Supplementary Figure S6A, right). To see how base pairing occurs with the nts in modules B and C, we compared the structures of the AGO2 with and without the seed of the RNA guide annealed to an mRNA target (PDB 4F3T and 4W5R). In the bound structure, we can observe that the PAZ domain pivots as a rigid body around the base of the α-7 helix (Ser371) by ∼13° (Figure 5A, angle θ). The PAZ-MID channel opens as the α-7 helix is displaced by 4–6 Å. The displacement is amplified at the 3′ end binding site of the PAZ domain to 9.3 Å (Figure 5A, top). Meanwhile, the number of hydrogen bonds between the PAZ domain and the 3′ end of the guide RNA is reduced from five to two, or even zero in some structures (Supplementary Figure S6B–D). This is indicative of a promoted release of the 3′ end of the guide RNA from AGO2, which is required for miRISC-mediated cleavage. With its 3′ end liberated, the guide strand is free to skip the central cleft of the AGO2, and progressively base pairs with the target strand in module-B, toward module-C, allowing for the formation of the RNA duplex beyond one turn (34).

We docked a guide of fifteen bps by performing a structural alignment between the AGO2 (PDB 4W5O) and a non-cleaving mutant of the *Thermus thermophilus* AGO (PDB 3HJF) (34) (Figure 5B; RMSD ∼ 1.15 Å). In our model, AGO2 is capable of accommodating the annealed duplex, consistent with the published model (32). Moderate clashes between the side chains of the α-7 helix and the RNA guide strand occurred. However, a rotation of the PAZ domain of a few degrees can remove the clashes. Indeed, the maximum rotation of the PAZ domain was evaluated to be ∼25° (23), which largely suffices. The coulomb potential surface of the AGO2 shows that the RNA-binding pocket has a natural tendency to open due to the presence of repulsive electrostatic charges lining its interior (Supplementary Figure S6A; blue regions). This view is also in agreement with a previous report stating that the AGO2 can recognize preformed duplexes and induce cleavage (61).

A possible model for non-seed nucleotide binding to AGO2

A recent structural study reported that base pairing at positions 9–11 is hindered regardless whether the 3′ end is released or not due to the location of the α-7 helix and the 600–609 loop in protein hAGO2 (36). Consequently, continued base pairing in the 5′ to 3′ direction along the guide RNA is interrupted at the central cleft of the hAGO2. Interestingly, the enhancement of cleavage activity by base pairing beyond the central loop indicates that some degree of base pairing is beneficial in the 3′ supplementary region (24). Yet no such intermediate structure of 3′ supplementary base pairing has been resolved, due to poor visibility. Slicing activity immediately following base pairing with the target may have made it difficult to observe a conformation bound to the target (36). Structural data eventually became available for AGO2 bound to a guide strand (31,36), a duplex of the guide RNA with a partial target (35,36), and a catalytic mutant AGO2 bound with a guide-target duplex of 15 nts (62).

Combining our experimental with the structural data, we suggest that the following sequence of events takes place for miB guide RNA to achieve AGO2-mediated silencing. As the seed of miB base pair with the target, the seed duplex is accommodated in the PAZ-MID channel. Suggested by previous studies, the narrowing of the channel forms a cleft and prohibits further base pairing immediately downstream of the seed. However, while the duplex pushes the α-7 helix outward and causes the PAZ domain to pivot, the 3′ end of the RNA guide becomes less tightly bound to the PAZ domain, and thus more prone to be released. The free 3′ end facilitates the base pairing process to 'skip' the cleft and resumes in module-B, and then propagates to module-C. Once the RNA guide is bound to the mRNA target between nts 12 and 18, the duplex is formed on both sides across the narrow cleft of the PAZ-MID channel. The nts around the scissile phosphate (nts 10–11) eventually anneal with the RNA guide and 'fit' into the cleavage site, either by further pivoting of the PAZ domain, which opens the channel, or with the help of twisting motions of the duplex formation on both flanks of the channel. When the bases are complementary in the central module-A, efficient cleavage occurs. On the other hand, in the presence of mismatches, the cleavage efficiency depends on the protein tolerance for them. Moving along the steps of the duplex formation, the mRNA target becomes less and less likely to dissociate from the AGO2 complex. As the 'dwell time' of the AGO2 complex on target gets greater, so as its chances to recruit protein factors for slicer-independent repression (Figure 6).

DISCUSSION

Mismatches introduced by a miRNA near the scissile phosphate of a target are generally tolerated in spite of a compromised endonucleolytic activity. In bilaterian animals, the miRISC mediates repression predominantly via slicer-independent pathways. We found that a number of mismatches further downstream of the central region can impair repression to a greater extent following a hierarchical pattern, which became apparent when triplet mismatches were made in the non-seed region. This unique experimental design explains why this pattern remained elusive in the past despite several systematic investigations.

From this pattern, we built a computational model that evaluates the importance of base pairing beyond the seed in AGO2-mediated repression. This model suggests that when the seed region is perfectly base paired with the target, then the bps in module-B (nts 12–14) play a decisive role. This rule applies recursively to modules C (nts 15–17), A (nts 9–11), and D (nts 18–21), in this order. The idea that miRNA/guide-RNA pairing to targets is modular has been proposed and tested through structural, computational, as well as reporter assays by many groups (described in the following section). Here, in addition to this growing body of work, our model suggests that base pairing follows a hierarchical decision-making process, which resolves the subtlety of the peculiar sequence of events occurring beyond the seed.

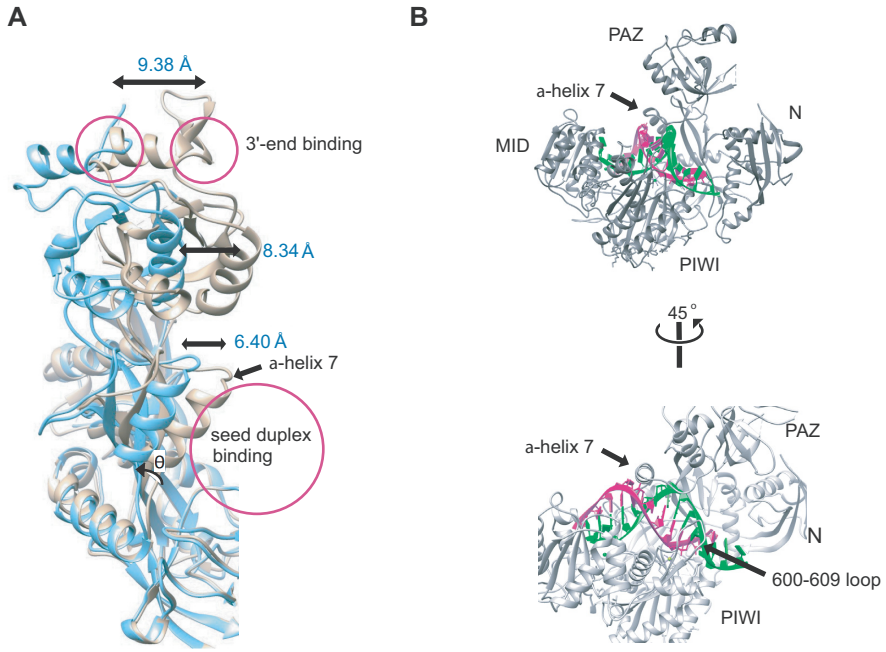


Figure 5. Structural analysis supports the proposed mechanism. (**A**) Amplified view of α-7 helix. The rotation originates at the base of α-7 helix, with a visible angle θ of about 13° between the structure before and after seed pairing. (**B**) The modeled accommodation of guide-target duplex of 15 bp in AGO2. The α-7 helix and the loop 600–609, which cause the narrowing of the central cleft of the AGO2, are interacting closely with the minor groove of the duplex. The guide strand is colored green; the target pink. The docking simulation was performed between a 15-bp guide-target duplex from *T. thermophilus* Argonaute protein (PDB 3HJF) and the human AGO2 structure (PDB 4W5T).

Simplicity and consistency of the sequential recognition model

Comparing with the nucleation-propagation model where the AGO2 is regarded as a two-state machine (34,63), our model also depicts it as a state machine with more states. State transitions take place in a stepwise fashion following a specific precedence of nt positions. Our results agree well with triplet mismatches made in a previous study of the cleavage activity of hAGO2 (25). The reason for the conservation signals detected around nt 13 (15) and the outstanding contributions of the 'supplementary' base pairing measured by Wee and coworkers (24) hence become clear: the base pairing in this module mechanically determines whether other non-seed bps should be accounted for silencing, rather than providing a large contribution in free energy of binding. More recently, the Pasquinelli group suggested that certain classes of miRNAs are capable of 3′ end pairing interactions (at nts 13-16) to outcompete miRNAs that support only seed pairing for a given site (64). This observation provides additional evidence for the proposed model and the role of the central loop.

Previous reports showed that elongating the central loop enhances repression of engineered miRNAs. Loop scores were then assigned to computational models to evaluate the effects of such central bulges (12). Our model suggests that the enlargement of the central loop, which corresponds to module-A, is likely to relax the central portion of the target so to bypass the protein structure blockage more easily and promote downstream base pairing. It has been shown that the target release was the rate-limiting step for AGO2 slicer activity (13,32,65) and that mismatches at the 3′ end of the guide RNA enhances slicer function. Our data corroborate these enhancing effects. However, we only observed enhancement under the premise that modules A, B, and C are paired. This may indicate that mismatches in these modules significantly slow down the formation of the pre-cleavage complex so that base pairing becomes the new rate-limiting step. With over 60 data points, including about half from third parties, the experimental measurements agree with the model with high confidence (Figure 4G; $P = 9.11\text{E}{-13}$), indicating that the order of base pairing beyond the seed can resolve differences in silencing efficiency.

Using microarray and SILAC data, this ability to determine silencing efficiency among over 2,500 targets became evident at the mRNA level. Using three bins, the repression levels of the top ranked targets by *MicroAlign* were consistently higher than those in the bottom bin. *TargetScan*, *miRanda*, *PITA* and *PicTar* were also able to bring similar enrichments. However, the scoring functions of these programs were derived in part from statistics including many additional factors, and in particular evolutionary conservation of the target sites. *MicroAlign* is solely based on our AGO2 mechanistic model. Its ability to enrich effectively repressed targets at the protein level was statistically less significant (Figure 4GH). This might be due to the fact that less data points were generated from SILAC. In support to this argument, the enrichment became statistically significant when we combined the protein levels from all three overexpressed miRNAs (Supplementary Figure S4H).

In total agreement with our model, it has recently been suggested that pairing in the miRNA:mRNA duplex does not move forward from the seed to the 3′ end, and the idea of

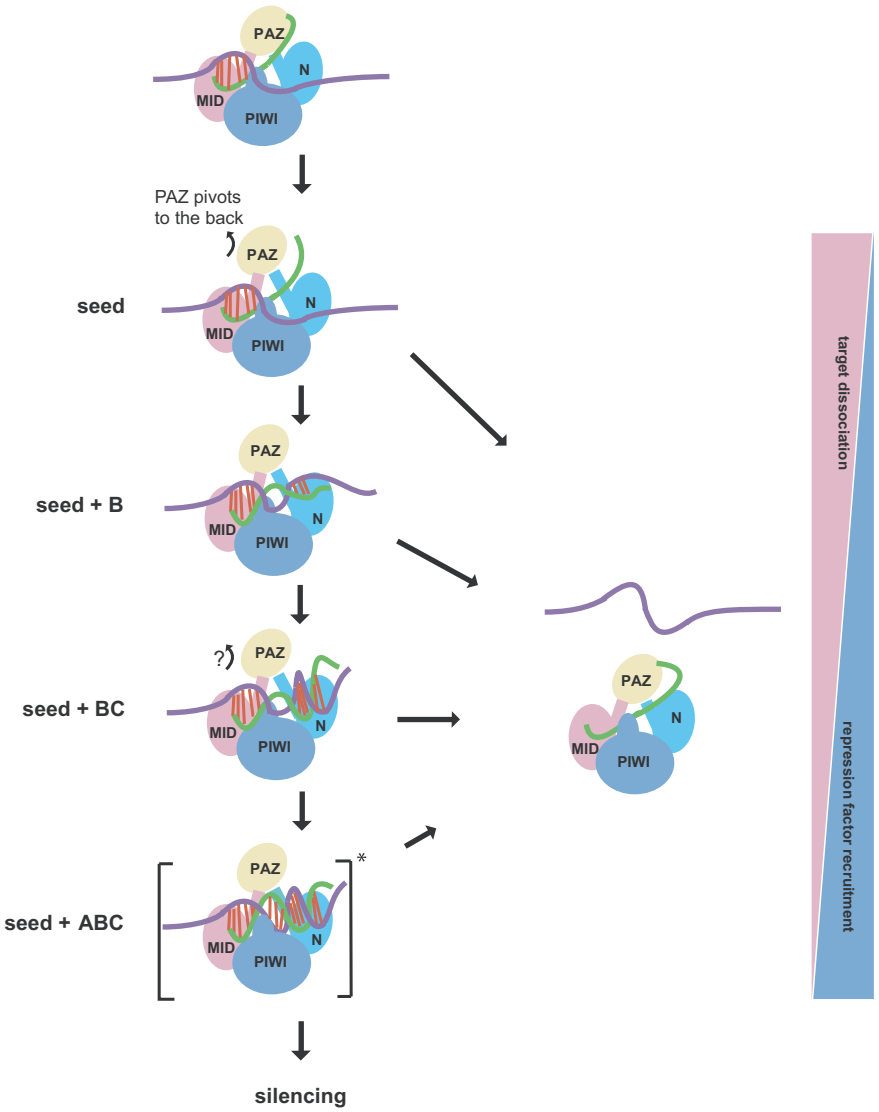


Figure 6. Summary of the skipped-propagation and coordinated annealing model. The step-wise nature of the AGO2 slicer activation process entails that it is a multi-state machine. In every step along its structural change, the efficiency of silencing becomes gradually enhanced and to the next complex structure (down arrows), or, in lack of base pairing, to the dissociation of the complex (right arrows). The greater the base pairing the more stable is the complex and longer the dwell time, which increases the probabilities to recruit repression factors (blue triangle) and decreases those of target dissociation (pink triangle).

a second nucleation site is defensible since miRNAs prefer 3′ supplementary pairing (66). Our experimental results and our structural modeling provide the first evidence of this 'skipping' model during target recognition by the miRISC. Moreover, our model further defines the order in which target recognition occurs beyond the seed, and module-B as the second nucleation site. Interestingly, a study using a massively parallel experiment reporter assay identified that miRNA position 14 (the last base in module-B) as the saturation point of mismatched bases that reduce repression (67).

Limitations of the current model

First, the goal of the *MicroAlign* program is not to predict miRNA targets, but rather to calculate the silencing efficiency of possible guide::target duplexes. *MicroAlign* does not consider extrinsic factors such as target site location, AU content, target site accessibility, abundance, and evolutionary conservation. Statistical training and combinations of these factors were shown by genome-wide analysis to have similar predictive power (50). Since *MicroAlign* was designed to evaluate 7–8 mer sites, its average performance of enrichment is upper-bounded by that of using these sites, which in the genome-wide analysis include false positives that introduce noise. Nevertheless, we observed enrichment of effectively regulated targets without combining or optimizing any additional factor. This enrichment was even greater than those obtained by prediction programs that use free energy without site conservation (50). This suggests that the order of base pairing beyond the seed plays a significant role in the regulation of the expression of the targets.

Though this feature alone is not sufficient to predict mRNA targets, it can be used in the design of effective RNA guide sequences to inhibit simultaneously multiple targets.

Second, contributions to silencing by other mechanisms, such as deadenylation, decapping and translational repression (5,24) could not be clearly discerned individually in this study. We observed that mismatches at the scissile phosphate (miRNA positions 10 and 11; module-A) impair the slicer activity, which can be made even worse if combined with mismatches in modules B or C. This suggests that the pattern we detected also reflects silencing efficiency related to slicer-independent mechanisms. For instance, by comparing miB-A, -AB, -AC and -AD, we observed the same hierarchical pattern of B > C > D, indicating that module-B directly contributes to slicer-independent repression as well. The precise mechanism by which these mismatches affect the slicer-independent pathways remains unclear. Perhaps the non-seed mismatches are capable of altering the AGO2's ability to recruit protein cofactors by changing its 'dwell time' spent on targets (68). Non-seed mismatches could thus produce different effects on the slicing and repression mechanisms. Simultaneously, they could reduce the slicing rate and define the time spent on a target, which determines the recruitment of slicer-independent repression factors. How this interplay produces the particular hierarchical pattern we observed is yet to be found.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Jerry Pelletier, Abba Malina, John Mills, Regina Cencic, Francis Robert, David Cotenoir-White, Khalid Hilmi and Justina Kulpa for discussions, cloning and cell culture materials, as well as assistance; Julie Pelloux, Jean Paquette and Angelique Bellmare-Pelletier for discussions; Sven Diederichs and Eric Cohen for constructs; and, André Laperrière for assistance.

FUNDING

Natural Sciences and Engineering Research Council of Canada (NSERC Discovery grant program to F.M.); Canadian Institutes of Health Research (CIHR) [MOP-93679]; National Institutes of Health [R01GM088813]. Funding for open access charge: NSERC.
Conflict of interest statement. None declared.

REFERENCES

1. Lewis,B.P., Shih,I.H., Jones-Rhoades,M.W., Bartel,D.P. and Burge,C.B. (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787–798.
2. John,B., Enright,A.J., Aravin,A., Tuschl,T., Sander,C. and Marks,D.S. (2004) Human MicroRNA targets. *PLoS Biol.*, **2**, e363.
3. Reyes-Herrera,P.H. and Ficarra,E. (2012) One decade of development and evolution of microRNA target prediction algorithms. *Genomics Proteomics Bioinformatics*, **10**, 254–263.
4. Saito,T. and Saetrom,P. (2010) MicroRNAs–targeting and target prediction. *Nano Biotechnol.*, **27**, 243–249.
5. Fabian,M.R., Sonenberg,N. and Filipowicz,W. (2010) Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.*, **79**, 351–379.
6. Doench,J.G. and Sharp,P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.*, **18**, 504–511.
7. Giraldez,A.J., Mishima,Y., Rihel,J., Grocock,R.J., Van Dongen,S., Inoue,K., Enright,A.J. and Schier,A.F. (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science*, **312**, 75–79.
8. Eulalio,A., Rehwinkel,J., Stricker,M., Huntzinger,E., Yang,S.F., Doerks,T., Dorner,S., Bork,P., Boutros,M. and Izaurralde,E. (2007) Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes Dev.*, **21**, 2558–2570.
9. Chen,K., Maaskola,J., Siegal,M.L. and Rajewsky,N. (2009) Reexamining microRNA site accessibility in Drosophila: a population genomics study. *PLoS One*, **4**, e5681.
10. Doench,J.G., Petersen,C.P. and Sharp,P.A. (2003) siRNAs can function as miRNAs. *Genes Dev.*, **17**, 438–442.
11. Ye,W., Lv,Q., Wong,C.K., Hu,S., Fu,C., Hua,Z., Cai,G., Li,G., Yang,B.B. and Zhang,Y. (2008) The effect of central loops in miRNA:MRE duplexes on the efficiency of miRNA-mediated gene regulation. *PLoS One*, **3**, e1719.
12. Kiriakidou,M., Nelson,P.T., Kouranov,A., Fitziev,P., Bouyioukos,C., Mourelatos,Z. and Hatzigeorgiou,A. (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.*, **18**, 1165–1178.
13. De,N., Young,L., Lau,P.W., Meisner,N.C., Morrissey,D.V. and MacRae,I.J. (2013) Highly complementary target RNAs promote release of guide RNAs from human Argonaute2. *Mol. Cell*, **50**, 344–355.
14. Friedman,R.C., Farh,K.K., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
15. Grimson,A., Farh,K.K., Johnston,W.K., Garrett-Engele,P., Lim,L.P. and Bartel,D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.
16. Boden,D., Pusch,O., Silbermann,R., Lee,F., Tucker,L. and Ramratnam,B. (2004) Enhanced gene silencing of HIV-1 specific siRNA using microRNA designed hairpins. *Nucleic Acids Res.*, **32**, 1154–1158.
17. Houzet,L., Klase,Z., Yeung,M.L., Wu,A., Le,S.Y., Quinones,M. and Jeang,K.T. (2012) The extent of sequence complementarity correlates with the potency of cellular miRNA-mediated restriction of HIV-1. *Nucleic Acids Res.*, **40**, 11684–11696.
18. Kamola,P.J., Nakano,Y., Takahashi,T., Wilson,P.A. and Ui-Tei,K. (2015) The siRNA Nnon-seed region and its target sequences are auxiliary determinants of off-target effects. *PLoS Comput. Biol.*, **11**, e1004656.
19. Du,Q., Thonberg,H., Wang,J., Wahlestedt,C. and Liang,Z. (2005) A systematic analysis of the silencing effects of an active siRNA at all single-nucleotide mismatched target sites. *Nucleic Acids Res.*, **33**, 1671–1677.
20. Saxena,S., Jonsson,Z.O. and Dutta,A. (2003) Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *J. Biol. Chem.*, **278**, 44312–44319.
21. Holen,T., Amarzguioui,M., Wiiger,M.T., Babaie,E. and Prydz,H. (2002) Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. *Nucleic Acids Res.*, **30**, 1757–1766.
22. Hibio,N., Hino,K., Shimizu,E., Nagata,Y. and Ui-Tei,K. (2012) Stability of miRNA 5′terminal and seed regions is correlated with experimentally observed miRNA-mediated silencing efficacy. *Sci. Rep.*, **2**, 996.
23. Robertson,B., Dalby,A.B., Karpilow,J., Khvorova,A., Leake,D. and Vermeulen,A. (2010) Specificity and functionality of microRNA inhibitors. *Silence*, **1**, 10.
24. Wee,L.M., Flores-Jasso,C.F., Salomon,W.E. and Zamore,P.D. (2012) Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell*, **151**, 1055–1067.
25. Lima,W.F., Wu,H., Nichols,J.G., Sun,H., Murray,H.M. and Crooke,S.T. (2009) Binding and cleavage specificities of human Argonaute2. *J. Biol. Chem.*, **284**, 26017–26028.

26. Agarwal,V., Bell,G.W., Nam,J.W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**, e05005.
27. Majoros,W.H., Lekprasert,P., Mukherjee,N., Skalsky,R.L., Corcoran,D.L., Cullen,B.R. and Ohler,U. (2013) MicroRNA target site identification by integrating sequence and binding information. *Nat. Methods*, **10**, 630–633.
28. Swarts,D.C., Makarova,K., Wang,Y., Nakanishi,K., Ketting,R.F., Koonin,E.V., Patel,D.J. and van der Oost,J. (2014) The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.*, **21**, 743–753.
29. Bohmert,K., Camus,I., Bellini,C., Bouchez,D., Caboche,M. and Benning,C. (1998) AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J.*, **17**, 170–180.
30. Willkomm,S., Zander,A., Gust,A. and Grohmann,D. (2015) A prokaryotic twist on argonaute function. *Life (Basel)*, **5**, 538–553.
31. Elkayam,E., Kuhn,C.D., Tocilj,A., Haase,A.D., Greene,E.M., Hannon,G.J. and Joshua-Tor,L. (2012) The structure of human argonaute-2 in complex with miR-20a. *Cell*, **150**, 100–110.
32. Willkomm,S. and Restle,T. (2015) Conformational dynamics of Ago-mediated silencing processes. *Int. J. Mol. Sci.*, **16**, 14769–14785.
33. Yuan,Y.R., Pei,Y., Ma,J.B., Kuryavyi,V., Zhadina,M., Meister,G., Chen,H.Y., Dauter,Z., Tuschl,T. and Patel,D.J. (2005) Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell*, **19**, 405–419.
34. Wang,Y., Juranek,S., Li,H., Sheng,G., Wardle,G.S., Tuschl,T. and Patel,D.J. (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature*, **461**, 754–761.
35. Schirle,N.T. and MacRae,I.J. (2012) The crystal structure of human Argonaute2. *Science*, **336**, 1037–1040.
36. Schirle,N.T., Sheu-Gruttadauria,J. and MacRae,I.J. (2014) Structural basis for microRNA targeting. *Science*, **346**, 608–613.
37. Wang,Y., Juranek,S., Li,H., Sheng,G., Tuschl,T. and Patel,D.J. (2008) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature*, **456**, 921–926.
38. Boden,D., Pusch,O., Lee,F., Tucker,L. and Ramratnam,B. (2003) Human immunodeficiency virus type 1 escape from RNA interference. *J. Virol.*, **77**, 11531–11535.
39. Malina,A., Mills,J.R., Cencic,R., Yan,Y., Fraser,J., Schippers,L.M., Paquet,M., Dostie,J. and Pelletier,J. (2013) Repurposing CRISPR/Cas9 for in situ functional assays. *Genes Dev.*, **27**, 2602–2614.
40. Lee,T., Di Paola,D., Malina,A., Mills,J.R., Kreps,A., Grosse,F., Tang,H., Zannis-Hadjopoulos,M., Larsson,O. and Pelletier,J. (2014) Suppression of the DHX9 helicase induces premature senescence in human diploid fibroblasts in a p53-dependent manner. *J. Biol. Chem.*, **289**, 22798–22814.
41. Mills,J.R., Malina,A., Lee,T., Di Paola,D., Larsson,O., Miething,C., Grosse,F., Tang,H., Zannis-Hadjopoulos,M., Lowe,S.W. *et al.* (2013) RNAi screening uncovers Dhx9 as a modifier of ABT-737 resistance in an Emu-myc/Bcl-2 mouse model. *Blood*, **121**, 3402–3412.
42. Dickins,R.A., Hemann,M.T., Zilfou,J.T., Simpson,D.R., Ibarra,I., Hannon,G.J. and Lowe,S.W. (2005) Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nat. Genet.*, **37**, 1289–1295.
43. De Guire,V., Caron,M., Scott,N., Menard,C., Gaumont-Leclerc,M.F., Chartrand,P., Major,F. and Ferbeyre,G. (2010) Designing small multiple-target artificial RNAs. *Nucleic Acids Res.*, **38**, e140.
44. Diederichs,S., Jung,S., Rothenberg,S.M., Smolen,G.A., Mlody,B.G. and Haber,D.A. (2008) Coexpression of Argonaute-2 enhances RNA interference toward perfect match binding sites. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 9284–9289.
45. Kiethega,G.N., Yan,Y., Turcotte,M. and Burger,G. (2013) RNA-level unscrambling of fragmented genes in Diplonema mitochondria. *RNA Biol.*, **10**, 301–313.
46. Mignacca,L., Saint-Germain,E., Benoit,A., Bourdeau,V., Moro,A. and Ferbeyre,G. (2016) Sponges against miR-19 and miR-155 reactivate the p53-Socs1 axis in hematopoietic cancers. *Cytokine*, **82**, 80–86.
47. Luo,X., Zhang,J., Wang,H., Du,Y., Yang,L., Zheng,F. and Ma,D. (2012) PolyA RT-PCR-based quantification of microRNA by using universal TaqMan probe. *Biotechnol. Lett.*, **34**, 627–633.
48. Zhang,J., Du,Y.Y., Lin,Y.F., Chen,Y.T., Yang,L., Wang,H.J. and Ma,D. (2008) The cell growth suppressor, mir-126, targets IRS-1. *Biochem. Biophys. Res. Commun.*, **377**, 136–140.
49. Burger,G., Yan,Y., Javadi,P. and Lang,B.F. (2009) Group I-intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. *Trends Genet.*, **25**, 381–386.
50. Baek,D., Villen,J., Shin,C., Camargo,F.D., Gygi,S.P. and Bartel,D.P. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64–71.
51. Pruitt,K.D., Tatusova,T., Klimke,W. and Maglott,D.R. (2009) NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.*, **37**, D32–D36.
52. Weill,N., Lisi,V., Scott,N., Dallaire,P., Pelloux,J. and Major,F. (2015) MiRBooking simulates the stoichiometric mode of action of microRNAs. *Nucleic Acids Res.*, **43**, 6730–6738.
53. Tan,X., Lu,Z.J., Gao,G., Xu,Q., Hu,L., Fellmann,C., Li,M.Z., Qu,H., Lowe,S.W., Hannon,G.J. *et al.* (2012) Tiling genomes of pathogenic viruses identifies potent antiviral shRNAs and reveals a role for secondary structure in shRNA efficacy. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 869–874.
54. Wilkinson,K.A., Gorelick,R.J., Vasa,S.M., Guex,N., Rein,A., Mathews,D.H., Giddings,M.C. and Weeks,K.M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.*, **6**, e96.
55. Gu,S., Jin,L., Zhang,F., Sarnow,P. and Kay,M.A. (2009) Biological basis for restriction of microRNA targets to the 3′ untranslated region in mammalian mRNAs. *Nat. Struct. Mol. Biol.*, **16**, 144–150.
56. Chi,S.W., Zang,J.B., Mele,A. and Darnell,R.B. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.
57. Bosson,A.D., Zamudio,J.R. and Sharp,P.A. (2014) Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Mol. Cell*, **56**, 347–359.
58. Watts,J.M., Dang,K.K., Gorelick,R.J., Leonard,C.W., Bess,J.W. Jr, Swanstrom,R., Burch,C.L. and Weeks,K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.
59. Long,D., Lee,R., Williams,P., Chan,C.Y., Ambros,V. and Ding,Y. (2007) Potent effect of target structure on microRNA function. *Nat. Struct. Mol. Biol.*, **14**, 287–294.
60. Broderick,J.A., Salomon,W.E., Ryder,S.P., Aronin,N. and Zamore,P.D. (2011) Argonaute protein identity and pairing geometry determine cooperativity in mammalian RNA silencing. *RNA*, **17**, 1858–1869.
61. Janas,M.M., Wang,B., Harris,A.S., Aguiar,M., Shaffer,J.M., Subrahmanyam,Y.V., Behlke,M.A., Wucherpfennig,K.W., Gygi,S.P., Gagnon,E. *et al.* (2012) Alternative RISC assembly: binding and repression of microRNA-mRNA duplexes by human Ago proteins. *RNA*, **18**, 2041–2055.
62. Wang,X., Ye,L., Hou,W., Zhou,Y., Wang,Y.J., Metzger,D.S. and Ho,W.Z. (2009) Cellular microRNA expression correlates with susceptibility of monocytes/macrophages to HIV-1 infection. *Blood*, **113**, 671–674.
63. Knott,S.R.V., Maceli,A.R., Erard,N., Chang,K., Marran,K., Zhou,X., Gordon,A., El Demerdash,O., Wagenblast,E., Kim,S. *et al.* (2014) A computational algorithm to predict shRNA potency. *Mol. Cell*, **56**, 796–807.
64. Broughton,J.P., Lovci,M.T., Huang,J.L., Yeo,G.W. and Pasquinelli,A.E. (2016) Pairing beyond the seed supports microRNA targeting specificity. *Mol. Cell*, **64**, 320–333.
65. Deerberg,A., Willkomm,S. and Restle,T. (2013) Minimal mechanistic model of siRNA-dependent target RNA slicing by recombinant human Argonaute 2 protein. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 17850–17855.
66. Bartel,D.P. (2018) Metazoan microRNAs. *Cell*, **173**, 20–51.
67. Vainberg Slutskin,I., Weingarten-Gabbay,S., Nir,R., Weinberger,A. and Segal,E. (2018) Unraveling the determinants of microRNA mediated regulation using a massively parallel reporter assay. *Nat. Commun.*, **9**, 529.
68. Chandradoss,S.D., Schirle,N.T., Szczepaniak,M., MacRae,I.J. and Joo,C. (2015) A dynamic search process underlies MicroRNA targeting. *Cell*, **162**, 96–107.