Molecular Therapy
Nucleic Acids
Original Article

HT-SELEX-based identification of binding pre-miRNA hairpin-motif for small molecules

Sanjukta Mukherjee,[1,3] Asako Murata,[1] Ryoga Ishida,[2] Ayako Sugai,[1] Chikara Dohno,[1] Michiaki Hamada,[2] Sudhir Krishna,[3] and Kazuhiko Nakatani[1]

[1]Department of Regulatory Bioorganic Chemistry, The Institute of Scientific and Industrial Research (SANKEN), Osaka University, 8-1 Mihogaoka, Ibaraki 567-0047, Japan; [2]Graduate School of Advanced Science and Engineering, Waseda University, 55N-06-10, 3-4-1 Okubo Shinjuku-ku, Tokyo 169-8555, Japan; [3]National Centre for Biological Sciences (NCBS), Tata Institute of Fundamental Research (TIFR), Bellary Road, Bangalore 560065, India

Selective targeting of biologically relevant RNAs with small molecules is a long-standing challenge due to the lack of clear understanding of the binding RNA motifs for small molecules. The standard SELEX procedure allows the identification of specific RNA binders (aptamers) for the target of interest. However, more effort is needed to identify and characterize the sequence-structure motifs in the aptamers important for binding to the target. Herein, we described a strategy integrating high-throughput (HT) sequencing with conventional SELEX followed by bioinformatic analysis to identify aptamers with high binding affinity and target specificity to unravel the sequence-structure motifs of pre-miRNA, which is essential for binding to the recently developed new water-soluble small-molecule CMBL3aL. To confirm the fidelity of this approach, we investigated the binding of CMBL3aL to the identified motifs by surface plasmon resonance (SPR) spectroscopy and its potential regulatory activity on dicer-mediated cleavage of the obtained aptamers and endogenous pre-miRNAs comprising the identified motif in its hairpin loop. This new approach would significantly accelerate the identification process of binding sequence-structure motifs of pre-miRNA for the compound of interest and would contribute to increase the spectrum of biomedical application.

INTRODUCTION

MicroRNAs (miRNAs) are a growing class of non-coding RNAs (ncRNAs) that have emerged as important regulators of gene expression and are predicted to regulate at least half of the human transcriptome at the posttranscriptional level.[1-4] miRNAs are generated from their primary transcripts (pri-miRNAs) by two successive enzymatic reactions: digestion of pri-miRNA by enzyme Drosha to give a 3′ overhang hairpin precursor RNA (pre-miRNA) and subsequent digestion of pre-miRNA by enzyme dicer to a mature miRNA of 18–25 nucleotides (nt) in length. miRNAs are assembled into the RNA-induced silencing complexes (RISCs) with argonaute proteins (Ago2), which bind to the 3′-untranslated region (3′-UTR) of target mRNAs and induces translational repression or mRNA degradation.[5-9] According to the miRNA database (miRBase: http://www.mirbase.org/), 1,917 miRNA precursors and 2,656 mature miRNAs have been so far identified in humans.[5] Misregulation of miRNA expression has been linked to variety of diseases,[10-15] such as diabetes,[16] cardiovascular,[17,18] neurological,[19,20] kidney,[21] infectious,[22] and particularly cancers.[23-29]

Due to the growing impact of regulatory effects of miRNAs associated with different diseases, it has become a challenge to develop strategies for regulating miRNA function.[30-32] There are two major tools to regulate miRNA function: one is antisense oligonucleotides (antagomirs) complementary to the mature miRNA and the other is small molecules that target miRNA biogenesis and/or the formation of RISCs.[30-32] The small-molecule modulators[33-36] may have advantages over antisense oligonucleotides because they are easily deliverable, their ease of use, and they are cost-effective for chemical alterations and manufacturing. However, the specificity of small molecules to the target sequence and/or secondary structure remains a concern and should be further investigated for application as RNA-targeting drugs. Targeting RNA by small molecules is at an emergent stage and encumbered with challenges due to the lack of fundamental understanding about the chemical motifs within small molecules capable of binding to a specific RNA. Most of the small-molecule modulators of miRNA biogenesis and/or function reported so far were discovered by library screening. Rational design of small molecules[37-42] to target and regulate a particular miRNA has been attempted; however, researchers are still far from being able to design novel and potent small-molecule modulators of miRNA function.

Systematic evolution of ligand by exponential enrichment (SELEX)[43-49] is a powerful *in vitro* method for the selection and identification of specific RNA binders for a given molecular target from an RNA library with a randomized sequence. We here describe our high-throughput (HT)-SELEX-based approach to identify hairpin loop sequence motifs

Received 11 May 2021; accepted 28 November 2021;
https://doi.org/10.1016/j.omtn.2021.11.021.

Correspondence: Kazuhiko Nakatani, Department of Regulatory Bioorganic Chemistry, The Institute of Scientific and Industrial Research (SANKEN), Osaka University, 8-1 Mihogaoka, Ibaraki 567-0047, Japan.
E-mail: nakatani@sanken.osaka-u.ac.jp
Correspondence: Sanjukta Mukherjee, National Centre for Biological Sciences (NCBS), Tata Institute of Fundamental Research (TIFR), Bellary Road, Bangalore 560065, India.
E-mail: msanjukta@ncbs.res.in

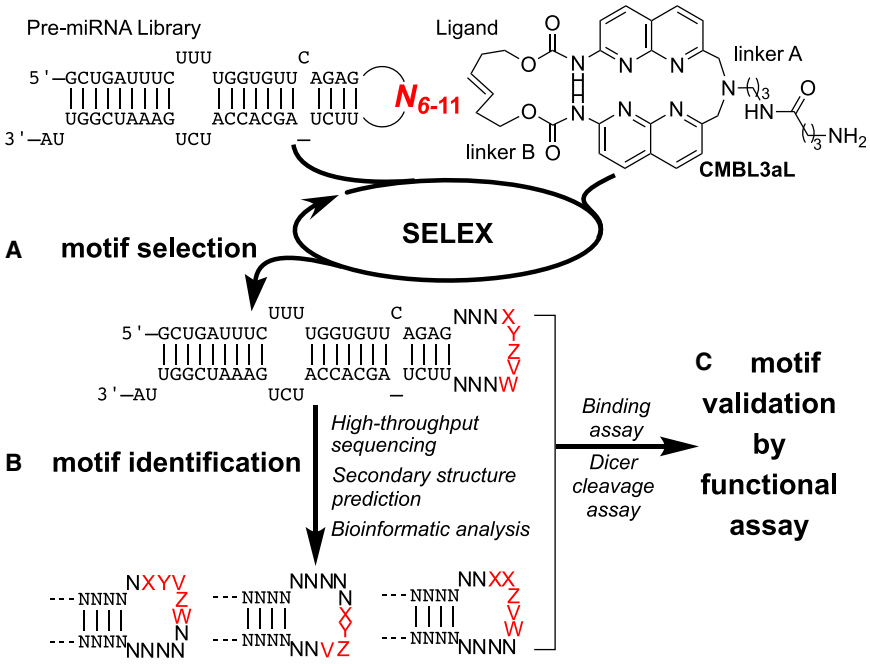

Figure 1. Conceptual illustration of the HT-SELEX-based identification of potential pre-miRNA motifs for a small molecular ligand, CMBL3aL

(A) motif selection by SELEX using a randomized hairpin loop library, (B) identification of characteristic sequence-structure motifs of aptamers by bioinformatic analysis using high-throughput (HT) sequencing data, and (C) validation of the interaction between the small molecule and the identified aptamers.

of pre-miRNA serving as the binding site for a small molecule of interest (Figure 1). A pre-miRNA contains a hairpin loop structure, which is relatively close to the dicer cleavage site. Therefore, small molecules binding to the hairpin loop might have potential to inhibit/promote pre-miRNA processing.[50] Furthermore, the hairpin loop of each pre-miRNA is comprised of unique sequences with different lengths, which can serve as a selective binding site for a small molecule. Identification of binding hairpin-sequence motifs of pre-miRNA for a small molecule of interest will provide an opportunity for the small molecule to regulate a specific pre-miRNA processing by interfering with the interaction between pre-miRNA and dicer.[33-36]

Our approach consists of three steps: (1) HT-SELEX against a small molecule using a pre-miRNA library with a randomized hairpin loop by integrating conventional SELEX[51] (Scheme S3) with HT sequencing, (2) identification of characteristic sequence-structure motifs for binding to the small molecule by bioinformatic analysis of HT sequencing data, and (3) validation of binding of the small molecule to the identified motifs, and the effect of the ligand's binding on the dicer cleavage reaction (Figure 1).

CMBLs (cyclic mismatch binding ligands) are recently reported compounds that were designed rationally to recognize an array of guanines (Gs) in nucleic acids by the complementary Watson-Crick hydrogen bonding surface.[52,53] Some of CMBLs were shown to bind to specific sequence-structure motifs in DNA/RNA;[52-55] however, binding RNA targets for our newly developed CMBL3aL have not been identified yet. The solubility of the drug candidate is an important parameter but also a major challenge for pharmacological researches.[56] Addition of a long linker possessing a terminal primary amine in CMBL3aL (Figure 1) makes it highly soluble in water, showing its future potential for any pharmacological evaluation. Herein, we report the synthesis of CMBL3aL (Scheme 1) and demonstrate the potential of our HT-SELEX-based strategy to identify binding hairpin-sequence motifs of pre-miRNA for CMBL3aL.

RESULTS

Motif selection: Identification of enriched sequence motif by SELEX

CMBL3aL has a long linker possessing a terminal primary amine, which was used for immobilization of CMBL3aL on the resin surface (Scheme S2) for SELEX study. Furthermore, for SELEX, two RNA libraries were used, one containing 11 nt (N_{11} library) and the other containing mixture of 6, 7, 8, 9, 10, and 11 nt (N_{6-11} library) randomized hairpin loops. Both RNA libraries were designed based on the secondary structure of pre-miR29a (Figure 1). After each selection cycle, eluted RNAs were reverse transcribed to cDNA and amplified by PCR (Figure S1), followed by sequencing using an Ion PGM sequencer.

Careful inspection of the HT sequencing data obtained from both N_{11} and N_{6-11} libraries revealed that most of the enriched aptamers at advanced selection cycles comprise consensus sequence motifs in the randomized region (Tables S1 and S2) with the array of Gs, particularly consecutive Gs (GG) flanked by one or two uracils (U). From the aptamers enriched in the N_{11} pools, we selected five most enriched aptamers (R-Seq1, 2, 3, 5, and 6) in the final selection cycle and one aptamer (R-Seq4) that was dominant in the 8^{th}–10^{th} selection cycles for further analysis (Figures 2A, 2B, and S2). The change in frequency of these aptamers over the selection cycles revealed a gradual enrichment of R-Seq4, which became dominant (frequency = 0.18) at the 9^{th} selection cycle. Prolonged selection cycle resulted in a decrease in the frequency of R-Seq4 and the appearance of elongated aptamers (R-Seq1, 2, 3, 5, and 6) (Figure 2A; Table S1). Formation of such an elongated by-product has already been reported, and several incidences can lead to the formation of an undesired by-product, such as random nature of the library sequences and their mispriming.[57,58] However, the underlying reasons for the elongation of aptamers in our study are not clear at this moment. Among these elongated aptamers, R-Seq1 was most significantly enriched after the 10^{th} selection cycle, and its frequency reached 0.82 at the 12^{th} cycle. For SELEX using

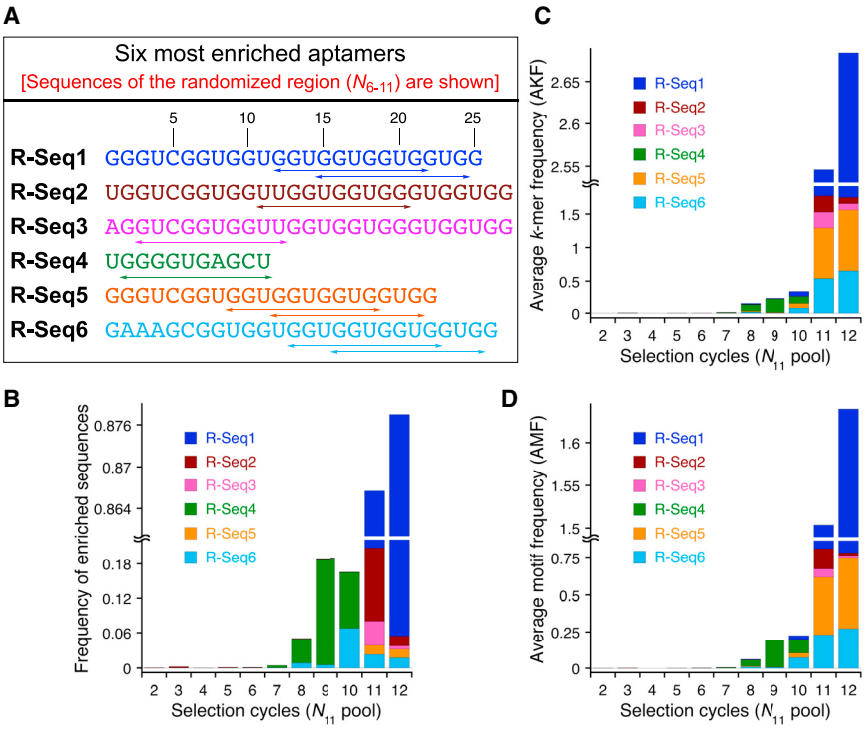

Figure 2. Analysis of sequencing data of SELEX selection pools (N_{11} library) by RaptRanker *in silico* method

(A) Sequences in the randomized region of selected six aptamers (R-Seq1–6). (B) Frequency of R-Seq1–6 in the 2nd–12th pools. Transition of (C) the averaged *k*-mer frequency (AKF) and (D) the averaged motif frequency (AMF) of 10-mer sub-sequences for R-Seq1–6 unique sequences across the selection cycles. Double-sided arrows below each sequence depict the sub-sequences used to investigate motif transition in the course of selection (Figures 3C–3F). Sub-sequences of R-Seq1 (positions 12 and 15), R-Seq5 (positions 9 and 11), and R-Seq6 (positions 13 and 16) were grouped into the same cluster.

the N_{6-11} library, we obtained essentially the same enriched sequences to those obtained from the N_{11} library (Table S2). R-Seq2 was preferentially enriched in the N_{6-11} pools during the 7th–10th selection cycles (frequency = 0.026, 0.26, 0.79, and 0.46). At the 10th selection cycle, R-Seq2 becomes diminished and this was associated with the appearance and gradual enrichment of R-Seq1 during the 8th–10th selection cycles (frequency = 0.003, 0.017, and 0.24) (Figure S2; Table S2).

Motif identification: Binding sequence-structure motifs identification using an *in silico* method

Next, we performed a more detailed analysis of the obtained sequences to determine sequence-structure motifs important for binding of these aptamers to CMBL3aL. We used recently reported RaptRanker,[59] an *in silico* method for identifying truly high binding affinity aptamers from HT-SELEX data. This method analyzes HT-SELEX data by evaluating not only the sequence but also the secondary structural motif, and thereby allows for identification of truly important sequence-structure motifs for binding to the target molecule. A brief description of the RaptRanker method mentioned here:[59] RaptRanker generates a set of sub-sequences with a defined window size (*k*-mer, 10-mer in this study) from the unique sequence of an aptamer, and computes their sub-sequence-structure profiles (SSSPs). According to the similarity of SSSPs, RaptRanker groups sub-sequences into a set of clusters and then calculates the motif frequency for each cluster, which is the sum of the frequency of sub-sequences in a cluster. To evaluate the enrichment of R-Seq1–6 at the sub-sequence level, we calculated the averaged *k*-mer frequency (AKF) and the averaged motif frequency (AMF) of sub-sequences for each of the six aptamers R-Seq1–6, and plotted them against selection cycles (Figures 2C and 2D). AKF and AMF were obtained by dividing the sum of the frequencies of sequences and motifs, respectively, of sub-sequences obtained from each unique sequence by the number of sub-sequences.

A significant increase in AKF and AMF was observed for R-Seq1, 2, 3, 5, and 6 after the 9th cycle, indicating the enrichment of the sequences and secondary structural motifs of sub-sequences obtained from the unique sequences of these aptamers. Since R-Seq1 was considerably enriched in the pool from the 10th to 12th cycles (Figure 2B), this made it difficult to observe the enrichment of the other aptamers. However, the analysis of SSSPs revealed a substantial enrichment of the other elongated aptamer, especially R-Seq5 and 6, at the sub-sequence level.

We finally visualized the transition of sequence-structure motif of the most enriched sub-sequence obtained from the six aptamers (Figures 3C–3F). The sub-sequences listed in Figure 3B had the highest motif frequency in the final selection cycle among the other sub-sequences obtained from the unique sequence of R-Seq1–6. Note that the five elongated aptamers (R-Seq1, 2, 3, 5, and 6) shared multiple sub-sequences (Figure 2B) and, among them, the sub-sequence "GGUGGUGGUG" of R-Seq1 (positions 12 and 15), R-Seq5 (positions 9 and 11), and R-Seq6 (positions 13 and 16) were grouped into the same cluster, which resulted in having the same motif frequency. As shown in Figure 3C, most of the nucleotide positions in the sub-sequence GGUGGUGGUG belong to "Stem" during the selection cycles. However, the structural context of 3′ end UG changed from Stem to "Hairpin" in the course of selection. Probability changes for some guanines at different positions between Stem and "Internal" were also observed. In the sub-sequences UUGGUGGUGG and GUCGGUGGUU, respectively, of R-Seq2 and 3 (Figures 3D and 3E), most of the nucleotides were categorized into Stem and Internal. Unlike the sub-sequences of R-Seq1, 5, and 6, the structural context did not change significantly at all

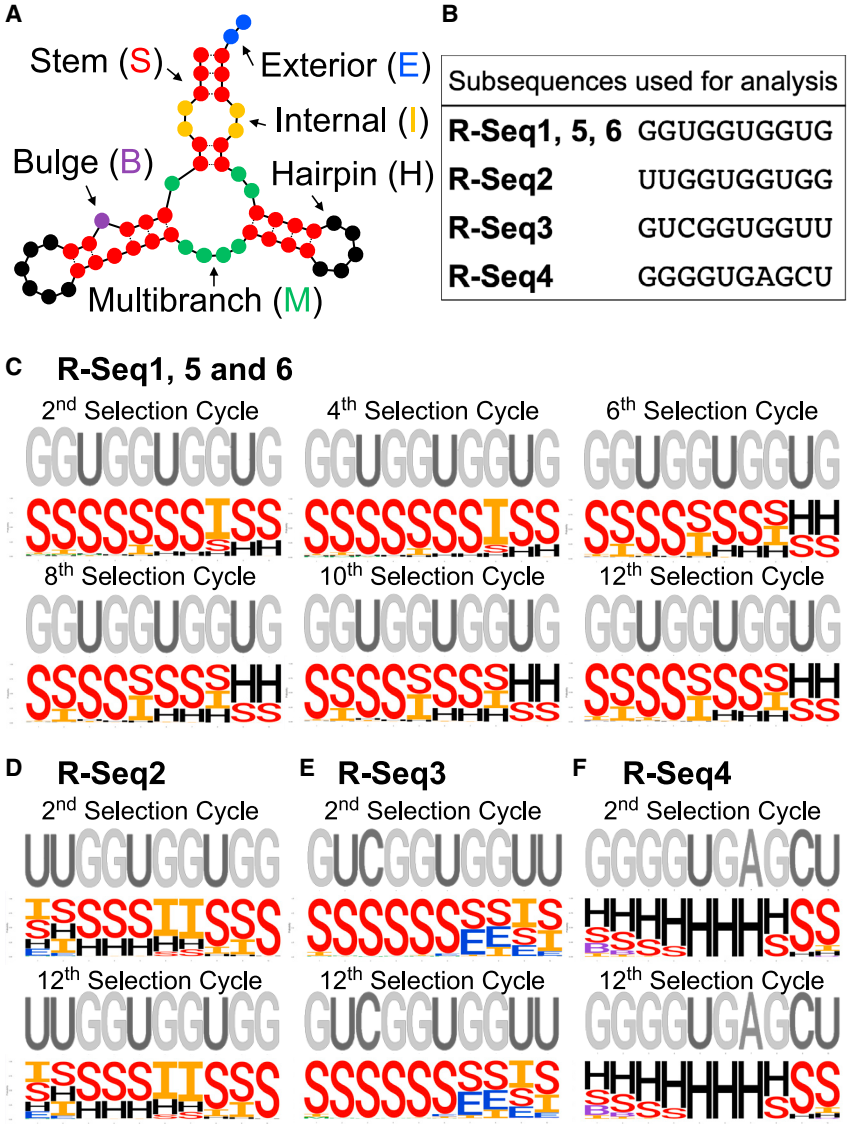

Figure 3. Sequence-structure motif transfer profile across selection cycles of the N_{11} library by RaptRanker analysis

(A) Six secondary structure features (B, E, H, I, M, and S) used to determine the structure profile of a sequence in the RaptRanker analysis. (B) The most enriched sub-sequences for R-Seq1–6 in the final selection cycle (12^{th} cycle). (C–F) The transition of the averaged sub-sequence motifs in the course of selection of the N_{11} library. All the sub-sequences are represented with the 5′ end to the left and the 3′ end to the right.

nucleotide positions. On the other hand, most of the nucleotide positions in the sub-sequence GGGGUGAGCU of R-Seq4 were categorized into Hairpin (Figure 3F) and the change of the structural context was not observed.

Considering the fact that a shift of structural context from Stem to Hairpin was observed in the enriched sub-sequences of R-Seq1, 5, and 6 in later selection cycles, Hairpin might be a more preferable structural feature as a binding site for CMBL3aL. To gain deeper insights into the binding sequence-structure motifs for CMBL3aL, we further investigated the sequence similarity of the hairpin loops formed in the enriched aptamers in advanced selection cycles (Figure 4A). Careful scrutiny of the overall structure of these aptamers revealed that those aptamers harbor one or more hairpin loop with conserved sequences of variable length (AL) (Figures 4A, 4B, S3, and S4). Those encompass characteristic sequence consensus of consecutive guanines flanked by one or two uracils (UGG/GGU/UGGU/UGGUG), indicating the importance of these characteristic hairpin-sequence motifs for CMBL3aL binding.

Motif validation: Binding assays using SPR spectroscopy

Next, we selected four aptamers R-Seq1–4 and investigated their binding to CMBL3aL (Figure 4A) by surface plasmon resonance (SPR) spectroscopy. The binding kinetics were determined in single-cycle kinetic mode using a 1:1 fitting model. CMBL3aL showed rapid association and slow dissociation kinetics toward R-Seq1, 3, and 4 immobilized surfaces (Figures 4C, 4E, and 4F), giving an apparent dissociation constant ($K_{d(app)}$) values of 218, 682, and 135 nM, respectively (Table S3). To the R-Seq2 immobilized sensor surface (Figure 4D), CMBL3aL showed faster dissociation rate, which resulted in a larger $K_{d(app)}$ of 1,310 nM. For R-Seq1–3, comparatively higher SPR responses were observed than in R-Seq4 (Figures 4C–4F; Table S3). Along with the core binding sequence motif, the flanking sequence context and overall structure of the substrate RNA plays a vital regulatory role on binding affinity as we observed among R-Seq1–4. R-Seq1, 2, and 3 comprise dumbbell-shaped secondary structural motifs, which consist of double-helical stem closed by two hairpin loops. On the other hand, R-Seq4 consists of hairpin loop structures. In R-Seq1, the expected binding sequence motif was present in both hairpins. However, in R-Seq2, 3, and 4 the expected binding sequence motif was present only in one hairpin loop. Furthermore, R-Seq4 differs in its flanking sequence context with respect to R-Seq1, 2, and 3. Next, we evaluated the binding affinity of CMBL3aL toward pre-miR29a, which does not contain a characteristic binding sequence motif. Very weak SPR response and hence low binding affinity was observed for CMBL3aL toward pre-miR29a, comprising the AU-rich hairpin-motif (Figure S5), clearly

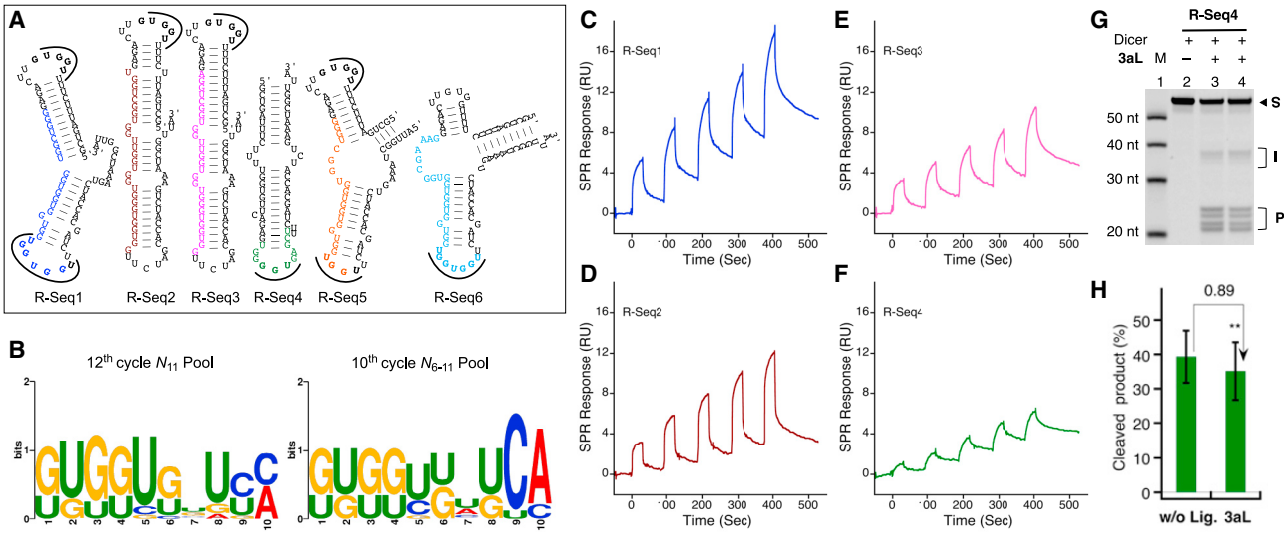

Figure 4. Predicted structure of enriched aptamers for CMBL3aL, their binding study, and regulatory activity on the dicer cleavage reaction

(A) Predicted secondary structure of the enriched RNA aptamers (R-Seq1–6) using CentroidFold (http://rtools.cbrc.jp/cgi-bin/index.cgi). (B) Consensus hairpin-sequence motifs in predicted secondary structure (Figures S3 and S4) of enriched aptamers (rank 1–10) at final selection cycles of N_{11} (12th cycle) and N_{6-11} (10th cycle) pool using TOMTOM in MEME suits (version 5.1.1). The letter thicknesses represent the probability of that nucleobase. The motif starts are represented numerically on the x axis. (C–F) SPR sensorgrams of CMBL3aL (100–500 nM) binding to R-Seq1–4. The binding kinetics were determined in single-cycle kinetic mode. (G) Gel image and (H) bar graph representation of the fraction of cleaved products for R-Seq4 in the absence and presence of CMBL3aL (250 μM) from the dicer cleavage reaction. Single cleavage intermediates and double cleavage products denoted by "I" and "P." Bars represent the average of triplicate reactions, and error bars represent standard deviations. **$p < 0.05$ is CMBL3aL-treated versus untreated control sample (paired t test).

demonstrating the importance of the identified sequence motif for CMBL3aL binding.

Functional assays: Effect of CMBL3aL binding on the dicer cleavage reaction of aptamers

Secondary structures in a pre-miRNA and the local structural environment of dicer binding sites have a profound impact on dicer-mediated cleavage and its efficiency.[60-67] To evaluate the effect of CMBL3aL on dicer cleavage of these aptamers, we performed an *in vitro* dicer cleavage reaction in the absence and presence of CMBL3aL using R-Seq1, 2, and 4 as the substrates. Formation of cleaved products (I + P) was observed for R-Seq1, 2, and 4, with a fraction of 15.8%, 16.0%, and 39.3%, respectively, in the absence of CMBL3aL (Figures 4G, 4H, and S6). Reduced formation of the cleaved products for R-Seq1 and 2 compared with R-Seq4 was observed, probably due to the dumbbell-shaped structures of R-Seq1 and 2 (Figure 4A). In the presence of CMBL3aL, no significant change (1.03-fold) in cleaved product formation was observed for R-Seq1 (Figure S6). Dumbbell-shaped siRNA mimics have been already reported to serve as a substrate for the dicer[66] but, in some mimics (mainly depending on the length of stem region), inefficient cleavage by the dicer was observed. In the presence of CMBL3aL, the fraction of cleaved products for R-Seq4 decreased by 0.89-fold and increased by 1.86-fold for R-Seq2. R-Seq4 form preferable hairpin loop structures, which are favorable binding substrates for dicers (Figures 4G and 4H).[67] Change in the conformation of R-Seq2 upon CMBL3aL binding might lead to a more favorable structure for dicer recognition, resulting in an increase in cleaved products.

Application: Effect of CMBL3aL binding on endogenous pre-miRNAs comprising partial binding motifs

To evaluate the scope of this method and expand its applicability on biologically significant endogenous pre-miRNA, we investigated the binding of CMBL3aL using endogenous pre-miR33a and pre-miR24-2 having a UGGU sequence motif in their hairpin loops, which we identified from the miRbase database using BLAST (Figures 5A, S7, and S8).[5,28,29] SPR assay revealed significant high SPR response (Figures 5B and S8) of CMBL3aL for pre-miR33a and pre-miR24-2 (Table S4). To demonstrate the importance of the identified motif for CMBL3aL binding, we performed the SPR assay with a pre-miR33a mutant by incorporating point mutations (pre-miR33a-G31A and pre-33a-G32A) in it (Figure S7A–S7D). Decreases in binding response were observed for both pre-miR33a-G31A and pre-33a-G32A with increases in $K_{d(app)}$ by 1.62- and 1.85-fold (Table S4), respectively. Moreover, *in vitro* dicer cleavage reaction of pre-miR33a showed dose-dependent decrease in the cleaved products (I + P) formation in the presence of CMBL3aL (Figures 5C and 5D). These data are an indication of the potential of this HT-SELEX-based approach for the identification of binding active hairpin-sequence motifs of pre-miRNA against a molecule of interest.

DISCUSSION

In this study, we uncovered binding sequence-structure motifs of pre-miRNA for our recently designed small-molecule CMBL3aL by utilizing a new HT-SELEX-based approach (Figure 1). Our SELEX study were continued up to 10–12 rounds of selection cycles using the

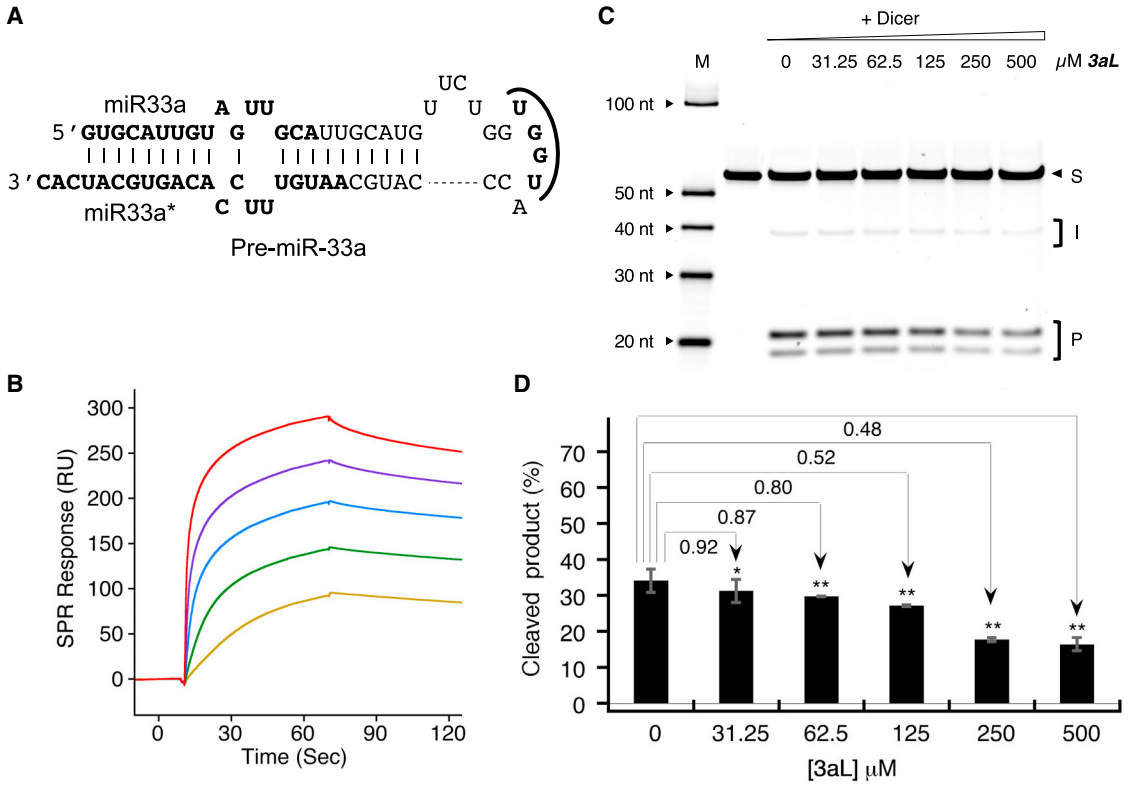

Figure 5. Binding and *in vitro* dicer cleavage reaction of pre-miR33a with CMBL3aL

(A) Secondary structure of human pre-miR33a (which contains the UGGU sequence). (B) SPR analysis of pre-miR33a (0.01 [yellow], 0.03 [green], 0.1 [blue], 0.3 [purple], and 1.0 [red] μM) with the CMBL3aL-immobilized sensor surface (immobilized amount: 3704 RU). (C) Denaturing PAGE analysis of the dicer-cleaved products from pre-miR33a in the absence and presence of CMBL3aL (0, 31.25, 62.5, 125, 250, and 500 μM) in a dose-dependent manner and (D) bar graph representation of densitometric analysis of the cleaved products. *p < 0.01 and **p < 0.05 versus the control sample without CMBL3aL addition (paired t test).

N_{6-11} and N_{11} randomized hairpin loop RNA libraries, followed by sequencing of each selection pool using an Ion PGM next-generation sequencer (Tables S1 and S2). HT sequencing data revealed the presence of a consensus sequence motif (consecutive Gs [GG] flanked by one or two Us) in the randomized region of the enriched aptamers (Figure 2A; Tables S1 and S2). Secondary and tertiary structures of aptamers play a crucial role on their affinity and selectivity. To unveil sequence-structure motifs of these high binding affinity aptamers important for CMBL3aL binding, we applied RaptRanker,[59] an *in silico* method, to our HT sequencing results. RaptRanker generates set of sub-sequences with a structural information and clusters the sub-sequences based on similarity in both nucleotide sequence and secondary structure features, thereby allowing identification of sequence-structure motifs truly significant for binding to the target molecule. This analysis showed most of the nucleotide positions for sub-sequences of R-Seq1, 5, and 6 belong to the stem loop (Figure 3C). For the sub-sequences of R-Seq2 and 3, most of the nucleotides were observed into Stem and Internal loops (Figures 3D and 3E). In contrast, for the sub-sequence of R-Seq4, most of the nucleotide were spotted into Hairpin (Figure 3F). By further inspection we unveiled sequence resemblances within the hairpin loops of most of the enriched aptamers, that embrace with the conserved hairpin-sequence motif comprising consecutive Gs (GG) flanked by one or two Us (UGG/GGU/UGGU) demonstrating its importance for CMBL3aL binding (Figures 4A, 4B, S3, and S4). CMBL analogs encompassed with two 2-amino-1,8-naphthyridine heterocycles, such as CMBL3aL, have been developed using a rational approach to recognize an array of Gs by using their complementary hydrogen bonding surface.[52,53] The characteristic binding hairpin-sequence motif obtained from this study indicate that the GG site within it would be the possible binding site for the CMBL3aL. We further elucidated the binding affinity of CMBL3aL toward four selected enriched aptamers (R-Seq1–4) by SPR assay (Figures 4C–4F). Significant SPR response and high binding affinity were observed for all four aptamers with different binding kinetics. In contrast, much lower binding affinity was observed toward pre-miR29a that comprises an AU-rich hairpin-sequence motif (Figure S5), indicating the importance of the identified hairpin-sequence motif for CMBL3aL binding. We further assessed the effect of CMBL3aL on dicer-mediated cleavage of selected aptamers (R-Seq1, 2, and 4) that differ in their overall secondary structure. No significant change in cleaved product (I + P) formation (Figures S6A and S6B) was observed for R-Seq1, which formed a dumbbell-shaped structure (Figure 4A). This kind of structure has already been reported not to be an appropriate substrate for

the dicer cleavage reaction. In contrast, R-Seq2, which also formed a dumbbell-shaped structure (Figure 4A), showed a 1.86-fold increase in cleaved product (I + P) formation (Figures S6C and S6D), which might be due to the formation of a preferable hairpin loop structure upon CMBL3aL binding. Moreover, a 0.89-fold decrease in cleaved product (I + P) formation (Figures 4G and 4H) was observed for R-Seq4, consisting of a hairpin structure (Figure 4A) favorable for the dicer cleavage reaction.

Several groups have been adopted various strategies[36] for the identification of small molecules that regulate miRNA biogenesis. For example, Davies and Arenz[68] and Bose et al.[69] reported a fluorescence resonance energy transfer (FRET)-based strategy, Tan et al.[70] developed a fluorescence polarization-based screening assay, and Shan et al.[71] and Connelly et al. [71] developed a luciferase-based system that contains complementary sequences of miRNAs in their 3′ UTR of the reporter genes. Furthermore, Disney et al.[73] developed a bioinformatic-based rational approach "Inforna" to identify small molecules that bind to Drosha and the dicer binding site of human miRNA precursors by integrating two-dimensional combinatorial screening with HT structure-activity relationships through sequencing (HiT-StARTS). Our group has also reported a fluorescent indicator displacement assay to screen xanthone derivatives against pre-miR29a and identified 21 hit compounds.[74] Furthermore, by utilizing conventional SELEX[44] as well as by advancement of the standard SELEX method, various approaches have been developed over three decades to identify RNA aptamers, such as cell-SELEX, *in vivo*-SELEX, immunoprecipitation-coupled (IP)-SELEX, capture-SELEX, microfluidic (M)-SELEX, and so on.[45-49] Herein, we described our HT-SELEX-based approach (Figure 1), where we integrated conventional SELEX (Scheme S3) with HT sequencing, followed by bioinformatic analysis. In this study, we were particularly interested in identifying sequence motifs in the pre-miRNA hairpin loop for CMBL3aL binding.

To validate the scope of this method we further evaluated the binding of CMBL3aL toward endogenous pre-miRNAs,[28,29] which comprises a partial active binding hairpin-sequence motif (Figures 5A and S8), by SPR assay. SPR data revealed significant binding responses for both endogenous pre-miR33a and pre-miR24-2 comprising characteristic binding motifs (Figures 5B and S8B). Further dose-dependent decrease in the cleaved products formation was also observed in the presence of CMBL3aL for the *in vitro* dicer cleavage reaction of pre-miR33a (Figures 5C and 5D). In conclusion, these data suggest the potential of this new HT-SELEX-based strategy that would accelerate the identification process of the sequence-structure motif of pre-miRNAs for the compound of interest selected from the chemical library and would contribute to increase the spectrum of biomedical application.

MATERIAL AND METHODS

Synthesis of CMBL3aL

CMBL3aL was synthesized from CMBL3a[53] in two steps (Scheme S1), by reductive amination followed by boc-deprotection. CMBL3a was prepared using a procedure reported previously.[53] Detail synthesis (Scheme S1) and characterization of CMBL3aL are given in the supplementary data.

In vitro selection (SELEX) procedure

(1) Preparation of CMBL3aL-immobilized resin: CMBL3aL was immobilized on NHS-activated Sepharose 4 Fast Flow resin (GE Healthcare) by amino coupling method according to the manufacturer's instructions. In brief, to NHS-activated Sepharose 4 Fast Flow resin (0.500 mL) was added a solution of CMBL3aL (50 nmoL) in 0.2 M $NaHCO_3$-0.5 M NaCl (pH 8–9) (0.5 mL) containing 5% DMSO. After shaking the suspension at room temperature for several hours, the solution was drained and the resin was further treated with 0.1 M Tris-HCl buffer (1 mL [pH 8.8]) for several hours to block unreacted carboxylic groups on the resin. The resulting resin was sequentially washed with 0.1 M Tris-HCl buffer (pH 8.8) and 0.1 M NaOAc-0.5 M NaCl solution (pH 5.2), which was repeated three times. The resin was washed with 20% aqueous EtOH, and three bed volumes of 20% aqueous EtOH was added to the resin to give a 25% slurry resin suspension. Tris-immobilized resin was also prepared for the preselection to remove RNA species that bind non-specifically to the matrix (Scheme S2).

(2) Preparation of DNA and RNA pools: the initial double-stranded DNA pools were prepared by PCR amplification of the oligonucleotide containing the N_{11} or N_{6-11} nt randomized region (5′-ACT GAT TTC TTT TGG TGT TCA GAG—the randomized region—TTC TAG CAC CAT CTG AAA TCG GTT A-3′) with Platinum *Pfx* DNA Polymerase (Invitrogen). The primers used in the PCR amplification were: 5′-TAATACGACTCACTATAGCTGATTTCTTTTGGTGTTCAGAG-3′ (forward primer, T7 promoter sequence is underlined) and 5′-TAACCGATTTCAGATGGTGCTAGAA-3′ (reverse primer). The dsDNA pool was transcribed *in vitro* with T7 RNA polymerase (MEGAshortscript T7 Transcription Kit, Ambion) to yield an RNA pool. The reaction mixture was treated with DNase at 37°C for 30 min to digest the template DNA. The mixture was diluted with RNase-free water, and passed through an NAP-5 column (GE Healthcare) to remove unincorporated NTPs. The RNA was precipitated by adding 7.5 M ammonium acetate and isopropanol, and pelleted by centrifugation. The pellet was dissolved in annealing buffer (10 mM Tris-HCl, 50 mM NaCl, 100 mM KCl [pH 7.5]). The RNA solution was denatured at 80°C for 3 min, and annealed by slow cooling to room temperature. $MgCl_2$ was added to a final concentration of 5 mM to the annealed RNA, and used for the next binding step.

(3) Selection step and regeneration of RNA pool: the RNA solution was preincubated with Tris-resin for 30 min at room temperature to remove non-specific binders to the resin. The flowthrough was applied to CMBL3aL-immobilized resin, and the suspension was incubated at room temperature for 30 min with vortexing every 10 min. The resin was drained and washed three times with washing buffer (10 mM Tris-HCl, 100 mM KCl, 50 mM NaCl, 5 mM $MgCl_2$ [pH 7.5]). Bound RNAs were eluted three times

with CMBL3aL solution (50 μL, 100 μM) and pooled in a 1.5 mL collection tube. The eluted RNAs were precipitated with 0.3 M sodium acetate and ethanol, and pelleted by centrifugation. The RNA pellet was dissolved in RNase free water and reverse transcribed with PrimeScript II Reverse Transcriptase (Takara) using the reverse primer (5′-TAACCGATTTCAGATGGTGCTAG AA-3′). The resulting cDNA was PCR-amplified with the forward and the reverse primers as described above. The DNA templates were transcribed *in vitro*, and the resulting RNAs were subjected to the next round of selection (Scheme S3).

SPR measurement

(1) Immobilization of RNAs to sensor chips: interaction between CMBL3aL and RNA-immobilized surfaces was measured using SPR with a Biacore T200 system (GE Healthcare). For preparation of the RNA-immobilized sensor surface, 5′-biotinylated RNA (R-Seq1, 2, 3, 4, or pre-miR29a) (GeneDesign) was diluted to 0.2 μM in 10 mM HEPES buffer containing 500 mM NaCl and injected over a streptavidin-coated sensor chip (Series S Sensor chip SA, GE Healthcare) at 5 μL/min to obtain an immobilization level of around 500 response unit (RU). The amount of RNA immobilized on sensor surface was 525.3 RU (R-Seq1), 562.1 RU (R-Seq2), 528.7 RU (R-Seq3), 503.4 RU (R-Seq4), and 546.2 RU (pre-miR29a) (used in Figures 4C–4F and S5).

(2) Immobilization of CMBL3aL to sensor chips: (1) for preparation of CMBL3aL-immobilized sensor surface, the carboxy groups of the dextran surface of a sensor chip CM5 (Series S Sensor chip CM5, GE Healthcare) were activated with NHS/EDC (600 s, 10 μL/min), and 200–500 μM CMBL3aL in borate buffer (10 mM sodium tetraborate, 1 M NaCl [pH 8.5]) was flowed onto the surface at 10 μL/min for 600 s to yield an immobilization level of 3704 RU (used in Figures 5 and S7). (2) The CMBL3aL-immobilized sensor surface used in Figure S8 was prepared by introducing a PEG spacer between CMBL3aL and the surface of the sensor chip to reduce steric hindrance at the sensor surface and to obtain a maximal response. In brief, the carboxy groups of the dextran surface of a sensor chip CM5 (Series S Sensor chip CM5, GE Healthcare) were activated with NHS/EDC, and 100 mM amino-dPEG$_4$ acid (Quanta BioDesign) in 1× HBS-N buffer (10 mM HEPES, 150 mM NaCl [pH 7.4]) was injected over the surface. After capping the residual activated ester with 1 M ethanolamine, the carboxy groups of amino-dPEG$_4$ acid were activated with NHS/EDC, and 1 mM CMBL3aL in borate buffer (10 mM sodium tetraborate, 1 M NaCl [pH 8.5]) was flowed onto the surface to obtain an immobilization level of 2764.8 RU for CMBL3aL (used in Figure S8).

(3) Preparation of pre-miRNA and its mutants: the dsDNA template was prepared by 3′ fill-in reaction with Klenow fragment using two oligonucleotides overlapping at their 3′ ends. The sequences of oligonucleotides used are listed at the end of this section. The reaction mixture was purified using the Nucleospin Gel and PCR clean up (Machery-Nagel) according to the manufacture's instruction. The purified dsDNA template was transcribed *in vitro* by T7 RNA polymerase (MEGAshortscript T7 Transcription Kit, Ambion) to produce the corresponding RNAs. After digestion of the template DNA with DNase I, the RNAs were purified on a 15% denaturing polyacrylamide gel (acrylamide:bisacrylamide = 19:1, 6 M urea). Oligonucleotides used to prepare the dsDNA templates for: pre-miR-33a (5′-TAATACGACTCACTATAGTGCATTG-TAGTTGCATTGCATG-3′/5′-GTGATGCACTGTGGAAACAT TGCATGGGTACCACCAGAACATGCAATGCAACTAC-3′), pre-miR-33a_G31A (5′-TAATACGACTCACTATAGTGCATT GTAGTTGCATTGCATG-3′/5′-GTGATGCACTGTGGAAACA TTGCATGGGTACTACCAGAACATGCAATGCAACTAC-3′), pre-miR-33a_G312A (5′-TAATACGACTCACTATAGTGCATT GTAGTTGCATTGCATG-3′/5′-GTGATGCACTGTGGAAACA TTGCATGGGTATCACCAGAACATGCAATGCAACTAC-3′) (T7 promoter sequence is underlined) (used in Figures 5 and S7).

(4) Binding analysis using RNA-immobilized sensor chip: kinetics of binding of CMBL3aL to the target RNA-immobilized sensor surface was determined by single-cycle kinetics. CMBL3aL in HBS-EP+ buffer (0.01 M HEPES, 0.15 M NaCl, 3.0 mM EDTA [pH 7.4], 0.005% [v/v] surfactant P20) containing 5% DMSO was injected in increasing concentrations (100, 200, 300, 400, and 500 nM) to the RNA-immobilized sensor surface without a regeneration step in between each concentration. Injections were done at 60 μL/min with a contact time of 30 s, followed by dissociation with running buffer for 120 s. All sensorgrams were corrected by subtraction of the blank flow cell response and the buffer injection response (used for Figures 4C–4F and S5).

(5) Binding analysis using CMBL3aL immobilized sensor chip: binding of pre-miR33a, pre-miR24-2 and its mutants to CMBL3aL was also investigated using the CMBL3aL-immobilized sensor chip prepared above. RNA (prepared by *in vitro* transcription) was diluted to 1 μM in 1× HBS-EP+ buffer, followed by annealing (80°C for 3 min, then slowly cooled down to room temperature. The annealed RNA was diluted in 1× HBS-EP+ buffer at concentrations of 0.01, 0.03, 0.1, 0.3, and 1.0 μM. Each RNA solution was injected over the CMBL3aL-immobilized sensor surface at 30 μL/min with a contact time of 60 s and a dissociation time of 120 s. For pre-miR24-2 contact time of 180 s and a dissociation time of 180 s were used. In between injections, the sensor surface was regenerated by injection of 50 mM NaOH. All sensorgrams were corrected by reference subtraction of the blank flow cell response (used in Figures 5B, S7, and S8).

In vitro dicer reaction

RNA stock solution (100 μM) was diluted in 1× annealing buffer (50 mM Tris-HCl [pH 7], 100 mM NaCl) and to obtain 20 μM RNA solution. The resulting solution was heated at 90°C for 1 min, and kept at 37°C for 1 h to anneal RNA. The annealed RNA can be stored at −80°C until use. For dicer cleavage reaction the annealed RNA solution (20 μM) was incubated with an in-house-prepared recombinant human dicer in the presence or absence of 250 μM CMBL3aL at 37°C for 3 h. A typical reaction mixture for dicer cleavage contained 2 μM RNA, 51.3 nM of the recombinant human dicer, 1 mM ATP, 10 mM $MgCl_2$, 0.05 mg/mL BSA, 5% (v/v) DMSO, and

1× Dicer Reaction Buffer (Genlantis) in a 5 μL reaction mixture. After requisite time of incubation, 1 μL of dicer stop solution (Genlantis) or 1 μL of 100 mM EDTA was added to the reaction mixture, followed by the addition of 14 μL of formamide. The resulting sample (20 μL) can be stored at −80°C; 10 μL of 20 μL of the sample was heated at 90°C for 3 min and cooled immediately on ice followed by analysis using denaturing polyacrylamide gel electrophoresis (PAGE).

Denaturing PAGE conditions: the denatured sample was analyzed by 15% denaturing PAGE (mono-acrylamide/bis-acrylamide = 19:1, 6 M urea) using 1× TBE as running buffer (running condition: 200 V, 60 min). Small RNA II Easy Load (DynaMarker) was used as an RNA marker. Gel was stained with SYBR Gold Nucleic Acid Gel Stain (Thermo Fischer Scientific). RNA bands were visualized using a LAS 4010 system, and quantified using ImageQuant (GE Healthcare) (Figures 4G, 4H, 5C, 5D, and S6).

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtn.2021.11.021.

ACKNOWLEDGMENTS

This work was supported by India-Japan (DST-JSPS) Cooperative Science Programme (IJCSP) to promote bilateral scientific collaboration, and JSPS KAKENHI Grant-in-Aid for Scientific Research (A) [19H00924] for KN. Funding for open access charge: JSPS KAKENHI Grant-in-Aid for Scientific Research (A) [19H00924].

AUTHOR CONTRIBUTIONS

K.N., A.M., and S.M. conceptualized the idea. S.M., A.M., R.I., and A.S. performed the experiments and analyzed the data. S.M. and A.M. prepared the figures. S.M. wrote the manuscript draft with the help of A.M. and R.I. K.N., A.M., C.D., M.H., S.K., and K.N. critically reviewed the manuscript. Data contribution, S.M.: Figures 1–5 and S1–S8, Schemes 1–3, Tables S1–'S4; A.M. and A.S.: Figures 4G–4F, 5, and S6–S8, Tables S1–S4; R.I. and M.H.: Figures 2B–2D, 3, and S2, Tables S1 and S2. All authors read and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Catalanotto, C., Cogoni, C., and Zardo, G. (2016). MicroRNA in control of gene expression: an overview of nuclear functions. Int. J. Mol. Sci. *17*, 1712.
2. Siomi, H., and Siomi, M.C. (2010). Posttranscriptional regulation of microRNA biogenesis in animals. Mol. Cell *38*, 323–332.
3. Cai, Y., Yu, X., Hu, S., and Yu, J. (2009). A brief review on the mechanisms of miRNA regulation. Genomics Proteomics Bioinformatics *7*, 147–154.
4. Bartel, D.P. (2009). MicroRNA target recognition and regulatory functions. Cell *136*, 215–233.
5. Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Res. *47*, D155–D162.
6. Das, S.S., Saha, P., and Chakravorty, N. (2018). MiRwayDB: a database for experimentally validated microRNA-pathway associations in pathophysiological conditions. Database *2018*, bay023.
7. Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. Nat. Rev. Genet. *16*, 421–433.
8. Orang, A.V., Safaralizadeh, R., and Kazemzadeh-Bavili, M. (2014). Mechanisms of MiRNA-mediated gene regulation from common downregulation to mRNA-specific upregulation. Int. J. Genomics *2014*, 970607.
9. Bartel, D.P. (2004). MicroRNAs: genomics, review biogenesis, mechanism, and function. Cell *116*, 281–297.
10. Paul, P., Chakraborty, A., Sarkar, D., Langthasa, M., Rahman, M., Bari, M., Singha, R.S., Malakar, A.K., and Chakraborty, S. (2018). Interplay between miRNAs and human diseases. J. Cell. Physiol. *233*, 2007–2018.
11. Ardekani, A.M., and Naeini, M.M. (2010). The role of microRNAs in human diseases. Avicenna J. Med. Biotechnol. *2*, 161–179.
12. Li, Y., and Kowdley, K.V. (2012). MicroRNAs in common human diseases. Gernomics Proteomics Bioinformatics *10*, 246–253.
13. Natarajan, S.K., Smith, M.A., Wehrkamp, C.J., Mohr, A.M., and Mott, J.L. (2013). MicroRNA function in human diseases. Med. Epigenet. *1*, 106–115.
14. Giza, D.E., Vasilescu, C., and Calin, G.A. (2014). Key principles of miRNA involvement in human diseases. Discoveries *2*, e34.
15. Dwivedi, S., Purohit, P., and Sharma, P. (2019). MicroRNAs and diseases: promising biomarkers for diagnosis and therapeutics. Ind. J. Clin. Biochem. *34*, 243–245.
16. Barutta, F., Bellini, S., Mastrocola, R., Bruno, G., and Gruden, G. (2018). MicroRNA and microvascular complications of diabetes. Int. J. Endocrinol. *2018*, 6890501.
17. Zhou, S.-S., Jin, J.-P., Wang, J.-Q., Zhang, Z.-G., Freedman, J.H., Zheng, Y., and Cai, L. (2018). MiRNAs in cardiovascular diseases: potential biomarkers, therapeutic targets and challenges. Acta Pharmacol. Sin. *39*, 1073–1084.
18. Barwari, T., Joshi, A., and Mayr, M. (2016). MicroRNAs in cardiovascular disease. J. Am. Coll. Cardiol. *68*, 2577–2584.
19. Abe, M., and Bonini, N.M. (2013). MicroRNAs and neurodegeneration: role and impact. Trends Cell Biol. *23*, 30–36.
20. Kamal, M.A., Mushtaq, G., and Greig, N.H. (2015). Current update on synopsis of miRNA dysregulation in neurological disorders. CNS Neurol. Disord. Drug Targets *14*, 492–501.
21. Ichii, O., and Horino, T. (2018). MicroRNAs associated with the development of kidney diseases in humans and animals. J. Toxicol. Pathol. *31*, 23–34.
22. Drury, R.E., O'Connor, D., and Pollard, A.J. (2017). The clinical application of microRNAs in infectious disease. Front. Immunol. *8*, 1182.
23. Peng, Y., and Croce, C.M. (2016). The role of microRNAs in human cancer. Signal Transduct. Target. Ther. *1*, 15004.
24. Jansson, M.D., and Lund, A.H. (2012). MicroRNA and cancer. Mol. Oncol. *6*, 590–610.
25. Reddy, K.B. (2015). MicroRNA (miRNA) in cancer. Cancer Cell Int. *15*, 38.
26. Davis-Dusenbery, B.N., and Hata, A. (2010). MicroRNA in cancer: the involvement of aberrant microRNA biogenesis regulatory pathways. Genes Cancer *1*, 1100–1114.
27. Adams, B.D., Kasinski, A.L., and Slack, F.J. (2014). Aberrant regulation and function of microRNAs in cancer. Curr. Biol. *24*, R762–R776.
28. Gao, C., Wei, J., Tang, T., and Huang, Z. (2020). Role of microRNA-33a in malignant cells (Review). Oncol. Lett. *20*, 2537–2556.
29. Wang, S., Liu, N., Tang, Q., Sheng, H., Long, S., and Wu, W. (2020). MicroRNA-24 in cancer: a double side medal with opposite properties. Front. Oncol. *10*, 553714.
30. Garzon, R., Marcucci, G., and Croce, C.M. (2010). Targeting microRNAs in cancer: rationale, strategies and challenges. Nat. Rev. Drug Discov. *9*, 775–789.
31. Shah, M.Y., Ferrajoli, A., Sood, A.K., Lopez-Berestein, G., and Calin, G.A. (2016). MicroRNA therapeutics in cancer-an emerging concept. EBioMedicine *12*, 34–42.
32. Kwok, G.T., Zhao, J.T., Weiss, J., Mugridge, N., Brahmbhatt, H., MacDiarmid, J.A., Robinson, B.G., and Sidhu, S.B. (2017). Translational applications of microRNAs in cancer, and therapeutic implications. Noncoding RNA Res. *2*, 143–150.

33. Wen, D., Danquah, M., Chaudhary, A.K., and Mahato, R.I. (2015). Small molecules targeting microRNA for cancer therapy: promises and obstacles. J. Control Release *219*, 237–247.

34. Monroig-Bosque, P.D.C., Chen, L., Zhanga, S., and Calin, G.A. (2016). Small molecule compounds targeting miRNAs for cancer therapy. Adv. Drug Deliv. Rev. *81*, 104–116.

35. Giorgio, A.D., Tran, T.P.A., and Duca, M. (2016). Small-molecule approaches toward the targeting of oncogenic miRNAs: roadmap for the discovery of RNA modulators. Future Med. Chem. *8*, 803–816.

36. Fana, R., Xiaob, C., Wanc, X., Chaa, W., Miaod, Y., Zhoua, Y., Qina, C., Cuie, T., Suf, F., and Shang, X. (2019). Small molecules with big roles in microRNA chemical biology and microRNA-targeted therapeutics. RNA Biol. *16*, 707–718.

37. Velagapudi, S.P., Gallo, S.M., and Disney, M.D. (2014). Sequence-based design of bioactive small molecules that target precursor microRNAs. Nat. Chem. Biol. *10*, 291–297.

38. Disney, M.D., Labuda, L.P., Paul, D.J., Poplawski, S.G., Pushechnikov, A., Tran, T., Velagapudi, S.P., Wu, M., and Childs-Disney, J.L. (2008). Two-dimensional combinatorial screening identifies specific aminoglycoside-RNA internal loop partners. J. Am. Chem. Soc. *130*, 11185–11194.

39. Disney, M.D. (2013). Rational design of chemical genetic probes of RNA function and lead therapeutics targeting repeating transcripts. Drug Discov. Today *18*, 1228–1236.

40. Disney, M.D., and Angelbello, A.J. (2016). Rational design of small molecules targeting oncogenic noncoding RNAs from sequence. Acc. Chem. Res. *49*, 2698–2704.

41. Costales, M.G., Haga, C.L., Velagapudi, S.P., Childs-Disney, J.L., Phinney, D.G., and Disney, M.D. (2017). Small molecule inhibition of microRNA-210 reprograms an oncogenic hypoxic circuit. J. Am. Chem. Soc. *139*, 3446–3455.

42. Murata, A., Otabe, T., Zhang, J., and Nakatani, K. (2016). BzDANP, a small-molecule modulator of pre-mir-29a maturation by dicer. ACS Chem. Biol. *11*, 2790–2796.

43. Hall, B., Micheletti, J.M., Satya, P., Ogle, K., Pollard, J., and Ellington, A.D. (2009). Design, synthesis, and amplification of DNA pools for *in vitro* selection. Curr. Protoc. Nucleic Acid Chem. *39*, 9.2.1–9.2.28.

44. Codrea, V., Hayner, M., Hall, B., Jhaveri, S., and Ellington, A. (2010). *In vitro* selection of RNA aptamers to a small molecule target. Curr. Protoc. Nucleic Acid Chem. *40*, 9.5.1–9.5.23.

45. Hoinka, J., Berezhnoy, A., Dao, P., Sauna, Z.E., Gilboa, E., and Przytycka, T.M. (2015). Large scale analysis of the mutational landscape in HT-SELEX improves aptamer discovery. Nucleic Acids Res. *43*, 5699–5707.

46. Caroli, J., Taccioli, C., Fuente, A.D.L., Serafini, P., and Bicciato, S. (2016). APTANI: a computational tool to select aptamers through sequence-structure motif analysis of HT-SELEX data. Bioinformatics *32*, 161–164.

47. Aquino-Jarquin, G., and Toscano-Garibay, J.D. (2011). RNA aptamer evolution: two decades of SELEction. Int. J. Mol. Sci. *12*, 9155–9171.

48. Terasaka, N., Futai, K., Katoh, T., and Suga, H. (2016). A human microRNA precursor binding to folic acid discovered by small RNA transcriptomic SELEX. RNA *22*, 1918–1928.

49. Zhang, Y., Lai, B.S., and Juhas, M. (2019). Recent advances in aptamer discovery and applications. Molecules *24*, 941.

50. Connelly, C.M., Boer, R.E., Moon, M.H., Gareiss, P., and Schneekloth, J.S., Jr. (2017). Discovery of inhibitors of microRNA-21 processing using small molecule microarrays. ACS Chem. Biol. *12*, 435–443.

51. Nakatani, K., Natsuhara, N., Mori, Y., Mukherjee, S., Das, B., and Murata, A. (2017). Synthesis of naphthyridine dimers with conformational restriction and binding to DNA and RNA. Chem. Asian J. *12*, 3077–3087.

52. Mukherjee, S., Dohno, C., Asano, K., and Nakatani, K. (2016). Cyclic mismatch binding ligand CMBL4 binds to the 5′-T-3′/5′-GG-3′ site by inducing the flipping out of thymine base. Nucleic Acids Res. *44*, 7090–7099.

53. Mukherjee, S., Dohno, C., and Nakatani, K. (2017). Design and synthesis of cyclic mismatch-binding ligands (CMBLs) with variable linkers by ring-closing metathesis and their photophysical and DNA repeat binding properties. Chem. Eur. J. *23*, 11385–11396.

54. Mukherjee, S., Błaszczyk, L., Rypniewski, W., Falschlunger, C., Micura, R., Murata, A., Dohno, C., Nakatani, K., and Kiliszek, A. (2019). Structural insights into synthetic ligands targeting A-A pairs in disease-related CAG RNA repeats. Nucleic Acids Res. *47*, 10906–10913.

55. Konieczny, P., Mukherjee, S., Stepniak-Konieczna, E., et al. (2021). Cyclic mismatch binding ligands interact with disease-associated CGG trinucleotide repeats in RNA and suppress their translation. Nucleic Acids Res. *49*, 9479–9495.

56. Ketan, T., Savjani, K.T., Gajjar, A.K., and Savjani, J.K. (2012). Drug solubility: importance and enhancement techniques. Int. Sch. Res. Notices *2012*, 195727.

57. Tolle, F., Wilke, J., Wengel, J., and Mayer, G. (2014). By-product formation in repetitive PCR amplification of DNA libraries during selex. PoLS One *9*, e114693.

58. Marshall, K.A., and Ellington, A.D. (1999). Molecular parasites that evolve longer genomes. J. Mol. Evol. *49*, 656–663.

59. Ishida, R., Adachi, T., Yokota, A., Yoshihara, H., Aoki, K., Nakamura, Y., and Hamada, M. (2020). RaptRanker: in silico RNA aptamer selection from HT-SELEX experiment based on local sequence and structure information. Nucleic Acids Res. *48*, e82.

60. Starega-Roslan, J., Koscianska, E., Kozlowski, P., and Krzyzosiak, W.J. (2011). The role of the precursor structure in the biogenesis of microRNA. Cell. Mol. Life Sci. *68*, 2859–2871.

61. Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H., and Tomari, Y. (2011). Recognition of the pre-miRNA structure by *Drosophila* dicer-1. Nat. Struct. Mol. Biol. *18*, 1153–1159.

62. Starega-Roslan, J., Galka-Marciniak, P., and Krzyzosiak, W.J. (2015). Nucleotide sequence of miRNA precursor contributes to cleavage site selection by dicer. Nucleic Acids Res. *43*, 10939–10951.

63. Gu, S., Jin, L., Zhang, Y., Huang, Y., Zhang, F., Valdmanis, P.N., and Kay, M.A. (2012). The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. Cell *151*, 900–911.

64. Auyeung, V.C., Ulitsky, I., Mcgeary, S.E., and Bartel, D.P. (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. Cell *152*, 844–858.

65. Gu, S., Jin, L., Zhang, F., Huang, Y., Grimm, D., Rossi, J.J., and Kay, M.A. (2012). Thermodynamic stability of small hairpin RNAs highly influences the loading process of different mammalian Argonautes. Proc. Natl. Acad. Sci. U S A *108*, 9208–9213.

66. Abe, N., Abe, H., and Ito, Y. (2007). Dumbbell-shaped nanocircular RNAs for RNA interference. J. Am. Chem. Soc. *129*, 15108–15109.

67. Luo, Q.-J., Jinsong Zhang, J., Li, P., Wang, Q., Zhang, Y., Roy-Chaudhuri, B., Xu, J., Kay, M.A., and Zhang, Q.C. (2021). RNA structure probing reveals the structural basis of dicer binding and cleavage. Nat. Commun. *12*, 3397.

68. Davies, B.P., and Arenz, C.A. (2006). Homogenous assay for microRNA maturation. Angew. Chem. Int. Ed. *45*, 5550–5552.

69. Bose, D., Jayaraj, G.G., Kumar, S., and Maiti, S. (2013). A molecular-beacon-based screen for small molecule inhibitors of miRNA maturation. ACS Chem. Biol. *8*, 930–938.

70. Tan, G.S., Chiu, C.-H., Garchow, B.G., Metzler, D., Diamond, S.L., and Kiriakidou, M. (2012). Small molecule inhibition of RISC loading. ACS Chem. Biol. *7*, 403–410.

71. Shan, G., Li, Y., Zhang, J., Li, W., Szulwach, K.E., Duan, R., Faghihi, M.A., Khalil, A.M., Lu, L., Paqroo, Z., et al. (2008). A small molecule enhances RNA interference and promotes microRNA processing. Nat. Biotechnol. *26*, 933–940.

72. Connelly, C.M., Thomas, M., and Deiters, A. (2012). High-throughput luciferase reporter assay for small-molecule inhibitors of microRNA function. J. Biomol. Screen. *17*, 822–828.

73. Disney, M.D., Winkelsas, A.M., Velagapudi, S.P., Southern, M., Fallahi, M., and Childs-Disney, J.L. (2016). Inforna 2.0: a platform for the sequence-based design of small molecules targeting structured RNAs. ACS Chem. Biol. *11*, 1720–1728.

74. Fukuzumi, T., Murata, A., Aikawa, H., Harada, Y., and Nakatani, K. (2015). Exploratory study on the RNA-binding structural motifs by library screening targeting pre-miRNA-29a. Chem. Eur. J. *21*, 16859–16867.