

Design of Effective Primary MicroRNA Mimics With Different Basal Stem Conformations

Fiona T van den Berg¹, John J Rossi², Patrick Arbutnot¹ and Marc S Weinberg^{1,3,4}

Primary microRNA (pri-miRNA) mimics are important mediators of effective gene silencing and are well suited for sustained therapeutic applications. Pri-miRNA mimics are processed in the endogenous miRNA biogenesis pathway, where elements of the secondary RNA structure are crucial for efficient miRNA production. Cleavage of the pri-miRNA to a precursor miRNA (pre-miRNA) by Drosha-DGCR8 typically occurs adjacent to a basal stem of ~11 bp. However, a number of pri-miRNA structures are expected to contain slightly shorter or longer basal stems, which may be further disrupted in predicted folding of the expressed pri-miRNA sequence. We investigated the function and processing of natural and exogenous RNA guides from pri-miRNAs with various basal stems (9–13 bp), where a canonical hairpin was predicted to be well or poorly maintained in predicted structures of the expressed sequence. We have shown that RNA guides can be effectively derived from pri-miRNAs with various basal stem conformations, while predicted guide region stability can explain the function of pri-miRNA mimics, in agreement with previously proposed design principles. This study provides insight for the design of effective mimics based on naturally occurring pri-miRNAs and has identified several novel scaffolds suitable for use in gene silencing applications.

Molecular Therapy—Nucleic Acids (2016) 5, e278; doi:10.1038/mtna.2015.53; published online 12 January 2016

Subject Category: siRNAs, miRNAs and shRNAs

Introduction

Primary microRNA (pri-miRNA) mimics have been shown to mediate effective gene silencing,^{1–3} while providing several advantages over conventional short hairpin RNAs (shRNAs) and avoiding pathway saturation⁴ and cellular toxicity.⁵ Pri-miRNA mimics derived from endogenous miRNA precursors and expressed from a polymerase II promoter^{6–8} can provide controlled, tissue-specific expression of exogenous guide RNAs, suitable for sustained posttranscriptional gene silencing. About 4,500 human miRNA precursors are available on the online miRNA repository,^{9,10} but mimic design has thus far relied on only several precursors.^{11–18} In particular, the model miR-30 backbone^{1,2} has been widely used in therapeutic applications against a number of diseases and pathogens, including the hepatitis B virus¹⁴ and human immunodeficiency virus,^{8,16,19} and is included in ongoing clinical trials against various cancers.^{20,21} The potential of thousands of other pri-miRNAs as scaffolds in mimic design has not yet been explored.

Pri-miRNA mimics, like endogenous pri-miRNAs, are processed in the miRNA biogenesis pathway.²² Pri-miRNAs fold into a characteristic hairpin structure, with a terminal loop, imperfect duplex stem region of ~33 bp and unstructured flanking sequences.²³ Pri-miRNA processing is typically modular in nature, where cleavage by the Drosha-DGCR8 microprocessor^{24–26} is reliant on both elements of the secondary RNA structure and pri-miRNA sequence, but to different extents for different pri-miRNAs.^{27–29} In productive processing,

Drosha recognizes the basal ssRNA-dsRNA junction and basal UG motif, cleaving as a molecular “ruler” at ~11 bp from the basal junction; and a DGCR8 dimer recognizes the stem, apical ssRNA-dsRNA junction, and loop UGU motif to facilitate processing fidelity.^{28,30} Noncanonical features within the stem also contribute to cleavage site selection,^{31,32} while structural distortions in proximity to cleavage sites may facilitate efficient cleavage.^{33,34} In addition, methylation motifs in the proximity of miRNA genes promote processing.³⁵ The precursor miRNA (pre-miRNA) produced by Drosha-DGCR8 cleavage is subsequently cleaved by the RNase III enzyme Dicer,^{36–39} to produce a staggered 5p/3p miRNA duplex with 3′ dinucleotide overhangs. One strand of the duplex is incorporated into the RNA-induced silencing complex⁴⁰ where it acts as a mature ~22 nt miRNA and directs posttranscriptional gene silencing.

Optimal properties of the pri-miRNA secondary RNA structure have been described that enable efficient recognition and processing. These include a large terminal loop (≥ 10 nt),⁴¹ moderate basal stem (≥ 8 bp) with an optimal length of ~11 bp,²⁷ and at least 40 nt of flanking sequence on either side of the pre-miRNA, that lack a strong secondary structure.^{23,42,43} These structural properties should be implemented in the design of effective pri-miRNA mimics. However, a number of pri-miRNA structures are expected to contain slightly shorter or longer basal stems, which are not necessarily maintained when mimics are generated from an expression construct. In particular, interactions between the hairpin and incorporated flanking sequences may potentially

¹Wits/SAMRC Antiviral Gene Therapy Research Unit, Department of Molecular Medicine and Haematology, School of Pathology, University of the Witwatersrand, Johannesburg, South Africa; ²Department of Molecular and Cellular Biology, Beckman Research Institute of City of Hope, Duarte, California, USA; ³HIV Pathogenesis Research Unit, Department of Molecular Medicine and Haematology, School of Pathology, University of the Witwatersrand, Johannesburg, South Africa; ⁴Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California, USA. Correspondence: Marc S Weinberg, Department of Molecular Medicine and Haematology, School of Pathology, The University of the Witwatersrand Medical School, 7 York Road, Parktown, Johannesburg 2193, South Africa. E-mail: marcow@scripps.edu or marc.weinberg@wits.ac.za

Keywords: artificial microRNAs; primary microRNA mimics; RNAi therapeutics

Received 17 October 2015; accepted 30 November 2015; advance online publication 12 January 2016. doi:10.1038/mtna.2015.53

support or disrupt the formation of a conventional basal stem and hairpin.

Here, we investigated the function and processing of miRNAs from expressed pri-miRNAs with various basal stem conformations (9–13 bp), where the canonical hairpin was either well or poorly maintained in the predicted secondary RNA structure of the expressed pri-miRNA sequence. The suitability of such pri-miRNAs as scaffolds was assessed through the design and characterization of corresponding pri-miRNA mimics. In their design, we have shown that minor variations in the basal stem length of pri-miRNA mimics were tolerated, while established stability features of the guide region were important and should be considered in mimic design. Several novel pri-miRNA scaffolds were characterized and are ideal for applications requiring exogenous guide sequences.

Results

Predicted secondary RNA structure of pri-miRNAs with various basal stem conformations

MicroRNA precursors with various basal stem conformations were selected from the online miRNA repository,^{9,10} which lists experimentally identified miRNAs and associated stem-loop structures, and five example precursors for human miRNAs were investigated further: miR-520b, miR-23a, miR-934, miR-34a, and miR-513a-2. To assess pri-miRNA folding conformations, predicted secondary RNA structures were generated for each naturally occurring pri-miRNA with 15–50 nt of sequence flanking the anticipated Drosha-DGCR8 cleavage site, using the mfold web server⁴⁴ with default parameters (**Supplementary Figures S1** and **S2**). Predicted secondary RNA structures were also generated for RNA sequences derived from the expression vector, to assess folding of the expressed pri-miRNA. Canonical hairpins were present in the most stable pri-miRNA structures generated with 15–20 nt of flanking sequence (**Figure 1a**). These hairpins were defined as structures with a terminal loop, an upper stem harboring the miRNA, a basal stem (8–13 bp), and at least 4 nt of unstructured residues flanking the expected Drosha-DGCR8 site on one side.²⁷ Stem regions were counted to include Watson-Crick pairs and G–U wobble pairs, while the potential contribution of single symmetrical mismatches was also considered. The basal stem region of pri-miR-520b, for example, was slightly shorter (7 bp), but two non-Watson-Crick pairs can also stack within the stem and contribute to its length (~9 bp). Interestingly, canonical hairpin structures were not necessarily maintained in predicted pri-miRNA conformations generated with 30–50 nt of sequence flanking the cleavage site or with full flanks of the expressed pri-miRNA (**Figure 1a,b**). Canonical hairpins were commonly generated for pri-miR-23a and pri-miR-934 structures with extended flanking sequences but were not generated at all for corresponding pri-miR-520b structures, where predicted interactions with the flanking sequences produced a truncated hairpin. Predicted structures for pri-miR-34a typically had an extended ~16–19 bp basal stem or a canonical ~9 bp basal stem lacking adjacent unstructured residues. In addition, predicted structures encompassing a canonical hairpin were not always the most stable (*), as noticed for pri-miR-34a and pri-miR-513a-2.

Functionality of endogenous pri-miRNAs with various basal stem conformations

Predicted secondary RNA structures are not a perfect indication of dynamic RNA conformations in the cell. However, in predicted structures of the expressed pri-miRNAs that included the extended flanking sequences, the ratio of canonical to noncanonical structures with similar levels of stability was lower for specific pri-miRNAs (**Figure 1b**). Therefore, in addition to variation of the basal stem length, the basal stem may be further disrupted in the expressed pri-miRNA sequence. Expressed pri-miRNAs with additional, noncanonical hairpin conformations could conceivably show reduced miRNA-mediated silencing, making these structures less suitable as mimic scaffolds. We therefore proceeded to functionally characterize selected pri-miRNAs.

The processing and silencing capabilities of the selected pri-miRNAs were investigated in cell culture through northern blot analyses and luciferase reporter assays. A shRNA designed to express each major guide sequence was included for each corresponding pri-miRNA as a positive control for guide production and function. Mature ~22 nt guides were detected for exogenously expressed pri-miRNAs with a range of basal stem lengths (9–13 bp) and efficient silencing (63–91 %) of corresponding targets was observed (**Figure 2a**). Pre-miRNAs (55–65 nt) were also detected, thus supporting cleavage at the anticipated processing sites. The most effective target silencing was observed for pri-miR-934, while target silencing of over 80% was observed for both pri-miR-34a and pri-miR-513a-2. Complementary strands of the miRNA duplex were detected for each pri-miRNA, but effective target silencing was only observed for miR-34a-3p (68%) and miR-513a-2-5p (92%) (**Figure 2b**), implying that these minor miRNAs can also be effectively loaded into RNA-induced silencing complex. Larger (75–120 nt) RNA products were also detected across all samples in several blots and were presumed to be endogenous background RNAs that bind nonspecifically to the probe. Reporter gene suppression was specific to individual guide–target interactions, as no silencing was observed in the absence of a target sequence (**Supplementary Figure S3**). Similar assay results were also obtained in the Huh-7 cell line (**Supplementary Figure S4**). These data suggest that minor variation in the basal stem length and additional, noncanonical folding of the basal stem were not restrictive in conventional pri-miRNA processing. Functional miRNAs were effectively derived from pri-miRNAs where the canonical hairpin was present in almost all (pri-miR-934), most (pri-miR-23a), or none (pri-miR-520b) of the predicted structures including the extended flanking sequence of the expressed pri-miRNA. However, as canonical hairpins were present in predicted structures generated with 15–20 nt of flanking sequence for all pri-miRNAs, these data also suggest that structures including the local flanking sequence were more useful in assessing the processive form.

Nonstructural determinants are also expected to contribute to the observed pri-miRNA function. Pri-miRNA sequences were analyzed for the presence of recently described sequence motifs, including the basal UG and CNNC motifs, the loop GUG motif, the mismatched GHG stem motif, and

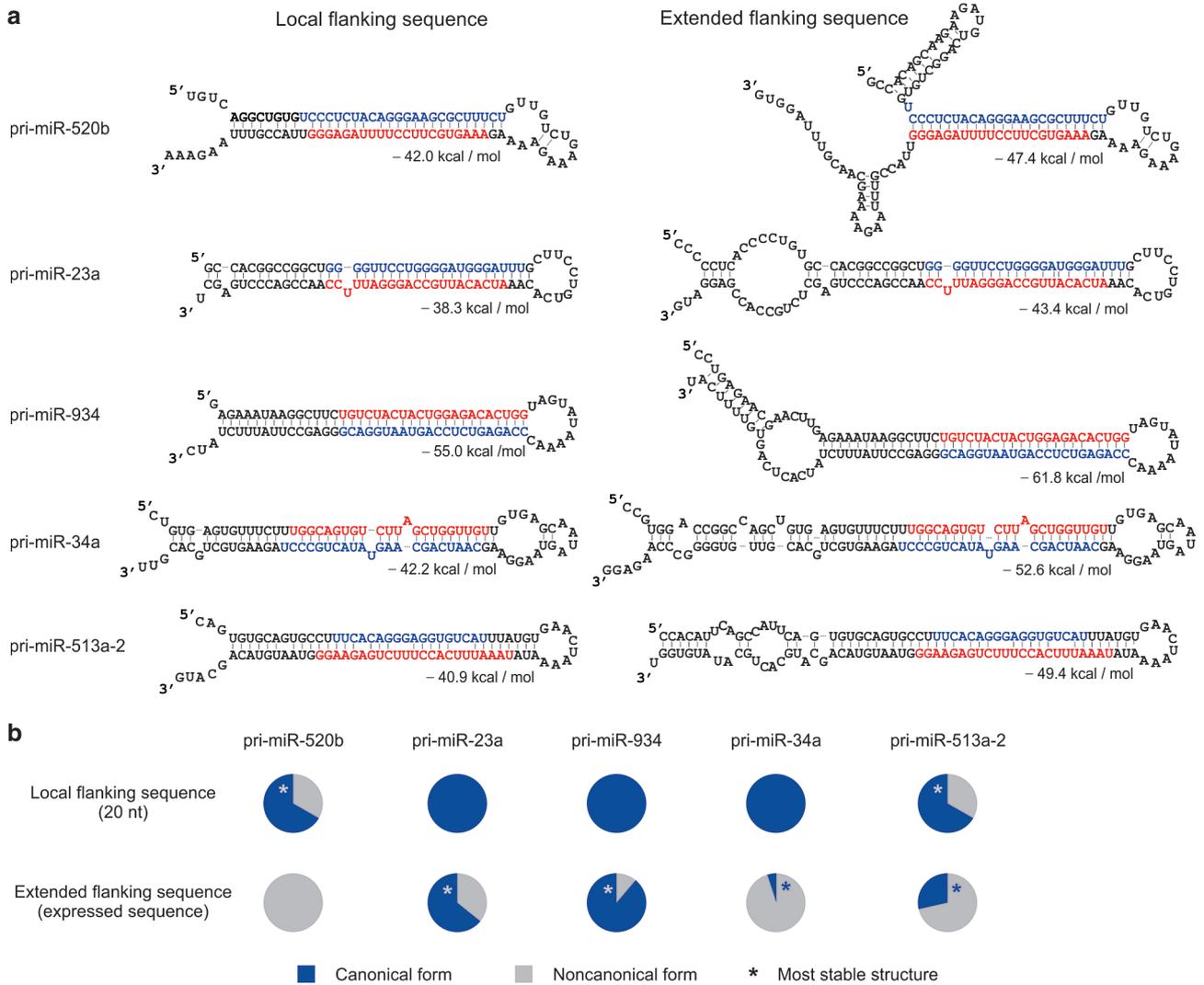


Figure 1 Predicted secondary RNA structures of selected pri-miRNAs. (a) (Left) Structures were generated with the local flanking sequence (15 nt) and are representative of the conventional hairpins predicted for each pri-miRNA. (Right) Structures were also generated with extended regions of the flanking sequence (30 nt) and show the additional alternative structures predicted for some pri-miRNAs. The anticipated (miRBase) major (red) or minor (blue) 5p or 3p miRNA guide sequences, and total free energy of the predicted structure (mfold ver. 3.5) are indicated. (b) The proportion of predicted secondary RNA structures with a canonical pri-miRNA hairpin generated using local (20 nt) or extended flanking sequences of the expressed pri-miRNA.

nearby METTL3 motifs^{27,29,35} (Figure 3a). Several potential motifs were identified for each pri-miRNA sequence and are expected to contribute to the observed function in a modular manner specific to each pri-miRNA. The contribution of individual pri-miRNA sequence motifs was not assessed, as these motifs were initially well characterized.^{27,29} However, it is interesting to note that pri-miR-520b potentially has all four pri-miRNA sequence motifs (Figure 3b) which are expected to compensate for the suboptimal basal stem length and permit conventional Drosha-DGCR8 processing.²⁹ Similarly, pri-miR-23a has all four potential pri-miRNA sequence motifs, which may permit processing in the context of various basal stem conformations (Supplementary Figure S1); while pri-miR-934 only has two potential pri-miRNA sequence motifs, but a well-maintained conventional basal

stem. Pri-miRNA sequences were also analyzed for the presence of the 5'-GGAC-3' METTL3 motif³⁵ in the context of both the expressed and genomic sequences (Figure 3c). Motifs occurring within 100 nt of the pre-miRNA in the genomic context were typically preserved in the cloned pri-miRNA sequence, with the exception of pri-miR-34a and pri-miR-513a-2 where a single motif was omitted. However, this was unlikely to have a major impact on function, as one or two original motifs were still retained in pri-miR-34a and pri-miR-513a-2 sequences, which were processed to give effective target silencing (>80%). In addition, an extra motif was present in the expressed pri-miRNA sequence of all five pri-miRNAs. In principle, these data suggest that effective miRNAs can be embedded and processed from within exogenously expressed pri-miRNAs with various basal stem lengths and

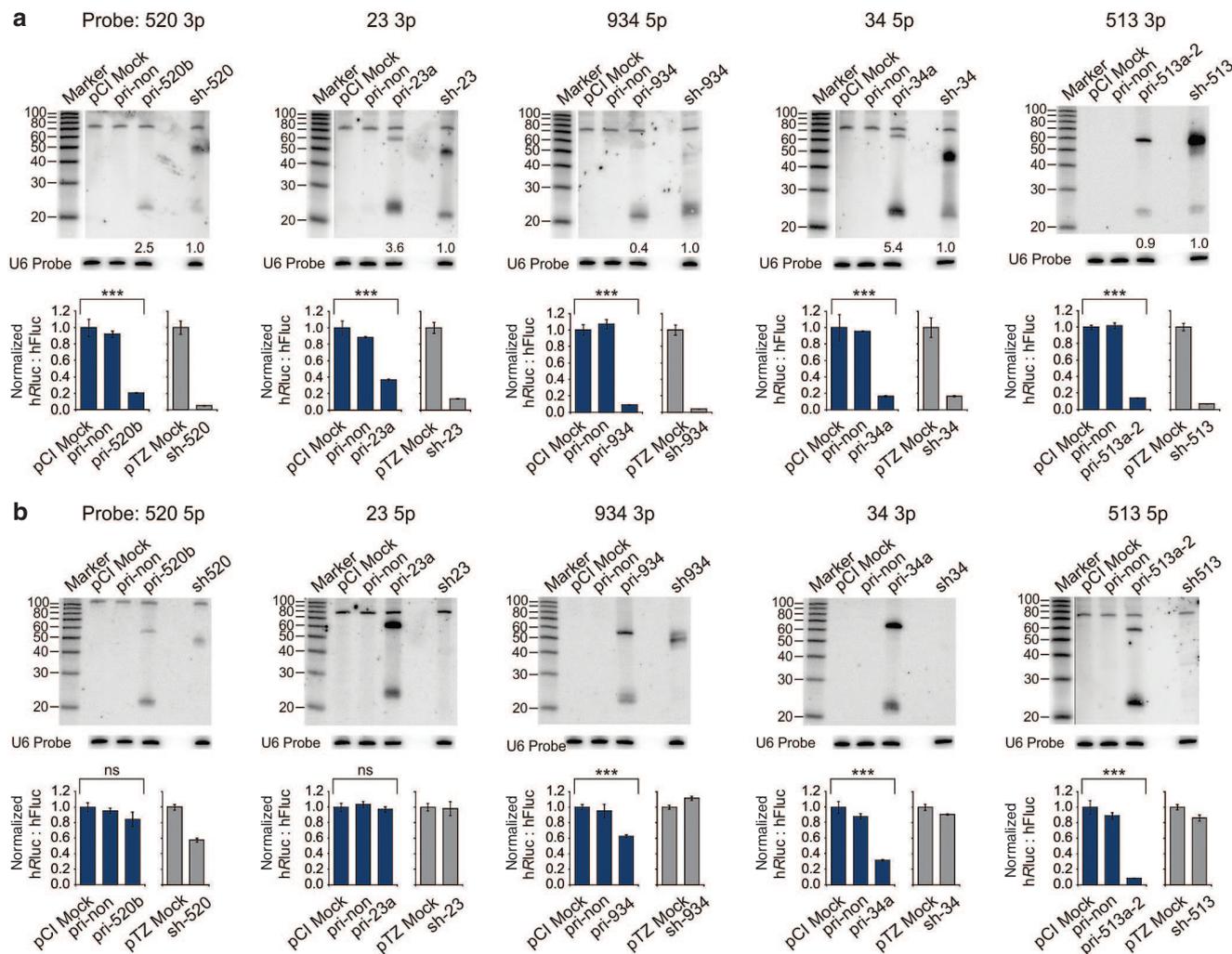


Figure 2 Processing and silencing efficacy of expressed pri-miRNAs. (a) (Upper panel) Major guide sequences were detected through northern blot analyses, following transfection of HEK293 cells with the corresponding pri-miRNA or shRNA (positive control) expression vectors. A mock expression vector (pCI/pTZ Mock) and nonspecific pri-miRNA expression vector (pri-non) were included as negative controls. Guide signal intensities were quantified relative to U6 signal intensity and are shown as a ratio of pri-miRNA-derived guide to shRNA-derived guide. (Lower panel) Target gene silencing by derived guides was measured by the mean ratio of *Renilla* luciferase: Firefly luciferase (hRluc: hFluc) ($n = 3$, \pm SD). Values were normalized with respect to the relevant mock sample. Sample means differed significantly (one-way analysis of variance (ANOVA), $***P < 0.0001$). (b) The processing and function of minor miRNA sequences was similarly assessed through northern blot analyses (upper panel) and target gene silencing (lower panel). Included shRNAs were only designed to express the major miRNA guide. ns: no significant difference between sample means (one-way ANOVA, $P > 0.05$); 934-3p: $P = 0.0003$.

motif combinations, while potentially disruptive interactions predicted to occur between the basal stem and flanking sequences appear to be of little consequence.

Design and functionality of pri-miRNA mimics

Since each characterized natural pri-miRNA appeared suitable for use as a scaffold into which functional exogenous miRNA/siRNA sequences can be embedded, we proceeded to determine the efficacy of corresponding pri-miRNA mimics. Pri-miRNA mimics were designed by replacing the endogenous miRNA guide sequences with one of four previously characterized therapeutic siRNA guides^{45–47} targeting the expression of *tat*, *int*, *gag*, or long terminal repeat (LTR) regions of the human immunodeficiency virus genome (Figure 4a). The original miRNA sequence was replaced

with 20–22 nt of a therapeutic guide, depending on the length of the original miRNA. Resulting minor variations in the 3'-terminal 1–3 nt of the miRNA were not expected to drastically influence the function of mimics within each guide sequence set, where there was a minimum of 20 nt of common 5' guide sequence, including the seed region,⁴⁸ and a minimum of 19 complementary guide–target interactions.

The predicted secondary structure (mfold)⁴⁴ of pri-miRNA mimics was kept as close as possible to that of the endogenous pri-miRNAs, as structural features, like internal bulges, may be crucial to maintain productive processing^{23,31,32,49} (Supplementary Figure S5). Alterations to the complementary strand of the guide region were used to maintain these structural features. To assess potential changes in the pri-miRNA conformation, the predicted secondary RNA structure

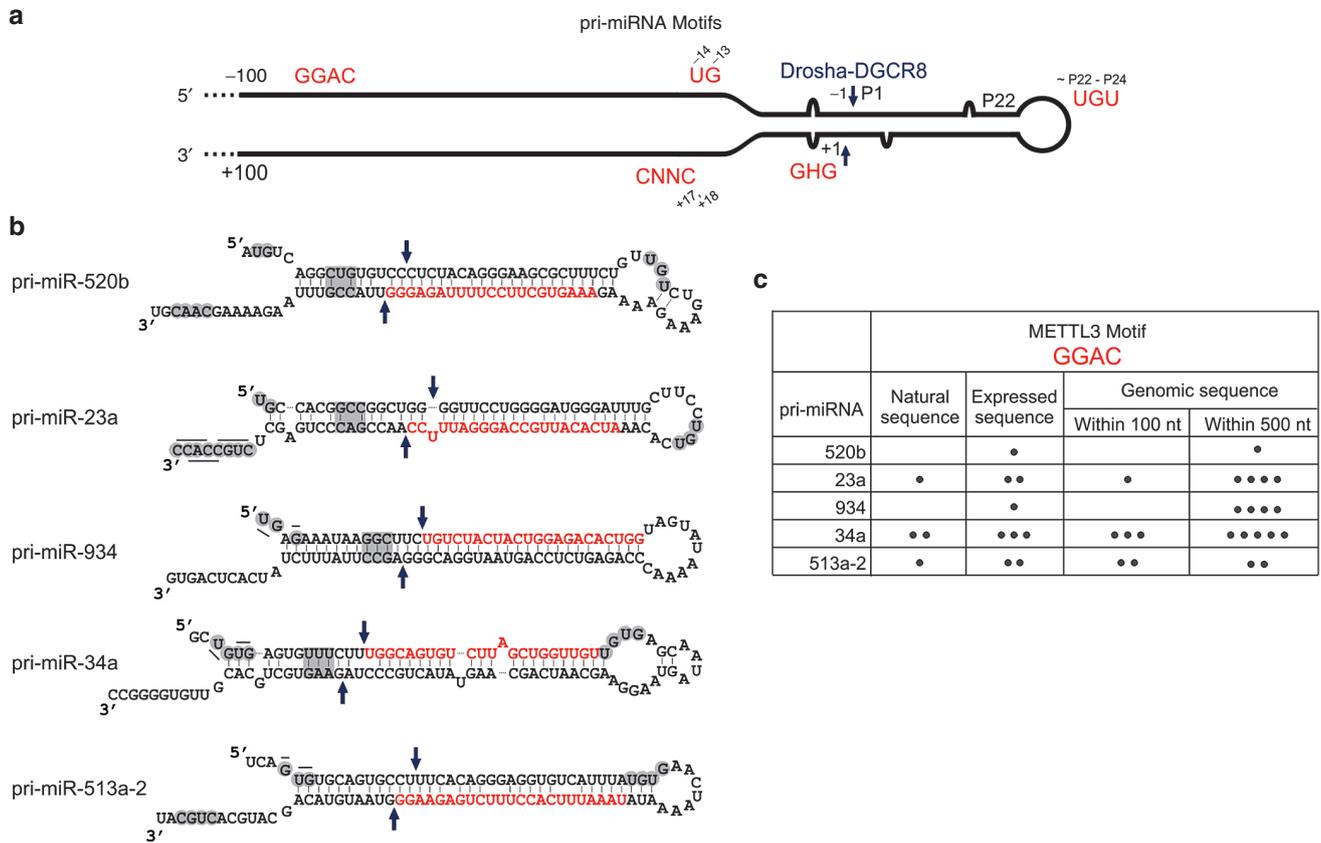


Figure 3 Potential pri-miRNA sequence motifs. (a) Schematic illustration of a typical pri-miRNA showing the location of potential sequence motifs. Primary sequence motifs²⁷ may occur in the 5p arm preceding (–) the Drosha-DGCR8 cleavage site (arrows), the 3p arm after (+) the Drosha-DGCR8 cleavage site or the terminal loop within the pre-miRNA (P) sequence. A mismatched GHG-type motif²⁹ may also be present at position 7–9 of the 3p basal stem. Methyltransferase-like 3 (METTL3) recognition motifs³⁵ may occur at nonspecific locations of the pri-miRNA sequence outside of the pre-miRNA. (b) Potential basal, stem, or loop motifs. (c) Potential METTL3 recognition motifs in the natural or expressed pri-miRNA sequence or within 100 or 500 nt of sequence flanking either side of the expected pre-miRNA in the genomic context.

of each mimic was again examined with both the local and extended sequence flanking the anticipated Drosha-DGCR8 site (**Supplementary Figure S6**). Although structural prediction results differed slightly for some pri-miRNA mimics, such as pri-miR-520-tat (local sequence) and pri-miR-23, -934 and 34a derivatives (extended sequence) (**Supplementary Figure S7**), the ratio of canonical to noncanonical forms was typically similar to that of the original pri-miRNA.

The processing and silencing efficacy of guides derived from pri-miRNA mimics was again assessed by northern blot analyses and luciferase reporter assays (**Figure 4b–e**). ShRNAs expressing each guide sequence, sh-tat,⁵⁰ sh-int,⁵¹ sh-gag,⁵² and sh-LTR,⁵² were included as positive controls. In contrast to the original pri-miRNAs, functional miRNAs were not effectively derived from all pri-miRNA mimics. Exogenous ~22 nt guides with effective target silencing were detected more frequently for certain pri-miRNA scaffolds. Target silencing of ~80% or more was observed for two mimics based on both the pri-miR-34a and 513a-2 scaffolds, but for only one mimic based on each of the pri-miR-520b, 23a, and 934 scaffolds. As a group, mimics based on the pri-miR-34a scaffold had the highest average target silencing, while mimics based on the pri-miR-520b scaffold had the lowest average target silencing. Effective therapeutic guides were

also detected more frequently for certain mimic sets. Target silencing of ~80% or more was observed for four of the LTR mimics, but none of the tat mimics. As a group, the LTR mimic set was generally the most successful with the greatest average decrease in target silencing, followed by the gag, int, and tat mimic sets, respectively. This ranking was reflected in target silencing by the set of pri-miR-520b mimics (pri-miR-520-tat: 0%, int: 12%, gag: 45%, ltr: 79%). Luciferase assay results were specific to individual guide–target interactions and no significant difference was observed between sample means (**Supplementary Figure S8a**). Guide sequences were detected for pri-miR-34a-tat and pri-miR-513a-tat despite poor target silencing. However, recalculating target silencing values relative to each respective endogenous pri-miRNA only produced slightly improved values for the tat mimic set (**Supplementary Figure 8b,c**). Minor variations in mimic design did not significantly affect target silencing results. While the exact number of complementary guide–target interactions was expected to vary, no correlation was observed between the total (19–22 bp) or consecutive (18–22 bp) number of guide–target interactions and target silencing (**Figure 4f**). Noncomplementary 5' guide–target interactions of gag and LTR mimics (1–2 nt) may be expected to slightly reduce target silencing but occurred uniformly within each set

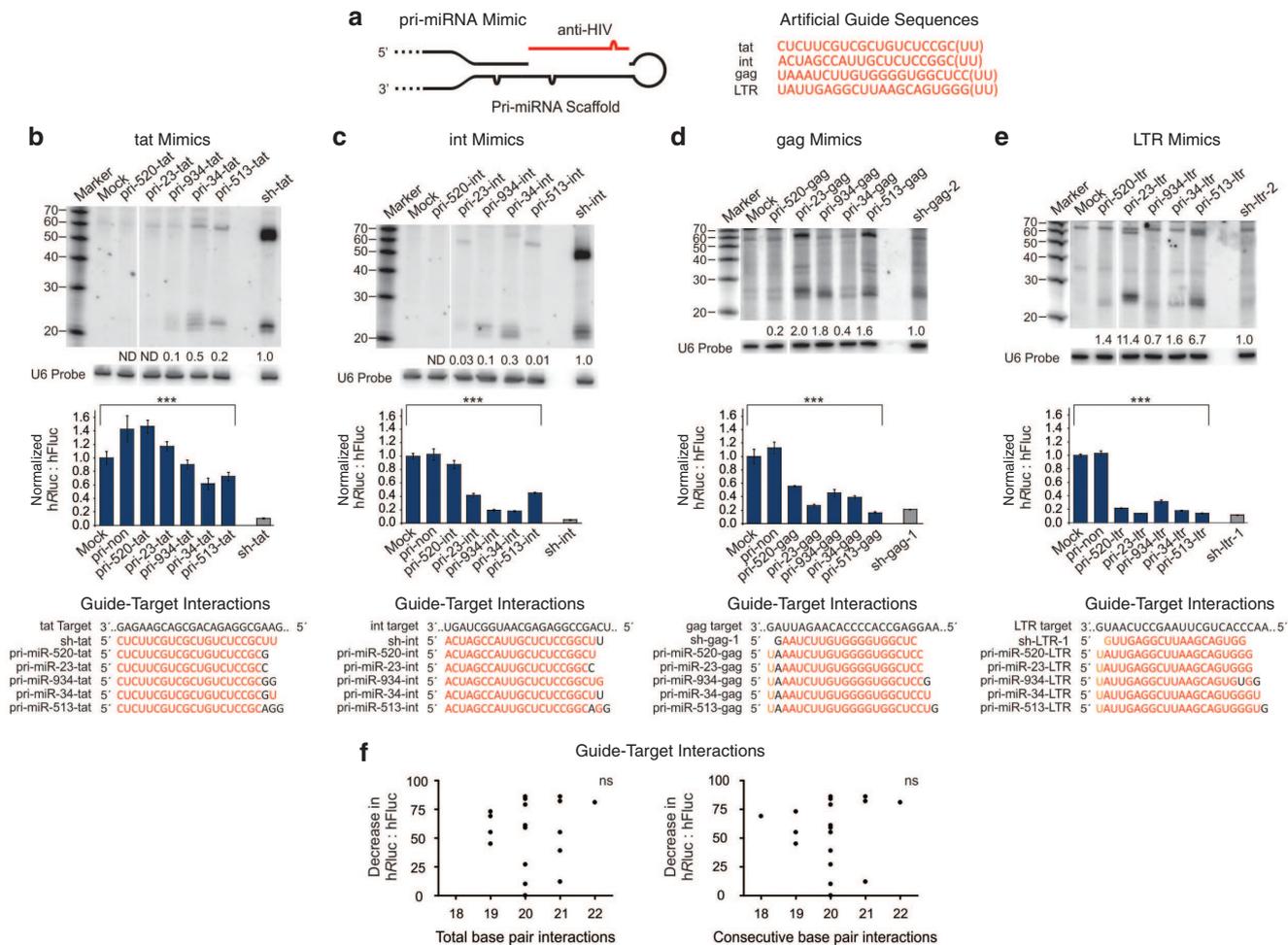


Figure 4 The design and functional characterization of expressed pri-miRNA mimics. (a) Each original guide sequence was replaced with a therapeutic guide against tat, int, gag, or LTR (long terminal repeat) regions of the HIV-1 (human immunodeficiency virus-1) genome. (b–e) The functional characterization of pri-miRNA mimic sets. (Upper panel) Processed antiviral guide sequences were detected in northern blot analyses, following transfection with the corresponding pri-miRNA mimic, shRNA (positive control) or mock (negative control) expression vectors. Guide signal intensities were again quantified relative to U6 signal intensity and are shown as a ratio of pri-miRNA-derived guide to shRNA-derived guide. ND: not detected. (Middle panel) Dual luciferase reporter assays show silencing of respective target sequences in cell culture, measured by mean hFluc: hFluc ($n = 3$, \pm SD). A nonspecific pri-miRNA (pri-non) expression vector was included as a negative control. Sample means differed significantly for each set of mimics (one-way analysis of variance, $***P < 0.0001$). (Lower panel) Expected guide–target interactions showing complementary base pairing (red). (f) The total or consecutive number of expected base pair interactions was plotted against the corresponding decrease in hFluc: hFluc ($n = 20$). No significant (ns) correlation was observed. Spearman (ranked) correlation coefficient ($r_s^{\text{TOTAL}} = 0.19$), ($r_s^{\text{CONS}} = 0.27$) (two-tailed P value > 0.2).

of mimics and were also expected for the respective shRNA controls. Pre-miRNAs (55–65 nt) were also detected for several mimics, again supporting cleavage at the anticipated Drosha-DGCR8 sites. Together, these results indicated that exogenous miRNAs can indeed be effectively derived from expressed pri-miRNAs with a variety of basal stem lengths and motifs, where the canonical hairpin was present in almost all, most, or none of the predicted structures including the extended flanking sequence, but the level of target silencing differed for mimics with the same scaffold or core guide sequence.

Stability of the pri-miRNA guide region

Differences in pri-miRNA mimic functionality could not be attributed to changes in the overall form of predicted

secondary RNA structure or known primary sequence motifs, which were unaltered in mimic design. However, we wondered whether these differences could be explained by the more subtle thermodynamic changes of the pri-miRNA stem incurred in guide replacement. We therefore investigated the stability of predicted pri-miRNAs over the guide region. A characteristic bias in pri-miRNA stability and common thermodynamic features of functional pri-miRNAs have been described in previous studies.^{23,33,49,53} The region corresponding to the 5' end of the mature miRNA is less stable than the region corresponding to its 3' end,⁴⁹ as revealed by thermodynamic profiling of pri-miRNAs using mfold values. In particular, the most unstable position of the stem corresponds to the first position of the expected miRNA.^{23,53} This approach relies on the predicted secondary RNA structures of pri-miRNAs,

which can differ from experimentally determined structures, especially at positions within the terminal loop and flanking sequences.⁴⁹ However, the percentage of correctly predicted base pairs tends to be high (88.4%),⁴⁹ supporting the predicted conformation of the upper pri-miRNA stem. We applied thermodynamic profiling to investigate stability of the pri-miRNA stem along the guide region for various pri-miRNA and mimic groups. In this approach, precise values cannot be assigned to positions within internal loops and we therefore estimated values at these positions using an average loop value, but similar results were also obtained using previously described methods^{23,49} (**Supplementary Figure S9**). As profiles differ for pri-miRNAs with 5p and 3p miRNAs in accordance with the miRNA orientation, we assessed each guide region from the 5' end of the expected miRNA. The guide region was also limited to the portion of the pri-miRNA stem corresponding to the first 20 nt of the exogenous guide, as this region was replaced for all mimics, with the exception of pri-mir-934-LTR where 19 nt of the first 20 nt were replaced. The difference in free energy between the position 1 and 20 (ddG (1–20)) was used to assess the bias in stability expected over the guide region.

Average thermodynamic profiles were constructed for functional groups of pri-miRNAs with effective (>80%), functional (>50%), or ineffective (<50%) target silencing (**Figure 5a**). Common features were identified, including a decrease in the relative stability at positions 4 and 9, as well as an increase in the relative stability at positions 11, 16, and 20. However, the stability at position 1 of the stem, which corresponds to the 5' end of the miRNA, decreased with increasing target silencing

of each pri-miRNA group. In turn, the expected bias in pri-miRNA stability, as indicated by the ddG (1–20) value, was highest for the group of effective pri-miRNAs (1.7 kcal/mol) and lowest for the group of ineffective pri-miRNAs (0.3 kcal/mol). A similar profile for the group of original pri-miRNAs had a ddG (1–20) value slightly lower than expected (0.7 kcal/mol, lower panel), as position 20 of the pri-miR-23a guide region was relatively unstable; but stable pairing at position 21 may facilitate a bias in guide region stability.

Average thermodynamic profiles were also constructed for groups of pri-miRNAs with the same exogenous guide (**Figure 5b**). Again, the ddG (1–20) signature increased with the average target silencing of each group and was pronounced for the LTR mimic set (1.8 kcal/mol). The stability of the guide region for each group was primarily defined by a common guide sequence, but small differences in base pairing were used to maintain the internal mismatches of individual scaffolds. To examine the extent of base pairing variation within each mimic set, we constructed base pairing profiles (**Figure 5c**). Common base pairing was observed with minor variation, thus supporting a common 5' to 3' bias in stable GC pairing for each mimic set, in agreement with the thermodynamic analysis. This bias was more pronounced for the gag and LTR mimic sets, where no GC base pairs were observed in the first four positions as a consequence of siRNA design principles used by McIntyre *et al.*^{47,54} Similarly, no GC base pairs were expected in the first position for the original, functional pri-miRNAs, as is common for natural miRNAs and supported by an Ago-2 preference for uracil in the first position of a mature miRNA.^{33,55}

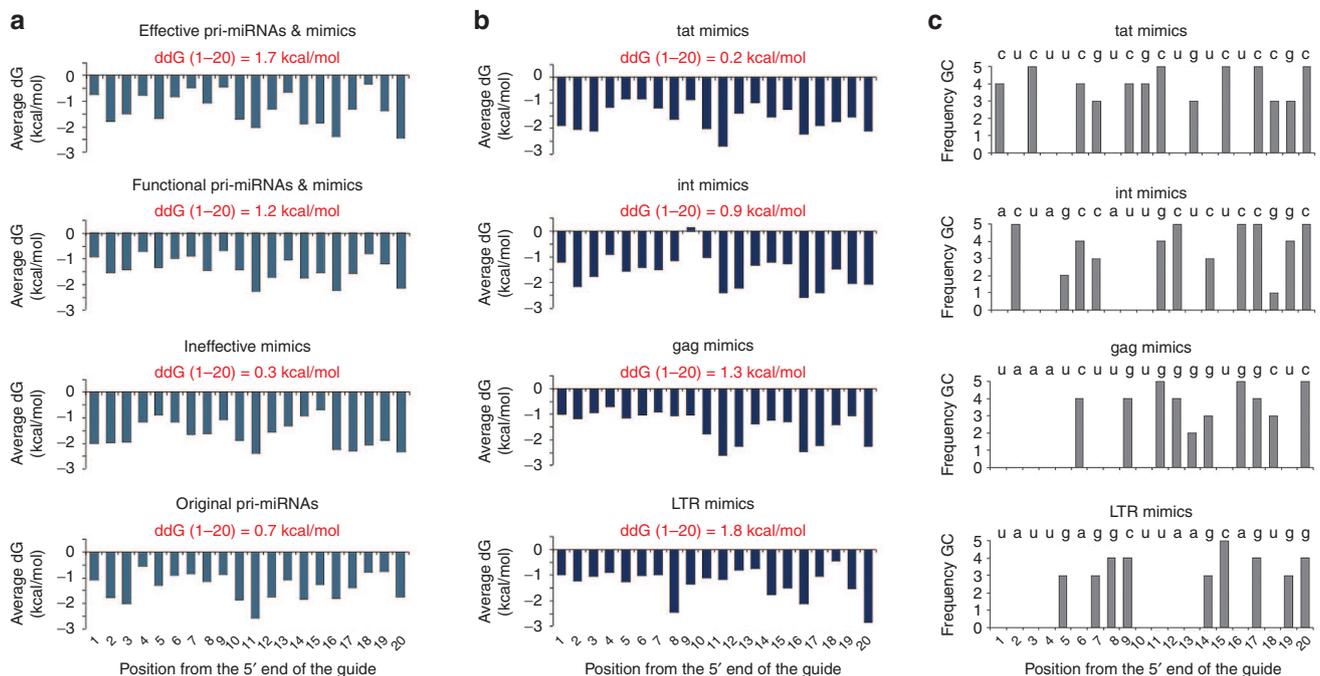


Figure 5 Predicted stability of the pri-miRNA or mimic guide region. (a) Average thermodynamic profiles of the guide region were constructed for pri-miRNA/mimic subsets with effective (>80%), functional (>50%), or ineffective (<50%) target silencing and for the group of original pri-miRNAs. The difference in average stability between positions 1 and 20 of the guide region (ddG (1–20)) is indicated (red). (b) Average thermodynamic profiles were also constructed for different pri-miRNA mimic subsets. (c) Base pairing profiles for the guide region of different pri-miRNA mimic subsets, indicating the frequency of GC base pairing.

Subtle thermodynamic changes of the pri-miRNA stem incurred in guide replacement may therefore explain the overall function of mimic sets. We could detect no difference in the form of the predicted secondary RNA structure, including flanking sequences, or in various motifs of pri-miRNA mimics; but the average target silencing of each mimic set increased with the magnitude of the bias in stability, as indicated by ddG (1–20). This confirms the importance of siRNA design and previously described thermodynamic features of the guide region in the generation of effective mimics. In fact, pri-miR-23-gag and pri-miR-23-LTR mimics had an increase in both ddG (1–20) and target silencing compared to pri-miR-23a. These results are in agreement with design principles suggested by Han *et al.*,²³ stating that the first position of an artificial guide should be unstable, while position ~20 should be relatively stable.

Duplex stability

To determine whether the expected miRNA duplexes also displayed thermodynamic asymmetry, we used conceptual dicing to assess the relative terminus stability.^{49,56} In most cases, the 5' end of the miRNA is expected to be less stable, affecting strand selection and RNA-induced silencing complex loading.^{53,56,57} The average terminus stability for pri-miRNA groups with effective (>80%), functional (>50%), or ineffective (<50%) target silencing was calculated using a 2 nt window, which has previously been described as the most informative window for miRNA identification⁴⁹ (Figure 6a). As expected, the average free energy was higher for the terminus corresponding to the 5' end of the major miRNA (red) than that of the minor miRNA (blue). The average free energy of both termini was lower for the group of ineffective pri-miRNAs; but, in contrast to the results of thermodynamic profiling, a similar difference in terminus stability was observed for all functional groups (~0.5 kcal/mol). Similarly, the difference in average terminus stability for duplexes derived from each mimic set did not increase with the average target silencing (Figure 6b). The average free energy of the miRNA 5' terminus was lower than expected for the group of original (wt) pri-miRNAs using a 2 nt window, as a result of a mismatched residue near the 5' terminus of the minor miRNA for pri-miR-934 and a noninternal mismatch at the 5' terminus of the major miRNA for pri-miR-34a. Similar results were obtained using a 1 nt or 3 nt window (Supplementary Figure S10).

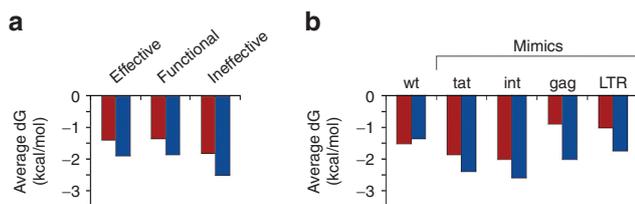


Figure 6 Terminus stability of expected miRNA duplexes. (a) The average terminus stability of duplexes derived from conceptual dicing of pri-miRNA/mimic subsets with effective (>80%), functional (>50%), or ineffective (<50%) target silencing. Average free energy of the major (red) and minor (blue) miRNA termini are shown. (b) The average terminus stability of duplexes derived from conceptual dicing of different pri-miRNA mimic subsets and the group of original, wild-type (wt) pri-miRNAs. LTR, long terminal repeat.

In contrast to thermodynamic profiling, the results of conceptual dicing did not clearly identify a link between the magnitude of thermodynamic asymmetry and the overall function of mimic sets. This may be because only 20 nt of the guide region were commonly replaced. However, stability of the both termini generally increased with a decrease in the average target silencing of different mimic sets. This suggests that while duplex asymmetry was observed for less effective pri-miRNA groups, terminus stability was generally unfavorable. This again supports the importance of relative instability at the 5' of the miRNA in mimic design.

Precise miRNA sequences

The results of characterization studies and guide region analyses rely on the successful derivation of the expected guide sequences and consistent processing of derived mimics. The expected miRNAs were detected in northern analysis, but we also sought to examine the derived miRNA species in more detail through deep sequencing. The top three reads aligning to pri-miR-34a, pri-miR-513a-2, and their respective int mimics are shown in Figure 7 (top 10: Supplementary Figure S11a). We expect that reads aligning to pri-miR-34a and pri-miR-513a-2 originate from both endogenous miRNA genes and the respective expression vectors. However, as the expressed pri-miRNAs were derived from genomic sequence, we expect the miRNA species to be processed in the same manner. Endogenous miRNA genes were expected to generate a smaller fraction of miRNAs, as relevant species were not detected in northern analysis (pCI Mock sample). Overall, the miRNA processing sites were well maintained between expressed pri-miRNAs and int mimics. In particular, Drosha-DGCR8 processing sites were more precisely maintained, as indicated by the consistent 5' ends of the 5p guides. A 1 nt 5' variation was observed at Dicer cleavage sites, although Dicer processing of a 22 nt miR-513-int-3p was more consistent. Evidence of 3' trimming and editing (uridylation and adenylation) of miRNAs was observed as is common for small RNA sequencing data.⁵⁸ The annotated miRNA guides of the original pri-miRNAs (Figure 1) were present in the top three reads, with the exception of miR-513a-2-5p, where several ~22 nt miRNAs were identified. This finding is in agreement with the collective deep sequencing data now available on miRBase (Supplementary Figure S11b,c). These data also suggest that more conventional miR-513a-2-5p species are in fact the major, functional miRNA products of pri-miR-513a-2, consistent with the anticipated processing sites of the 3p miRNA and characterization studies (Figure 2). An effective mimic may therefore also be designed by replacement of the 5p miRNA in the pri-miR-513a-2 scaffold.

Discussion

In this study, we have characterized several novel and effective naturally occurring pri-miRNA scaffolds, ideal for use in applications requiring exogenous guide sequences for posttranscriptional gene silencing. Our results indicated that suboptimal basal stem lengths and potentially disruptive interactions of the basal stem region in the expressed pri-miRNA did not appear to restrict pri-miRNA function or

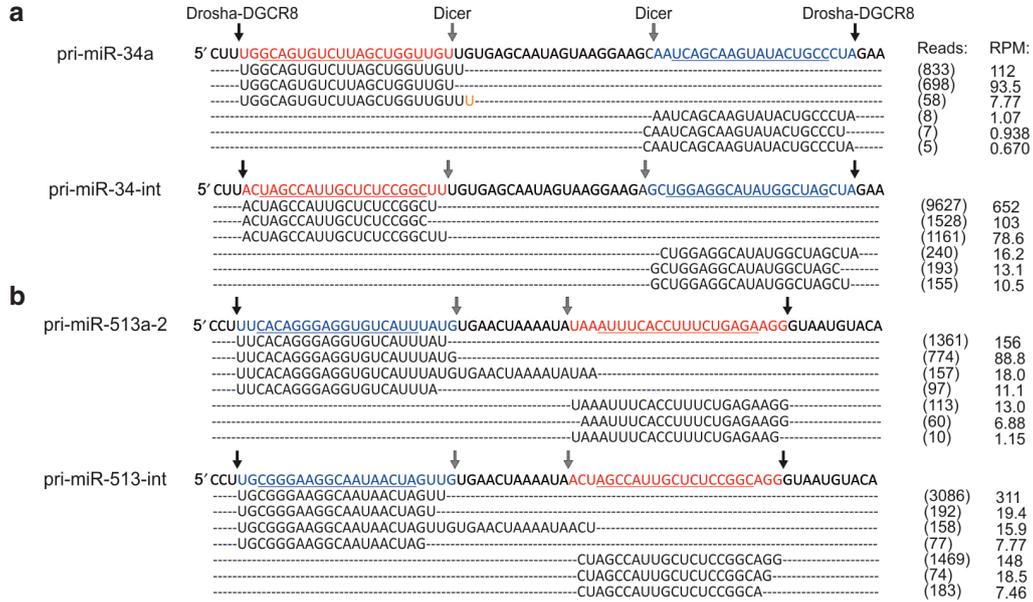


Figure 7 Exact miRNA species and implied processing sites of pri-miRNA mimics. Deep sequencing was used to assess guide production from expressed pri-miR-34a, pri-miR-513a-2 and corresponding int mimics. The top three reads aligned to each 17 nt identifier (underlined) for major miRNA (red) and minor miRNA (blue) sequences are shown and were quantified in reads per million (RPM, three significant figures). Nontemplated nucleotides are indicated (orange).

conventional processing. Therefore, these factors do not appear of great concern or consequence in the design of effective pri-miRNA mimics. This is presumably because a suboptimal basal stem length may be compensated for by an optimal total stem length. The total stem length of pri-miR-23a, -934, -34a, and -513a-2 is 35 ± 1 bp which has recently been described by Fang and Bartel²⁹ as optimal for pri-miRNA processing by Drosha-DGCR8. Sequence motifs may also compensate for suboptimal or varied basal stem conformations. In the case of a pri-miR-520b with a suboptimal basal (9 bp) and total (32 bp) predicted stem length, the presence of all four potential pri-miRNA sequence motifs is expected to facilitate conventional Drosha-DGCR8 recognition and processing.^{27,29} Our data also suggested that hairpin structures including the local flanking sequence were more indicative of the conventional, processive form. Naturally occurring pri-miRNAs with a range of basal stem lengths and motif combinations were therefore suitable for mimic design, provided that the original sequence and secondary RNA structural features supported pri-miRNA processing, in agreement with the current modular understanding of pri-miRNA processing.^{27,29}

Our results also indicated that while naturally occurring pri-miRNAs were effective at silencing relevant targets, derived mimics were not necessarily as effective, or even functional. Thermodynamic profiling showed that the most effective mimic groups were those where the guide sequence permitted a decrease in relative stability at the first position of the miRNA sequence and an increased bias in pri-miRNA stability over the first 20 nt of the guide region. A zero or positive change in ddG (1–20) was also observed for 11 out of 13 individual functional mimics when compared to the original pri-miRNA. These results are in agreement with a number of previous studies describing both common features of

functional pri-miRNAs,^{23,49,53} including a U at the first position of a mature miRNA,⁵⁵ and principles of effective siRNA design.^{54,59,60} Guide sequences selected with respect for siRNA design principles (gag, LTR) were generally more effective in the context of a pri-miRNA scaffold than those selected solely on the basis of target accessibility and conservation (tat, int), despite the fact that all four guides were effective when derived from a pol III shRNA format. Suboptimal properties of a processed RNA duplex may also be more evident in a less robust pol II expression system. In a previous study by Aagaard *et al.*,⁶¹ an effective tat mimic based on the miR-106b precursor was created, where the core guide sequence was positioned within the scaffold to generate a ~21 nt guide with a U at position 1 and an additional 3' GC. The function of this mimic contrasts with the generally poor performance of the tat mimics created here, again highlighting the importance of these sequence features in mimic design.

This work provides a practical example of effective mimic design using naturally occurring pri-miRNA scaffolds with a range of predicted basal stem conformations and sequence motif combinations. Effective pri-miRNA mimics satisfied basic requirements of the secondary RNA structure in folding of the local sequence, had features commonly observed for functional pri-miRNAs, and guide sequences selected with regard for established siRNA design principles. In addition, ddG (1–20) proved to be a useful parameter in pri-miRNA mimic design. Several highly effective pri-miRNA mimics have recently been generated by utilizing artificial guide region features within a natural scaffold¹⁸ and a *de novo* design approach to optimize mimic function,²⁹ but our work illustrates that the use of naturally occurring scaffolds with original guide region features remains a facile and effective approach for the design of pri-miRNA mimics.

Materials and methods

pri-miRNA and mimic expression plasmids. Stem-loop precursors available on the miRBase repository^{9,10} were inspected manually and five were selected as examples of precursors with exceptional basal stems. The Ensembl genome browser⁶² was used to extend the miRBase genomic coordinates by ± 50 nt to incorporate the expected pri-miRNA flanking regions. Sequences for pri-miR- 520b, -23a, -934, and -34a were amplified by PCR from genomic DNA (HEK116 extract), using primers listed in **Supplementary Table S1**. Sequences for pri-miR-513a-2, and all pri-miRNA mimics were constructed from a two-step PCR with oligonucleotides (**Supplementary Table S2**). After initial TA cloning (InsTAclone PCR Cloning Kit, Thermo Fisher Scientific, Waltham, MA), sequences were inserted into the pCI-neo Mammalian Expression Vector (Promega, Madison, WI) and expressed from the pol II cytomegalovirus immediate-early enhancer/promoter. All oligonucleotides and primers were from Integrated DNA Technologies (IDT, Coralville, IA). The sequences of all RNA expression and target constructs were verified by standard dideoxy chain termination DNA sequencing (Inqaba Biotec, Pretoria, South Africa).

Short hairpin RNA expression plasmids. shRNAs were constructed for each pri-miRNA as a positive control of guide function, using a two-step PCR method⁶³ with primers listed in **Supplementary Table S3**. ShRNA expression cassettes included a U6 pol III promoter sequence and a transcription termination signal of five deoxythymidines. These expression cassettes were inserted into a TA cloning vector (InsTAclone PCR Cloning Kit, Thermo Fisher Scientific). Antiviral shRNAs were previously constructed and characterized by our lab: sh-tat,⁵⁰ sh-int,⁵¹ sh-gag, and sh-LTR.⁵²

Target reporter plasmids. Target sequences were fully complementary to the anticipated ~22 nt miRNA, as well as 2 nt before and after the anticipated guide to accommodate minor processing variation, and cloned into the 3' UTR of the *Renilla* luciferase gene of the psiCHECK-2 Vector (Promega). Cleavage or translational suppression mediated by an effective miRNA results in decreased *Renilla* luciferase expression, while expression of Firefly luciferase is unaffected. Target duplexes were constructed by annealing complementary oligonucleotides (**Supplementary Table S4**). Antiviral reporter plasmids were constructed by others: psiCHECK-tat,⁵⁰ psiCHECK-int,⁵¹ psiCHECK-gag,⁵² and psiCHECK-LTR.⁵²

Cell culture and transfections. HEK293, Huh-7, and HeLa cells were cultured in Dulbecco's modified Eagle medium (Biowhittaker, Lonza, Walkersville, MD), supplemented with 10% heat-inactivated fetal calf serum, penicillin (100 U/ml), and streptomycin (0.1 mg/ml), at 37 °C, 5% CO₂. Antibiotic-free media was used for transfections. Cells for northern blot analysis were seeded in a 6-cm dish ($\sim 1.1 \times 10^6$ cells) 24 hours prior to cotransfection with 10 μ g total pDNA (9 μ g pri-/shRNA expression plasmid; 1 μ g pCI-eGFP) using Lipofectamine 2000 Reagent (Invitrogen, Carlsbad, CA) according to manufacturer's protocol. Cells for the dual luciferase assay were seeded in a 24-well format ($\sim 1.2 \times 10^5$ cells/well) 24 hours prior to triplicate cotransfections with 1 μ g total

pDNA (750 ng pri-/shRNA expression plasmid; 150 ng target reporter plasmid; 100 ng pCI-eGFP). Cells for deep sequencing were seeded in a 6-well format ($\sim 4.8 \times 10^5$ cells/well) 24 hours prior to cotransfection with 5 μ g total pDNA (4.5 μ g pri-/shRNA expression plasmid; 0.5 μ g pCI-eGFP). pCI-eGFP was constructed previously by our lab⁶⁴ to express an enhanced green fluorescent protein and was included as a reporter of transfection efficiency visualized by fluorescence microscopy. Cell culture medium was replaced 24 hours after transfection. At 48 hours after transfection, cells were lysed for the dual luciferase assay or total RNA was extracted for northern blot or deep sequencing samples by phase separation using TRI Reagent (Sigma-Aldrich, St. Louis, MO) according to the manufacturers' protocol for monolayer cells.

Northern blot analysis. Total RNA (30 μ g) was resolved using 15% denaturing PAGE with Decade Marker (Ambion, Austin, TX). RNA was transferred to a Hybond N+ membrane (Amersham, GE Healthcare, Buckinghamshire, UK) using a semi-dry blotter (Sigma-Aldrich, Dorset, UK). Blots were UV cross-linked at 254 nm (0.2 J/cm²) (UVP Crosslinker, Upland, CA) and baked for 1 hour at 80 °C. Probes (**Supplementary Table S5**) were 5' radiolabeled by T4 polynucleotide kinase with 1 μ l γ -³²P ATP (PerkinElmer, Waltham, MA) and purified using a Sephadex-G25 spin column. Hybridization was performed for 16 hours at 42 °C for standard probes and at 50 °C for locked nucleic acid-modified probes (Exiqon, Woburn, MA) using Rapid-hyb buffer (Amersham, GE Healthcare). Membranes were washed as described in the Rapid-hyb protocol or elsewhere.⁶⁵ Signals were quantified using a phosphor-imager (FLA-7000 Image Reader, Fuji Film, Kanagawa, Japan) and accompanying software (Fuji Film Multi Gauge, ver. 3.0, Science Lab 2005).

Dual luciferase reporter assay. Assays were performed in triplicate using the Dual-Luciferase Reporter Assay System (Promega) and a Veritas dual injection Luminometer (Turner Biosystems, Sunnyvale, CA). Silencing was measured as a ratio of target-specific *Renilla* luciferase signal to background Firefly luciferase signal (hFluc: hFluc). The average ratio of each sample was normalized with respect to the average ratio of a control sample (empty expression plasmid).

Thermodynamic profiling and conceptual dicing. Average thermodynamic stability profiles were constructed in a similar manner to previous studies^{23,33,49,53} using free energy values (kcal/mol) from the mfold web server (version 3.5)⁴⁴ with default parameters at 37 °C. In this approach, precise values cannot be assigned to positions within internal loops as the values for loop features are assigned to the preceding closing pair. We therefore estimated values at these positions based on previously described methods. Positions within internal loops were assigned an average value for the loop feature; assigned the average value of a 3 nt window, including one position on either side, where all positions within a loop are assigned a single value⁴⁹; or designated as blank positions and omitted from the average free energy value at that position.²³ Conceptual dicing was performed as previously described,⁴⁹ where the stability of each duplex terminus was estimated from the predicted mfold structure of a corresponding hairpin. Terminal mismatches were assigned the value of the external loop feature.

Deep sequencing. Samples were sequenced using the Illumina Genome Analyzer II platform. The 3'-adapter of raw reads were trimmed and reads were aligned to the full pri-miRNA reference sequence using Novoalign (<http://www.novocraft.com>). Aligned reads were converted to .bam format and indexed using SAMtools⁶⁶ and visualized using Integrative Genomics Viewer (IGV, ver. 2.3).⁶⁷ Further quantification analysis was performed using miRDeep2.⁶⁸ Identifier sequences of 17 nt, starting with the third nucleotide from the 5' end of each expected 22 nt miRNA sequence, were used to identify RNA species aligning to the core of each miRNA. The expected miRNA sequences were adjusted in relation to the Drosha-DGCR8 cleavage site of the major characterized miRNA species. Reads were ranked according to frequency and quantified in reads per million (RPM) for a single experiment.

Supplementary material

- Figure S1.** Summary of folding results for pri-miRNAs.
Figure S2. Predicted secondary RNA structures of selected pri-miRNAs.
Figure S3. Target gene silencing of pri-miRNAs: control assays.
Figure S4. Target gene silencing in a second cell line.
Figure S5. Predicted secondary RNA structures of pri-miRNA mimics.
Figure S6. Summary of folding results for pri-miRNA mimics.
Figure S7. The proportion of canonical hairpins predicted for each set of pri-miRNA mimics.
Figure S8. Target gene silencing of mimics: control assays.
Figure S9. Predicted stability of the pri-miRNA or mimic guide region.
Figure S10. Terminus stability of expected miRNA duplexes.
Figure S11. Deep sequencing results of pri-miRNAs and mimics.
Table S1. Primers for PCR amplification of pri-miRNA sequences from genomic DNA.
Table S2. Oligonucleotides for two-step PCR construction of pri-miRNAs and pri-miRNA mimics.
Table S3. Primers for PCR construction of shRNAs.
Table S4. Oligonucleotides for construction of target duplexes.
Table S5. Probes for detection of guide RNAs in northern analyses.

Acknowledgments We gratefully acknowledge grant support to M.S.W. by the South African National Research Foundation (NRF) and the South African Medical Research Council (MRC). F.T.v.d.B. acknowledges scholarship support from the NRF, Andrew W Mellon Foundation, the South African Poliomyelitis Research Foundation, the Freda Lawenski Foundation, and the Professional Provident Society of South Africa Chairman's Bursary. We thank Sheena Saayman and Neli Nhlabatsi for antiviral shRNA constructs used as positive controls and corresponding target reporter plasmids; Samantha Barichievy for the nonspecific pri-miRNA construct used as a negative control; and Jiehua Zhou, Xiwei Wu, and Jonathan Hart for assistance with the generation and analysis of deep sequencing data.

1. Zeng, Y, Wagner, EJ and Cullen, BR (2002). Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol Cell* **9**: 1327–1333.
2. Zeng, Y, Cai, X and Cullen, BR (2005). Use of RNA polymerase II to transcribe artificial microRNAs. *Methods Enzymol* **392**: 371–380.
3. Silva, JM, Li, MZ, Chang, K, Ge, W, Golding, MC, Rickles, RJ *et al.* (2005). Second-generation shRNA libraries covering the mouse and human genomes. *Nat Genet* **37**: 1281–1288.
4. Grimm, D, Streetz, KL, Jopling, CL, Storm, TA, Pandey, K, Davis, CR *et al.* (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441**: 537–541.
5. Boudreau, RL, Martins, I and Davidson, BL (2009). Artificial microRNAs as siRNA shuttles: improved safety as compared to shRNAs *in vitro* and *in vivo*. *Mol Ther* **17**: 169–175.
6. Zhou, H, Xia, XG and Xu, Z (2005). An RNA polymerase II construct synthesizes short-hairpin RNA with a quantitative indicator and mediates highly efficient RNAi. *Nucleic Acids Res* **33**: e62.
7. Stegmeier, F, Hu, G, Rickles, RJ, Hannon, GJ and Elledge, SJ (2005). A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. *Proc Natl Acad Sci USA* **102**: 13212–13217.
8. Lo, HL, Chang, T, Yam, P, Marcovecchio, PM, Li, S, Zaia, JA *et al.* (2007). Inhibition of HIV-1 replication with designed miRNAs expressed from RNA polymerase II promoters. *Gene Ther* **14**: 1503–1512.
9. Griffiths-Jones, S (2004). The microRNA Registry. *Nucleic Acids Res* **32**(Database Issue): D109–D111.
10. Kozomara, A and Griffiths-Jones, S (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**(Database Issue): D68–D73.
11. Chung, KH, Hart, CC, Al-Bassam, S, Avery, A, Taylor, J, Patel, PD *et al.* (2006). Polycistronic RNA polymerase II expression vectors for RNA interference based on BIC/miR-155. *Nucleic Acids Res* **34**: e53.
12. Liu, YP, Haasnoot, J, ter Brake, O, Berkhout, B and Konstantinova, P (2008). Inhibition of HIV-1 by multiple siRNAs expressed from a single microRNA polycistron. *Nucleic Acids Res* **36**: 2811–2824.
13. Ely, A, Naidoo, T, Mufamadi, S, Crowther, C and Arbutnot, P (2008). Expressed anti-HBV primary microRNA shuttles inhibit viral replication efficiently *in vitro* and *in vivo*. *Mol Ther* **16**: 1105–1112.
14. Ely, A, Naidoo, T and Arbutnot, P (2009). Efficient silencing of gene expression with modular trimeric Pol II expression cassettes comprising microRNA shuttles. *Nucleic Acids Res* **37**: e91.
15. Yang, X, Haurigot, V, Zhou, S, Luo, G and Couto, LB (2010). Inhibition of hepatitis C virus replication using adeno-associated virus vector delivery of an exogenous anti-hepatitis C virus microRNA cluster. *Hepatology* **52**: 1877–1887.
16. Zhang, T, Cheng, T, Wei, L, Cai, Y, Yeo, AE, Han, J *et al.* (2012). Efficient inhibition of HIV-1 replication by an artificial polycistronic miRNA construct. *Viral J* **9**: 118.
17. Huang, X and Jia, Z (2013). Construction of HCC-targeting artificial miRNAs using natural miRNA precursors. *Exp Ther Med* **6**: 209–215.
18. Choi, JG, Bharaj, P, Abraham, S, Ma, H, Yi, G, Ye, C *et al.* (2015). Multiplexing seven miRNA-based shRNAs to suppress HIV replication. *Mol Ther* **23**: 310–320.
19. 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.
20. Rao, DD, Maples, PB, Senzer, N, Kumar, P, Wang, Z, Pappen, BO *et al.* (2010). Enhanced target gene knockdown by a bifunctional shRNA: a novel approach of RNA interference. *Cancer Gene Ther* **17**: 780–791.
21. Senzer, N, Barve, M, Kuhn, J, Melnyk, A, Beitsch, P, Lazar, M *et al.* (2012). Phase I trial of “bi-shRNAi(furin)/GMCSF DNA/autologous tumor cell” vaccine (FANG) in advanced cancer. *Mol Ther* **20**: 679–686.
22. Bartel, DP (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281–297.
23. Han, J, Lee, Y, Yeom, KH, Nam, JW, Heo, I, Rhee, JK *et al.* (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**: 887–901.
24. Denli, AM, Tops, BB, Plasterk, RH, Ketting, RF and Hannon, GJ (2004). Processing of primary microRNAs by the microprocessor complex. *Nature* **432**: 231–235.
25. Gregory, RI, Yan, KP, Amuthan, G, Chendrimada, T, Doratotaj, B, Cooch, N *et al.* (2004). The microprocessor complex mediates the genesis of microRNAs. *Nature* **432**: 235–240.
26. Han, J, Lee, Y, Yeom, KH, Kim, YK, Jin, H and Kim, VN (2004). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**: 3016–3027.
27. Auyeung, VC, Ulitsky, I, McGeary, SE and Bartel, DP (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell* **152**: 844–858.
28. Nguyen, TA, Jo, MH, Choi, YG, Park, J, Kwon, SC, Hohng, S *et al.* (2015). Functional anatomy of the human microprocessor. *Cell* **161**: 1374–1387.
29. Fang, W and Bartel, DP (2015). The menu of features that define primary microRNAs and enable *de novo* design of microRNA genes. *Mol Cell* **60**: 131–145.
30. Quick-Cleveland, J, Jacob, JP, Weitz, SH, Shoffner, G, Senturia, R and Guo, F (2014). The DGCR8 RNA-binding heme domain recognizes primary microRNAs by clamping the hairpin. *Cell Rep* **7**: 1994–2005.

31. Ma, H, Wu, Y, Choi, JG and Wu, H (2013). Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. *Proc Natl Acad Sci USA* **110**: 20687–20692.
32. Burke, JM, Kelenis, DP, Kincaid, RP and Sullivan, CS (2014). A central role for the primary microRNA stem in guiding the position and efficiency of Drosha processing of a viral pri-miRNA. *RNA* **20**: 1068–1077.
33. Warf, MB, Johnson, WE and Bass, BL (2011). Improved annotation of *C. elegans* microRNAs by deep sequencing reveals structures associated with processing by Drosha and Dicer. *RNA* **17**: 563–577.
34. Quarles, KA, Sahu, D, Havens, MA, Forsyth, ER, Wostenberg, C, Hastings, ML *et al.* (2013). Ensemble analysis of primary microRNA structure reveals an extensive capacity to deform near the Drosha cleavage site. *Biochemistry* **52**: 795–807.
35. Alarcón, CR, Lee, H, Goodarzi, H, Halberg, N and Tavazoie, SF (2015). N6-methyladenosine marks primary microRNAs for processing. *Nature* **519**: 482–485.
36. Bernstein, E, Caudy, AA, Hammond, SM and Hannon, GJ (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363–366.
37. Grishok, A, Pasquinelli, AE, Conte, D, Li, N, Parrish, S, Ha, I *et al.* (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**: 23–34.
38. Hutvagner, G, McLachlan, J, Pasquinelli, AE, Bálint, E, Tuschl, T and Zamore, PD (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**: 834–838.
39. Ketting, RF, Fischer, SE, Bernstein, E, Sijen, T, Hannon, GJ and Plasterk, RH (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev* **15**: 2654–2659.
40. Hammond, SM, Bernstein, E, Beach, D and Hannon, GJ (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**: 293–296.
41. Zeng, Y, Yi, R and Cullen, BR (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J* **24**: 138–148.
42. Chen, CZ, Li, L, Lodish, HF and Bartel, DP (2004). MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**: 83–86.
43. Zeng, Y and Cullen, BR (2005). Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J Biol Chem* **280**: 27595–27603.
44. Zuker, M (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.
45. Lee, NS, Dohjima, T, Bauer, G, Li, H, Li, MJ, Ehsani, A *et al.* (2002). Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nat Biotechnol* **20**: 500–505.
46. Nishitsuji, H, Kohara, M, Kannagi, M and Masuda, T (2006). Effective suppression of human immunodeficiency virus type 1 through a combination of short- or long-hairpin RNAs targeting essential sequences for retroviral integration. *J Virol* **80**: 7658–7666.
47. McIntyre, GJ, Groneman, JL, Yu, YH, Jaramillo, A, Shen, S and Applegate, TL (2009). 96 shRNAs designed for maximal coverage of HIV-1 variants. *Retrovirology* **6**: 55.
48. Lewis, BP, Shih, IH, Jones-Rhoades, MW, Bartel, DP and Burge, CB (2003). Prediction of mammalian microRNA targets. *Cell* **115**: 787–798.
49. Krol, J, Sobczak, K, Wilczynska, U, Drath, M, Jasinska, A, Kaczynska, D *et al.* (2004). Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *J Biol Chem* **279**: 42230–42239.
50. Saayman, S, Barichiev, S, Capovilla, A, Morris, KV, Arbutnot, P and Weinberg, MS (2008). The efficacy of generating three independent anti-HIV-1 siRNAs from a single U6 RNA Pol III-expressed long hairpin RNA. *PLoS One* **3**: e2602.
51. Saayman, S, Arbutnot, P and Weinberg, MS (2010). Deriving four functional anti-HIV siRNAs from a single Pol III-generated transcript comprising two adjacent long hairpin RNA precursors. *Nucleic Acids Res* **38**: 6652–6663.
52. Nhlabatsi, N (2014). *Comparing the Silencing Efficacy of Dicer-Independent and Dependent shRNAs*. Dissertation, Department of Molecular Medicine and Haematology, University of the Witwatersrand: Johannesburg, South Africa.
53. Khvorova, A, Reynolds, A and Jayasena, SD (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**: 209–216.
54. Reynolds, A, Leake, D, Boese, Q, Scaringe, S, Marshall, WS and Khvorova, A (2004). Rational siRNA design for RNA interference. *Nat Biotechnol* **22**: 326–330.
55. Frank, F, Sonenberg, N and Nagar, B (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**: 818–822.
56. Schwarz, DS, Hutvagner, G, Du, T, Xu, Z, Aronin, N and Zamore, PD (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**: 199–208.
57. Gregory, RI, Chendrimada, TP, Cooch, N and Shiekhattar, R (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* **123**: 631–640.
58. Landgraf, P, Rusu, M, Sheridan, R, Sewer, A, Iovino, N, Aravin, A *et al.* (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**: 1401–1414.
59. Ui-Tei, K, Naito, Y, Takahashi, F, Haraguchi, T, Ohki-Hamazaki, H, Juni, A *et al.* (2004). Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res* **32**: 936–948.
60. Amarzguioui, M and Prydz, H (2004). An algorithm for selection of functional siRNA sequences. *Biochem Biophys Res Commun* **316**: 1050–1058.
61. Aagaard, LA, Zhang, J, von Eije, KJ, Li, H, Saetrom, P, Amarzguioui, M *et al.* (2008). Engineering and optimization of the miR-106b cluster for ectopic expression of multiplexed anti-HIV RNAs. *Gene Ther* **15**: 1536–1549.
62. Cunningham, F, Amode, MR, Barrell, D, Beal, K, Billis, K, Brent, S *et al.* (2015). Ensemble 2015. *Nucleic Acids Res* **43**(Database Issue): D662–D669.
63. Castanotto, D, Li, H and Rossi, JJ (2002). Functional siRNA expression from transfected PCR products. *RNA* **8**: 1454–1460.
64. Passman, M, Weinberg, M, Kew, M and Arbutnot, P (2000). *In situ* demonstration of inhibitory effects of hammerhead ribozymes that are targeted to the hepatitis Bx sequence in cultured cells. *Biochem Biophys Res Commun* **268**: 728–733.
65. Várallyay, E, Burgyn, J and Havelda, Z (2008). MicroRNA detection by northern blotting using locked nucleic acid probes. *Nat Protoc* **3**: 190–196.
66. Li, H, Handsaker, B, Wysoker, A, Fennell, T, Ruan, J, Homer, N *et al.*; 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
67. Robinson, JT, Thorvaldsdóttir, H, Winckler, W, Guttman, M, Lander, ES, Getz, G *et al.* (2011). Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26.
68. Friedländer, MR, Mackowiak, SD, Li, N, Chen, W and Rajewsky, N (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res* **40**: 37–52.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Supplementary Information accompanies this paper on the Molecular Therapy–Nucleic Acids website (<http://www.nature.com/mtna>)