Published in final edited form as:

Nano Lett. 2016 March 09; 16(3): 1726–1735. doi:10.1021/acs.nanolett.5b04651.

Multistrand Structure Prediction of Nucleic Acid Assemblies and Design of RNA Switches

Eckart Bindewald^{#†}, Kirill A. Afonin^{#‡§}, Mathias Viard[†], Paul Zakrevsky[‡], Taejin Kim[‡], and Bruce A. Shapiro^{*‡}

[†] Basic Science Program, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, Maryland 21702, United States

[‡] Gene Regulation and Chromosome Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, Maryland 21702, United States

[§] Department of Chemistry, University of North Carolina at Charlotte, 9201 University City Boulevard, Charlotte, North Carolina 28223, United States

[#] These authors contributed equally to this work.

Abstract

RNA is an attractive material for the creation of molecular logic gates that release programmed functionalities only in the presence of specific molecular interaction partners. Here we present HyperFold, a multistrand RNA/DNA structure prediction approach for predicting nucleic acid complexes that can contain pseudoknots. We show that HyperFold also performs competitively compared to other published folding algorithms. We performed a large variety of RNA/DNA hybrid reassociation experiments for different concentrations, DNA toehold lengths, and G+C content and find that the observed tendencies for reassociation correspond well to computational predictions. Importantly, we apply this method to the design and experimental verification of a two-stranded RNA molecular switch that upon binding to a single-stranded RNA toehold disease-marker trigger mRNA changes its conformation releasing an shRNA-like Dicer substrate structure. To demonstrate the concept, connective tissue growth factor (CTGF) mRNA and enhanced green fluorescent protein (eGFP) mRNA were chosen as trigger and target sequences, respectively. In vitro experiments confirm the formation of an RNA switch and demonstrate that the functional unit is being released when the trigger RNA interacts with the switch toehold. The designed RNA switch is shown to be functional in MDA-MB-231 breast cancer cells. Several other switches were also designed and tested. We conclude that this approach has considerable potential because, in principle, it allows the release of an siRNA designed against a gene that differs from the gene that is utilized as a biomarker for a disease state.

GRAPHICAL ABSTRACT

*Corresponding Author Phone 301-846-5536; fax 301-846-6210; shapirbr@mail.nih.gov.

Notes

The authors declare no competing financial interest.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.nanolett.5b04651.

All sequences used in this project and experimental and computational data supplementary data (PDF)

Keywords

RNA switch; RNA/DNA hybrid; RNA interference; Dicer; secondary structure

The mechanism of conventional therapeutics is frequently the inhibition of one gene product, typically the inhibition of a protein by a small molecule.^{1,2} This approach is effectively a combination of a diagnosis step (binding to a molecule that acts as a biomarker) and a treatment step (inhibiting the function of the same biomarker molecule). While this approach is successful, more therapeutic options become available if the diagnosis and treatment steps can be separated, if for example the binding of a drug to one molecule triggers the inhibition of the function of a different molecule. A second motivation for this approach is that it potentially allows for drugs to only exert their function in the targeted cells. Such improvements in drug targeting could be used for the development of drugs with fewer side effects. Targeted treatment is already possible; however, current nanoparticle developments typically utilize differences in cell surface properties. The targeting strategy envisioned here exploits differences in the cytoplasm in general, and different expression patterns of transcribed RNA in particular to release a designed functionality unrelated to the utilized molecular marker. A third motivation is that the activatable drug approach could be used to utilize highly cytotoxic drugs safely, because their therapeutic function is only activated in cells with certain properties. Notice that in the case of cancers, this cell-distinction capability is far from trivial.

A variety of self-cleaving ribozyme-based logic gates have been reported that release an RNA oligonucleotide upon binding to a trigger DNA oligonucleotide.³ By incorporating such RNA-based logic devices into the untranslated regions of mRNAs, the “output” can also be the increased translation of a protein.^{4,5} Synthetic riboswitches can make either the transcription⁶ or the translation of an mRNA conditional to the binding of a ligand.^{7,8} Several micro-RNA based synthetic regulatory circuits have been reported.^{9–11}

Here, we report the design and experimental verification of two-stranded RNA molecular switches that upon binding to trigger mRNAs changes conformation and release their single-stranded shRNA-like structures that then can be processed by Dicer resulting in production of siRNA.

The design concept is schematically depicted in Figure 1. The idea is that an shRNA-like designed RNA (called antitarget strand) is initially in an inactive conformation, while it is bound to another RNA strand (called the antitrigger strand). This antitrigger strand is

composed of sequence regions that are complementary to parts of the RNA antitarget strand and parts of the trigger mRNA. In the presence of the trigger mRNA, the antitarget strand and the trigger mRNA strand form a duplex, leading to the dissociation of the antitarget strand from the RNA switch antitarget strand. The antitarget strand now changes its conformation into an shRNA-like Dicer substrate RNA structure that is recognized and cleaved by Dicer, ultimately leading to the down-regulation of a target mRNA. Ideally, a trigger RNA strand would be highly expressed or only present in all targeted cells and not in other nontargeted cells.

In the presented design eGFP is chosen as the target gene, triggered by the presence of connective tissue growth factor (CTGF/CCN2/IGFBP8) mRNA. We chose CTGF mRNA as a trigger because CTGF is characterized by low expression in normal adult (human) tissue, while its expression in adult tissue is associated with a variety of diseases including certain cancers.^{12–17} The RNA switch presented here is not microRNA but siRNA-based and is, unlike previously published systems, not dependent on transcription factor proteins.

The approach of utilizing interacting RNA strands with non-nested base pairings required in silico secondary structure prediction capabilities that are presented here. The novel computational approach (named HyperFold) is able to predict base pairing of RNA/RNA, DNA/DNA, and RNA/DNA hybrid interactions. A variety of secondary structure prediction programs for multiple RNA strands have been reported.^{18–22} The NUPACK software is in addition to RNA/RNA interactions also able to predict DNA/DNA interactions.^{23,24} Secondary structure predictions of RNA/DNA hybrid duplexes have been made available via the RNA Vienna package.¹⁸ The software described here is unique in the sense that it allows for secondary structure predictions of multiple nucleic acids strands simultaneously taking into account possible RNA/RNA, RNA/DNA and DNA/DNA interactions which may include pseudoknots. RNA structures with non-nested base pairings (in other words secondary structures whose circular diagram representations contain “crossing arcs”) are referred to as pseudoknotted.²¹ An example of a pseudoknotted secondary structure is depicted in Figure S3: the arc-representation in Figure S3b represents each base pair as an arc, while the nucleotide sequences are arranged in circular fashion. Because the shown secondary structure contains intersecting arcs (in other words non-nested base pairs), the shown secondary structure is pseudoknotted.

The presented prediction approach does consider base pairs that are non-nested with respect to each other even if the involved nucleotides involve more than one nucleotide strand. Lastly, a novel distance-based approach is presented that is a unified framework for the estimation of entropic contributions for folding events involving single or multiple strands and pseudoknots.

To evaluate the RNA/DNA hybrid capabilities of the approach, experimental and computational results of RNA/DNA hybrid structures with different toehold lengths and compositions are presented and compared. These hybrid complexes display interesting tendencies in terms of measured amounts of reassociation: reassociation is ineffective for short toehold lengths and low concentrations. The computational predictions made by HyperFold are consistent with these tendencies. The algorithm is furthermore evaluated and

compared to other single-strand secondary structure prediction programs using a large set of RNA structures with known base pairing.

Methods.

Computational Approaches. Algorithm for Prediction of Multistrand RNA Secondary Structures

A variety of approaches have been used to predict the secondary structure of RNA strands. Because the RNA switch in its inactive conformation (antitarget strand bound to the antitrigger) corresponds to a pseudoknotted RNA structure, a new computational method is implemented for the prediction of pseudoknotted multistrand RNA secondary structures. While some elements of the algorithm were recently described for the prediction of RNA–DNA hybrid structures, the approach described here is based on a novel search algorithm as well as a novel way to ascertain entropic contributions and kinetic accessibility.²⁵

A complete enumeration of all secondary structures is not feasible for longer nucleotide sequences because of a combinatorial explosion of possible secondary structures. Because of that, the algorithm performs an enumeration of secondary structures that are generated by choosing helices from a set of “core helices”. Those “core helices” are defined as helices that consist of at least six base pairs and that cannot be extended further on either end by Watson–Crick or GU base pairs.

Representation of Secondary Structures.

Each secondary structure is internally represented by a matrix of maximum and minimum distances between nucleotides. For nucleotide pairs of single-stranded regions of the same strand, their maximum distance is computed as the product of the number of separating nucleotides times a maximum distance per nucleotide. If both nucleotides are part of the same base pair, their maximum and minimum distances are set to a constant value. If two nucleotides belong to different nucleotide strands without interstrand base pairing, their maximum distance is given through the inverse of the strand concentrations. For each folded base pair, all maximum and minimum nucleotide-nucleotide distances are updated to fulfill these rules. The maximum and minimum distances between residues define an estimated volume of conformational freedom that is used to compute the conformational entropy of a secondary structure. Importantly, the approach provides estimated entropy contributions for potentially pseudoknotted structures consisting of multiple strands. Details of the approach for estimating the conformational entropy are described in the Supporting Information. Free energy contributions of nucleotide base pairing are computed using base pair stacking parameters for RNA/RNA, DNA/DNA and RNA/DNA interactions.^{26–28} Because this matrix data structure in “distance space” is memory-intensive, another more compact representation is utilized in form of a vector of integer numbers that indicate which of the core helices are part of the structure. An example of this representation is shown schematically in Figure 2a. A “0” at position i in that integer vector indicates that core helix i is not part of the structure; a “1” at position i indicates that a core helix i is in its full length present in the secondary structure. Indices greater than one describe different partially placed variants of a helix. The matrix representation of secondary structures facilitates

entropy estimations, while the integer vector representation is memory-efficient. The program is able to interconvert between the distance matrix representation and the integer vector representations thus making the distance matrix available only when needed.

Search Algorithm.

The search algorithm is depicted schematically in Figure 2b. Briefly, this folding algorithm works as follows: Two priority queues A and B for storing free energy-ranked secondary structures are generated. To prevent the combinatorial explosion of the computational search, the implementation of the priority queue data structure is such that it automatically deletes the lowest-priority elements, if the number of stored elements exceeds a specified maximum size (this maximum size is set to 1000 elements). The two priority queues hold the integer vector representations of the partially folded nucleic acid complexes such that the contained structures are at any given time sorted by the estimated free energy of folding. The algorithm starts by generating an array of “core helices” (defined above) that is sorted in order of free energy. In addition, a second array of shorter noncore “short helices” is generated. Next, the integer vector representation of an unfolded structure (containing zero base pairs) is added to priority queue A. If there are n core helices, the search consists of n rounds. In the first round, the top-ranking element of queue A (which happens to be the previously added unfolded structure) is removed from queue A. Four variants of that structure are generated and added to queue B. These four variants are (i) core helix 1 being unplaced, (ii) core helix 1 fully placed, (iii) and (iv) core helix 1 placed to half its length from either end. Helix candidates can be shortened in order to not conflict with already placed helices. Structures in which a helix cannot be placed are not stored. In other words, in round 1 the decision is made to what extent core helix 1 is placed. For each structure composed of a combination of core helices, another structural variant is generated by placing the previously generated short helices in the order of free energy, leading for the first round to up to $4 \times 2 = 8$ stored structural variants. In the first round, queue B has the role of a “receiving” queue, and queue A has the role of an “emitting” queue. In the second round, queue A becomes the receiving queue and queue B the emitting queue. For a general round i , and for each structure removed from the emitting queue, the different variants with respect to the placement of core helix i are generated and the resulting secondary structures are added to the receiving queue (one version without additional small helices and one version with additional small helices). This process is continued until the emitting queue is empty and all core helices have been explored.

In this fashion, all combinations of core helices are searched quasi-exhaustively while placing additional small helices in order of free energy for each examined combination of core helices. Helices can be pseudoknotted with the constraint that each helix can be non-nested with respect to at most one other helix. This approach thus combines entropy estimations and a tunable search strategy that avoids redundant structure evaluations and is thus applicable (and used) for estimating the partition function.

Kinetic Considerations.

The computational approach identifies pairs of secondary structures that have a different strand connectivity and differ by only one additionally placed helix. The minimum energetic

cost to place one additional helix to transition from one interstrand connectivity to a different interstrand connectivity is used to estimate the free energy of activation. We show for the case of RNA/DNA hybrids that the estimated free energy of activation (going from the nonre-associated state to an intermediate complex consisting of four strands) can lead to a qualitative prediction where the RNA/ DNA hybrids can readily reassociate but are kinetically trapped due to short toehold lengths leading to a high free energy of activation of the transition state (Afonin et al. The use of minimal RNA toeholds to trigger the activation of multiple functionalities. Accepted to this Journal).

Sequence Design.

An eGFP Dicer Substrate RNA (DS RNA) sequence was chosen as a starting point for the sequence design for the CTGF/eGFP switch.^{29,30} The design of the antitarget strand is based on the idea of energetically outcompeting the relatively stable part of the antitarget strand that forms a DS RNA duplex. CTGF mRNA was chosen as the trigger RNA, intended to bind the antitarget strand, thus releasing the functionality of the antitarget strand. A known siRNA binding site of the target CTGF mRNA was chosen in order to ensure that the CTGF mRNA binding site is accessible.¹⁵

The design of the two-stranded switch is based on an envisioned scaffold shown in Figure 3. The idea is to design two strands, such that upon removal of the antitarget strand the antitarget folds into a functional shRNA-like structure. The challenge is to design the two-stranded complex (here called switch), such that the energetically stable Dicer substrate helix of the shRNA is not present. This is accomplished in the proposed design by introducing several “decoy” helices that energetically outcompete the Dicer substrate helix. These decoy helices are shown in Figure 3 as helices H1, H2, and H6. The complex structure is further stabilized by cases in which A-U base pairs of the switch structure replace G-U base pairs of the Dicer substrate helix (indicated with arrows in Figure 3) without changing the sequence region corresponding to the siRNA antisense strand. One important constraint with this strategy is to keep the antisense strand of the Dicer substrate helix unmodified (the 21 nucleotides at the 3' end of the antitarget strand). The complex structure is further stabilized by helices H3, H4 and H5. Helix H3 is introduced to stabilize a 3way junction, in which helices H2 and H4 form a sharp angle. The antitarget strand consists to a large extent of the trigger CTGF mRNA reverse-complementary sequences (highlighted in dark blue in Figure 3). In order to facilitate the folding process of the three-dimensional structure, several three-nucleotide single-stranded RNA stretches are introduced. Importantly, almost all nucleotides of the switch are predetermined by the given siRNA sequences, the CTGF mRNA binding site and the introduced linker regions. A flow chart of these steps is shown in Figure S1. The computationally designed and experimentally tested sequences are given in the Supporting Information.

Creation of a Three-Dimensional (3D) Model.

Starting from a model secondary structure, the Nanotiler program was used to create a set of three-dimensional all-atom representations of helices without steric clashes. This set of helices was then augmented with single-stranded linker regions.³¹ RNA-Composer was used to generate an initial 3D structure of the 3' terminal hairpin loop (nt positions 64 to 96 of

the antitarget strand).³² Energy minimization was applied to the structure using the ff10 force field and the Amber12 molecular dynamics package.^{33,34} The energy-minimized structure was obtained when the convergence criterion for the energy gradient was less than 1.0×10^{-5} kcal/mol Å.

RNA Test Set.

To test the ability of the algorithm to fold single-sequence RNAs a secondary structure test set based on the RNA Strand (version 2.0) Database was used.³⁵ This database is a compilation of RNA secondary structures that were derived using a variety of source databases with RNA structural information. For evaluating prediction qualities, RNA secondary structures originating from a PDB structure and that have a length of up to 250 nt were used. These criteria corresponded to 741 pseudoknot-free secondary structures and 110 pseudoknotted RNA structures.

Experimental Approaches. Preparation and Reassociation of Various RNA/DNA Hybrids.

Sets of RNA/DNA hybrid sequences with different DNA toehold lengths (2 nt, 4 nt, 6 nt, 8 nt, and 10 nt) and different toehold G+C content (60% and 25% of G+C) were prepared, using a previously described approach.³⁶ Reassociation experiments were performed at 37 °C for strand concentrations of 30 nM, 250 nM, 1000 nM, 5000 nM, and 10 000 nM. Concentrations of reassociated DNA duplexes were measured after 6 h through analysis of the FRET established between the Alexa 488 and the Alexa 546 labeled RNA strand upon their reassociation.

Preparation of RNA Switches.

All RNA molecules were prepared by transcription of PCR amplified DNA templates followed by treatment with RNaseH to remove the starting sequences.³⁷ Synthetic DNA templates containing T7 RNA polymerase promoters and one additional sequence downstream from the promoter and prior to the 5' side of the designed RNAs were purchased from IDT DNA and amplified by PCR using primers. PCR products were purified using the DNA Clean & concentrator-5 (Zymo Research), and RNA molecules were prepared enzymatically by in vitro transcription using wildtype T7 RNA polymerase. Samples were incubated at 37 °C for 4 h in a buffer containing 5 mM MgCl₂, 0.5 mM MnCl₂, 2 mM spermidine, 50 mM Tris buffer (pH 7.5), 2.5 mM NTPs, 10 mM DTT, 0.1 µg/µL IPP, and 0.8 u/µL RNasin. Two methods were used to cleave off the starting sequences required for good transcription yields. In one method, KCl (75 mM final), RNase H, and short DNAs (5'-gctcttccttccc) complementary to the starting sequence were added to the transcription mixture following the 4 h transcription and incubated for 30 min at room temperature. Samples were purified on a denaturing urea gel (PAGE) (8% or 10% acrylamide, 8 M urea). The RNA was eluted from gel slices overnight at 4 °C into a buffer containing 300 mM NaCl, 10 mM Tris pH 7.5, 0.5 mM EDTA. After precipitating the RNA in two volumes of 100% ethanol, samples were rinsed twice with 90% ethanol, vacuum-dried, and dissolved in TE buffer. In the second approach, full length transcription products were initially gel purified, precipitated, and reconstituted in water, after which these full length RNAs were supplemented with the short DNA oligo (5'-gctcttccttccc), RNase H and supplied 10× reaction buffer (New England Biolabs). The reaction was allowed to progress

for 45 min at 37 °C before the cleavage products were gel purified, precipitated, and reconstituted in water.

The RNA units at concentrations specified in the text were mixed in double-deionized water and incubated in a heat block at 95 °C for 2 min. The samples were removed from heat and snap-cooled to 55 °C over a period of 20 min. Assembly buffer (89 mM Tris, 80 mM boric acid (pH 8.3), 2 mM magnesium acetate) was added to the mixtures either prior or after the heating. Assembled RNA switches were gel-purified using native-PAGE (see below). The bands were visualized as described above, excised from the gel, and eluted in 89 mM Tris-borate, pH 8.3, 2 mM Mg(OAc)₂ for 5–12 hours at 4 °C.

Nondenaturing PAGE Experiments.

All constructs were annealed as per the desired protocol described earlier. Typically,^{38,39} all assembly experiments reported were analyzed at 10 °C in 7% (37:1) or 10% (19:1) native polyacrylamide gels in the presence of 89 mM Tris-borate, pH 8.3, and 2 mM Mg(OAc)₂. Once assembled, the RNA solutions were placed on ice and combined with an equal volume of loading buffer consisting of 2 mM Mg(OAc)₂, 50% glycerol, 0.01% bromophenol. Gels were run for either 4 h at 25 W (larger gels) or 8 W (small gels) with the temperature set to 10 °C. Native-PAGE experiments were stained using SYBR Gold or ethidium bromide and visualized using a Hitachi FMBIO II Multi-View Imager.

Recombinant Human Dicer Assay.

Samples containing different sets of RNAs (specified in text) were prepared at a final concentration of 3 μM and incubated for 3 h at 37 °C with recombinant human turbo Dicer enzyme kit (Genlantis), according to the manufacturer's suggested protocol. Dicing reactions were quenched by adding Dicer stop solution (provided by the manufacturer) prior to analysis on native-PAGE.

Transfection Experiments in Cultured Human Breast Cancer Cells.

MDA-MB-231 cells that stably express enhanced green fluorescent protein (eGFP) were used in all cell culture experiments. Cells were cultured in DMEM/high glucose media (HyPure) supplemented with 10% FBS and 1% penicillin–streptomycin in a 5% CO₂ incubator at 37 °C. All transfections were conducted using Lipofectamine 2000 (L2K) purchased from Invitrogen. For a typical transfection experiment, cells were dissociated from the flask using Cell Stripper (Corning) dissociation buffer and 24-well plates were seeded with 20 000 cells per well. Transfection was performed 24 h after plating. 50× [RNA] solutions were prepared and preincubated with L2K at 25 °C for 20 min, after which the solutions were diluted to 1× concentration using Opti-MEM media (Gibco BRL). Cell growth media was then aspirated from wells and replaced with the 1× [RNA] solution and incubated for 4 h, after which the transfection solution was replaced with the previous DMEM/high glucose-FBS-pen/ strep media. Cells were allowed to grow for an additional 3 days, at which time analysis was performed. For experiments examining the kinetics of switch mediated eGFP knockdown 1, 2, and 3 days post-transfection, cells were plated in 24-well plates at 40 000 cells per well to ensure an adequate number of cells were available for analysis at 1 day post-transfection.

To examine the effect of trigger transcript expression levels on switch function, cells were subjected to two sequential transfections. 24-well plates were seeded with 5000 cells per well. Twenty-four hours after plating, cells were transfected as described above with a Dicer substrate RNA (DS RNA) against CTGF mRNA (5 or 50 nM final concentration). This anti-CTGF DS RNA was derived from previously used shRNA sequences against CTGF and targets a region of the CTGF mRNA that is upstream of the region that interacts with the antitrigger strand of the RNA switch.^{15,40} After an additional 2 days, cells were then transfected with the RNA switch or its component strands (50 nM final concentration) targeting eGFP. Cells were analyzed 3 days following the second transfection.

Flow Cytometry.

Flow cytometry was used to quantitatively analyze eGFP expression levels following the transfection experiments. Cells were washed with 1× PBS and lifted from wells using cell dissociation buffer. eGFP fluorescence was measured using a BD Acurri C6 flow cytometer. In general, at least 10 000 events were measured and analyzed using BD Accuri C6 software.

Microscopy.

Fluorescent microscopy was used to qualitatively assess the level of eGFP expression following transfection experiments. Images were acquired of cells in 24-well plates in growth media using an Olympus IX70 microscope, Lambda DG-4 light source, and using MetaMorph v2 software.

Results and Discussion.

Computational Results for Test Set of Single-Sequence RNAs.

The computational approach was utilized in order to predict the base pairing of a large set of 851 reference RNA structures obtained from the RNA Strand Database.³⁵ Our HyperFold program as well as the secondary structure prediction programs pknotsRG and UnaFold were applied to the RNA sequences using default parameters.^{41,42} The results are shown in the form of a box-whisker plot Figure 4, for pseudoknotted and pseudoknot free structures. One can see that, for RNA sequences with less than 50 nt length, the quality of prediction results obtained from HyperFold is comparable to UnaFold and pknotsRG, while also comparing favorably for longer sequences. One can view as the main advantage of HyperFold is its ability to provide predictions for multiple nucleic acid strands (RNA, DNA including RNA/DNA hybrid interactions), using strand concentrations while allowing for pseudoknots.

Reassociation of RNA/DNA Hybrids.

Figure S2 depicts the experimentally determined relative extents of reassociation for different toehold lengths, G+C content, and individual hybrids concentrations. The computational structure prediction approach was used to predict equilibrium concentrations of complexes consisting of RNA/DNA hybrid sequences. For all toehold lengths and concentrations, it was found computationally that the reassociated RNA and DNA duplexes correspond to the lowest free energy conformation. In other words, the computational results

suggest, that experimentally observed dramatic differences in reassociation are due to kinetic effects. This is also confirmed by experimental results where reassociation is observed for RNA/DNA hybrids without toeholds provided they have high concentration (Figure S2). Because of that, the free energy change that corresponds to a transition of placing a single helix from the lowest-energy nonreassociated state to an intermediate state consisting of a complex consisting of one copy of each of the four strand species is used as free energy of activation for a transition state. If several helices fulfill this restriction, the free energy change of the lowest energy transition state is used. These computationally estimated values are plotted together with the experimentally determined reassociation rates in Figure 5. One can see that the data, even though it corresponds to different concentrations, toehold lengths and G+C contents, follows fairly well a sigmoidal curve as one would expect for a chemical system that contains a kinetic barrier. Note that the same computational algorithm has also been applied to a different set of RNA/DNA hybrid sequences with RNA toeholds (Afonin et al. The use of minimal RNA toeholds to trigger the activation of multiple functionalities. Accepted to this *Nano Letters*).

Computational Results for RNA Switches.

Secondary structure predictions using the described approach were performed for the three involved RNA strands (antitarget strand, antitrigger strand, and trigger strand). It was, using this program, computationally verified that the antitarget strand in isolation is predicted to fold into the shRNA-like structure exposing a DS RNA helix. The structure with the lowest identified free energy is shown in Supplementary Figure S3. As a proof of concept, several switches designed to release shRNA-like constructs containing PLK1 DS RNA upon recognition of binding to eGFP mRNA were also tested computationally and experimentally (Figure S9). Three different sites of GFP mRNAs were used (121, 238, and 633) as triggers. As in the other designed switches, the chosen trigger sites on the trigger mRNA correspond to known siRNA binding sites. In addition, information about secondary-structure accessibility of the mRNA binding site was ascertained using a base-pairing probability prediction generated by the RNAfold program. Each binding site was “scanned” by offsetting the trigger/antitrigger binding site by up to 10 nt upstream or downstream of the siRNA binding site. Predictions and sequences are shown in Figures S4–S6.

One can see, that in all cases, the antitarget and antitrigger strand are predicted to form a complex with a scaffold similar to the intended design and a structure that does not contain the Dicer-substrate helix. Furthermore, the addition of a trigger strand fragment is predicted in all cases to lead to the release of the antitarget strand. The active shRNA-like conformation of the antitarget strand is predicted to correspond to the lowest free energy in the cases shown in Figure S3, but not in S4–S6. The predictions indicate participation of the toehold regions in additional base pairing interactions with the antitarget and antitrigger strand.

A three-dimensional model of the CTGF–eGFP switch duplex has been generated as described in Materials and Methods and is shown in Figure 3 (right). Shown in red is the siRNA sequence region, dark-blue corresponds to the part of the antitrigger strand that is reverse-complementary to the trigger mRNA, light-blue regions indicate loop regions of the

antitrigger strand. One can see that the three-dimensional core is relatively compact, suggesting that the single-stranded regions connecting those helices should not be much shorter than 3 nts.

If, additionally, the trigger strand is present, the thermodynamic folding algorithm predicts the formation of a duplex between the trigger strand and the antitrigger strand (see Figure S3c) and the antitarget strand that undergoes a conformational change leading to the formation of a Dicersubstrate helix.

In Vitro Studies of RNA Switches.

One particular challenge of this project was synthesizing the individual strands entering the composition of the switches. The yield of in vitro transcription with wild type T7 RNA polymerase is highly dependent on the sequence in the +1 to +6 region of the promoter.⁴³ Due to the nature of the designed switches, the 5'-ends of the RNA strands were predefined. Therefore, we used RNase H to process the transcribed RNAs carrying the expanded promoter regions (Figure S7).⁴⁴ The two-stranded RNA switches were assembled in one-pot and then gel-purified as described in Methods. The ability of the switches to interact with the mRNA fragment was assessed by native-PAGE experiments (Figure 6). The incubation of the switches (1 μ M final) with an excess of the CTGF mRNA fragment (3 μ M final) at 37 °C resulted in the formation of antitrigger/trigger-mRNA duplex (waste) and release of the refolded antitarget strand that then becomes an active RNA (Figure 6a). Higher order bands appearing in the switch+mRNA lanes are in part likely due to complexes formed consisting of three strands (one copy of the trigger strand, antitarget strand and antitrigger strand). These bands disappear at higher incubation temperature (55 °C) (Figure S8). Interestingly, melting and snap cooling of the two-stranded switch resulted only in partial release of the antitarget and antitrigger strands while ~34% of the switch remains intact (Figure S8, lane "switch (remelted)"). These results also confirm the expected two-stranded composition of the switch. The release of a conventional (21-mer) siRNA from refolded antitarget strand was confirmed by Dicer assays (Figure 6b), while incubation with Dicer of the nonactivated two-stranded switch (no interaction with mRNA trigger) led to the complete degradation of the switch. To confirm the importance of the single-stranded toehold in the two-stranded switches for mRNA recognition and the reassociation process, a switch without any toehold was tested (Figure 6c). The results show that the absence of toehold completely prevents any interaction and reassociation between the truncated switch and the mRNA. Also, the switches with permuted "dummy" toeholds show no interactions with the mRNA (Figure 6d). This suggests the importance of the sequence specific single-stranded RNA toehold for switch activation. The native PAGE results corresponding to the additionally designed eGFP/PLK1 tested switches are shown in supporting Figure S9. Each switch was tested against three different GFP mRNA trigger fragments (121, 238, and 633). However, given the presence of the corresponding mRNA fragments only one of the switches, the one designed to bind GFP mRNA at position 633, releases the shRNA-like structure. The other switches show some binding to the corresponding mRNA fragments without the correct conformational changes. These results correspond to the different amounts of base pairing that the toeholds of the RNA switches are involved in only the PLK1 switch "633" is

predicted to have a toehold that is not involved in base pairing (Figure S6) as opposed to the two other switches (Figures S4– S5).

Cell Culture Studies of CTGF-eGFP Switch.

We performed a variety of experiments to ascertain the functionality of the CTGF-eGFP switch in human breast cancer cells that express both CTGF and eGFP (MDA-MB-231 cells that stably express eGFP). First, we confirmed that the RNA switch leads in this cell line to the knockdown of eGFP (Figure 7a,b). This is expected, because the cells express the CTGF mRNA trigger. Next, we measured using flow cytometry eGFP fluorescence for cells that have first been transfected with anti-CTGF Dicer substrate RNA (DS RNA), followed by transfection with the RNA switch (Figures 7c and S10). Rewardingly, we find that the eGFP fluorescence signal is higher for these cells compared to cells that have only been transfected with the switch but not with anti-CTGF DS RNA. This indicates that the knockdown of CTGF via anti-CTGF DS RNA, and the thus resulting lack of CTGF mRNA trigger, leads to a higher amount of RNA switch that remains in its inactive conformation, resulting in higher eGFP expression. In comparison, if instead of the complete switch only the antitarget strand is provided, the eGFP expression changes only marginally with respect to antiCTGF DS RNA concentration (Figure 7c). This disproportionally high response of eGFP expression as a function of CTGF expression can that be viewed as a hallmark of an RNA logic gate. Furthermore, we verified that a modified RNA switch where that CTGF-mRNA binding toehold has been removed leads to reduced amounts of eGFP knockdown (higher eGFP fluorescence) (Figure S11). The differences in measured eGFP expression are statistically significant, but are relatively small. Also, the differences between the RNA switch and the no-toehold control become smaller as a function of time after switch transfection (compare Figure S11, days 1, 2, and 3). This is evidence that the RNA switch may undergo degradation. Computationally, we found that simulated strand breaks of the antitriple strand at the loop regions leads to dissociation of antitriple strand fragments and to the formation of a complex exhibiting active Dicer–substrate helix (Figure S12). This suggests that some switch degradation products may lead to a false positive knockdown of the target gene. We surmise that this property can be improved upon by utilizing nuclease-resistant modified nucleotides in parts of the RNA switch, in particular loop regions.

Even though the development of “smart drugs” that act only in the presence of potentially complex biochemical patterns is still in its infancy, the potential of this approach is evident. The idea not only goes beyond the concept of drugs that act in both diseased and nondiseased cells; it also goes beyond the concept of cell-surface receptor-based targeting, by potentially allowing the distinguishing of cells by the cytoplasmic milieu. Much progress has been made with reported designed and verified RNA-based scaffold structures such as rings and polyhedra.^{21,36,45–51} The challenge will now be to combine the advantages of static scaffolds (such as programmable multifunctionality as well as resistance to nucleotide degradation) with the advantages of molecular logic gates (molecular functions are only exercised in the presence of programmed biochemical environments).⁵² Taken together, we present a general approach for the computational prediction of potentially pseudoknotted nucleic acid complexes consisting of RNA and/or DNA strands. The algorithm suggests a kinetic interpretation of a plethora of RNA/DNA hybrid reassociation experiments. We

demonstrated the computational design and experimental verification of a two-stranded RNA switch that upon binding to a trigger RNA releases the programmed RNAi functionality that then leads to the silencing of a gene unrelated to the trigger. This approach has many potential uses, and may be an early proof-of-concept of a new generation of “smart drugs” that only act in the presence of certain cellular environments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Computational support by the NCI Advanced Biomedical Computing Center (ABCC) as well as the NIH Helix/BioWulf facility is highly appreciated. This work has been funded in whole or in part with Federal funds from the Frederick National Laboratory for Cancer Research, National Institutes of Health, under Contract No. HHSN261200800001E. This research was supported [in part] by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

REFERENCES

- (1). Ou-Yang SS; Lu JY; Kong XQ; Liang ZJ; Luo C; Jiang H *Acta Pharmacol. Sin* 2012, 33 (9), 1131–40. [PubMed: 22922346]
- (2). Bernardo PH; Tong JC *Methods Mol. Biol* 2012, 800, 25–31. [PubMed: 21964780]
- (3). Penchovsky R; Breaker RR *Nat. Biotechnol* 2005, 23 (11), 1424–33. [PubMed: 16244657]
- (4). Win MN; Smolke CD *Science* 2008, 322 (5900), 456–60. [PubMed: 18927397]
- (5). Liang JC; Chang AL; Kennedy AB; Smolke CD *Nucleic Acids Res.* 2012, 40 (20), e154. [PubMed: 22810204]
- (6). Green AA; Silver PA; Collins JJ; Yin P *Cell* 2014, 159 (4), 925–939. [PubMed: 25417166]
- (7). Wachsmuth M; Findeiss S; Weissheimer N; Stadler PF; Morl M *Nucleic Acids Res.* 2013, 41 (4), 2541–51. [PubMed: 23275562]
- (8). Muranaka N; Yokobayashi Y *Chem. Commun. (Cambridge, U. K.)* 2010, 46 (36), 6825–7.
- (9). Xie Z; Wroblewska L; Prochazka L; Weiss R; Benenson Y *Science* 2011, 333 (6047), 1307–11. [PubMed: 21885784]
- (10). Xie Z; Liu SJ; Bleris L; Benenson Y *Nucleic Acids Res.* 2010, 38 (8), 2692–701. [PubMed: 20194121]
- (11). Rinaudo K; Bleris L; Maddamsetti R; Subramanian S; Weiss R; Benenson Y *Nat. Biotechnol* 2007, 25 (7), 795–801. [PubMed: 17515909]
- (12). Cicha I; Goppelt-Strube M *BioFactors* 2009, 35 (2), 200–208. [PubMed: 19449449]
- (13). Pickles M; Leask AJ *Cell Commun. Signal* 2007, 1 (2), 85–90.
- (14). Kwon S; Munroe X; Crawley SC; Lee HY; Spong S; Bradham D; Gum JR, Jr.; Sleisenger MH; Kim YS *Int. J. Oncol* 2007, 31 (4), 693–703. [PubMed: 17786299]
- (15). Bennewith KL; Huang X; Ham CM; Graves EE; Erler JT; Kambham N; Feazell J; Yang GP; Koong A; Giaccia AJ *Cancer Res.* 2009, 69 (3), 775–84. [PubMed: 19179545]
- (16). Wang MY; Chen PS; Prakash E; Hsu HC; Huang HY; Lin MT; Chang KJ; Kuo ML *Cancer Res.* 2009, 69 (8), 3482–91. [PubMed: 19351859]
- (17). Chen PS; Wang MY; Wu SN; Su JL; Hong CC; Chuang SE; Chen MW; Hua KT; Wu YL; Cha ST; Babu MS; Chen CN; Lee PH; Chang KJ; Kuo ML *J. Cell Sci* 2007, 120 (12), 2053–2065. [PubMed: 17550972]
- (18). Lorenz R; Bernhart SH; Honer Zu Siederdissen C; Tafer H; Flamm C; Stadler PF; Hofacker IL *Algorithms Mol. Biol* 2011, 6, 26. [PubMed: 22115189]

- (19). Cao S; Xu X; Chen SJ RNA 2014, 20 (6), 835–45. [PubMed: 24751648]
- (20). Bernhart SH; Tafer H; Muckstein U; Flamm C; Stadler PF; Hofacker IL Algorithms Mol. Biol 2006, 1 (1), 3. [PubMed: 16722605]
- (21). Bindewald E; Afonin K; Jaeger L; Shapiro BA ACS Nano 2011, 5 (12), 9542–51. [PubMed: 22067111]
- (22). Andronescu M; Zhang ZC; Condon AJ Mol. Biol. 2005, 345 (5), 987–1001.
- (23). Dirks RM; Pierce NA J. Comput. Chem 2003, 24 (13), 1664–77. [PubMed: 12926009]
- (24). Zadeh JN; Steenberg CD; Bois JS; Wolfe BR; Pierce MB; Khan AR; Dirks RM; Pierce NA J. Comput. Chem 2011, 32 (1), 170–3. [PubMed: 20645303]
- (25). Afonin K; Desai R; Viard M; Kireeva M; Bindewald E; Case CL; Maciag AE; Kasprzak WK; Kim T; Sappe A; Stepler M; Kewal Ramani VN; Kashlev M; Blumenthal R; Shapiro BA Nucleic Acids Res. 2014, 42, 2085. [PubMed: 24194608]
- (26). Sugimoto N; Nakano S; Katoh M; Matsumura A; Nakamuta H; Ohmichi T; Yoneyama M; Sasaki M Biochemistry 1995, 34 (35), 11211–6. [PubMed: 7545436]
- (27). Mathews DH; Sabina J; Zuker M; Turner DH J. Mol. Biol 1999, 288 (5), 911–40. [PubMed: 10329189]
- (28). SantaLucia J, Jr. Proc. Natl. Acad. Sci. U. S. A 1998, 95 (4), 1460–5. [PubMed: 9465037]
- (29). Rose SD; Kim DH; Amarzguioui M; Heidel JD; Collingwood MA; Davis ME; Rossi JJ; Behlke MA Nucleic Acids Res. 2005, 33 (13), 4140–56. [PubMed: 16049023]
- (30). Afonin KA; Grabow WW; Walker FM; Bindewald E; Dobrovolskaia MA; Shapiro BA; Jaeger L Nat. Protoc 2011, 6 (12), 2022–2034. [PubMed: 22134126]
- (31). Bindewald E; Grunewald C; Boyle B; O'Connor M; Shapiro BA J. Mol. Graphics Modell 2008, 27 (3), 299–308.
- (32). Popena M; Szachniuk M; Antczak M; Purzycka KJ; Lukasiak P; Bartol N; Blazewicz J; Adamiak RW Nucleic Acids Res. 2012 40, (14), e112. [PubMed: 22539264]
- (33). Case DA; Cheatham TE, 3rd; Darden T; Gohlke H; Luo R; Merz KM, Jr.; Onufriev A; Simmerling C; Wang B; Woods RJ J. Comput. Chem 2005, 26 (16), 1668–88. [PubMed: 16200636]
- (34). Case DA; Darden TA; Cheatham TE; Simmerling CL; Wang J; Duke RE; Luo R; Walker RC; Zhang W; Merz KM; Roberts B; Hayik S; Roitberg A; Seabra G; Swails J; Götz AW; Kolossvary I; Wong KF; Paesani F; Vanicek J; Wolf RM; Liú J; Wu X; Brozell SR; Steinbrecher T; Gohlke H; Cai Q; Ye X; Wang J; Hsieh M-J; Cui G; Roe DR; Mathews DH; Seetin MG; Salomon-Ferrer R; Sagui C; Babin V; Luchko T; Gusarov S; Kovalenko A; Kollman PA AMBER 12; University of California: San Francisco, 2012.
- (35). Andronescu M; Bereg V; Hoos HH; Condon A BMC Bioinf. 2008, 9, 340.
- (36). Afonin KA; Lindsay B; Shapiro BA RNA Nanotechnol 2013, 1, 1–15.
- (37). Zhu X; Li T; Dang Y; Feng Y; Huang P Mol. Biotechnol 2005, 31 (3), 187–92. [PubMed: 16230768]
- (38). Afonin KA; Danilov EO; Novikova IV; Leontis NB ChemBioChem 2008, 9 (12), 1902–5. [PubMed: 18655086]
- (39). Afonin KA; Lin YP; Calkins ER; Jaeger L Nucleic Acids Res. 2012, 40 (5), 2168–80. [PubMed: 22080507]
- (40). Lai D; Ho KC; Hao Y; Yang X Cancer Res. 2011, 71 (7), 2728–38. [PubMed: 21349946]
- (41). Reeder J; Steffen P; Giegerich R Nucleic Acids Res. 2007, 35, W320–W324. [PubMed: 17478505]
- (42). Markham NR; Zuker M Methods Mol. Biol 2008, 453, 3–31. [PubMed: 18712296]
- (43). Milligan JF; Groebe DR; Witherell GW; Uhlenbeck OC Nucleic Acids Res. 1987, 15 (21), 8783–98. [PubMed: 3684574]
- (44). Lapham J; Crothers DM RNA 1996, 2 (3), 289–96. [PubMed: 8608452]
- (45). Afonin KA; Bindewald E; Yaghoubian AJ; Voss N; Jacovetty E; Shapiro BA; Jaeger L Nat. Nanotechnol 2010, 5 (9), 676–82. [PubMed: 20802494]

- (46). Grabow WW; Zakrevsky P; Afonin KA; Chworos A; Shapiro BA; Jaeger L Nano Lett. 2011, 11 (2), 878–87. [PubMed: 21229999]
- (47). Guo P; Sheng Wu X; Xue H; Sheng Wu Wu Yu; Bao LX Acta Biochim. Biophys. Sin. (Shanghai) 2002, 34 (5), 533–43.
- (48). Shu D; Shu Y; Haque F; Abdelmawla S; Guo P Nat. Nanotechnol 2011, 6 (10), 658–67. [PubMed: 21909084]
- (49). Xiao F; Zhang H; Guo P Nucleic Acids Res 2008, 36 (20), 6620–32. [PubMed: 18940870]
- (50). Chworos A; Severcan I; Koyfman AY; Weinkam P; Oroudjev E; Hansma HG; Jaeger L Science 2004, 306 (5704), 2068–72. [PubMed: 15604402]
- (51). Severcan I; Geary C; Chworos A; Voss N; Jacovetty E; Jaeger L Nat. Chem 2010, 2 (9), 772–9. [PubMed: 20729899]
- (52). Afonin KA; Viard M; Martins AN; Lockett SJ; Maciag AE; Freed EO; Heldman E; Jaeger L; Blumenthal R; Shapiro BA Nat. Nanotechnol 2013, 8 (4), 296–304. [PubMed: 23542902]

Figure 1.

Schematic (a) and 2D (b) representation of the interaction between the functional antitarget strand (red), an antitrigger strand (blue) and a trigger strand (yellow). The antitarget strand and the antitrigger strand form a duplex that represents a nonfunctional unit. The trigger mRNA strand (yellow) interacts with the RNA toehold (blue); this reaction leads to the formation of a duplex between the antitrigger strand and the trigger strand (“waste”) as well as a reformed antitarget strand that represents a functional Dicer substrate RNA (“functional unit”).

Figure 2.

Schematic representation of the secondary structure search space (a) and search algorithm (b). The example shown in (a) corresponds to a hypothetical secondary structure that consists of two fully extended helices. The folding status of each helix is represented by one integer number that can be 0 (helix is unfolded), 1 (helix is fully base paired), 2 or 3 (helix is halfway folded from one end). The completely unfolded secondary structure is for two hypothetical helices represented as “0,0” (first row), while the structure “1,1” (see red arrow) corresponds to a secondary structure with both helices fully base paired. This leads for the case of two helices to 16 possible structural alternatives. The search algorithm depicted in (b) consists of two priority queues where the lowest-free energy structures have the highest priority for exploring additional folding events. During one stage structures are removed from search queue 1 (the “emitting queue”), and multiple structures with additional helices are generated and placed into search queue 2 (the “receiving queue”). This process is continued until search queue 1 is empty, after which search queue 2 becomes the emitting queue and search queue 1 becomes the receiving queue. This process is continued until no more helices can be placed. The search queues are “leaky” in the sense that when a maximum number of containing structures is exceeded, the structures with the least favorable free energy are discarded. During the search process, the free energies of the found structures are analyzed in order to compute concentrations and in order to identify the structure with the lowest free energy.

Figure 3.

Schematic representation of the secondary structure model (a) and three-dimensional model (b) of the duplex between the antitarget strand (red) and the antitrigger strand (blue). A predicted interaction between the toehold region of the antitrigger and the 3' end of the antitarget strand (shown in Figure S3) is not part of the 3D model.

Figure 4. Box-whisker plots of the prediction quality (Matthews Correlation Coefficients) of HyperFold, pknotsRG, and UnaFold corresponding to a set of 110 pseudoknotted RNAs (top) and 741 nonpseudoknotted RNAs (bottom) obtained from the RNA Strand Database. The box plots group data corresponding to RNA lengths of 1–50, 51–100, 101–150, 151–200, and 201–250 nt.

Figure 5.

Scatter plot of experimentally measured fraction of reassociating RNA/DNA hybrid complexes (“Reassociation”) as a function of computationally predicted free energy of toehold binding. The size, color, and shape of the plotted symbols depict the toehold length, G+C content, and strand concentration, respectively. The plotted relative amounts of reassociation are means over experimental data values (for the same strand concentration and toehold length) as depicted also in Figure S2. The reason for the choice of predicted free energy of toehold binding as coordinate axis is that the initial duplex formation of cognate toeholds acts as transition state. As expected, the higher the free-energy barrier of the transition state (right side of diagram), the lower the reaction rate and (for a given fixed time point) the lower the relative amount of the reaction products (y -axis). An analysis of the kinetics of RNA-toehold formation has been performed in a related publication (Afonin et al. The use of minimal RNA toeholds to trigger the activation of multiple functionalities. Accepted to this Journal).

Figure 6.

Experimental verification of RNA switch formation and function by total RNA staining native-PAGE experiments. (a) Results demonstrate formation of a two-stranded switch and its further interaction with the mRNA fragment leading to the release of DS RNA containing shRNA-like structure (refolded antitarget strand). (b) Results for Dicer experiments carried out for switch and released antitarget strand respectively with recombinant Dicer using the human turbo Dicer enzyme kit (Genlantis). (c and d) Results show that the communication between the switch and mRNA is initiated through the toehold interaction. As negative controls, switches without a toehold (c) or with a mutated toehold (d) were tested. Red boxes indicate the bands corresponding to the released Dicer substrate RNA.

Figure 7.

RNA switch mediated gene silencing in human breast cancer cells. Enhanced green fluorescent protein (eGFP) knockdown assays were performed using MDA-MB-231 cells that stably express eGFP. Cells were transfected with the CTGF-eGFP RNA switch, or the individual antitarget strand or antitrigger strand that compose the switch. All transfections were performed at 50 nM [RNA]. Cells transfected with Lipofectamine 2000 only (no RNA) served as a negative control. Levels of eGFP expression were assessed 3 days post-transfection by (a) flow cytometry and (b) fluorescence microscopy. (c) Mean and standard error for flow cytometry fluorescence experiments for cells transfected with 0 nM and 5 nM CTGF siRNA 2 days prior to transfection with the RNA switch. Black: eGFP fluorescence for switch (50 nM); Red: eGFP fluorescence for antitarget strand (50 nM). One can see that the complete switch leads to a higher response in eGFP expression as a function of CTGF trigger knockdown.