ACCOUNTS of chemical research

pubs.acs.org/accounts Article

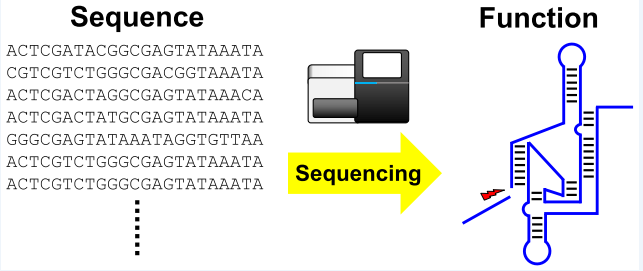
High-Throughput Analysis and Engineering of Ribozymes and Deoxyribozymes by Sequencing

Yohei Yokobayashi*

Cite This: *Acc. Chem. Res.* 2020, 53, 2903−2912 Read Online

ACCESS | Metrics & More | Article Recommendations

CONSPECTUS: Ribozymes and deoxyribozymes are catalytic RNA and DNA, respectively, that catalyze chemical reactions such as self-cleavage or ligation reactions. While some ribozymes are found in nature, a larger variety of ribozymes and deoxyribozymes have been discovered by in vitro selection from random sequences. These catalytic nucleic acids, especially ribozymes, are of fundamental interest because they are crucial for the RNA world hypothesis, which suggests that RNA played a central role in both the propagation of genetic information and catalyzing metabolic reactions in primordial life prior to the emergence of proteins and DNA. On the practical side, catalytic nucleic acids have been extensively engineered for various applications, such as biosensors and genetic devices for synthetic biology. Therefore, it is important to gain a deeper understanding of the sequence−function relationships of ribozymes and deoxyribozymes.



Mutational analysis, or measurements of activities of catalytic nucleic acid mutants, is one of the most fundamental approaches for that purpose. Mutations that abolish, reduce, retain, or even increase activity provide useful information about nucleic acid catalysts for engineering and other purposes. However, methods for mutational analysis of ribozymes and deoxyribozymes have not evolved much for decades, requiring tedious and low-throughput assays (e.g., gel electrophoresis) of individually prepared mutants. This has prevented researchers from performing quantitative mutational analysis of ribozymes and deoxyribozymes on a large scale.

To address this limitation, we developed a massively parallel ribozyme and deoxyribozyme assay strategy that allows >10^4 assays using high-throughput sequencing (HTS). We used HTS to literally count the number of cleaved (or ligated) and uncleaved (or unligated) ribozyme (or deoxyribozyme) sequences and calculated the activities of each mutant in a reaction mixture. This simple yet powerful strategy was applied to analyze the mutational effects of various natural and synthetic ribozymes and deoxyribozymes at scales impossible for conventional mutational analysis. These large-scale sequence−function data sets were used to better understand the functional consequences of mutations and to engineer ribozymes for practical applications. Furthermore, these newly available data are motivating researchers to employ more rigorous computational methods to extract additional insights such as structural information and nonlinear effects of multiple mutations. The new HTS-based assay strategy is distinct from and complementary to a related strategy that uses HTS to analyze ribozyme and deoxyribozyme populations subjected to in vitro selection. Postselection sequencing can cover a larger sequence space, although it does not directly quantify the activities of ribozyme and deoxyribozyme mutants. With further advances in DNA sequencing technologies and computational methods, there should be more opportunities to harness the power of HTS to deepen our understanding of catalytic nucleic acids and enhance our ability to engineer them for even more applications.

KEY REFERENCES

- Kobori, S.; Nomura, Y.; Miu, A.; Yokobayashi, Y. High-throughput assay and engineering of self-cleaving ribozymes by sequencing. *Nucleic Acids Res.* **2015**, *43*, e85.[1] High-throughput sequencing was used to assay in parallel >1000 self-cleaving ribozyme mutants to analyze a tertiary contact in a twister ribozyme and to identify allosteric ribozymes.
- Kobori, S.; Yokobayashi, Y. High-throughput mutational analysis of a twister ribozyme. *Angew. Chem., Int. Ed.* **2016**, *55*, 10354−10357.[2] Comprehensive mutational analysis of all single and double mutants within a 42-nt region of a twister ribozyme was performed, elucidating functionally relevant base−base interactions.

Received: August 27, 2020
Published: November 9, 2020

https://dx.doi.org/10.1021/acs.accounts.0c00546
Acc. Chem. Res. 2020, 53, 2903−2912

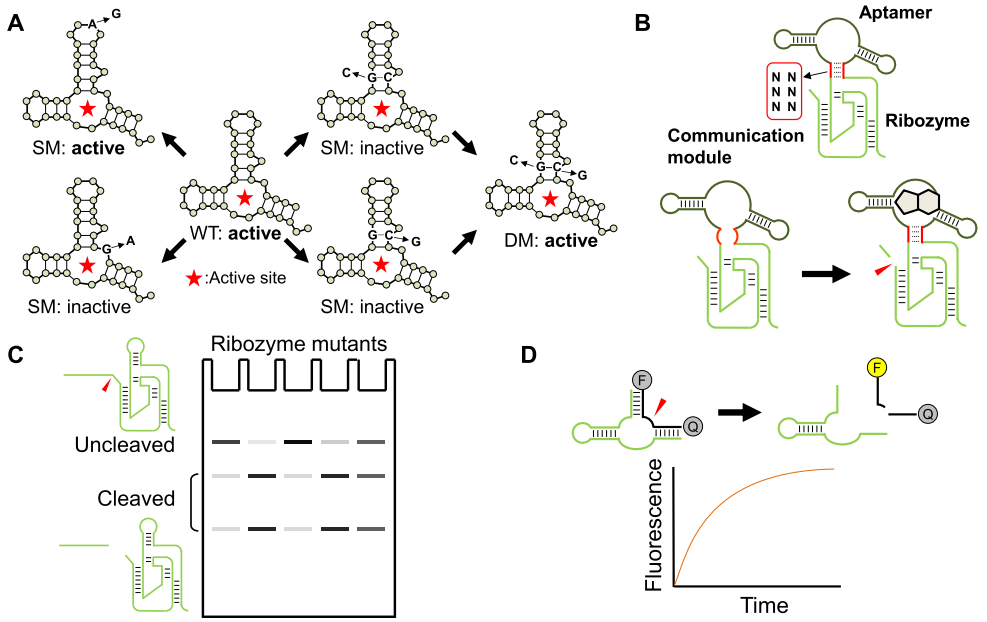


Figure 1. Ribozyme mutational analysis. (A) Activities of single mutants reveal critical and noncritical nucleotides. If a double mutant that combines two deleterious single mutations restores activity, a base−base interaction between the two positions is inferred. WT, wild-type; SM, single mutant; DM, double mutant. (B) Randomization of a communication module that connects an RNA aptamer with a ribozyme to search for allosteric ribozymes (aptazymes). (C) Gel electrophoresis assay of self-cleaving ribozymes. (D) Kinetic assay of ribozymes using a fluorescently labeled substrate. Fluorophore F is quenched by a quencher moiety Q in the substrate but becomes fluorescent upon cleavage.

- Dhamodharan, V.; Kobori, S.; Yokobayashi, Y. Large scale mutational and kinetic analysis of a self-hydrolyzing deoxyribozyme. *ACS Chem. Biol.* **2017**, *12*, 2940−2945.[3] Parallel kinetic assay of self-cleaving deoxyribozyme mutants yielded quantitative insights into nonlinear effects of multiple mutations.
- Nomura, Y.; Yokobayashi, Y. Systematic minimization of RNA ligase ribozyme through large-scale design-synthesis-sequence cycles. *Nucleic Acids Res.* **2019**, *47*, 8950−8960.[4] High-throughput sequencing and oligomer pool synthesis were used to explore the minimal catalytic core of an RNA ligase ribozyme.

1. INTRODUCTION

While storage and propagation of genetic information are the most well-known functions of nucleic acids, chemists and biologists have uncovered an impressive array of functions that DNA and RNA (and their synthetic analogs) exhibit in nature and the laboratory, including molecular recognition, self-assembly, computation, and gene regulation. Chemical catalysis is yet another known function of nucleic acids, making them arguably the most functionally versatile class of molecules in chemistry. The ability of nucleic acids, particularly RNA, to catalyze chemical reactions is also of fundamental importance because this suggests that RNA can both propagate genetic information and catalyze metabolic reactions that are necessary to maintain life in the absence of proteins.[5,6] While modern organisms exploit protein enzymes for almost all metabolic reactions, the fact that catalytic functions of ribosomes[7] (protein translation) and RNase P (tRNA processing)[8] are essentially executed by RNA and also the discovery of bona fide RNA that self-splices within an mRNA[9] strongly support the hypothesis that an RNA world preceded life as we know it today.[5,6] More recently, small self-cleaving RNAs have been found to be broadly distributed in the genomes of viruses and organisms in all kingdoms of life.[10,11]

However, nucleic acid catalysis in nature is limited to those functions mentioned above by RNA (ribozymes). Even if ribozymes once catalyzed other essential (or nonessential) metabolic reactions, their roles have long been taken over by protein enzymes, which are more efficient and robust. Therefore, exploration of the intrinsic potential of RNA to catalyze diverse chemical reactions has captivated researchers for decades since natural ribozymes were discovered. Such explorations are made possible by in vitro selection, which allows sequential enrichment of active sequences from a large pool of random RNA sequences through selection and amplification cycles.[12] This has led to discoveries of ribozymes that catalyze reactions that are potentially relevant to prebiotic metabolism, such as phosphorylation,[13] RNA ligation,[14] and aminoacylation.[15] Interestingly, in vitro selection yields diverse classes of catalytic DNAs (deoxyribozymes) that facilitate various reactions rivaling those catalyzed by ribozymes.[16] The realization that nucleic acids can be engineered de novo to catalyze various reactions has inspired researchers to design synthetic ribozymes and deoxyribozymes for new applications. For example, by combining chemical catalysis with molecular recognition by DNA and RNA (i.e., aptamers or hybridization with complementary sequences), ribozymes and deoxyribozymes have been engineered to function as biosensors with signal amplification capability.[17,18] Similarly, chemically regulated ribozymes have been used to control gene expression in living cells.[19,20]

Consequently, elucidating and understanding sequence−function relationships of catalytic nucleic acids have both fundamental and practical implications. Questions about the probability of RNA sequences exhibiting catalytic function and functional consequences of mutations in ribozymes have direct implications for the RNA world hypothesis. Engineering of

https://dx.doi.org/10.1021/acs.accounts.0c00546
Acc. Chem. Res. 2020, 53, 2903−2912

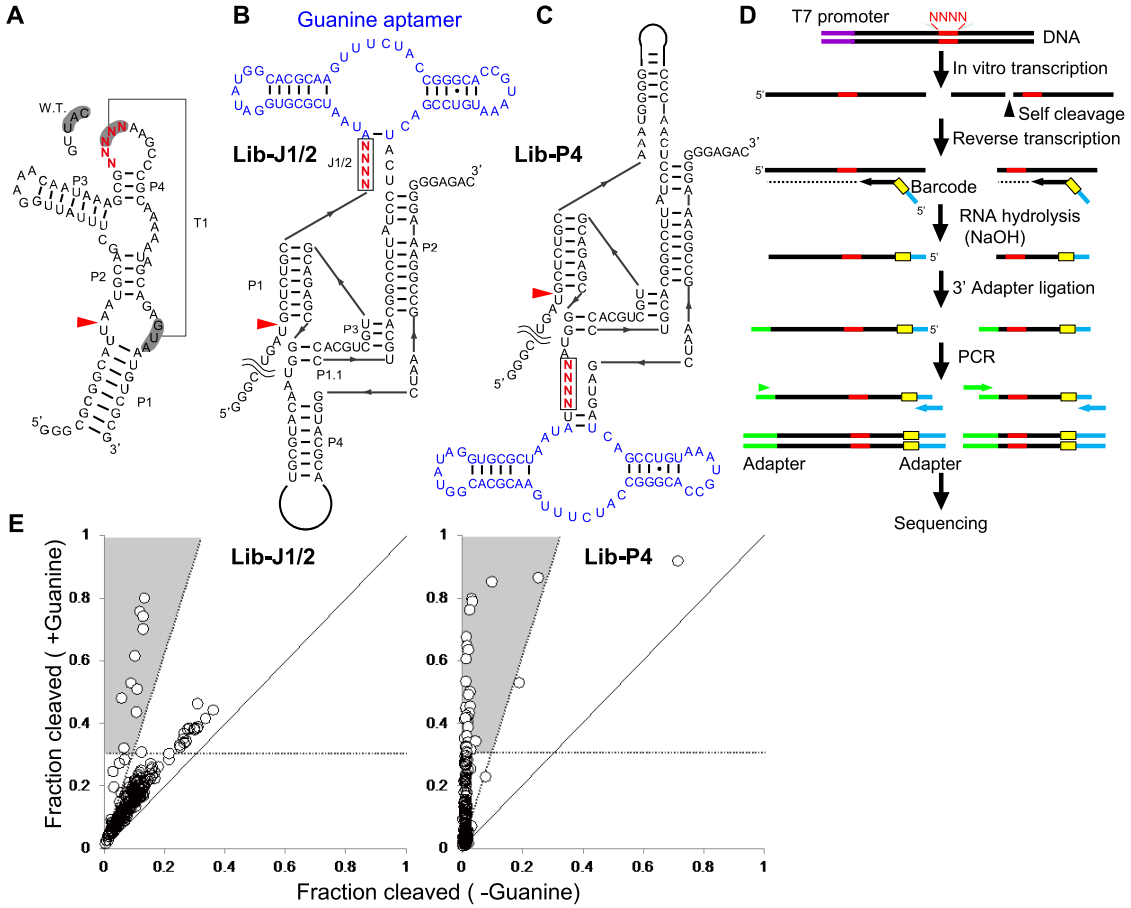


Figure 2. Ribozyme assay by HTS. (A) Twister ribozyme library in which five nucleotides (red) that include the pseudoknot T1 were randomized. (B) Guanine aptamer was inserted in the J1/2 junction of the HDV-like ribozyme drz-Agam-2-1 via a 4-nt (red) randomized stem (Lib-J1/2). (C) Guanine aptamer was inserted in the P4 stem of the ribozyme via a 4-nt randomized stem (Lib-P4). (D) Sequencing library construction process. The barcode sequence in the reverse transcription primer identified the library and the reaction conditions (±guanine). (E) Ribozyme activities of the aptazyme mutants in the presence and absence of guanine. The shaded regions indicate mutants that were activated >3-fold by guanine with FC > 0.30. Adapted with permission from ref 1. Copyright 2015 Oxford University Press.

nucleic acid enzymes would benefit from large-scale sequence−function relationship data that can lead to improved design strategies. Therefore, mutational analysis has always been an important part of both basic and applied research of ribozymes and deoxyribozymes. To understand how the sequence defines its function, the first instinct is to mutate it and measure its catalytic activity. Taking a ribozyme as an example, a researcher could mutate one nucleotide in the ribozyme suspected to play a critical role in catalysis. If the single mutation (e.g., in the active site) results in an inactive ribozyme, the nucleotide is likely to play a direct or an indirect role in the ribozyme function. Conversely, if the mutation results in an active ribozyme, the nucleotide is not likely to be directly involved in catalytic activity (Figure 1A).

Like proteins, catalytic nucleic acids fold into well-defined secondary and tertiary structures. Base−base interactions comprise the basic forces that define such structures. These functionally relevant base−base interactions in ribozymes can be inferred by mutational analysis. When mutations at two positions are deleterious as single mutations but the compensatory double mutation restores ribozyme activity, it strongly suggests that those two bases interact (Figure 1A). Therefore, the mutational analysis of single and double mutants provides basic knowledge about ribozyme structure and function.

Engineering of ribozymes and deoxyribozymes often requires a different type of mutational analysis. We and others have engineered allosteric self-cleaving ribozymes for applications in synthetic biology.[19,20] To design an allosteric ribozyme, a ribozyme is fused with an RNA aptamer (resulting in an "aptazyme") that recognizes a small molecule via a "communication module."[21] The rationale is that aptamer−ligand interaction induces a structural reorganization of the aptamer that is "communicated" to the ribozyme structure to either enhance or diminish ribozyme activity. In most cases, the aptamer and ribozyme sequences are fixed, and several defined positions within the communication module are randomized to identify optimal variants that most efficiently couple the two RNA elements (Figure 1B). Upon large-scale screening or selection of communication module mutants, the few sequences that exhibit the desired properties are analyzed.[22,23]

2. MEASURING RIBOZYME AND DEOXYRIBOZYME ACTIVITY BY SEQUENCING

2.1. Classical Biochemical Assay of Ribozymes and Deoxyribozymes

For decades, mutational analysis of ribozymes and deoxyribozymes has not evolved significantly. The most conventional method, especially when involving RNA cleavage or ligation, is

https://dx.doi.org/10.1021/acs.accounts.0c00546
Acc. Chem. Res. 2020, 53, 2903−2912

gel assay. A ribozyme and its mutants are individually reacted in separate test tubes, and the reacted and unreacted sequences are separated by gel electrophoresis (Figure 1C). For cleavage reactions, a fluorescently labeled substrate can be used to monitor a reaction continuously (Figure 1D), but this also requires individual reactions for each mutant. While these conventional assays, when sampled at multiple time points to make kinetic measurements, provide accurate kinetic parameters (k_{cat}, K_m, etc.) of ribozyme mutants, the number of mutants that can be tested is limited (~100) due to low throughput. Considering that a small ribozyme with 50 nt has 150 single mutants and 11025 double mutants, the choice of mutants that are tested is highly susceptible to researcher bias.

2.2. Combinatorial Mutational Assays of Ribozymes and Deoxyribozymes

Several important methods have been developed to comprehensively analyze single mutants of (deoxy)ribozymes and other functional RNAs. Strobel and Shetty randomly introduced a nucleotide analog with a phosphorothioate linkage during in vitro transcription of the *Tetrahymena* group I intron. After radioactively labeling active ribozyme mutants, the phosphorothioate linkage was selectively cleaved by iodine. Gel analysis of the ribozyme pool revealed the precise positions where a substitution by the nucleotide analog resulted in loss of activity.[24] This strategy of nucleotide analog interference mapping (NAIM) was adapted to deoxyribozymes by Höbartner and co-workers as dNAIM[25] and combinatorial mutation interference analysis (CoMA).[26] Meanwhile, the mutate-and-map strategy developed by the Das group, while not primarily used for ribozymes, provides detailed structural insights from selective 2′ hydroxyl acylation with primer extension (SHAPE) mapping of every complementary single mutant in the sequence.[27] The strategy does require laborious preparation and SHAPE assays of all single mutants to be tested. These combinatorial mutational assays, while highly useful, are still limited to providing indirect and qualitative information on the activities of single nucleotide mutants.

2.3. Parallel Assay of Ribozyme Activity by Sequencing

We sought to address the throughput limitation of ribozyme and deoxyribozyme activity assays by applying high-throughput sequencing (HTS) technology. As a demonstration, we first prepared three self-cleaving ribozyme libraries in which four to five consecutive nucleotides were randomized.[1] In one library, five nucleotides of a twister ribozyme[28] that included a tertiary pseudoknot contact were randomized (Figure 2A). In the other two (aptazyme) libraries, four nucleotides connecting a guanine aptamer[29] and a hepatitis delta virus (HDV)-like ribozyme[10] were randomized but with different positions at which their respective aptamer was fused with the ribozyme (Figure 2B,C). The three ribozyme libraries were prepared separately but sequenced in one session (Figure 2D). First, the ribozyme libraries were transcribed in vitro from their DNA templates. For the aptazyme libraries, separate reactions with and without guanine were prepared. The in vitro transcription products contained all library mutants, of which some were cleaved and some uncleaved. The RNA mixtures were then reverse-transcribed to yield cDNAs that were either long or short depending on the cleavage status of the template RNA. The primer used in the reverse transcription reaction also contained a barcode in the 5′ part of the sequence, which was used to indicate the presence or absence of guanine in the reaction, in addition to an adapter sequence for polymerase chain reaction (PCR) and sequencing. After ligating another adapter sequence to the 3′ end of the cDNAs, the libraries were amplified by PCR to yield samples for HTS. The assumption here is that each sequencing read yields the following information: the mutant sequence, whether the original RNA was cleaved or uncleaved, and (for aptazyme libraries) whether the reaction was performed with or without guanine. By counting the number of cleaved and uncleaved reads ($N_{cleaved}$ and $N_{uncleaved}$, respectively) for each mutant, one can calculate a "fraction cleaved" (FC) value for each mutant as follows: $FC = N_{cleaved}/(N_{cleaved} + N_{uncleaved})$. There are several caveats, which are discussed below, but this method allowed for an equivalent of 2048 assays (4^5 twister and $2 \times 2 \times 4^4$ each aptazyme library ± guanine) in one sequencing run.

The large and quantitative sequence−function data set obtained in one experiment can be analyzed to gain useful insights. The twister library shows that the randomized region is not highly tolerant to mutations, with only the wild-type and eight single mutants showing more than 30% cleavage. The ribozyme activities of the two aptazyme libraries each consisting of 256 mutants assayed in the presence and absence of guanine can be plotted as shown in Figure 2E to visualize the distribution of guanine-responsiveness (or lack thereof) among the mutants. Importantly, this method quantifies ribozyme activities (FC values) of all variants in the library whether or not they are functional. This contrasts with the conventional biomolecular selection or screening experiments in which sequence information can only be acquired for the functional variants ("hits"). In this experiment, we observe that there is optimal stability of the communication module that renders the aptazyme responsive to guanine.

This initial demonstration of a HTS-based ribozyme assay[1] leaves a few caveats. First, the FC values based on sequencing are subject to experimental bias that can be introduced during library preparation. For example, efficiencies of reverse transcription and adapter ligation reactions may differ between cleaved and uncleaved RNAs. Consequently, it is important to validate the FC values derived from HTS (FC_{HTS}) against those measured by conventional gel assay (FC_{Gel}). This can be done by sampling a few variants that span the observed sequencing-based FC values and measuring their activities by the gel-based assay individually. If there is a systematic bias, the FC_{HTS} vs FC_{Gel} plot can be used for calibration as long as there is a linear correlation.

Second, the FC values measured in this work represent single time-point ribozyme cleavage status in a cotranscriptional reaction. This means that the absolute FC values cannot be used to quantitatively differentiate ribozyme kinetics that occur significantly faster or slower than the sampled time point. Finally, the sequencing library preparation procedure is not applicable to ribozyme mutants if the cleavage site is located downstream of the mutated bases because the cDNA no longer carries the RNA cleavage status information. As discussed below, some of these limitations can be addressed by modifying the sequencing library preparation method. It is also possible to extend the sequencing-based assay strategy to (deoxy)-ribozymes with activities other than self-cleavage, as discussed in section 3.2.

Scalability is another issue. To calculate the FC value of a ribozyme sequence, the cleaved or uncleaved condition must be read multiple times. Assuming an average of 500 reads per variant, a typical Illumina MiSeq run that yields 2×10^7 reads can measure FC values of $\sim 4 \times 10^4$ variants. Using a current state-of-

https://dx.doi.org/10.1021/acs.accounts.0c00546
Acc. Chem. Res. 2020, 53, 2903−2912

the-art sequencer (e.g., NovaSeq 6000) with up to 10^{10} reads per run, the possible number of variants that can be assayed can be as many as ~2×10^7. This is enough to cover 12 randomized bases. Complexity of the library and the nature of mutants (e.g., local vs global mutations) to be analyzed must be adjusted for each experiment. Thus far, we have applied this method to assay an aptazyme library containing seven randomized positions (16 384 variants) measured in the presence and absence of guanine (32 768 assays).[30] In this work, we designed a new aptazyme architecture in which an aptamer and a ribozyme were fused in tandem such that the two RNA structures were mutually exclusive (Figure 3A). The key design criterion was the stability of the overlapping $P1_{Rz}$ and $P1_{Apt}$ stems. It was anticipated that when the relative stabilities of the two stems were optimal, $P1_{Rz}$ (active ribozyme) would form preferably in the absence of the aptamer ligand, but $P1_{Apt}$ (inactive ribozyme) would become favored in the presence of the ligand. By randomizing seven bases upstream of the aptamer that would constitute $P1_{Apt}$, we were able to comprehensively map the sequence−function relationship of the aptazyme architecture from low to high $P1_{Apt}$ stability. It was also revealed that the stability of $P1_{Apt}$ closer to the aptamer is important for aptazyme function.[30]

We also developed a new sequencing library preparation protocol that allowed for a ribozyme cleavage site downstream of the mutated nucleotides. After the ribozyme reaction, the cleaved and uncleaved RNA fragments were separated by gel electrophoresis and reverse-transcribed separately using a primer containing a barcode sequence to indicate if the original RNA template was cleaved or uncleaved (Figure 3B).[30] We also used the template-switching method to add the 3′ adapter sequence during reverse transcription.[31] These technical improvements significantly expand the scope of ribozyme libraries that can be analyzed by HTS.

2.4. Large-Scale Mutational Analysis

The examples discussed above analyzed mutants in a localized region of a ribozyme. Conventional mutational analysis of ribozymes, however, requires single and double mutations across all parts of a ribozyme to identify functionally critical nucleotides and base−base interactions. However, unbiased and exhaustive mutational analysis had been practically impossible due to the large number of possible mutants. HTS-based ribozyme assay allows complete coverage of single and double mutants for the first time. This was first demonstrated using a twister ribozyme found in the genome of rice (Osa-1−4) (Figure 4A).[2] This ribozyme is compact (54 nt) yet folds into a complex structure that includes three helices, two pseudoknots, and three loops.[32] Ten positions are conserved in more than 97% of the naturally found twister ribozyme variants.[28] We sought to comprehensively analyze the single and double mutants of Osa-1−4 in the 48 nucleotides downstream of the cleavage site (between U6 and A7). The mutant library was prepared by "doped oligo" synthesis in which the oligo-DNA was synthesized with 97% WT nucleotide and 1% each of the remaining nucleotides at each position. This resulted in statistically distributed mutants that include all single and double mutants, as well as some higher-order mutants (Figure 4B).

Analysis of the sequence−function data set of the wild-type and its 10296 mutants (Figure 4B) reveals some interesting insights.[2] For example, only four positions in the ribozyme were found where a mutation to any other base would result in >80% reduction in relative activity. In general, the ribozyme was found

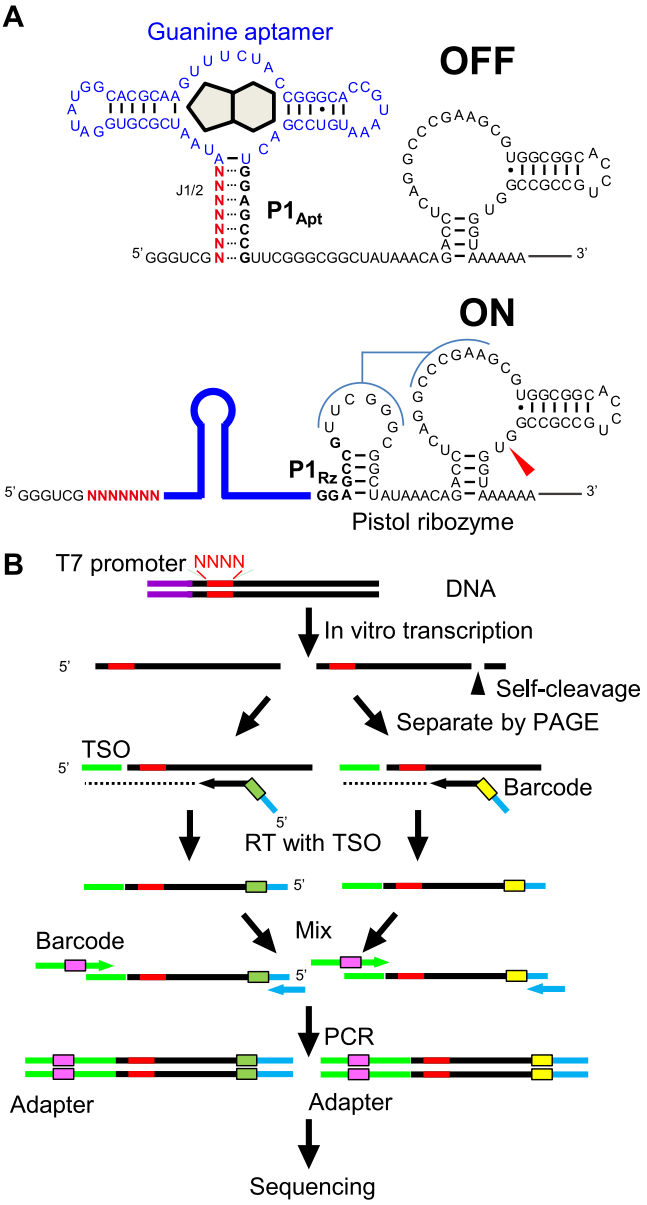


Figure 3. Tandem aptamer−ribozyme architecture for aptazymes. (A) A guanine aptamer (blue) was fused upstream of a pistol ribozyme, and a 7-nt randomized region (red) was placed upstream of the guanine aptamer. The aptamer and the ribozyme structures were mutually exclusive. When the relative stability of $P1_{Apt}$ and $P1_{Rz}$ was optimal, guanine binding triggered the aptamer structure, thereby inhibiting ribozyme cleavage. (B) Sequencing library preparation steps. The barcode inserted during reverse transcription (RT) encodes cleaved/uncleaved status, and the second barcode introduced during PCR encodes presence/absence of guanine. A 3′ adapter sequence was added during reverse transcription via template-switching process.[31] TSO, template-switching oligonucleotide. Adapted with permission from ref 30. Copyright 2017 American Chemical Society.

to be relatively robust against single and double mutations, while the sensitivity to mutations varied among the different structural elements. Single and double mutants within the two pseudoknots showed epistasis in which single mutants that disrupted the pseudoknot contact were deleterious but double mutants that restored the contact regained activity (Figures 1A and 4B). We believe this is the first comprehensive and quantitative mutational analysis of a nucleic acid enzyme.

https://dx.doi.org/10.1021/acs.accounts.0c00546
Acc. Chem. Res. 2020, 53, 2903−2912

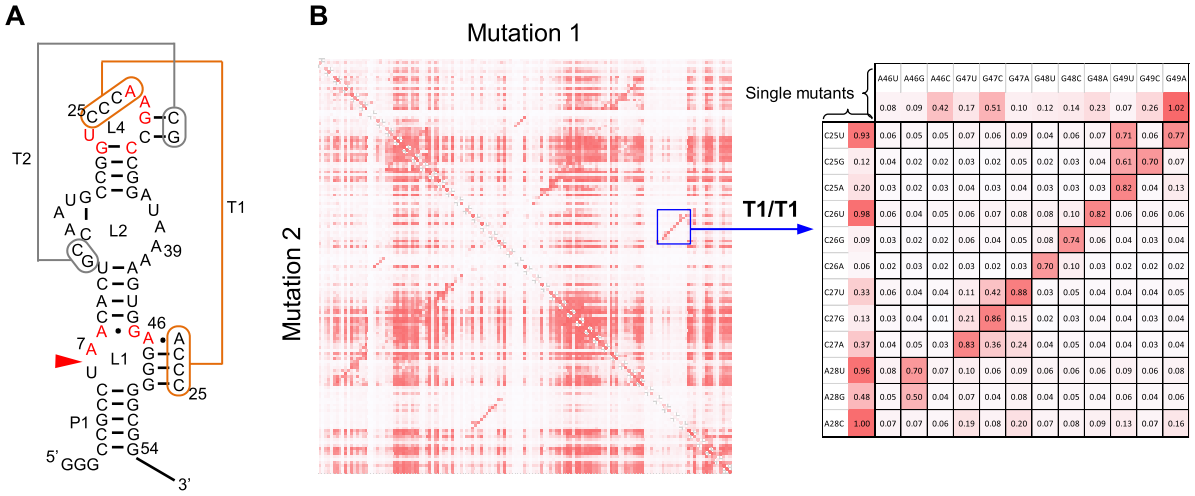


Figure 4. Mutational analysis of a twister ribozyme. (A) Secondary structure of the ribozyme based on the crystal structure by Liu et al.[32] The red nucleotides indicate positions that are conserved in 97% of the twister ribozyme family. (B) Relative activity (to that of the wild-type) of single and double mutants measured by HTS. The diagonal region (from upper left to lower right) represents single mutants. The region that represents the T1 pseudoknot is magnified for closer inspection. Color intensity and numbers indicate relative activities to the wild-type. Adapted with permission from ref 2. Copyright 2016 Wiley.

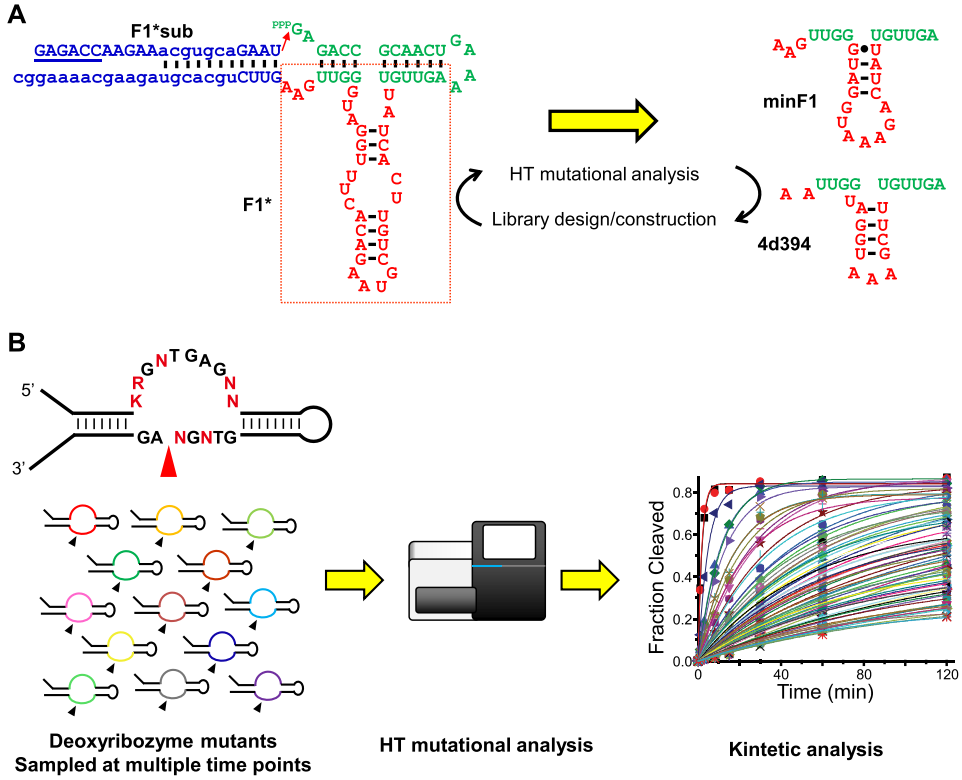


Figure 5. Applications of mutational analysis by HTS. (A) An RNA ligase ribozyme was analyzed by HTS to minimize its catalytic core. Adapted with permission from ref 4. Copyright 2019 Oxford University Press. (B) Zn^{2+}-dependent DNA self-hydrolase was assayed at multiple time points to obtain kinetic rate constants for 533 variants. Adapted with permission from ref 3. Copyright 2017 American Chemical Society.

Mutational analysis by HTS was expanded to an in vitro-selected RNA ligase ribozyme (Figure 5A).[4] This artificial ribozyme discovered by the Joyce group[33] catalyzes the native 5′−3′ ligation of two RNA fragments in a template-directed manner, with one fragment activated by a 5′-triphosphate group. The ligation chemistry is chemically equivalent to that of the natural RNA polymerases. Consequently, this and other RNA ligase ribozymes are of interest for the RNA world hypothesis. In fact, Joyce and co-workers have engineered the RNA ligase ribozyme to construct self-replicating systems that display exponential replication and genetic selection.[33,34]

The application of HTS-based mutational analysis to the RNA ligase ribozyme was straightforward.[4] Gel purification of the ligated and the unligated fragment and subsequent barcoding during reverse transcription allowed comprehensive mutational analysis of all single and double mutants of the catalytic core. We were also able to analyze deletion mutants by exploiting custom on-chip oligo pool synthesis that provided a

https://dx.doi.org/10.1021/acs.accounts.0c00546
Acc. Chem. Res. 2020, 53, 2903−2912

mixture composed of an arbitrary set of DNA sequences. Starting with the F1* ligase that was based on the sequence optimized by Robertson and Joyce,[33] the first round of comprehensive mutational analysis led us to a significantly downsized catalytic core of minF1 (Figure 5A). While the observed rate constant (k_{obs}) was reduced by ∼15-fold compared to F1*, minF1 was still quite active with k_{obs} = 0.48 min^{-1}. Another round of HTS-based mutational analysis identified a further minimized 18-nt catalytic core (4d394) that showed slower (k_{obs} = 0.49 h^{-1}) but clearly detectable activity (Figure 5A). Identification of minimal catalytic motifs is an important goal in the context of the RNA world research because smaller (but not necessarily more active) functional RNA motifs are more likely to have emerged compared to large and complex motifs. The expected low fidelity of the primitive replication machinery would also likely limit the size of primordial ribozymes. While in vitro selection is a powerful strategy for de novo discovery of catalytic nucleotide sequences, minimization of an existing (natural or synthetic) catalytic motif by selection is challenging due to the strong bias in selection for higher activity. The HTS-based ribozyme assay has no such bias and therefore is well suited for studying minimal catalytic motifs.

While the examples discussed above provided large sequence−function data sets of ribozymes, the activity values obtained were those measured at one time point. In principle, more quantitative information, such as observed rate constants (k_{obs}), can be acquired by HTS if the measurements are taken at multiple time points. We performed such high-throughput kinetic measurements[3] using a self-hydrolyzing deoxyribozyme originally discovered by the Breaker group.[35] I-R3 is a small Zn^{2+}-dependent deoxyribozyme that contains a catalytic loop flanked by two stems. Initially, the catalytic core (loop) was statistically mutated by a synthetically doped oligonucleotide to analyze all single and double mutations upstream of the cleavage site (15-nt). This mutational analysis resulted in a consensus sequence of the active variants as shown in Figure 5B. A second library was constructed according to this consensus sequence that included 4096 variants. Using this consensus library, the HTS-based deoxyribozyme assay was performed at multiple time points up to 2 h. The results allowed us to calculate the rate constants (k_{obs}) for 533 variants that showed >20% cleavage after 2 h. Rate constants provide more quantitative insights into the deoxyribozyme fitness landscape compared to catalytic activities measured at one time point. Using the Arrhenius equation, the activation energy of the reaction of each variant can be calculated, and the epistatic effects (if any) of multiple mutations can be assessed quantitatively. We analyzed epistatic effects of 105 double mutants whose k_{obs} were measurable and found relatively few cases where the two single mutations showed significant epistasis.[3]

2.5. Applications

In addition to providing rich sequence−function data sets that can deepen our understanding of how sequence variation affects the catalytic activity of nucleic acid enzymes, the HTS-based ribozyme and deoxyribozyme assays can be a valuable method for some applications. One of our motivations to develop this method was to engineer synthetic ribozymes to control gene expression in mammalian cells. Inserting an active self-cleaving ribozyme in the 3′ untranslated region (UTR) of an mRNA has been known to downregulate protein expression because the poly(A) tail is detached.[20] Controlling the ribozyme activity by a small molecule that binds to an RNA aptamer further allows chemical regulation of gene expression.

We randomized 3-bp proximal to the ribozyme core of the P4 stem in a HDV-like ribozyme to analyze the effect of the stem stability on the ribozyme activity (Figure 6).[36] We observed a

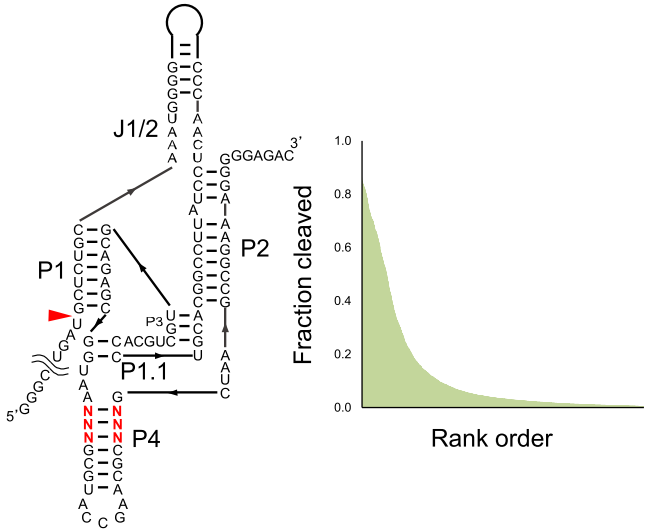


Figure 6. Varying the P4 stem stability to tune the ribozyme activity of an HDV-like ribozyme (drz-Agam-2-1). Adapted with permission from ref 36. Copyright 2018 American Chemical Society.

correlation between the base-pairing pattern within the mutated positions and the ribozyme activity, resulting in an extensive list of ribozyme mutants with very low to very high activity. It was then shown that, by selecting various mutants with different levels of activity and inserting them into the mammalian mRNA, we could tune the protein expression level in the cells without altering the promoter sequence.[36] Consequently, these ribozyme variants can collectively serve as a compact and interchangeable RNA device that can be used to modulate gene expression levels in mammalian cells.

Similarly, in the study of a HDV-like ribozyme fused with a guanine aptamer (aptazyme) described in section 2.2,[1] some of the guanine-activated ribozymes were inserted in the 3′-UTR of an EGFP-coding mRNA. The functional aptazymes identified by the HTS-based assay worked as mammalian riboswitches where the addition of guanine in the cell culture medium resulted in repression of EGFP expression.

It has been known that ribozyme activity in various cells is not easily predictable or is not highly correlated to in vitro activity.[37] This is understandable considering the differences in cellular environments in which ribozymes evolved. It is also possible for surrounding sequences to interfere with ribozyme folding and function. Therefore, we adapted the HTS-based assay to screen and identify self-cleaving ribozyme variants that are active in mammalian cells (Figure 7).[38]

To exploit the diversity of natural ribozymes discovered in the sequenced genomes, we synthesized 375 variants of pistol ribozymes computationally discovered by the Breaker group.[11] In addition, we generated 2625 synthetic variants by shuffling the structural elements (e.g., stems, pseudoknots) of natural pistol ribozymes. DNA templates encoding these natural and synthetic pistol ribozymes were transfected into mammalian cells, and the RNAs corresponding to the cleaved and uncleaved fragments were extracted and sequenced (Figure 7). We

https://dx.doi.org/10.1021/acs.accounts.0c00546
Acc. Chem. Res. 2020, 53, 2903−2912

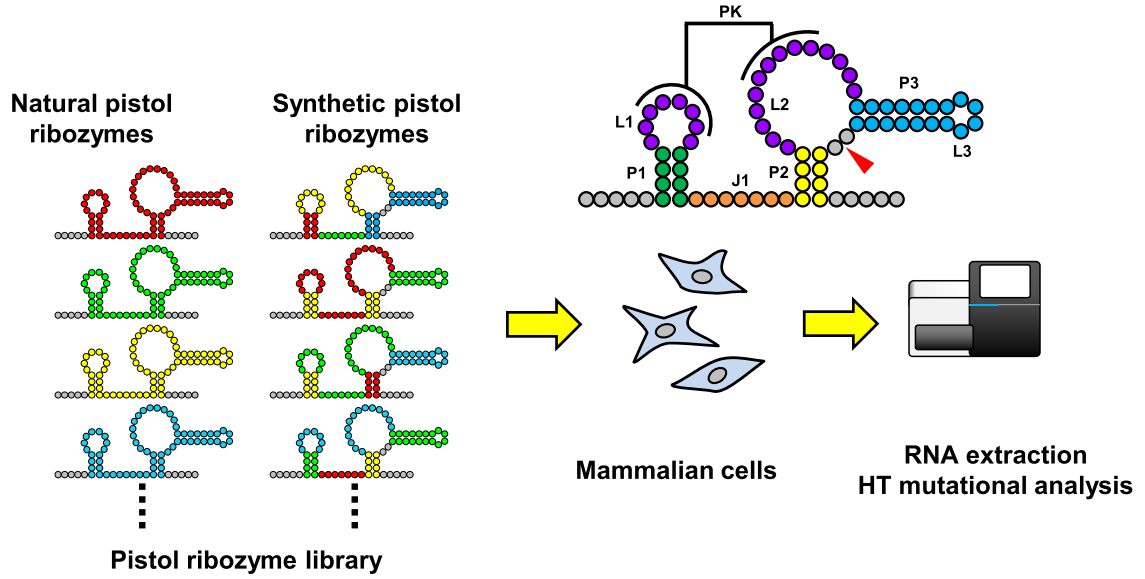


Figure 7. Discovery of ribozymes active in mammalian cells. DNA templates encoding natural and synthetic variants of the pistol ribozyme class were transfected into mammalian cells. The transcribed RNAs were extracted, converted to sequencing templates, and analyzed by HTS for intracellular activity. Adapted with permission from ref 38. Copyright 2017 Royal Society of Chemistry.

identified 13 ribozymes that were highly active in mammalian cells, which were among the first examples of pistol ribozymes demonstrated to be functional in mammalian cells.[38] Interestingly, eight out of the top 10 sequences were synthetic variants, which suggests that shuffling structural elements may be a useful strategy to expand the repertoire of ribozyme sequences.

2.6. Postselection Sequencing

Ribozymes (and deoxyribozymes) are amenable to in vitro selection.[12] A small fraction of (more) active ribozyme variants can be enriched from a large pool of inactive or less active variants, for example, by size selection or selective capturing of active variants on solid phase. Therefore, HTS analysis of the ribozyme population before and after selection can elucidate active sequences and their relative activities. It can be expected that more active sequences are enriched more efficiently than those that are less active. Pitt and Ferré-D'Amaré were the first to apply HTS to analyze ribozyme mutants subjected to in vitro selection to gain broader insights into the fitness landscape of ribozymes.[39] They heavily mutagenized (21% per position in 45 nt) the class II RNA ligase ribozyme and enriched the active mutants using immobilized substrate before analyzing by HTS. More recently, Pressman et al. analyzed almost all sequences within a 21-nt region of a self-aminoacylating ribozyme by combining in vitro selection and kinetic sequencing.[40] Such mutate-select-and-sequence experiments can provide a broader overview of the fitness landscape surrounding an active ribozyme sequence compared to our mutate-and-sequence strategy. On the other hand, the mutate-and-sequence approach offers much more quantitative sequence−function data sets for smaller but significantly large (potentially up to $\sim 10^7$) populations of arbitrary variants. Therefore, the two strategies are complementary in terms of the sequence space that can be explored and the quantitative nature of the acquired data.[41]

3. NEW DIRECTIONS

3.1. Data Analysis

HTS-based ribozyme assay can provide quantitative sequence−function data sets for potentially up to $\sim 10^7$ desired sequences. In our work on the mutational analysis of a twister ribozyme,[2] we made some qualitative observations based on visual inspection of the relative activities of single and double mutants. However, more rigorous statistical and computational tools can be applied to extract additional information on the local fitness landscapes of ribozymes from such sequence−function data sets. The Hayden group analyzed our twister ribozyme data[2] (along with other noncoding RNAs) and observed that negative epistasis predominated the fitness landscape.[42] Recently, Zhang et al. and Rollins et al. used our twister ribozyme data[2] (along with other RNA and protein mutational scan data) as an example to demonstrate their computational strategies to infer secondary and tertiary structures from large-scale mutational analysis data.[43,44] Further improvements in both HTS-based ribozyme and deoxyribozyme assays (e.g., higher sensitivity and accuracy) and computational analysis methods should enhance our ability to estimate the structures of catalytic nucleic acids without crystallography or NMR.

3.2. Other Ribozyme and Deoxyribozyme Activities

We have only applied the HTS-based assay to a handful of ribozymes and deoxyribozymes with self-cleavage and ligation activities. Of these, all of them were ribozymes except for one deoxyribozyme with self-hydrolase activity. The method, however, is not limited to nucleic acid catalysts with self-cleavage or ligation activities. If the ribozyme (or deoxyribozyme) is self-modified with a substrate and the reacted and unreacted sequences can be separated (e.g., by gel electrophoresis), it is possible to apply the HTS-based assay in a manner similar to what we demonstrated previously. Since most in vitro-evolved ribozymes and deoxyribozymes were selected via self-modification with a substrate, many catalytic nucleic acids should be amenable to sequencing-based assay. Sequence−function relationships of ribozymes and deoxyribozymes with diverse catalytic activities should deepen our understanding of the effects of mutations in catalytic nucleic acids.

https://dx.doi.org/10.1021/acs.accounts.0c00546
Acc. Chem. Res. 2020, 53, 2903−2912

3.3. In Vivo Ribozymes for Riboswitch Applications

Engineered self-cleaving ribozymes that respond to small molecules (aptazymes) are emerging as useful tools for controlling gene expression, for example, as riboswitches in eukaryotic cells.[19,20] Such engineering endeavors often depend on high-throughput screening of a large number of variants to identify those with optimal switching capability. While in vitro sequencing-based ribozyme assay has been used to generate functional riboswitches in mammalian cells,[1] it has been observed that ribozyme activities in vitro and in bacteria and yeast often do not translate to activities in mammalian cells[37] in which high-throughput riboswitch assay is technically challenging. Recently, two groups independently reported high-throughput riboswitch screening based on RNA sequencing of cells transfected with riboswitch (aptazyme) libraries.[45,46] In these strategies, HTS is used to measure the depletion of active ribozyme sequences in living cells due to the rapid degradation of the cleaved mRNAs. This RNA-seq approach can be a more direct and useful tool for engineering functional ribozymes in living cells.

4. CONCLUSIONS

As was the case with Sanger sequencing, chemistry has played a major role in the development of new technologies to sequence massive numbers of DNA sequences that have in turn transformed biology and medicine, enabling researchers to precisely map genomes, quantitatively analyze transcriptomes, and discover genotype-to-phenotype relationships. Perhaps chemists deserve to reap some benefit of the technologies they have helped to develop. In this Account, we have briefly summarized our recent efforts to harness the power of HTS to measure ribozyme and deoxyribozyme activities at scales that are impossible by conventional methods that have been in use for decades. HTS has also been creatively adapted for high-throughput binding assay of RNA aptamers,[47−50] further demonstrating the potential of sequencing technologies to yield information relevant to chemistry. The ability to generate large data sets that were previously unavailable generates new opportunities to explore the sequence−function relationships of nucleic acid catalysts using various computational tools. With rapid progress in sequencing technologies, we can expect more opportunities to harness the power of HTS to advance nucleic acid chemistry.

AUTHOR INFORMATION

Corresponding Author

Yohei Yokobayashi − *Nucleic Acid Chemistry and Engineering Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan;* orcid.org/0000-0002-2417-1934; Email: yohei.yokobayashi@oist.jp

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.accounts.0c00546

Notes

The author declares no competing financial interest.

Biography

Yohei Yokobayashi is a Professor heading the Nucleic Acid Chemistry and Engineering Unit at Okinawa Institute of Science and Technology Graduate School (OIST). After receiving B.Eng. and M.Eng. degrees from the University of Tokyo, he completed his Ph.D. at The Scripps Research Institute under the guidance of Prof. M. Reza Ghadiri in 2001 working on self-replicating peptide networks. He then received postdoctoral training under Prof. Frances H. Arnold at Caltech working on directed evolution of genetic switches. After appointment as a faculty member at University of California, Davis, in 2004, he has been focusing on engineering RNA for synthetic biology applications. In 2015, he moved his group to OIST to expand the research program to include functional nucleic acids other than RNA.

ACKNOWLEDGMENTS

The author thanks the current and past members of the research group who have contributed to the work described in this Account. Pei Chen King provided editorial assistance. OIST DNA Sequencing Section is also acknowledged with gratitude for its support in HTS experiments. Research funding was provided by OIST and Japan Society for the Promotion of Science KAKENHI (18K19944 and 19H02855).

REFERENCES

(1) Kobori, S.; Nomura, Y.; Miu, A.; Yokobayashi, Y. High-throughput assay and engineering of self-cleaving ribozymes by sequencing. *Nucleic Acids Res.* **2015**, *43*, e85.
(2) Kobori, S.; Yokobayashi, Y. High-Throughput Mutational Analysis of a Twister Ribozyme. *Angew. Chem., Int. Ed.* **2016**, *55*, 10354−10357.
(3) Dhamodharan, V.; Kobori, S.; Yokobayashi, Y. Large Scale Mutational and Kinetic Analysis of a Self-Hydrolyzing Deoxyribozyme. *ACS Chem. Biol.* **2017**, *12*, 2940−2945.
(4) Nomura, Y.; Yokobayashi, Y. Systematic minimization of RNA ligase ribozyme through large-scale design-synthesis-sequence cycles. *Nucleic Acids Res.* **2019**, *47*, 8950−8960.
(5) Chen, X.; Li, N.; Ellington, A. D. Ribozyme Catalysis of Metabolism in the RNA World. *Chem. Biodiversity* **2007**, *4*, 633−655.
(6) Pressman, A.; Blanco, C.; Chen, I. A. The RNA World as a Model System to Study the Origin of Life. *Curr. Biol.* **2015**, *25*, R953−963.
(7) Cech, T. R. The ribosome is a ribozyme. *Science* **2000**, *289*, 878−879.
(8) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **1983**, *35*, 849−857.
(9) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **1982**, *31*, 147−157.
(10) Webb, C. H.; Riccitelli, N. J.; Ruminski, D. J.; Lupták, A. Widespread occurrence of self-cleaving ribozymes. *Science* **2009**, *326*, 953.
(11) Weinberg, Z.; Kim, P. B.; Chen, T. H.; Li, S.; Harris, K. A.; Lünse, C. E.; Breaker, R. R. New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nat. Chem. Biol.* **2015**, *11*, 606−610.
(12) Joyce, G. F. Forty Years of In Vitro Evolution. *Angew. Chem., Int. Ed.* **2007**, *46*, 6420−6436.
(13) Lorsch, J. R.; Szostak, J. W. In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature* **1994**, *371*, 31−36.
(14) Bartel, D. P.; Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences. *Science* **1993**, *261*, 1411−1418.
(15) Morimoto, J.; Hayashi, Y.; Iwasaki, K.; Suga, H. Flexizymes: Their Evolutionary History and the Origin of Catalytic Function. *Acc. Chem. Res.* **2011**, *44*, 1359−1368.
(16) Silverman, S. K. Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes. *Trends Biochem. Sci.* **2016**, *41*, 595−609.
(17) Rossetti, M.; Porchetta, A. Allosterically regulated DNA-based switches: From design to bioanalytical applications. *Anal. Chim. Acta* **2018**, *1012*, 30−41.
(18) Vinkenborg, J. L.; Karnowski, N.; Famulok, M. Aptamers for allosteric regulation. *Nat. Chem. Biol.* **2011**, *7*, 519−527.

https://dx.doi.org/10.1021/acs.accounts.0c00546
Acc. Chem. Res. 2020, 53, 2903−2912

(19) Berens, C.; Groher, F.; Suess, B. RNA aptamers as genetic control devices: the potential of riboswitches as synthetic elements for regulating gene expression. *Biotechnol. J.* **2015**, *10*, 246−257.
(20) Yokobayashi, Y. Aptamer-based and aptazyme-based riboswitches in mammalian cells. *Curr. Opin. Chem. Biol.* **2019**, *52*, 72−78.
(21) Silverman, S. K. Rube Goldberg goes (ribo)nuclear? Molecular switches and sensors made from RNA. *RNA* **2003**, *9*, 377−383.
(22) Felletti, M.; Stifel, J.; Wurmthaler, L. A.; Geiger, S.; Hartig, J. S. Twister ribozymes as highly versatile expression platforms for artificial riboswitches. *Nat. Commun.* **2016**, *7*, 12834.
(23) Nomura, Y.; Zhou, L.; Miu, A.; Yokobayashi, Y. Controlling Mammalian Gene Expression by Allosteric Hepatitis Delta Virus Ribozymes. *ACS Synth. Biol.* **2013**, *2*, 684−689.
(24) Strobel, S. A.; Shetty, K. Defining the chemical groups essential for Tetrahymena group I intron function by nucleotide analog interference mapping. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2903−2908.
(25) Wachowius, F.; Höbartner, C. Probing Essential Nucleobase Functional Groups in Aptamers and Deoxyribozymes by Nucleotide Analogue Interference Mapping of DNA. *J. Am. Chem. Soc.* **2011**, *133*, 14888−14891.
(26) Wachowius, F.; Javadi-Zarnaghi, F.; Höbartner, C. Combinatorial Mutation Interference Analysis Reveals Functional Nucleotides Required for DNA Catalysis. *Angew. Chem., Int. Ed.* **2010**, *49*, 8504−8508.
(27) Kladwang, W.; VanLang, C. C.; Cordero, P.; Das, R. A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat. Chem.* **2011**, *3*, 954−962.
(28) Roth, A.; Weinberg, Z.; Chen, A. G.; Kim, P. B.; Ames, T. D.; Breaker, R. R. A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat. Chem. Biol.* **2014**, *10*, 56−60.
(29) Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C.; Breaker, R. R. Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell* **2003**, *113*, 577−586.
(30) Kobori, S.; Takahashi, K.; Yokobayashi, Y. Deep Sequencing Analysis of Aptazyme Variants Based on a Pistol Ribozyme. *ACS Synth. Biol.* **2017**, *6*, 1283−1288.
(31) Zhu, Y. Y.; Machleder, E. M.; Chenchik, A.; Li, R.; Siebert, P. D. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *BioTechniques* **2001**, *30*, 892−897.
(32) Liu, Y.; Wilson, T. J.; McPhee, S. A.; Lilley, D. M. Crystal structure and mechanistic investigation of the twister ribozyme. *Nat. Chem. Biol.* **2014**, *10*, 739−744.
(33) Robertson, M. P.; Joyce, G. F. Highly Efficient Self-Replicating RNA Enzymes. *Chem. Biol.* **2014**, *21*, 238−245.
(34) Sczepanski, J. T.; Joyce, G. F. Synthetic Evolving Systems that Implement a User-Specified Genetic Code of Arbitrary Design. *Chem. Biol.* **2012**, *19*, 1324−1332.
(35) Gu, H.; Furukawa, K.; Weinberg, Z.; Berenson, D. F.; Breaker, R. R. Small, Highly Active DNAs That Hydrolyze DNA. *J. Am. Chem. Soc.* **2013**, *135*, 9121−9129.
(36) Kobori, S.; Yokobayashi, Y. Analyzing and Tuning Ribozyme Activity by Deep Sequencing To Modulate Gene Expression Level in Mammalian Cells. *ACS Synth. Biol.* **2018**, *7*, 371−376.
(37) Wurmthaler, L. A.; Klauser, B.; Hartig, J. S. Highly motif- and organism-dependent effects of naturally occurring hammerhead ribozyme sequences on gene expression. *RNA Biol.* **2018**, *15*, 231−241.
(38) Nomura, Y.; Chien, H.-C.; Yokobayashi, Y. Direct screening for ribozyme activity in mammalian cells. *Chem. Commun.* **2017**, *53*, 12540−12543.
(39) Pitt, J. N.; Ferré-D'Amaré, A. R. Rapid construction of empirical RNA fitness landscapes. *Science* **2010**, *330*, 376−379.
(40) Pressman, A. D.; Liu, Z.; Janzen, E.; Blanco, C.; Müller, U. F.; Joyce, G. F.; Pascal, R.; Chen, I. A. Mapping a Systematic Ribozyme Fitness Landscape Reveals a Frustrated Evolutionary Network for Self-Aminoacylating RNA. *J. Am. Chem. Soc.* **2019**, *141*, 6213−6223.
(41) Yokobayashi, Y. Applications of high-throughput sequencing to analyze and engineer ribozymes. *Methods* **2019**, *161*, 41−45.
(42) Bendixsen, D. P.; Østman, B.; Hayden, E. J. Negative Epistasis in Experimental RNA Fitness Landscapes. *J. Mol. Evol.* **2017**, *85*, 159−168.
(43) Rollins, N. J.; Brock, K. P.; Poelwijk, F. J.; Stiffler, M. A.; Gauthier, N. P.; Sander, C.; Marks, D. S. Inferring protein 3D structure from deep mutation scans. *Nat. Genet.* **2019**, *51*, 1170−1176.
(44) Zhang, Z.; Xiong, P.; Zhang, T.; Wang, J.; Zhan, J.; Zhou, Y. Accurate inference of the full base-pairing structure of RNA by deep mutational scanning and covariation-induced deviation of activity. *Nucleic Acids Res.* **2020**, *48*, 1451−1465.
(45) Xiang, J. S.; Kaplan, M.; Dykstra, P.; Hinks, M.; McKeague, M.; Smolke, C. D. Massively parallel RNA device engineering in mammalian cells with RNA-Seq. *Nat. Commun.* **2019**, *10*, 4327.
(46) Strobel, B.; Sporing, M.; Klein, H.; Blazevic, D.; Rust, W.; Sayols, S.; Hartig, J. S.; Kreuz, S. High-throughput identification of synthetic riboswitches by barcode-free amplicon-sequencing in human cells. *Nat. Commun.* **2020**, *11*, 714.
(47) Abdelsayed, M. M.; Ho, B. T.; Vu, M. M. K.; Polanco, J.; Spitale, R. C.; Lupták, A. Multiplex Aptamer Discovery through Apta-Seq and Its Application to ATP Aptamers Derived from Human-Genomic SELEX. *ACS Chem. Biol.* **2017**, *12*, 2149−2156.
(48) Buenrostro, J. D.; Araya, C. L.; Chircus, L. M.; Layton, C. J.; Chang, H. Y.; Snyder, M. P.; Greenleaf, W. J. Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat. Biotechnol.* **2014**, *32*, 562−568.
(49) Tapsin, S.; Sun, M.; Shen, Y.; Zhang, H.; Lim, X. N.; Susanto, T. T.; Yang, S. L.; Zeng, G. S.; Lee, J.; Lezhava, A.; Ang, E. L.; Zhang, L. H.; Wang, Y.; Zhao, H.; Nagarajan, N.; Wan, Y. Genome-wide identification of natural RNA aptamers in prokaryotes and eukaryotes. *Nat. Commun.* **2018**, *9*, 1289.
(50) Tome, J. M.; Ozer, A.; Pagano, J. M.; Gheba, D.; Schroth, G. P.; Lis, J. T. Comprehensive analysis of RNA-protein interactions by high-throughput sequencing-RNA affinity profiling. *Nat. Methods* **2014**, *11*, 683−688.

https://dx.doi.org/10.1021/acs.accounts.0c00546
Acc. Chem. Res. 2020, 53, 2903−2912