Methods 161 (2019) 46–53

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Optimization of fluorogenic RNA-based biosensors using droplet-based microfluidic ultrahigh-throughput screening

Alexis Autour, Farah Bouhedda, Roger Cubi, Michael Ryckelynck*

Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, F-67000 Strasbourg, France

ARTICLE INFO

Keywords:
Fluorogenic biosensors
Light-up aptamer
High-throughput screening
Next generation sequencing
RNA
Aptasensors

ABSTRACT

Biosensors are biological molecules able to detect and report the presence of a target molecule by the emission of a signal. Nucleic acids are particularly appealing for the design of such molecule since their great structural plasticity makes them able to specifically interact with a wide range of ligands and their structure can rearrange upon recognition to trigger a reporting event. A biosensor is typically made of three main domains: a sensing domain that is connected to a reporting domain *via* a communication module in charge of transmitting the sensing event through the molecule. The communication module is therefore an instrumental element of the sensor. This module is usually empirically developed through a trial-and-error strategy with the testing of only a few combinations judged relevant by the experimenter. In this work, we introduce a novel method combining the use of droplet-based microfluidics and next generation sequencing. This method allows to functionally characterize up to a million of different sequences in a single set of experiments and, by doing so, to exhaustively test every possible sequence permutations of the communication module. Here, we demonstrate the efficiency of the approach by isolating a set of optimized RNA biosensors able to sense theophylline and to convert this recognition into fluorescence emission.

1. Introduction

Structure-switching aptamers hold great promise for the conception of biosensors aiming at detecting small molecules (e.g. metabolites, chemicals) or biological polymers (e.g. proteins) with applications ranging from bio-surveillance to gene expression monitoring [1,2]. A typical allosteric biosensor is made of a sensing domain that specifically recognizes a target molecule and transmits the information, *via* a communication module (CM), to an effector domain reporting the sensing event [3]. Nucleic acids are particularly interesting for the development of such sensors since they are cheap to produce in large scale by chemical synthesis, they can easily be expressed in living cells, specific sensing aptamers can be rapidly identified using SELEX *in vitro* selection processes [4,5] and their engineering is greatly facilitated by the existence of structure prediction algorithms [6,7]. The reporting domain can consist of a catalytic moiety [3] or another domain that converts the sensing event into an electrochemical, luminescent or fluorescent signal [8].

Light-up RNA aptamers constitute a new appealing class of reporting domain. These RNAs are nucleic acid scaffolds able to specifically recognize and activate the fluorescence of small fluorogenic molecules (also called fluorogens) that are poorly emissive in their free state but become highly fluorescent upon complex formation [9,10]. Many aptamers (e.g. MG [11], Spinach [12], Mango [13], DIR2s [14]) able to specifically activate the fluorescence of their cognate fluorogen (Malachite green, 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), Thiazole Orange 1-biotin and dimethylindole red (DIR) respectively) have been developed over the past decade. The crystal structure of most of these aptamer/fluorogen complexes has also been determined [15–21], making possible to further engineer the aptamer by rational design. For instance, weakening the stability of a key structural element (e.g. a helix located close to the fluorogen-binding site) can be used to suppress the capacity of the aptamer to interact with its fluorogen. However, this inactivation can be conditionally restored by connecting the destabilized aptamer to a structure-switching aptamer acting as a sensing domain. In this scheme, target detection by the sensing aptamer should lead to a structural switching transmitted to the light-up aptamer that recovers its capacity to form a fluorescent complex with its fluorogen (Fig. 1A). The communication module (CM) is the instrumental element of such a molecule since high fluorescence should be emitted only in the presence of the target whereas background fluorescence should be kept as low as possible. CM are usually empirically developed through a trial-and-error strategy and their design would strongly benefits from ultrahigh-throughput screening

* Corresponding author.
E-mail address: m.ryckelynck@unistra.fr (M. Ryckelynck).

https://doi.org/10.1016/j.ymeth.2019.03.015
Received 2 February 2019; Received in revised form 17 March 2019; Accepted 18 March 2019
Available online 19 March 2019

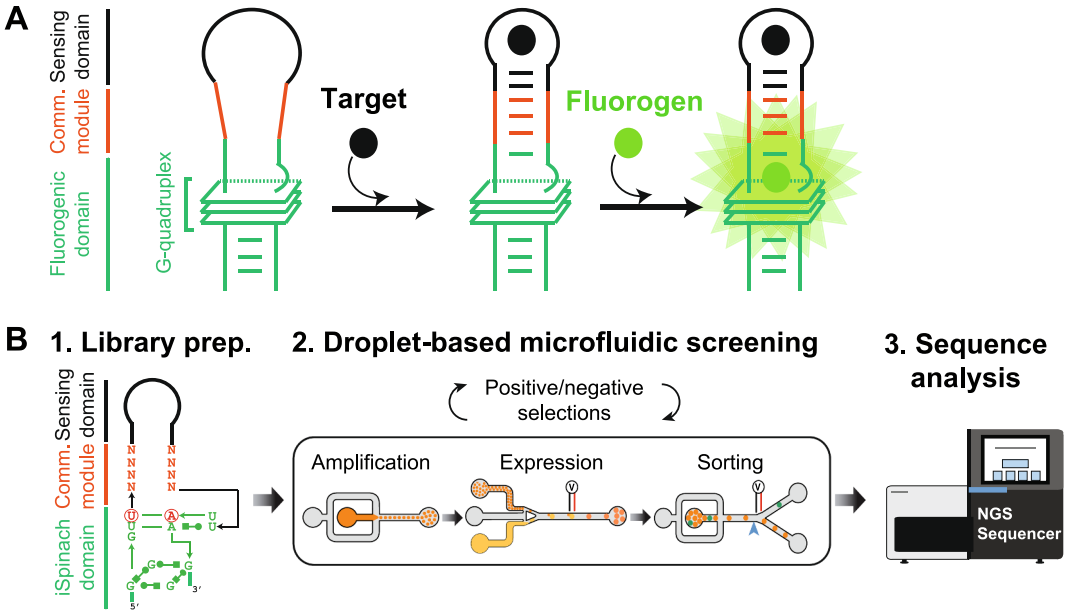

Fig. 1. Principle and development of a fluorogenic RNA biosensor. A Working principle of a fluorogenic RNA biosensor. Upon detection of a target ligand (black circle) by the sensing domain (black), the information is transduced to the fluorogenic domain *via* the communication module (orange). As a consequence, the fluorogenic aptamer (in green) gets stabilized, recovers the capacity to bind its fluorogen (green circle) and forms a fluorescent complex. The G-quadruplex domain of iSpinach is outlined by squares. B Strategy proposed in this work for high-throughput development of RNA biosensors. A library is designed with a randomized communication module (orange) connecting a sensing aptamer to a fluorogenic one, herein iSpinach (in green). Mutations introduced to weaken the apical part of iSpinach are circled in red. The library is then subjected to a droplet-based microfluidic screening pipeline made of 3 major steps: digital droplet PCR (gene individualization/amplification), droplet fusion (gene expression) and droplet sorting (selection of droplets of interest) as described in [23]. The screening operates in two modes: i) either a positive selection in which the expression mixture is supplemented in theophylline and the fluorescent droplets are recovered or ii) a negative selection in which the expression mixture is free of theophylline and non-fluorescent droplets are recovered. Finally, all collected sequences are analyzed by Next Generation Sequencing (NGS) and bioinformatics. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

technologies [22].

In this work, we combined the use of droplet-based microfluidic screening platforms with Next Generation Sequencing (NGS) to devise an ultrahigh-throughput analytical pipeline allowing for efficiently and exhaustively screening large RNA libraries while minimizing the labor-intensive and time-consuming post-screening analysis work. We demonstrate the efficiency of this method by developing a theophylline fluorogenic biosensor in which a theophylline-binding aptamer is linked to a light-up RNA aptamer by an optimized CM allowing efficient extinction of the probe in the absence of target, whereas the construct becomes brighter than the parental light-up aptamer in the presence of the theophylline.

2. Material and methods

2.1. Reagents

Theophylline, caffeine, and guanine were purchased from Sigma-Aldrich and the corresponding solution were prepared at 20 mM in DEPC-treated water (Roth) or HCl for guanine. 10 mM c-diGMP (Sigma-Aldrich) solution was prepared in water and 32 mM S-adenosylmethionine (SAM) solution was obtained from New England Biolabs.

2.2. Preparation of the biosensor library

The library (Fig. S1) was designed using iSpinach [23] as starting point. A double mutant of the aptamer was fused to theophylline aptamer *via* 2 stretches of 4 randomized residues. The corresponding iSpinachUA-lib-Theo template (5′-GCGACTACGGTGAGGGTCGGGTT**NNNN**ataccagccgaaaggcccttggcag**NNNN**TTAAGTAGAGTGTGGGCTCCG TAGTCGC-3′, with the randomized positions bolded, the mutations introduced in iSpinach underlined and theophylline aptamer shown in lower case) was ordered from Integrated DNA Technologies using "hand-mixing" option to adjust the insertion rate of each nucleotide to 25%. Constant regions and T7 RNA polymerase promoter were added by PCR by mixing 2 nM of iSpinachUA-lib-Theo template diluted into 200 µg/mL yeast total RNA solution (Ambion), 0.2 µM of Fwd-add-ExtiSpi (5′-GC*TAATACGACTCACTATA*GGAAGACGTAGCAAGGCGACTAC GGTGAGGGTCGGG-3′, with the constant region underlined and the T7 RNA polymerase promoter italicized), 0.2 µM of Rev-add-ExtiSpi (5′-TGTGTTCTCTGTCATCTCTCTGGCGACTACGGAGCCCACACTC-3′, with the constant region underlined), 0.2 mM of each dNTPs (Fermentas), 2 units (U) of Q5 DNA polymerase (New England Biolabs) and the corresponding buffer at recommended concentration.

Critical points:

- *throughout the procedure, the addition of yeast total RNA is highly recommended to prevent the loss of DNA by adsorption on tube walls.*
- *ordering the template DNA using an 'hand-mixing' option is of prime importance here to limit important sequence biases that may occur in solid-phase synthesis and that may challenge the proper coverage of the library during the screening step.*

The PCR mixture was thermocycled starting with an initial denaturation step of 30 s at 95 °C followed by 25 repeats of the two-step cycle: 95 °C for 5 s and 60 °C for 30 s. PCR products were purified using "Wizard® SV Gel and PCR Clean-up System " kit (Promega).

2.3. Unique Droplet Identifier addition

A Unique Droplet Identifier (UDI) was appended to each DNA molecule of the library by introducing 1 pmol of the iSpinachUA-lib-Theo template extended by constant regions in 100 µL of PCR mixture containing 0.2 µM of Bracode-T7 primer (5′- TCGTCGGCAGCGTCAGATG

TGTATAAGAGACAGAA**NNNNNNNNNNNNNNNNNNNN**AATATC*TAATACGACTCACTATA*GGAAGACGTAGCAAG -3′ with the UDI sequence bolded, the T7 RNA polymerase promoter italicized and the sequence complementary to iSpinachUA-lib-Theo template underlined) and 0.2 µM of Rev-add-ExtiSpi primer (see above), 0.2 mM of each dNTPs (Fermentas), 2 U of Q5 DNA polymerase (New England Biolabs) and the corresponding buffer at the recommended dilution.

The PCR mixture was thermocycled starting with an initial denaturation step of 1 min at 98 °C followed by 30 cycles of: 10 s at 98 °C, 30 s at 55 °C and 30 s at 72 °C, and the PCR products were purified using a "Wizard® SV Gel and PCR Clean-up System " kit (Promega). Upon PCR amplification, an aliquot of this amplification mixture was introduced into a fresh one of similar composition but containing 0.2 µM of Fwd-AmpliBarcode (5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′) and Rev-add-ExtiSpi primers instead. PCR products were again purified using a "Wizard® SV Gel and PCR Clean-up System " kit (Promega). Note that all the PCR amplification mixtures used during this work were prepared using a laminar flow PCR hood (Captair®Bio by Erlarb) isolated in a dedicated room to prevent contaminations.

Critical point: The use of a laminar flow hood to prepare PCR reaction mixtures is highly recommended to prevent contaminations.

2.4. Droplet-based microfluidic screening

Microfluidic chips were molded into polydimethylsiloxane (PDMS) and electrodes were fabricated as described in [24].

2.4.1. Droplet digital PCR

The libraries were diluted in 200 µg/mL yeast total RNA solution (Ambion) to obtain the desired droplet occupancy. 1 µL of this dilution was then introduced in 100 µL of PCR mixture containing 0.2 µM of Fwd-AmpliBarcode and Rev-add-ExtiSpi primers, 0.2 mM of each dNTPs (Fermentas), 10 µM Dextran-Texas Red 70 kDa (Molecular Probes), 0.1% Pluronic F68 (Sigma), 2 U of Q5 DNA polymerase (New England Biolabs) and the corresponding buffer at the recommended dilution.

Critical point: Adding Pluronic F68 is critical to prevent unwanted adsorption of reagents (especially DNA) at the surface of the tubing or at inner droplet surface.

The mixture was loaded into a length of PTFE tubing (Thermo) and infused into a droplet generator microfluidic chip where it was dispersed in 2.5 pL droplets carried by an HFE 7500 fluorinated oil (3 M) supplemented with 3% of a fluorosurfactant [25].

Droplet production frequency (~12,000 droplets per second) was monitored in real time using an optical device and a software developed by the team [24] and used to determine droplet volume. 2.5 pL droplets were generated by adjusting pumps flowrates (MFCS, Fluigent). The emulsion was collected into 0.2 mL tube and subjected to an initial denaturation step of 30 s at 95 °C followed by 30 PCR cycles of: 5 s at 95 °C, 30 s at 60 °C.

2.4.2. Addition of in vitro expression mixture by droplet fusion

PCR droplets were reinjected into a droplet fusion device at a rate of ~1500 droplets per second as described previously [24]. PCR droplets were spaced by a stream of HFE 7500 fluorinated oil (3 M) supplemented with 2% (w/w) of a fluorosurfactant prior to being synchronized and paired one-to-one with a 16 pL *in vitro* transcription (IVT) droplet containing 2 mM of each NTP (Larova), 25 mM $MgCl_2$, 44 mM Tris-HCl pH 8.0 (at 25 °C), 50 mM KCl, 5 mM DTT, 1 mM Spermidine, 0.1% of Pluronic F68 (Sigma), 1 µg of inorganic pyrophosphatase (Roche), 20 µM DFHBI-1 T (Lucerna), 0.5 µM of Dextran-Texas Red 70 kDa (Molecular Probes), 17.5 µg/mL T7 RNA polymerase (purified in the laboratory) and optionally supplemented with 1 mM theophylline in the case of the positive selection.

IVT mixture was loaded in a length of PTFE tubing and kept on ice during all the experiment and IVT droplets were produced using an HFE 7500 fluorinated oil (3 M) stream supplemented with 2% (w/w). Flowrates (MFCS, Fluigent) were adjusted to generate 16 pL IVT droplets and to maximize the synchronization of 1 PCR droplet with 1 IVT droplet. Pairs of droplets were then fused with an AC field (400 V at 30 kHz) and the resulting emulsion was collected off-chip in a tube prior to being incubated for 1 h 30 min at 37 °C.

2.4.3. Droplet fluorescence analysis and sorting

The emulsion was finally re-injected into an analysis and sorting microfluidic device [26] mounted onto a Thermo plate (Tokai Hit) holding the temperature at 37 °C. Droplets were infused at a frequency of ~200 droplets per second and spaced with a stream of surfactant-free HFE 7500 fluorinated oil (3 M). Green (DFHBI-1T in complex with competent fluorogenic aptamer) and orange (dextran Texas-Red) fluorescence of each droplet was analyzed. In rounds of positive selection (i.e. performed in the presence of theophylline), the 1 to 10% greenest droplets displaying an orange fluorescence corresponding to single fused droplets were gated (Fig. S2). During the rounds of negative selection (i.e. performed in absence of theophylline), the ~25% dimmest green droplets displaying an orange fluorescence corresponding to single fused droplets were gated (Fig. S2). In both cases, gated droplets were deflected into collection channel by applying an AC field (1000 V 30 kHz) and collected into a 1.5 mL tube. Sorted droplets were recovered from the collection tubing by flushing 200 L of HFE 7500 fluorinated oil (3 M). 100 L of 1H, 1H, 2H, 2H-perfluoro-1-octanol (Sigma-Aldrich) and 200 µL of 200 µg/mL yeast total RNA solution (Ambion) were then flushed and the droplets were broken by vortexing the mixture. DNA-containing aqueous phase was finally transferred a new tube.

Critical point: the sequential addition of the three reagents (i.e. HFE 7500, 1H, 1H, 2H, 2H-perfluoro-1-octanol and yeast total RNA) flushed through the collection tubing is important to maximize DNA recovery from the droplets.

An aliquot (2 µL) of DNA-containing aqueous phase was treated as described in 2.3 to both amplify the material and reset the UDI carried by each molecule before using these molecules to prime a new round of screening.

2.5. High-throughput sequencing and bioinformatic analysis

Upon each round of screening, an aliquot of DNA-containing aqueous phase was prepared for Next Generation Sequencing analysis. In parallel, an aliquot of the starting library was treated the same way. 2 µL of aqueous phase were introduced in a PCR mixture containing 0.2 µM of both Fwd-AmpliBarcode and Tag-iSpiTA-Lib-Theo (5-GTCT CGTGGGCTCGGAGATGTGTATAAGAGACAGTGTGTTCTCTGTCATCTC TCTG-3′) primers, 0.2 mM of each dNTP (Fermentas), 2 U of Q5 DNA polymerase (New England Biolabs) and the corresponding buffer at the recommended dilution. This mixture was subjected to an initial denaturation step of 1 min at 98 °C followed by 30 cycles of: 10 s at 98 °C, 30 s at 55 °C and 30 s at 72 °C. The recovered DNA was then indexed using Nextera Indexing kit (Illumina) following supplier recommendations. Finally, the indexed library was purified on a 1% agarose gel and the DNA recovered using a Monarch® purification kit (New England Biolabs). The quality of the DNA was then assessed on a Bioanalyzer (Agilent technologies) device and the DNA sequenced on MiSeq device (Illumina) using a V3-150 sequencing cartridge.

Critical point: proper quantification of the library on a Bioanalyzer platform is critical to adjust the concentration of the DNA loaded on the sequencer and avoid over-clustering.

Upon sequencing, 1 to 3 million reads were obtained for each library. Reads with a Q-score > 30 were then analyzed using a custom Python bioinformatic pipeline including 4 major steps (Fig. S3). First, identical sequences with the expected length were clustered together, and their occurrence measured. Second, an occurrence cut-off was manually set for each library (Fig. S4). Sequences with an occurrence below that threshold were likely mutants (raised from PCR or sequencing errors) and were no longer considered for the rest of the analysis. Moreover, sequences displaying mutations outsides of randomized regions (i.e. the UDI and the communication module) were also filtered out. Third, counting the number of different UDIs associated with each "communication module" sequence, allowed for counting the number of droplets containing this particular sequence, so its occurrence. Finally, occurrence frequency of each sequence was computed and used to compute the enrichment factor (defined as the ratio of the occurrence frequency of the sequence after a given round over that in the starting library) of each sequence. All fastq files containing raw sequencing data generated in this study are available in the EMBL-EBI European Nucleotide Archive [ENA: PRJEB31612]: http://www.ebi.ac.uk/ena/data/view/PRJEB31612.

2.6. Characterization of library responsiveness to theophylline

An aliquot (1 µL) of DNA-containing aqueous phase recovered at the end of each round of screening (or from the starting library) was amplified into 100 µL of PCR mixture containing 0.1 µM of Fwd-add-ExtiSpi and Rev-add-ExtiSpi primers, 0.2 mM of each dNTP (Fermentas), 0.05 U/µL Q5 DNA polymerase (New England Biolabs) and the corresponding buffer at recommended concentration. The mixture was thermocycled with an initial denaturation step of 30 s at 95 °C followed by 25 repeats of the two-step cycle: 95 °C for 5 s and 60 °C for 30 s. PCR products were finally purified using a "Wizard® SV Gel and PCR Clean-up System " kit (Promega) and quantified with a Nanodrop (Thermo Scientific). 1.5 pmol of purified PCR product was introduced in an *in vitro* transcription (IVT) mixture containing 40 mM Tris-HCl pH 8.0, 22 mM $MgCl_2$, 5 mM DTT, 1 mM Spermidine, 1.6 mM of each NTP, 20 µM DFHBI-1 T (Lucerna), 50 mM KCl, 5 ng/µL inorganic pyrophosphatase (Roche), 1400 U T7 RNA polymerase (purified in the laboratory). Moreover, this mixture was optionally supplemented with 1 mM theophylline. The green fluorescence (ex: 492 nm/em: 516 nm) of activated DFHBI-1 T was then monitored every minute for 2 h 30 min at 37 °C on a real-time thermocycler (Mx 3005P, Agilent).

2.7. Ligand sensing by purified RNA

RNAs were synthetized and purified following the same protocol described in [23]. Prior to fluorescence measurement, purified RNAs were diluted in water and renatured by warming them 2 min at 85 °C prior to letting them cool at 25 °C for 5 min. Then, 1 vol of renatured RNA (at desired concentration, herein 1 µM) was added to 1 vol of a mixture containing: 80 mM Tris-HCl pH 8.0, 20 mM $MgCl_2$, 20 µM DFHBI-1 T, 200 mM KCl and optionally 100 µM of theophylline or another ligand (i.e. guanine, S-adenosyl methionine or c-di-GMP) for 1 h at 25 °C. The green fluorescence (ex: 492 nm/em: 516 nm) was then monitored for 5 min at 25 °C on a real-time thermocycler (Mx 3005P, Agilent).

2.8. Affinity measurement

Affinity between the biosensor and target ligands was determined by incubating 500 nM of the biosensor renatured using the same procedure described above (see §2.7) with 0.05 µM–100 µM theophylline or caffeine in a mixture containing 40 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 10 µM DFHBI-1 T, 100 mM KCl. The green fluorescence (ex: 480 nm/em: 520 nm) of the activated DFHBI-1 T was then measured at room temperature on a real-time thermocycler (Mx 3005P, Agilent) and the K_D was defined as the value of ligand leading to half the maximal fluorescence and was calculated by fitting the data to a Hill equation (OriginLab).

3. Results

The goal of this work was to develop a fluorogenic RNA biosensor of theophylline using a 3-step methodology (Fig. 1B) in which: i) we designed and prepared a library of biosensors whom the CM region was randomized; ii) we screened this library in ultrahigh-throughput regime using the droplet-based microfluidic pipeline we previously used to improve the function of ribozymes [24] and light-up aptamers [23,27]; iii) we characterized the screening process using Next Generation Sequencing (NGS).

3.1. Design of an iSpinach-based biosensor library

We choose to start the development of our biosensor using iSpinach, an aptamer able the light-up the fluorogen DFHBI and its derivative DFHBI-1T [28] and improved for *in vitro* applications [23]. Indeed, we previously found that the molecule is endowed with very good folding properties, especially at the level of its basal stem [21]. Then, we weakened the structure of iSpinach apical helix by mutating the C_{38}-G_{59} base pair located just above the DFHBI-binding pocket (Figs. 2 and S5). Introducing an A_{38}-U_{59} base pair or deleting it had a severe effect on complex formation (Fig. S5). Conversely, using a U_{38}-A_{59} pair led to a weak < 20% fluorescence reduction indicative of the expected structure weakening, while not compromising aptamer function. We next randomized two stretches of four nucleotides located directly on the top of the U_{38}-A_{59} pair and connected them to the theophylline-binding aptamer (Fig. 2), a well characterized structure-switching aptamer [29,30]. Randomizing stretches of four nucleotides made possible to explore a large sequence space, increasing therefore the likelihood to isolate an efficient communication module, while still being able to exhaustively screen the library with a good coverage (see below). Considering the current upper number of droplets (5 millions) that our technology allows to analyze per round and the reduced droplet occupancy (~10%) that one needs to use to limit multiple encapsulation events, using longer stretches of nucleotides would have compromised the possibility to exhaustively screen the starting library. The molecule was flanked with constant regions later used as primer-binding sites for PCR amplification and one of these regions was further also flanked with the T7 RNA polymerase promoter sequence (Fig. S1). Finally, a random sequence of 20 nucleotides serving as Unique Droplet Identifier (UDI) was appended to the construct, upstream the transcription promoter. By analogy with the Unique Molecular Identifiers (UMIs) used to precisely establish transcript copy-number in transcriptomic analyses [31], here UDIs served to cluster together the genes coming from the same droplet and ultimately to count the number of droplets containing a given sequence upon sorting (see below).

3.2. Ultrahigh-throughput screening

The theophylline biosensor library was made of ~65,500 (4^8) different sequences and was exhaustively screened with a ~5-fold coverage using droplet-based microfluidics (Table S1). Briefly, each round of µIVC screening consisted of 3 main steps (Fig. 1B). First, DNA molecules were diluted into a PCR mixture prior to being distributed into 2.5 pL water-in-oil droplets. Adjusting this dilution to have an average of 1 gene per 10 droplets, a value also known as λ (here λ was ~0.1, Table S1) allowed limiting the rate of multiple encapsulation events. Indeed, since DNA molecules distribute into the droplets according to Poisson statistics [32], knowing λ makes possible to precisely calculate the fraction of droplets initially occupied by one or more genes (less than 0.5% at λ ~ 0.1). Water-in-oil droplets were then produced by

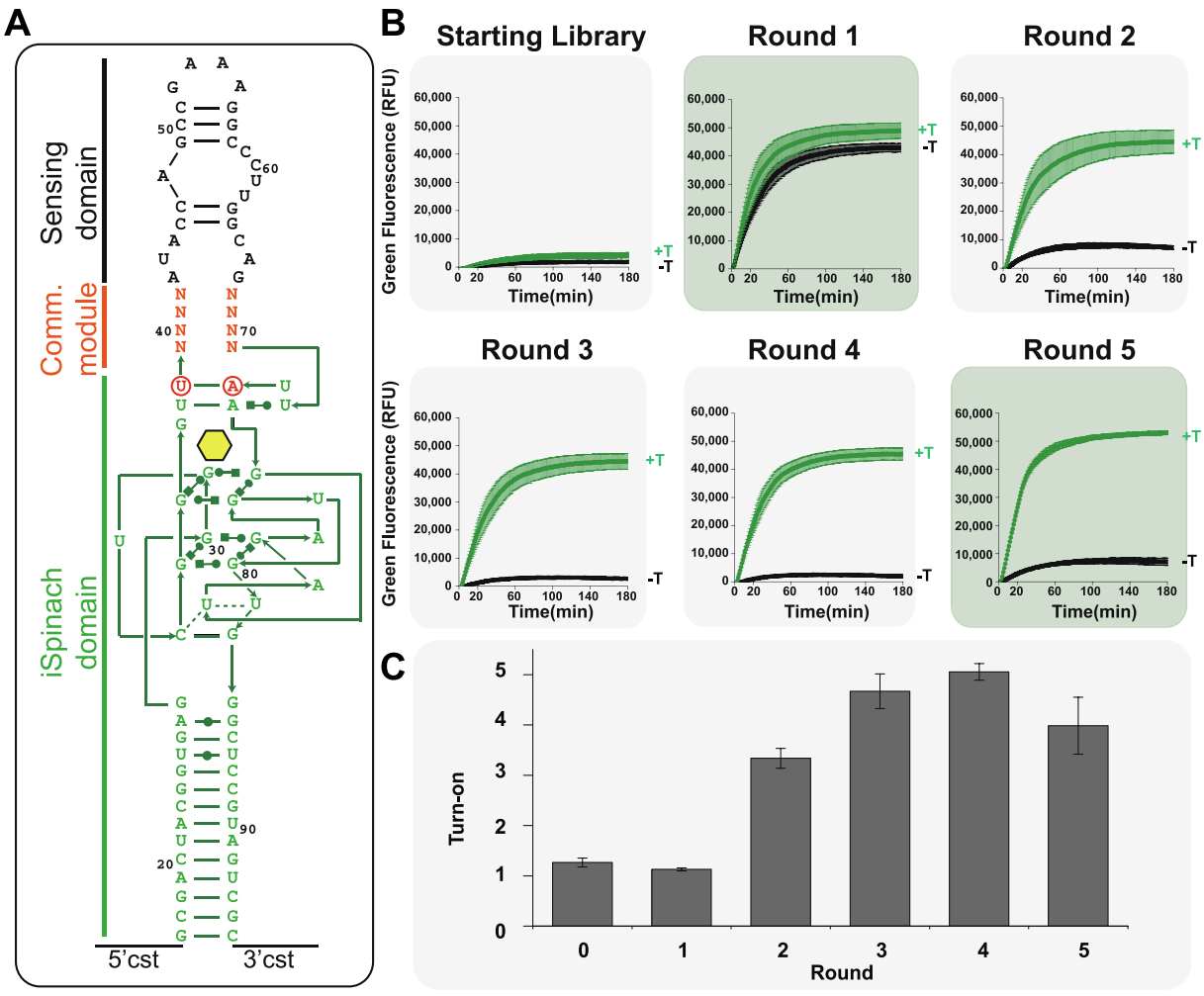

Fig. 2. Microfluidic-assisted functional screening of fluorogenic RNA biosensors of theophylline. A Secondary structure of the RNA biosensors encoded by the gene library. The molecule is made of the theophylline specific aptamer (black) described in [30] linked *via* a randomized region of twice four nucleotides (orange) to a fluorogenic domain (green) derived from iSpinach [23]. Mutations introduced to weaken the apical part of iSpinach are circled in red, cst stands for constant region and Comm. module for Communication module. B Responsiveness of the starting and the selected libraries to theophylline. The enrichment in functional fluorogenic RNA biosensor was monitored by monitoring in real-time the transcription of the library in presence (green line) or in absence (in black) of theophylline. Plots shadowed in dark green represent the profile of the libraries enriched upon a positive selection (in presence of theophylline). C Turn-on capacity of the library along the screening process. The values were obtained from the ratio of the fluorescence in the presence over that in the absence of theophylline. The values are the mean of three independent experiments and error bars correspond to ± 1 standard error. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

infusing the aqueous phase into a microfluidic chip together with a fluorinated oil phase supplemented in a fluorosurfactant [25] allowing for stabilizing droplets and preventing their uncontrolled coalescence. Upon production, droplets were thermocycled to PCR-amplify each gene into ~120,000 copies. Then, droplets were re-injected into a droplet fusion device [32] in which they were spaced by a stream of fluorinated oil (supplemented with 2% fluorosurfactant) and synchronized one-to-one with larger (16 pL) droplets generated on-chip and containing an *in vitro* transcription (IVT) mixture supplemented with DFHBI-1T as well as theophylline in case of positive selection (see below). Pairs of droplets were then fused together when passing between a pair of electrodes energized by an AC field. One-to-one pairing was maximized by monitoring the orange fluorescence of dextran-Texas Red added into both sets of droplets. Indeed, PCR droplets contained a high concentration (10 µM) of dextran-Texas Red and were highly orange fluorescent. On the contrary, IVT mixture was supplemented with a low concentration (0.5 µM) of dextran-Texas Red making the resulting droplets poorly orange fluorescent. However, upon droplet merging, fluorescence was expected to average, and a population of intermediate orange fluorescence appeared (Fig. S2). Using this strategy, it was possible to fuse more than 85% of IVT droplets mostly with one PCR droplet (Table S1). Upon fusion, droplets were collected and incubated for 1 h 30 min at 37 °C to allow the genes to be transcribed. Droplets were re-injected into a sorting device [26], in which they were spaced by a surfactant-free oil stream prior to getting their fluorescence analyzed. Those droplets displaying an orange fluorescence corresponding to single-fused droplets as well as the expected green fluorescence from the DFHBI/aptamer complex were gated and specifically sorted from the bulk. Finally, upon sorting, droplets were broken, their DNA content recovered by PCR, their original UDI was exchanged for a new one to be able to track individual droplets during the next rounds and the process restarted.

To select molecules able to efficiently activate DFHBI-1T fluorescence only in the presence of theophylline, we performed two rounds of positive selection (i.e. recovering those droplets highly fluorescent in the presence of theophylline) interspaced by three rounds of negative selection (i.e. recovering those droplets poorly fluorescent in the absence of theophylline) selections (Fig. 2). The efficiency of each round of screening was then assessed by *in vitro* transcribing libraries (the starting one and those enriched after the round) in the presence or in

the absence of theophylline while monitoring DFHBI-1T fluorescence over the time (Fig. 2B). The first round of positive screening (round 1) allowed to significantly enrich the library in molecules able to activate DFHBI-1T fluorescence. However, the library also displayed a strong fluorescence in the absence of theophylline, indicating that a significant fraction of the aptamers activated DFHBI-1T in a theophylline-independent manner. To further enrich the library in theophylline-responsive aptamers, we performed three rounds of negative selection in which the expression mixture was no longer supplemented in theophylline and we selected the non-fluorescent droplets (i.e. containing aptamer that do not activate DFHBI-1T in the absence of theophylline). Along these three rounds, the population progressively enriched in theophylline-responsive molecules. Indeed, whereas a high fluorescence was still observed in the presence of theophylline, almost no fluorescence was observed in the absence of the ligand (Fig. 2B). To further enrich the library in highly fluorescent aptamers, we performed a last round of positive selection in the presence of theophylline. Computing the turn-on factor (defined as the ratio of the fluorescence in the presence over that in the absence of theophylline) allowed quantifying the responsiveness of each library and gave an overview of the screening process (Fig. 2C). We found that the value rapidly increased over the first rounds prior to reaching a plateau at round 4. We therefore stopped the screening process at round 5 and more deeply analyzed selected RNAs using high-throughput sequencing.

3.3. High-throughput sequence analysis

Upon Illumina sequencing (MiSeq, V3-150 chemistry) of the different libraries (the starting library and the 5 enriched ones), reads were QC-filtered using a custom-made Python pipeline (Fig. S3). This process allowed to compute the enrichment of each sequence throughout the whole screening process (Fig. S6).

This analysis allowed us to confirm that the starting library was properly randomized at each position since each nucleotide was found present at an equivalent frequency at each position (Fig. S7). After the first round of positive selection, more than 97% of the sequence diversity was lost (diversity reduced from 65,536 to 1715 sequences, Fig. 3A), highlighting the great selection efficient of our microfluidic screening. This diversity kept on decreasing along the four next rounds to finally reach 151 different sequences (0.23% of the starting diversity) at the end of the process. As expected, the variants discarded between round 1 and 2 (Discarded R1 on Fig. 3C) displayed high fluorescence but were poorly responsive to theophylline. Higher responsiveness to theophylline was observed with variants discarded from round 4 but, as expected, these molecules tended to be dimmer than those conserved at the end of the process (round R5). Indeed, among all the aptamers tested, the five most abundant conserved after the fifth round displayed, in general, the best theophylline responsiveness (Fig. 3C). Interestingly, the 10 most abundant biosensors were also found to be continuously enriched throughout the different rounds (Fig. S6), indicating that they properly replied to each selection condition, an observation in agreement with their good general responsiveness to theophylline, whereas those less enriched sequences tend also to display a lower responsiveness to theophylline (Fig. S8).

Finally, when looking closer at the sequence composition of the most enriched variants, we found a strong bias against C residues that were almost completely absent whereas Us represented at least half of the occurrences (Figs. 3D and S7). Therefore, the sensors tend to prefer communication module made of weak base pairs, with a strong preference for UG (or UU) pairs at the second and the third position. Interestingly, a similar preference for non-Watson-Crick base pairs and bias against strong GC base pairs was also described for the communications modules found in artificial on-switch aptazymes isolated *in vitro* [33] or *in vivo* [34], further reinforcing the conclusion of the present study. Moreover, similar sequence composition is also typically observed in thermosensor RNAs, a set of temperature-dependent structure switching RNA controlling gene expression [35,36].

3.4. Functional validation of the theophylline biosensor

The most abundant variant found at the end of our analysis had a communication module (5′-AUUU-3′/5′-AUUU-3′) made of 2 non-canonical (UU) base pairs surrounded by 2 canonical (AU) ones. This sensor displayed a nice 5-fold activation amplitude in the presence of theophylline (Fig. S8) while preserving a good μM-range affinity for theophylline and intact discrimination of caffeine (Fig. 4A), a close homologue of theophylline. Interestingly, when looking at the ten most enriched variants, we found that the eighth most enriched variant displayed a very close sequence (5′-ACUU-3′/5′-AUUU-3′) together with similar activation amplitude, affinity for theophylline and specificity (Fig. S9), suggesting that 5′-A(U/C)UU-3′/5′-AUUU-3′might be the optimal communication module for our construct. Even though they are efficient, both biosensors display affinities for theophylline an order of magnitude below that reported for the wild type theophylline aptamer [30]. Such a decrease can partly be explained by the exchange of the basal stem of the aptamer for a less stable communication module. Moreover, the use of a theophylline concentration (i.e. 1 mM) far exceeding the K_D (320 nM) of the aptamer for its ligand did not allow to maintain a selection pressure for high affinity biosensors during the screening process. Therefore, one cannot rule out the possibility that biosensors of an even higher affinity could be selected using the same workflow but with lower concentration of theophylline.

Moreover, to estimate to what extend the 5′-AUUU-3′/5′-AUUU-3′ communication module was universal, we tested several constructs in which the theophylline aptamer was exchanged for others known to specifically sense S-adenosyl methionine [37], guanine [37] or cyclic di-GMP [38,39]. We were glad to observe that all the tested constructs were responsive to their cognate ligand (Fig. 4B), even though some constructs were less responsive. However, this was not surprising since a communication module is not a stand-alone element but, on contrary, it is embedded between two aptamers and, due to its function, it has to be sensitive to the sequences directly surrounding it.

4. Conclusions

In this work, we showed that droplet-based microfluidics could be used to screen aptamer-coding gene libraries by performing rounds of positive and negative selection. To the best of our knowledge, this is the first time that the successful development of a fluorogenic RNA biosensor is reported using such a high throughput. The possibility of quantitatively analyzing and sort millions of variants per day makes possible to exhaustively screen large libraries of molecules in which long regions were randomized. Moreover, characterizing selection products using NGS and bioinformatics further increase the efficiency of the method by making possible to track every sequence throughout the process and to rapidly identify those sequences enriching the best at each step, therefore those being very likely to be best adapted to the selection pressures, so having the wished properties. Whereas this work allowed isolating an efficient communication module (CM), this sequence was not completely universal, should a true universal CM (i.e. a CM optimally working in any possible sequence context) could really be developed. Consequently, the development of any new biosensor may need some refinement of the CM sequence to work optimally. However, the robustness of the present technology and its capacity to predict to some extend the best adapted molecules directly from sequencing data strongly reduces the experimental labor while providing deep information since every sequence permutation can be tested in a single set of experiments. It is noteworthy that the use of NGS for the characterization of selected nucleic acids already proved its efficiency when used in tandem with SELEX technology [40–42] as well as for the characterization of self-cleaving ribozymes (see the review by Yohei Yokobayashi in this issue [43]). This list of NGS-assisted analytical

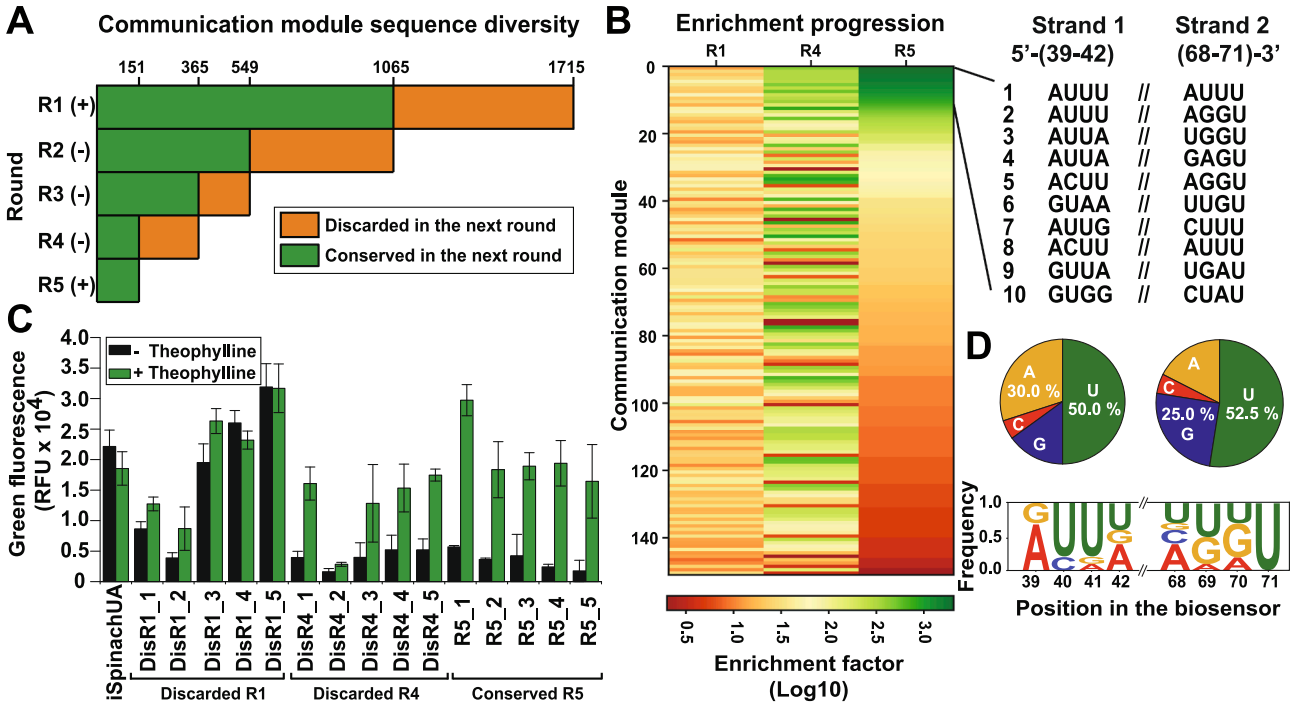

Fig. 3. Sequence analysis of the starting and the enriched libraries. A Sequence diversity of the communication modules. The number of different sequences found in the enriched libraries after each round of screening is given by the length of the both boxes. For each round the green box represents the number of sequences conserved at the end of the next round, whereas the discarded fraction is indicated by the orange box. Rounds in which positive selection was performed are indicated by the (+) whereas those where a negative selection was used are indicated by the (−). B Enrichment progression throughout the screening process. The enrichment factor of the 151 communication modules isolated at the end of R5 was computed from the ratio of the occurrence frequency of the sequence in R1, R4 or R5 over that of the sequence in R0. The heatmap (expressed in Log10) summarizes the progression of this factor over the screening process. The sequence of the ten most abundant communication modules is given on the side of the heatmap. C Biosensor responsiveness to theophylline. 500 nM of purified biosensor were incubated in presence of 10 µM DFHBI-1T and 50 µM of theophylline (in green) or in the absence of theophylline (in black) at 25 °C. Values are the mean of three independent experiments and error bars correspond to ± 1 standard error. D Sequence composition of the ten most abundant biosensors isolated upon round 5. The global composition of each strand is shown as pie charts (top panel) and the composition at each position as a Logo representation (bottom panel). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

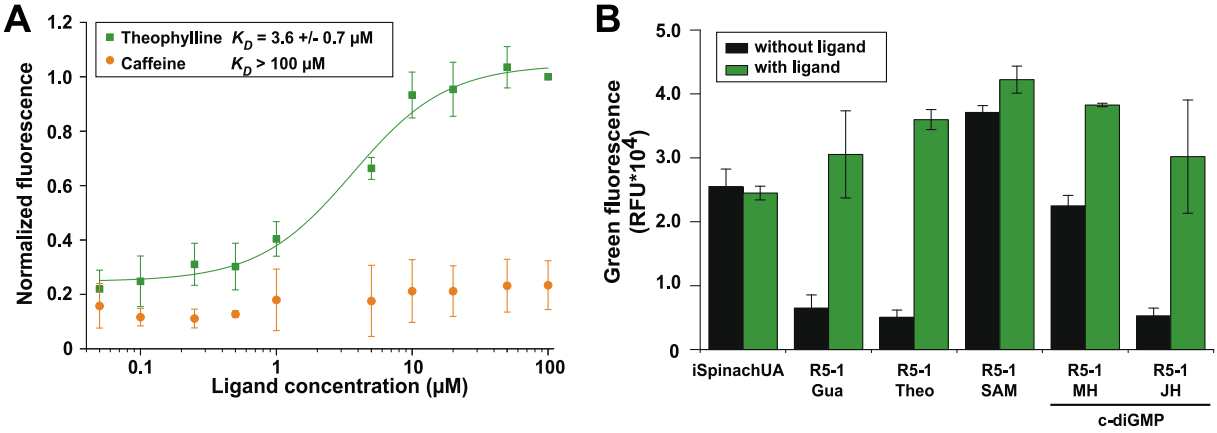

Fig. 4. Functionality of the 5′-AUUU-3′/5′-AUUU-3′ communication module. A Affinity and specificity of the most abundant variant (R5-1) possessing the 5′-AUUU-3′/5′-AUUU-3′ communication module. The RNA was gel purified and incubated with a range of concentration of theophylline or caffeine at 28 °C and the fluorescence of the biosensor/DFHBI-1T complex was measured. A K_D was measurable only for the theophylline (no signal observed in the presence of caffeine) and was calculated by fitting the data to a Hill equation. B The sensing aptamer of R5-1 was exchanged for alternative aptamers sensing Guanine [37], S-adenosyl methionine [37] as well as two aptamer sensing c-diGMP and previously described by the group of Ming Hammond (MH) [39] and Jörg Hartig (JH) [38]. The RNA was incubated with 10 µM DFHBI-1T and in the absence or in presence of 50 µM of ligand prior to fluorescence measurement. The values are the mean of three independent experiments and error bars correspond to ± 1 standard error.

technologies of nucleic acids is therefore now completed by the technology we introduced in this work. Moreover, its application range is not limited to light-up aptamers as it can easily be extended to the selection of other structure-switching aptamer (e.g. riboswitches or aptazymes [44]) provided a fluorescence-based assay can be set-up.

Acknowledgements

We thank Sandrine Koechler and Abdelmalek Alioua from the IBMP Gene Expression Analysis facility (funded as of the LabEx NetRNA) for technical assistance with high-throughput sequencing. We also thank

Eric Westhof for stimulating scientific discussions.

Funding

This work received financial support from Agence Nationale de la Recherche (BrightRiboProbes, ANR-16-CE11-0010-01/02) and from the University of Strasbourg Institute of Advanced Study (USIAS, program Translatomix). This work has also been published under the framework of the LabEx: ANR-10-LABX-0036_NETRNA and benefits from a funding from the state managed by the French National Research Agency as part of the Investments for the future program. It was also supported by the Université de Strasbourg and the Centre National de la Recherche Scientifique.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymeth.2019.03.015.

References

[1] A. Karunanayake Mudiyanselage, R. Wu, M.A. Leon-Duque, K. Ren, M. You, "Second-generation" fluorogenic RNA-based sensors, Methods (2019) This issue.
[2] Z.F. Hallberg, Y. Su, R.Z. Kitto, M.C. Hammond, Engineering and in vivo applications of riboswitches, Annu. Rev. Biochem. 86 (2017) 515–539.
[3] G.A. Soukup, R.R. Breaker, Nucleic acid molecular switches, Trends Biotechnol. 17 (12) (1999) 469–476.
[4] C. Tuerk, L. Gold, Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, Science 249 (4968) (1990) 505–510.
[5] A.D. Ellington, J.W. Szostak, In vitro selection of RNA molecules that bind specific ligands, Nature 346 (6287) (1990) 818–822.
[6] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, Nucl. Acids Res. 31 (13) (2003) 3406–3415.
[7] J.N. Zadeh, C.D. Steenberg, J.S. Bois, B.R. Wolfe, M.B. Pierce, A.R. Khan, R.M. Dirks, N.A. Pierce, NUPACK: Analysis and design of nucleic acid systems, J. Comput. Chem. 32 (1) (2011) 170–173.
[8] F. Pfeiffer, G. Mayer, Selection and biosensor application of aptamers for small molecules, Front. Chem. 4 (2016) 25.
[9] F. Bouhedda, A. Autour, M. Ryckelynck, Light-Up RNA aptamers and their cognate fluorogens: from their development to their applications, Int. J. Mol. Sci. 19 (1) (2018).
[10] S. Neubacher, S. Hennig, RNA structure and cellular applications of fluorescent light-up aptamers, Angew. Chem. 58 (5) (2019) 1266–1279.
[11] D. Grate, C. Wilson, Laser-mediated, site-specific inactivation of RNA transcripts, Proc. Natl. Acad. Sci. U S A 96 (11) (1999) 6131–6136.
[12] J.S. Paige, K.Y. Wu, S.R. Jaffrey, RNA mimics of green fluorescent protein, Science 333 (6042) (2011) 642–646.
[13] E.V. Dolgosheina, S.C. Jeng, S.S. Panchapakesan, R. Cojocaru, P.S. Chen, P.D. Wilson, N. Hawkins, P.A. Wiggins, P.J. Unrau, RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking, ACS Chem. Biol. (2014).
[14] X. Tan, T.P. Constantin, K.L. Sloane, A.S. Waggoner, M.P. Bruchez, B.A. Armitage, Fluoromodules consisting of a promiscuous RNA aptamer and red or blue fluorogenic cyanine dyes: selection, characterization, and bioimaging, J. Am. Chem. Soc. (2017).
[15] C. Baugh, D. Grate, C. Wilson, 2.8 A crystal structure of the malachite green aptamer, J. Mol. Biol. 301 (1) (2000) 117–128.
[16] K.D. Warner, M.C. Chen, W. Song, R.L. Strack, A. Thorn, S.R. Jaffrey, A.R. Ferre-D'Amare, Structural basis for activity of highly efficient RNA mimics of green fluorescent protein, Nat. Struct. Mol. Biol. 21 (8) (2014) 658–663.
[17] R.J. Trachman 3rd, N.A. Demeshkina, M.W.L. Lau, S.S.S. Panchapakesan, S.C.Y. Jeng, P.J. Unrau, A.R. Ferre-D'Amare, Structural basis for high-affinity fluorophore binding and activation by RNA Mango, Nat. Chem. Biol. 13 (7) (2017) 807–813.
[18] R. Trachman, A. Abdolahzadeh, A. Andreoni, R. Cojocaru, J.R. Knutson, M. Ryckelynck, P.J. Unrau, A. Ferre-D'Amare, Crystal structures of the Mango-II RNA aptamer reveal heterogeneous fluorophore binding and guide engineering of variants with improved selectivity and brightness, Biochemistry (2018).
[19] S.A. Shelke, Y. Shao, A. Laski, D. Koirala, B.P. Weissman, J.R. Fuller, X. Tan, T.P. Constantin, A.S. Waggoner, M.P. Bruchez, B.A. Armitage, J.A. Piccirilli, Structural basis for activation of fluorogenic dyes by an RNA aptamer lacking a G-quadruplex motif, Nat. Commun. 9 (1) (2018) 4542.
[20] H. Huang, N.B. Suslov, N.S. Li, S.A. Shelke, M.E. Evans, Y. Koldobskaya, P.A. Rice, J.A. Piccirilli, A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore, Nat. Chem. Biol. 10 (8) (2014) 686–691.
[21] P. Fernandez-Millan, A. Autour, E. Ennifar, E. Westhof, M. Ryckelynck, Crystal structure and fluorescence properties of the iSpinach aptamer in complex with DFHBI, RNA 23 (12) (2017) 1788–1795.
[22] T.A. Feagin, N. Maganzini, H.T. Soh, Strategies for creating structure-switching aptamers, ACS Sens 3 (9) (2018) 1611–1615.
[23] A. Autour, E. Westhof, M. Ryckelynck, iSpinach: a fluorogenic RNA aptamer optimized for in vitro applications, Nucleic Acids Res. 44 (6) (2016) 2491–2500.
[24] M. Ryckelynck, S. Baudrey, C. Rick, A. Marin, F. Coldren, E. Westhof, A.D. Griffiths, Using droplet-based microfluidics to improve the catalytic properties of RNA under multiple-turnover conditions, RNA 21 (3) (2015) 458–469.
[25] C. Holtze, A.C. Rowat, J.J. Agresti, J.B. Hutchison, F.E. Angile, C.H.J. Schmitz, S. Koster, H. Duan, K.J. Humphry, R.A. Scanga, J.S. Johnson, D. Pisignano, D.A. Weitz, Biocompatible surfactants for water-in-fluorocarbon emulsions, Lab Chip 8 (10) (2008) 1632–1639.
[26] J.C. Baret, O.J. Miller, V. Taly, M. Ryckelynck, A. El-Harrak, L. Frenz, C. Rick, M.L. Samuels, J.B. Hutchison, J.J. Agresti, D.R. Link, D.A. Weitz, A.D. Griffiths, Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity, Lab Chip 9 (13) (2009) 1850–1858.
[27] A. Autour, S.C.Y. Jeng, A.D. Cawte, A. Abdolahzadeh, A. Galli, S.S.S. Panchapakesan, D. Rueda, M. Ryckelynck, P.J. Unrau, Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells, Nat. Commun. 9 (1) (2018) 656.
[28] W. Song, R.L. Strack, N. Svensen, S.R. Jaffrey, Plug-and-play fluorophores extend the spectral properties of spinach, J. Am. Chem. Soc. 136 (4) (2014) 1198–1201.
[29] G.R. Zimmermann, R.D. Jenison, C.L. Wick, J.P. Simorre, A. Pardi, Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA, Nat. Struct. Biol. 4 (8) (1997) 644–649.
[30] R.D. Jenison, S.C. Gill, A. Pardi, B. Polisky, High-resolution molecular discrimination by RNA, Science 263 (5152) (1994) 1425–1429.
[31] S. Islam, A. Zeisel, S. Joost, G. La Manno, P. Zajac, M. Kasper, P. Lonnerberg, S. Linnarsson, Quantitative single-cell RNA-seq with unique molecular identifiers, Nat. Methods 11 (2) (2014) 163–166.
[32] L. Mazutis, A.F. Araghi, O.J. Miller, J.C. Baret, L. Frenz, A. Janoshazi, V. Taly, B.J. Miller, J.B. Hutchison, D. Link, A.D. Griffiths, M. Ryckelynck, Droplet-based microfluidic systems for high-throughput single DNA molecule isothermal amplification and analysis, Anal. Chem. 81 (12) (2009) 4813–4821.
[33] A. Kertsburg, G.A. Soukup, A versatile communication module for controlling RNA folding and catalysis, Nucl. Acids Res. 30 (21) (2002) 4599–4606.
[34] M. Felletti, A. Bieber, J.S. Hartig, The 3'-untranslated region of mRNAs as a site for ribozyme cleavage-dependent processing and control in bacteria, RNA Biol. 14 (11) (2017) 1522–1533.
[35] J. Rossmanith, F. Narberhaus, Exploring the modular nature of riboswitches and RNA thermometers, Nucl. Acids Res. 44 (11) (2016) 5410–5423.
[36] J. Kortmann, F. Narberhaus, Bacterial RNA thermometers: molecular zippers and switches, Nat. Rev. Microbiol. 10 (4) (2012) 255–265.
[37] J.S. Paige, T. Nguyen-Duc, W. Song, S.R. Jaffrey, Fluorescence imaging of cellular metabolites with RNA, Science 335 (6073) (2012) 1194.
[38] C. Wang, M. Sinn, J. Stifel, A.C. Heiler, A. Sommershof, J.S. Hartig, Synthesis of all possible canonical (3'-5'-Linked) cyclic dinucleotides and evaluation of riboswitch interactions and immune-stimulatory effects, J. Am. Chem. Soc. 139 (45) (2017) 16154–16160.
[39] X.C. Wang, S.C. Wilson, M.C. Hammond, Next-generation RNA-based fluorescent biosensors enable anaerobic detection of cyclic di-GMP, Nucl. Acids Res. 44 (17) (2016) e139.
[40] N. Nguyen Quang, G. Perret, F. Duconge, Applications of high-throughput sequencing for in vitro selection and characterization of aptamers, Pharmaceuticals (Basel) 9 (4) (2016).
[41] S. Ameta, M.L. Winz, C. Previti, A. Jaschke, Next-generation sequencing reveals how RNA catalysts evolve from random space, Nucl. Acids Res. 42 (2) (2014) 1303–1310.
[42] M. Popovic, P.S. Fliss, M.A. Ditzler, In vitro evolution of distinct self-cleaving ribozymes in diverse environments, Nucl. Acids Res. 43 (14) (2015) 7070–7082.
[43] Y. Yokobayashi, Applications of high-throughput sequencing to analyze and engineer ribozymes, Methods (2019) This issue.
[44] M. Sack, J. Stifel, S.G. Kreft, E. Deuerling, J.S. Hartig, Neomycin-dependent hammerhead ribozymes for the direct control of gene expression in Saccharomyces cerevisiae, Methods (2019) This issue.