UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

UPR 9002 du CNRS : Architecture et Réactivité de l'arN (ARN)

THÈSE présentée par :

Alexis AUTOUR

Soutenue le **17 Septembre 2018**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**
Discipline/ Spécialité : Science de la vie, biologie, biochimie

Amélioration et criblages de propriétés d'ARN aptamères fluorogènes en systèmes microfluidiques

THÈSE dirigée par :	
Dr RYCKELYNCK Michaël	Maître de conférences, IBMC, France
RAPPORTEURS :	
Dr SUESS Beatrix	Group leader, Technische Universität Darmstadt, Allemagne
Dr DESCROIX Stéphanie	Directrice de recherches, Institut Curie, France
AUTRES MEMBRES DU JURY :	
Dr RIVELINE Daniel	Directeur de recherches, IGBMC, France

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

UPR 9002 du CNRS : Architecture et Réactivité de l'arN (ARN)

THÈSE présentée par :

Alexis AUTOUR

Soutenue le **17 Septembre 2018**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Science de la vie, biologie, biochimie

Amélioration et criblages de propriétés d'ARN aptamères fluorogènes en systèmes microfluidiques

THÈSE dirigée par :

Dr RYCKELYNCK Michaël	Maître de conférences, IBMC, France

RAPPORTEURS :

Dr SUESS Beatrix	Group leader, Technische Universität Darmstadt, Allemagne
Dr DESCROIX Stéphanie	Directrice de recherches, Institut Curie, France

---

AUTRES MEMBRES DU JURY :

Dr RIVELINE Daniel	Directeur de recherches, IGBMC, France

Remerciements:

J'aimerai remercier tous les gens qui ont participé à la réalisation de cette thèse à différents niveaux.
En premier lieu les membres du jury qui ont accepté d'évaluer ces travaux : Dr Suess Beatrix, Dr Descroix Stéphanie et Dr Riveline Daniel.

Il me paraît important aussi de remercier Eric Westhof et Michaël Ryckelynck pour m'avoir accueilli au sein de l'unité et de cette équipe, ainsi que m'avoir laisser ma chance de réaliser cette thèse, ce qui n'aurait pas été possible sans vous.

Petite mention spéciale pour certain membres de l'équipe : Ketty, Stéphanie, Farah qui m'ont apporté leur aide en temps voulu pendant ces années (et surtout de m'avoir supporté… pour les nouveaux aussi).

J'aimerai souligner le rôle important de mon directeur de thèse Michaël Ryckelynck, sans qui cette formation n'aurait été complète *via* son support continu et de son accompagnement scientifique. Je suis vraiment reconnaissant qu'il m'ait pris sous son aile et l'investissement qu'il a porté a mon égard (l'apport de la rigueur, un esprit critique, et l'objectivité…) en plus d'être toujours très disponible, réactif et à l'écoute durant toutes ces années.

Sans oublier ma famille, que ce soit mes parents ou bien mes deux sœurs et mes amis qui m'ont soutenu aussi durant cette thèse.

J'aimerai finir mes remerciements par un générique comprenant toutes les personnes ayant participé de plus ou moins près à cette thèse :

The team

Michaël Ryckelynck..............Le Patron

Stéphanie Baudrey................La Maman
Ketty Pernod.....................Le feu follet
Farah Bouhedda.......................«Hodor»
Natacha Dentz..........................Watasha
Kristina Hintz..........................Le phare
Emilie Geersens...........La Marseillaise
Christian Rick.........................Le Barbu
Rémi Leblay..............................Reblay
Pablo Fernandez-Millan....Le Cristallo

Assistants production

Duygu Yilmaz...........Première stagiaire
Neus Mestre-Farras..............la catalane
Gauthier Lieber...........................Bergor

Special Guests

Sunny Jeng..................Testeur de burger
Ignazio Geraci........................Tutti lingui

The extras

Olivier Terenzi.........................Mr.T
Paula Lopez.................PP-Sunshine
Olivier Petitjean.............Schengele
Antoine Creugny.........Blacktoinou

Yacine Dahman...............Le Daron
Antinea Barbarit...................la M2
Ludovic Enkler.............«Rainbow»
Lauriane Gross............la crazy girl
Florian Pierre...................Flopinou
Elena Gonzalez................La latine
Marta Cela ....L'amérindienne
Agnès Gaudry........La mort-de-faim

Matthieu Sawaf.............La Machine
Andrien Franchet...........et ses cafés
Redmond Smyth....Testeur de bières
Anne Sophie Gribling....Maman bis
Melodie Duval.......«La vie est dure»
Semih Cetin.......................Ikea guy
Loukman Karim...Mitochondrie guy
Matteo Negroni.........Le basketteur
Marine Kanja................Support n°1
Safi Coulibaly...............Support n°2

Financial & emotional support

Les amis:
Rémy, Maxime, Kuku, Gerald...

La famille:
Mon père, ma mère
Mes deux soeurs

Others

et tous les autres....

Table of content

I. INTRODUCTION **9**

1. GENE EXPRESSION, REGULATION AND CELL-TO-CELL VARIABILITY 9
  - *1.1. Gene expression* *9*
  - *1.2. Gene expression regulation* *10*
    - 1.2.1. Regulation of transcription initiation 11
      - 1.2.1.1. Operon-specific transcription factors 11
      - 1.2.1.2. Regulon-specific transcription factors 12
      - 1.2.1.3. Alternative sigma factors 12
    - 1.2.2. Transcriptional modulators 13
      - 1.2.2.1. Protein-mediated termination 14
      - 1.2.2.2. Riboswitches and ribosensors 14
    - 1.2.3. Post-transcription regulation 17
      - 1.2.3.1. Modulating RNA stability 17
      - 1.2.3.2. Translation initiation regulation 18
    - 1.2.4. Regulation in eukaryotic cells 19
  - *1.3. Cell-to-cell variability* *20*
    - 1.3.1. Evidence and origin of cell-to-cell phenotypic heterogeneity in clonal populations 20
    - 1.3.2. Consequences of cell-to-cell phenotypic heterogeneity 22
2. CHARACTERIZING GENE EXPRESSION WITH SINGLE-CELL RESOLUTION 24
  - *2.1. In situ hybridization-based RNA imaging* *25*
  - *2.2. RNA imaging using genetically encoded probes* *26*
    - 2.2.1. Fluorescent protein-based approaches 27
      - 2.2.1.1. Properties and engineering of the green fluorescent protein 27
      - 2.2.1.2. Imaging RNA using fluorescent proteins 29
    - 2.2.2. RNA-based fluorogenic modules 30
      - 2.2.2.1. RNA-based fluorogenic modules and their applications 30
      - 2.2.2.2. Development strategies of light-up RNA aptamers 56
3. DROPLET-BASED MICROFLUIDICS: PRINCIPLE AND APPLICATIONS IN BIOLOGY 58
  - *3.1. Impact of microfluidics on biology* *58*
  - *3.2. Droplet-based microfluidics* *59*
    - 3.2.1. Droplet generation and manipulation 59
    - 3.2.2. Analysing nucleic acids with droplet-based microfluidics 60
      - 3.2.2.1. Digital droplet PCR 61
      - 3.2.2.2. Sequencing-based gene expression analysis with single-cell resolution 62
    - 3.2.3. Droplet-based microfluidics screening applications 63
      - 3.2.3.1. From IVC to µIVC 64
      - 3.2.3.2. Searching for new or improved catalysts using µIVC 64

3.2.3.3. Evolving RNA using µIVC 89

4. Thesis objectives 90

II. RESULTS & DISCUSSION 93

5. Improvement, characterization and application of Spinach-derived aptamers 93

5.1. A brief story of Spinach RNA aptamer 93

5.2. Development of an improved Spinach (iSpinach) RNA aptamer 95

5.2.1. iSpinach development 95

5.2.2. Additional data and discussion 109

5.2.2.1. iSpinach versatility in several buffers 109

5.2.2.2. Comparative analysis of DFHBI-binding aptamer sequences 109

5.2.2.3. iSpinach imaging in prokaryotes 111

5.3. iSpinach structural characterization 113

5.4. Engineering iSpinach in a semi-automated way 126

5.4.1. RNA-based sensors of metabolites 126

5.4.1.1. Ribozyme-based sensors 126

5.4.1.2. RNA-based fluorogenic sensors 127

5.4.1.3. Nucleic acids selection assisted by Next Generation Sequencing 128

5.4.2. iSpinach allosteric library design 129

5.4.2.1. Optimizing iSpinach sequence for sensor discovery 129

5.4.2.1. Experimental strategy 131

5.4.2.2. iSpinach-based sensor library screening 132

5.4.2.3. Sequence analysis 134

5.4.2.4. Validation of the selected fluorogenic theophylline responsive sensors 135

5.4.3. Conclusion on sensor selection 136

5.5. Conclusive remarks on Spinach chapter 137

6. Improvement, characterization and application of Mango RNA aptamers 138

6.1. The original Mango/TO1-Biotin module 138

6.2. Isolation of improved Mango aptamers 139

6.2.1. Setting-up R12 library screening 139

6.2.2. Isolating new Mango variants with improved performances 140

6.3. Structural characterization and structure-guided improvement of Mango aptamers 157

6.3.1. Structural characterization of Mango II 157

6.3.2. Structural characterization and optimization of Mango III 158

6.3.2.1. Crystal structure of Mango III/TO1-Biotin complex 158

6.3.2.2. Screening Mango III-derived mutant library 159

6.3.2.3. Sequence analysis 161

6.4. Conclusive remarks on the Mango chapter 165

III. GENERAL DISCUSSION, CONCLUSION AND OUTLOOK 169

7. Screening technology 169

7.1. Benefits of µIVC over competitive screening technologies 169
7.2. Improvement of the µIVC procedure 170
7.2.1. Increasing the screening throughput 172
7.2.2. Using an integrated device to better tailor optimized aptamers 172
7.2.3. Optimizing the post-screening analysis by systemizing NGS analysis 173
8. RNA-BASED FLUOROGENIC MODULES: CURRENT AND FUTURE APPLICATIONS 174
8.1. Live-cell imaging applications 174
8.1.1. Imaging messenger RNAs 174
8.1.2. Imaging small non-coding RNAs 175
8.1.3. Toward the development of a new generation of efficient fluorogenic modules 175
8.2. Other application of the optimized light-up aptamers 175
8.2.1. Developing optimized trans-acting RNA sensors 176
8.2.2. Developing new biosensors of small molecules 176
8.2.3. Developing synthetic riboswitches 177
IV. ANNEX 1 – MATERIALS AND METHODS 181
9. FLUORESCENCE MEASUREMENTS OF RNA PURIFIED 181
9.1. iSpinach, fluorescence measurement into different buffers. 181
9.2. Mango fluorescence measurements 181
10. *IN VIVO E.COLI* IMAGING 182
11. GEL-IMAGING AND STAINING 182
12. BARCODING AND µIVC SELECTIONS FOR ALLOSTERIC ISPINACH AND MANGO III 182
12.1. Barcoding 182
12.2. Digital PCR 182
12.3. Fusion and sorting for theophylline-responsive iSpinach sensor selection 183
12.4. Fusion and sorting for Super Mango III selection 184
12.5. NGS analysis 184
13. ENRICHMENT MEASUREMENT BY IVT MONITORING 184
14. SEQUENCES OF THE PRIMERS USED 185
V. ANNEX 2 – FRENCH SUMMARY 189
VI. REFERENCES 195

List of abbreviations:

asRNA	Anti-sense RNA
Cbl	Cobalamin
CRP	cAMP receptor protein
ddPCR	Digital Droplet PCR
DFHBI	3,5-difluoro-4-hydroxybenzylidene imidazolinone
DNA	DeoxyriboNucleic Acid
dPCR	Digital PCR
eGFP	enhanced GFP
FACS	Fluorescence-Active Cell Sorting
FADS	Fluorescence-Active Droplet Sorting
FISH	Fluorescent *In situ* hybridization
FIT	Forced intercalation probes
FMN	Flavin mononucleotide
FP	Fluorescent Protein
FRET	Förster Resonance Energy Transfer
GFP	Green fluorescent protein
HBDI	4'-hydroxybenzylidene-2,3-dimethylimidazolinone
IPTG	Isopropylthio-β-galactoside
ISH	*In situ* hybridization
IVC	*in vitro* compartmentalization
IVT	*In vitr*o Transcription
lncRNA	Long non-coding RNA
LSI	Large Scale Integration
MB	Molecular Beacon
MBS	MS2 Binding Site
MCP	Membrane Coat Protein
miRNA	microRNA
mRNA	messenger RNA
ncRNA	non-coding RNA
NGS	Next Generation Sequencing
NMM	N-methyl mesoporphyrin IX
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase Chain Reaction
qPCR	quantitative PCR
QY	Quantum Yield
RBP	RNA Binding Protein
RBS	Ribosome Binding Site

RNA	RiboNucleic Acid
RNAP	RNA polymerase
RNAT	RNA thermometers
SAM	S-adenosylmethionine
scRNA-seq	single-cell RNA sequencing
SELEX	Systemactic Evolution of Ligands by Exponential enrichment
smFISH	Single molecule Fluorescent *In situ* hybridization
sRNA	Small RNA
TO	Thiazole Orange
tRNA	transfer RNA
UCI	Unique Cell Identifier
UDI	Unique Droplet Identifier
UMI	Unique Molecular Identifier
UTR	Untranslated Region
μIVC	Microfluidic-assisted IVC

Introduction:

1. Gene expression, regulation and cell-to-cell variability
2. Characterizing gene expression with single-cell resolution
3. Droplet-based microfluidics: principle and applications in biology

I. Introduction

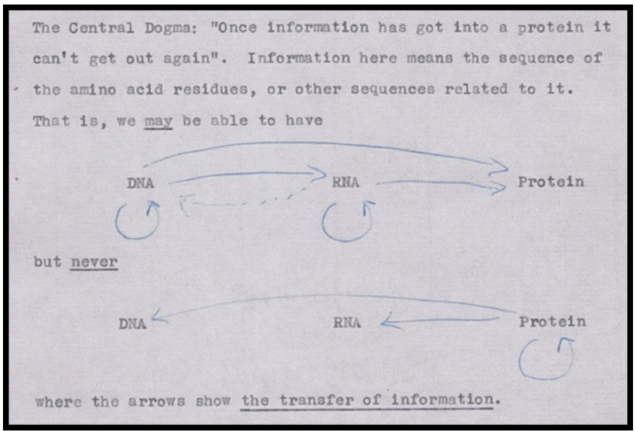
In a metaphoric and (over)simplified view, the cell can be seen as a computer made of software and hardware elements: the former corresponding to the genetic information stored into the DNA and the latter being made of the machineries allowing for expressing this information (transcription and translation machinery) as well as a power supply, the metabolism, that fuels the whole system in energy. Further extending the cell/computer analogy, the information is not simply constitutively expressed but, like in an informatic script, this expression is rather subjected to conditions and adapted accordingly. Therefore, even though understanding how the different machineries (e.g. RNA polymerase, ribosome…) work on a mechanistic point of view is of prime importance, understanding regulatory events controlling this expression as well as their dynamics is even more important. However, given the complexity of these systems and their high degree of integration/interconnection, such characterization requires sensitive and efficient tools and technologies to be developed. In this view, this PhD thesis worked aimed at developing such tools in the form of new fluorogenic RNA modules that would ultimately allow monitoring gene expression at various levels.

Prior to presenting the results and developments I performed during this thesis, I will first introduce it by giving a rapid overview of how genes expression is modulated/regulated (especially through RNA-based mechanisms), how such regulations are characterized and why new technologies need to be developed.

1. Gene expression, regulation and cell-to-cell variability

1.1. Gene expression

According to the central dogma of molecular biology enounced by Francis Crick in 1958 (Crick, 1958), gene expression flows from the DNA, in which the information is stored, to proteins *via* a transient intermediate, the messenger RNA (mRNA). Therefore, the DNA is first transcribed into a mRNA by an RNA polymerase machinery and this mRNA is then translated into protein by the translation machinery, especially the ribosome (Figure 1). This vision is nevertheless slightly reductionist since the protein itself can also be viewed as an intermediate and one can consider the resulting phenotype as being the true gene expression product. Moreover, not all the genes encode for a protein and the phenotype-expressing molecule can directly be the RNA. For instance, in bacteria the 23S ribosomal RNA (28S in eukaryotes) carries the peptidyl-transferase activity (Nissen et al., 2000). Similarly, in most RNase Ps (the enzyme maturing the 3' of transfer RNA (tRNA) molecules) the catalytic activity is also carried by the RNA moiety of the ribonucleoprotein (Guerrier-Takada et al., 1983). Beside these examples of ribozymes, non-coding RNAs (ncRNA) can also be endowed with additional functions such as decoding (tRNAs) and regulation (see below).



The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it. That is, we may be able to have

DNA RNA Protein

but never

DNA RNA Protein

where the arrows show the transfer of information.

Figure 1: The central dogma in molecular biology from unpublished Crick's note. Taken from Matthew Cobb 2017.

Whatever the type of gene, only the subset of the so-called housekeeping genes is constitutively expressed, whereas the expression of the other genes is tightly controlled and adapted to cell needs and environmental conditions. As this will be presented below, this regulation involves complex intertwined pathways, interconnecting the whole system with the net result of ensuring cell homeostasis.

1.2. Gene expression regulation

The effectors of regulatory pathways are as diverse as protein, RNA or even metabolites and they may be either dedicated molecules or the expression product of the controlled gene/pathway (or of another one). For simplification reasons I will restrict this chapter to a few well-studied prokaryotic cases and I will only occasionally extend it to eukaryotes.

In 1961, Jacob and Monod provided a first glimpse into gene regulation with their famous publication on the mechanism controlling the expression of genes involved in lactose consumption in *Escherichia coli* (Jacob and Monod, 1961). A first important concept introduced in this work was the idea of bacterial genes being organized in operon; i.e. several genes involved in the same function, are placed under the control of a unique promoter and therefore co-transcribed into a single polycistronic RNA (Figure 2 model II). A second important outcome of this work was the proposition that cellular factors (interestingly in the original paper these molecules were drawn as being RNAs) could regulate gene expression following two models: i) either by a mechanism involving a direct binding to the genomic DNA (Model I) which would define them as transcription factors or ii) by an action at the mRNA level (Model II) and would make them co/post-transcriptional modulators. Nowadays, it is known that both models and several others are indeed at work in the cell to modulate gene expression.

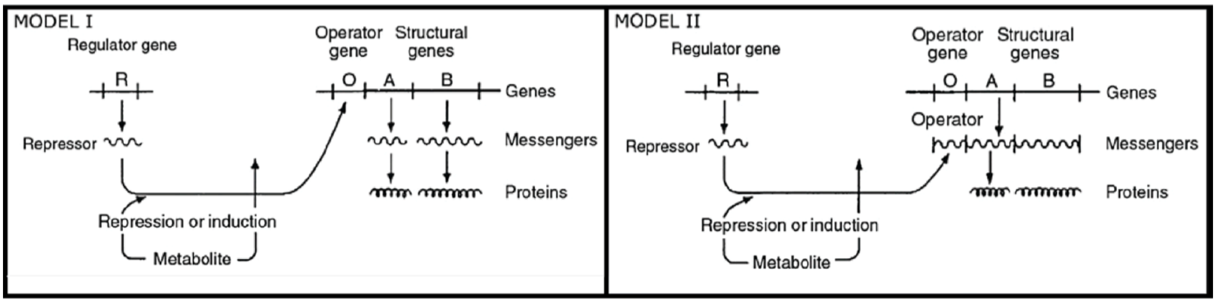

MODEL I
Regulator gene
R
Repressor
Repression or induction
Metabolite
Operator gene
Structural genes
O
A
B
Genes
Messengers
Proteins
MODEL II
Regulator gene
R
Repressor
Repression or induction
Metabolite
Operator gene
Structural genes
O
A
B
Operator
Genes
Messengers
Proteins

Figure 2: Regulation models proposed by Jacob and Monod in 1961. Model I depicts a gene regulation by a regulatory molecule binding the DNA operator region of the gene. In Model II, the effector (repressor or activator) acts at mRNA level.

1.2.1. Regulation of transcription initiation

Transcription factors are proteins that modulate gene transcription by direct interaction with the DNA. Upon binding to DNA, these factors will either favour (activator) of prevent (repressor) the recruitment of the RNA polymerase (RNAP). Moreover, these factors can either be specific to a gene (or an operon) or act on larger sets of genes (or operons) distributed throughout the genome and that forms a regulon. Finally, instead of direct binding to DNA, the factor can also guide the RNAP to a subset of transcription promotors as it is the case with sigma factors.

1.2.1.1. Operon-specific transcription factors

The most famous operon-specific bacterial transcription factor is the repressor of the lactose (lac) operon that controls the expression of genes coding for a β-galactosidase (LacZ), a galactose permease (LacY) and a thiolgalactoside transacetylase (LacA). These enzymes are involved in lactose metabolism and are respectively in charge of lactose cleavage, entry and detoxification allowing for converting lactose into monosaccharide used as alternative carbon source. In the absence of lactose, the operon is poorly expressed because of the binding of an allosteric protein, the Lac^{i} repressor, to the operator region next to the transcription promoter of the lac operon (Figure 3) and that interferes with the recruitment of the RNAP. However, this repression can be suppressed by the addition of lactose (or an analogue like the Isopropylthio-β-galactoside or IPTG), which enters the cell *via* the lactose permease. The lactose is isomerized into allolactose that binds to the allosteric repressor. The latter losses its affinity for the promoter and leaves it, allowing the RNAP to be recruited and the operon to be expressed.

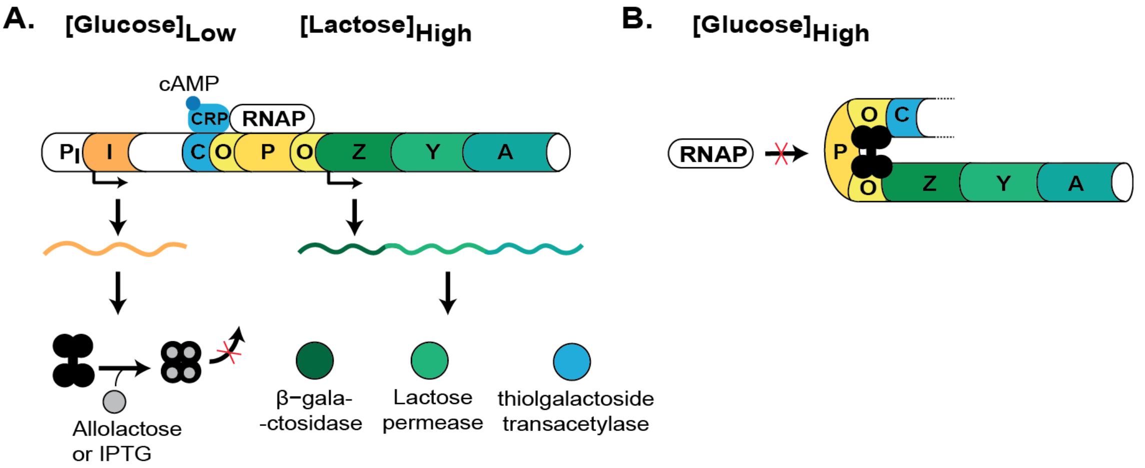


Figure 3: Operon lactose regulation. **(A)** In the presence of lactose, allolactose (in grey) is formed and binds to Lac^i allosteric repressor (in black) encoded by *lacI* locus (in orange). At low glucose concentration, cAMP concentration rises and activated CRP protein, which helps RNA polymerase (RNAP) recruitment onto lactose promoter. Both conditions increase the transcription rate of Lac operon transcription and lead the synthesis of enzymes involved lactose catabolism. **(B)** However, when glucose concentration is high, Lactose repressor (in black) binds to DNA operator region (Lac^O) and closes lactose promotor, preventing RNAP recruitment.

1.2.1.2. Regulon-specific transcription factors

Operon-specific transcription factors are not so frequent and, most of the time, transcription factors regulate several genes (or operons) belonging to the same regulon. Besides the obvious economy of energy made possible by the use of common transcription factors, the main advantage of a sharing them between several different pathways is the possibility of interconnecting and regulating all of them simultaneously upon the sensing by a single stimulus. These interconnections and the interplay of different stimuli allow cell homeostasis to take place and to be preserved.

Lac operon expression regulation is not exclusively under the control of the operon-specific repressor since its transcription is also activated by the cAMP receptor protein (CRP). CRP senses the cell concentration in cAMP, which accumulates upon glucose starvation (Pastan and Perlman, 1970). The CRP-cAMP complex then binds a DNA region upstream the lac transcription promoter and helps the recruitment of the RNAP *via* its α subunit. Moreover, CRP-cAMP also enhances the transcription of a large set of genes involved mainly in sugar catabolic enzymes (Kolb et al., 1993) and, this way, allows transducing the sensing of low glucose concentration to the whole system. CRP is only one of the numerous transcription factors known today and many other proteins like, for instance, ArgR (which controls a regulon of ~ 5 genes involved in arginine metabolism pathways (Maas, 1994)) or LexA (the factors expression of ~ 40 genes involved in the SOS response (Butala et al., 2009) have also been identified as global regulators.

1.2.1.3. Alternative sigma factors

In addition to the transcription factors introduced above, adapting the composition of the RNAP can also be used to modulate gene expression. Indeed, in bacteria, the RNAP is made of five subunits (two α, one β, one β' and one γ) that compose the core enzyme, which can further form a holoenzyme upon association with a sigma (σ) factor that guides the RNAP to the promoter. In normal laboratory growth conditions, the RNAP is associated with the housekeeping factor RpoD (or σ^{70}). However, various stress conditions triggered by environmental changes (e.g. nutrient depletion, pH or temperature shock) leads to the synthesis of alternative sigma factors that compete with and eventually replace RpoD on the RNAP (Österberg et al., 2011). This exchange reprograms the RNAP toward alternative promoters and changes the transcriptional landscape of the cell. As this was the case above, each σ factor will drive the expression of its own regulon. For instance, RpoF controls the regulon involved in the cell motility, RpoN controls the regulon of response to nitrogen starvation, RpoE controls the regulon of response stress at the membrane.

To date, RpoS is the most studied and the best characterized σ factor (Österberg et al., 2011). This factor is the master regulator of the general stress response in *E. coli*. Upon stress (e.g. heat shock, nutrient starvation, acidic shock, entry in stationary growth phase…) sensing, the factor is loaded onto the RNAP and triggers a massive change in the transcriptional landscape as it controls the expression of up to 500 of the 4500 genes of the bacterium. Due to the key role of RpoS in cell adaptation, the expression of its coding gene (*rpos*) has been extensively studied and revealed that it is subjected to variety of mechanisms acting at every post-transcriptional step (Figure 4). These different mechanisms will be introduced below to illustrate the different class of mechanisms at work.

1.2.2. Transcriptional modulators

The transcription and σ factors introduced above allow modulating the recruitment efficiency of the RNAP at a given promoter. Therefore, they act at the transcription initiation level. However, even though the transcription has been started, several regulatory mechanisms interfering with the elongation have been identified. Moreover, whereas transcription factors were mainly proteins, post-transcription regulators involve both proteins and RNAs.

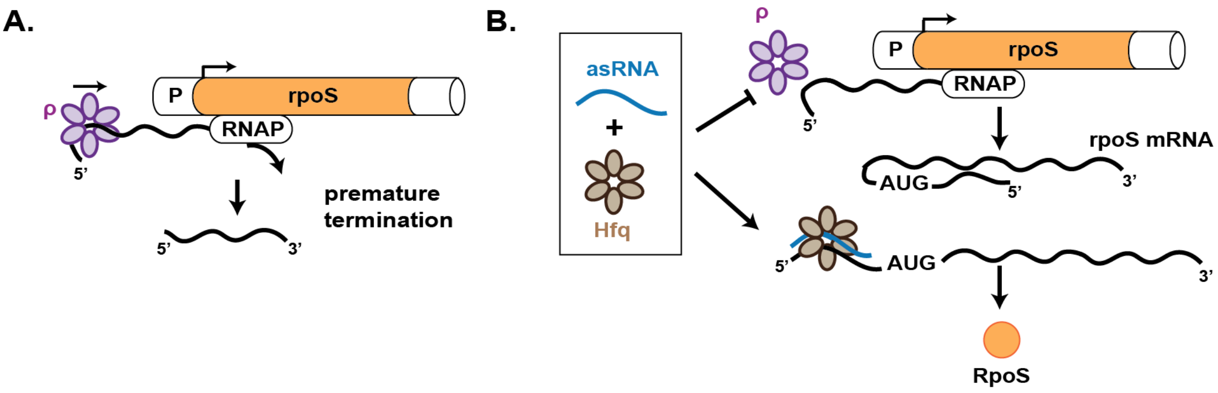


Figure 4: RpoS expression regulation. (A) In regular growth condition, Rho helicase (ρ) is loaded onto nascent *rpos* mRNA and climbs up mRNA to reach RNAP and triggering premature termination **(B)** In stress conditions, other factors such as chaperone protein (Hfq) and small antisense RNA (asRNA) bind nascent RNA preventing Rho helicase binding. Hence, it allows the synthesis of a full-length *rpos* transcript. In addition, those factors help to free AUG codon start that is embedded into mRNA structure, allowing for the mRNA to be translated.

1.2.2.1. Protein-mediated termination

During the transcription, the ATP-dependent Rho helicase can be recruited at the level of long unstructured regions usually fund after the coding region. Upon loading on the mRNA, Rho proceeds along the transcript until reaching the RNAP. The contact between Rho and RNAP unwinds RNA-DNA hybrids, which ultimately leads to transcription termination and mRNA release.

It was recently shown that Rho could also be loaded in the 5' untranslated region (UTR) of ~ 300 mRNAs in *E. coli* and triggers a premature Rho-dependent transcription termination (Sedlyarova et al., 2016). This was nicely exemplified by the case of *rpoS* expression regulation. Indeed, in normal laboratory growth conditions, a strong Rho-dependent termination occurs at the beginning of *rpoS* coding region, preventing the synthesis of RpoS. However, under stress conditions three small RNAs (sRNA) are produced and anneal to *rpos* mRNA on a site between the Rho loading region and the termination site. The presence of these sRNAs correlates with a drastic increase in full-length *rpos* transcripts, suggesting that Rho-dependent termination constitutes a first level of action of small RNAs.

1.2.2.2.Riboswitches and ribosensors

Beside the Rho-dependent mechanisms, transcription termination can also take place in a Rho-independent manner. This typically involves the formation of a GC-rich stem-loop followed by a stretch of Us that destabilizes the transcriptional complex leading to its dissociation and transcription arrest. This terminator stem-loop structure can exist in equilibrium with an anti-terminator structure that does not affect polymerase processivity and allow the transcription to further proceed (Figure 5). Such terminator/anti-terminator equilibrium is actually exploited in several RNA-based regulatory strategies like, for instance, riboswitches and ribosensors.

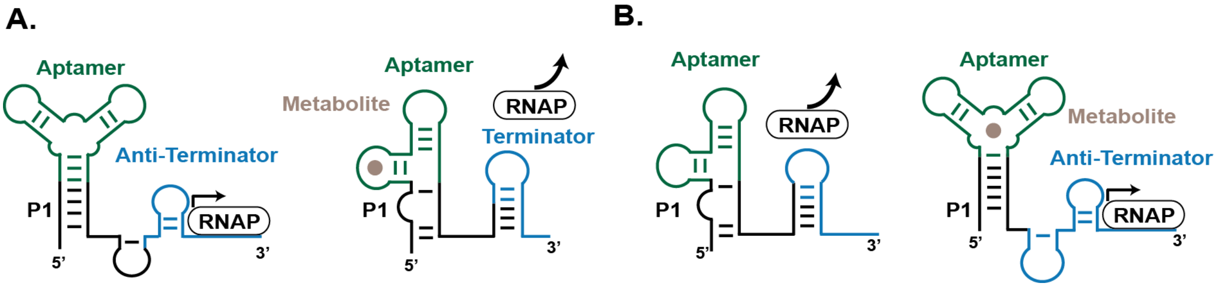


Figure 5: Simplified view of riboswitch-mediated transcription regulation. A riboswitch is composed of a sensing platform (aptamer in green) and a communication module (in black) linked to an expression platform (transcription terminator or anti-terminator in light blue). **(A)** Negative regulation of transcription. The presence of the metabolite (brown circle) leads to riboswitch conformation change and the formation of terminator structure triggering premature transcription termination. **(B)** Conversely, in positive regulation the presence of metabolite leads to change the formation of an anti-terminator structure allowing the transcription to proceed.

Riboswitches are RNA folds that regulate gene expression in *cis* (part of the sequence of the regulated RNA) *via* the sensing of small molecules and that are organized in three functional domains: the aptamer, the communication module and the regulatory platform (Figure 5). An aptamer is a short nucleotide sequence folding into a three-dimensional structure able to specifically recognize a molecule with an affinity that may approach (and sometimes even exceed) that of antibodies (Iliuk et al., 2011). In the case of riboswitches, the affinity of the aptamer for its ligand can vary from millimolar down to nanomolar range and is adapted to be relevant of the sensed molecule and the cellular context. For instance, metabolites are typically found at millimolar concentration in the cell and the decision to turn-on or off genes coding for the enzymes involved in the corresponding pathway has be taken when metabolite concentration pass a critical millimolar concentration. Therefore, it makes sense that the aptamer part of these riboswitches has limited affinity (but high selectivity) for their ligand, since a much higher affinity would constantly saturate aptamer making it non-responsive to cell-relevant concentrations. Other riboswitches display a much higher affinity for the sensed molecule. This is typically illustrated by the recently described ppGpp-responsive riboswitch (Sherlock et al., 2018). This aptamer displays nanomolar affinity for the alarmone ppGpp, a secondary metabolite produced in the bacterium in response to stress conditions and that is the main actor of the stringent response (Harms et al., 2016). In this second case, it makes sense for the aptamer to be able to sense its ligand in a fast and efficient way to rapidly adjust bacterial gene expression and enable a fast stress adaptation. Another important feature of riboswitch-embedded aptamer is their structural plasticity. Indeed, the binding event of the ligand triggers a local structural modification (e.g. remodelling, stabilization…) of the aptamer that is instrumental for the regulatory event. This structural change is then transmitted to the regulatory platform by the second key element of the riboswitches: the communication module. This module is usually a loose helix (e.g. P1 on Figure 5) that gets stabilized upon complex formation between the aptamer and the ligand. Alternatively, ligand binding can lead to a strand displacement event within the communication (Figure 5). Finally, upon stabilization (or remodelling) the communication module induces a structural change of the regulatory platform (the third important component) and, in transcriptional riboswitches, switch the anti-terminator structure to a terminating one and *vice versa*.

Transcriptional riboswitches are typically found in Gram-positive bacteria such as *Bacillus* where they have been extensively studied (Mandal and Breaker, 2004a). For instance, in *B. subtilis*, the *rib* operon encoding enzymes involved in flavin biosynthesis contains a riboswitch in its 5' UTR, that leads to premature transcription termination of this operon (Winkler et al., 2002). On a mechanistic point of view, when the concentration of flavin mononucleotide (FMN), the final product of the pathway, exceeds a critical concentration, it binds to the aptamer contained in *rib* mRNA 5' UTR and triggers the formation of a terminator hairpin leading to a premature transcription arrest and the extinction of *rib* gene expression.

Beside gene expression inhibition, metabolite sensing can also positively act on gene expression. For instance, the presence of adenine switches the *pbuE* riboswitch to an "ON" conformation, leading to the expression of a gene coding for a purine efflux pump that is been proposed to protect the cell, especially against toxic purine base analogs (Mandal and Breaker, 2004b).

Instead of direct aptamer-based detection, metabolite sensing can also be mediated by larger molecules, such as transfer RNA (tRNA). This is typically observed in Gram-positive bacteria, where a T-box element can be found in 5’ UTR region of genes involved in amino acid metabolism (Mandal and Breaker, 2004a). This regulation mode relies on the fact that nutrient starvation leads to the reduction of the cellular amino acid pool and, consequently, to the reduction of the aminoacylated fraction of tRNA. Non-aminoacylated tRNAs are then recruited at the level of T-box elements to which they pair *via* their anti-codon and 3’ free amino-acceptor stem (Figure 6). This RNA-RNA complex stabilizes the regulatory platform of the T-box into an anti-terminator structure leading to down-stream gene transcription and gene expression.

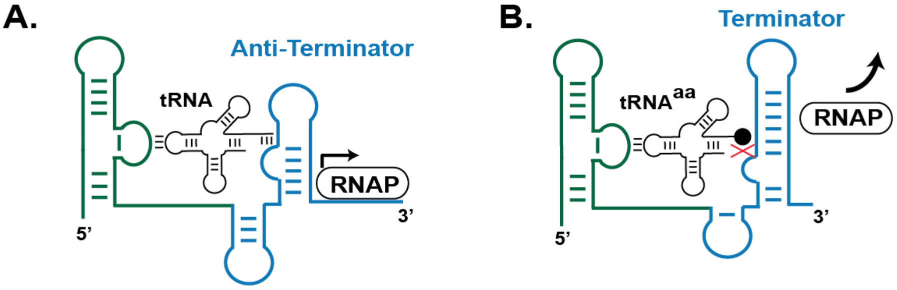

A.
Anti-Terminator
tRNA
RNAP
5’
3’
B.
Terminator
tRNA^{aa}
RNAP
5’
3’

Figure 6: Schematic T-box regulation. (A) A non-aminoacylated tRNA binds to the T-box element *via* its anticodon and its 3’ free acceptor stem leading to the stabilization of an anti-terminator transcription structure. **(B)** Aminoacylated tRNA can no longer bind to T-box and the terminator structure persists.

Interestingly, more complex scenarios exist like in *Clostridium acetobutylicum* that regulates the transcription of some of its genes by combining two sensing strategies (Figure 7). This is typically the case of the *ubiG* operon, which encodes enzymes (a cystathionine γ-lyase and a cystathionine β-synthase) involved in methionine to cysteine conversion (André et al., 2008). The transcription of this operon is first controlled by a T-box element contained in the 5’ UTR leader sequence and that is activated by non-aminoacylated tRNA^{Cys}.

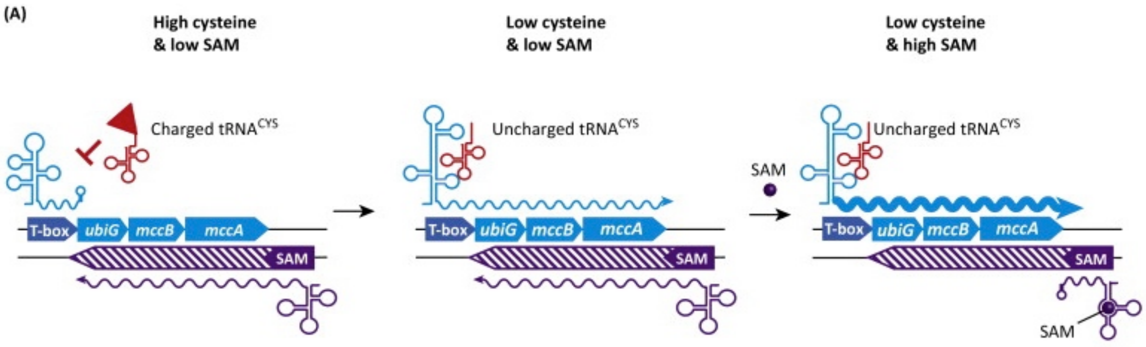

(A)
High cysteine
& low SAM
Charged tRNA^{CYS}
T-box
ubiG
mccB
mccA
SAM
Low cysteine
& low SAM
Uncharged tRNA^{CYS}
T-box
ubiG
mccB
mccA
SAM
Low cysteine
& high SAM
SAM
Uncharged tRNA^{CYS}
T-box
ubiG
mccB
mccA
SAM
SAM

Figure 7: Example of dual regulation. At high cysteine concentration, a T-box (in light-blue) negatively regulates the expression downstream genes. When cysteine concentration is low, uncharged tRNA (in red) elicits operon transcription. However, a second layer of regulation (hashed purple) generates an anti-sense RNA of ubiG operon that prevents ubiG expression. This second layer is under the SAM riboswitch negative regulation control. When SAM concentration is high, the SAM riboswitch leads to premature termination of this anti-sense RNA and the operon is expressed. Figure adapted from Mellin and Cossart, 2015.

However, in this system, a reduced level of cysteine is not enough to trigger gene expression due to a second layer of regulation. Indeed, the expression of this sulfur metabolic operon is also under the dependence of S-adenosylmethionine (SAM), an intermediate of the methionine to cysteine conversion. A SAM-responsive riboswitch found downstream *ubiG* attenuates the expression of an antisens RNA that negatively regulates the transcription of *ubiG*. Therefore, the gene is subjected to a dual regulation and require both a reduced concentration of cysteine (T-box) and an elevated concentration of SAM (SAM riboswitch) to get expressed. *ubiG* is not an isolated case and, in other systems, in which non-coding regulatory RNAs act in tandem with riboswitches have been described (Mellin and Cossart, 2015).

1.2.3. Post-transcription regulation

Upon synthesis, the mRNA still needs to persist in the cell and to be translated into proteins. These two features offer additional possibilities of regulating gene expression at a post-transcriptional level.

1.2.3.1. Modulating RNA stability

Due to the high RNase activity cell content, bacterial RNAs have an average half-life of ~ 3 minutes (Rauhut and Klug, 1999). However, this is just an average value and RNA lifetime strongly depends on the structure of the molecule and the action of external factors. Among the cellular RNA pool, mRNAs are rather labile as expected for their transient function and are subjected to a much faster turn-over. However, the lifespan of given mRNA is still variable and can be modulated by several factors. First, the presence of highly structured domains within the 5' or the 3' UTR is known to strongly impact mRNA half-life (Rauhut and Klug, 1999). Second, the pairing of antisense RNA (asRNAs) or protein to the 3'UTR has been shown to impede ribonuclease to initiate mRNA decay(Ren et al., 2017), therefore offering a way of regulating gene expression through the modulation of RNA stability.

An interesting case of regulation is found at work in the control of the expression of the gene coding for the glutamine-fructose-6-phosphate amidotransferase (*glmS*) of most gram-positive bacteria. This enzyme is in charge of Glucosamine-6-Phosphate (GlcN6P) synthesis. Like in the examples introduced above (§ 1.2.2.2.), *glmS* expression is controlled by the presence of the metabolite at the level of a riboswitch present in the 5' UTR of the mRNA (Klein and Ferré-D'Amaré, 2006). However, the originality of this system relies on the fact that the switch is not conformational (like the others known cases) but it is instead an activity switch. Indeed, the aptamer moiety of *glmS* specifically recognizes the GlcN6P and uses it as a co-factor in a self-cleaving reaction that ultimately leads to the rapid clearance of the cleaved mRNA.

1.2.3.2. Translation initiation regulation

In bacteria, mRNA translation starts with the proper positioning of the small ribosomal subunit 30S on the Ribosome Binding Site (RBS) also known as the Shine-Dalgarno sequence (Shine and Dalgarno, 1974). Modulating the accessibility of this sequence represents a very efficient way of regulating mRNA translation. A wide variety of mechanisms controlling RBS accessibility has been described (Kozak, 2005) and can involve regulatory RNAs acting in *cis* or in *trans*.

1.2.3.2.1. Translation regulatory RNAs acting in cis

Most of the riboswitches found in Gram-positive bacteria belong to the class of transcriptional riboswitches introduced above. In Gram-negative bacteria, the situation is different since most of the riboswitches act at the translational level. These riboswitches display in same architecture than their transcriptional counterparts, except that instead of having a regulatory platform switching between a transcriptional terminator and an anti-terminator, their platform acts by sequestering or releasing the RBS (Figure 8.A). The riboswitch controlling the expression of *thiM* in the Gram-negative *E. coli* is one of the best-studied systems. In absence of thiamin pyrophosphate (TPP), the structure is in an "ON" conformation in which the RBS is accessible for translation initiation, however upon TPP binding to the aptamer a structural change leads to RBS sequestration.

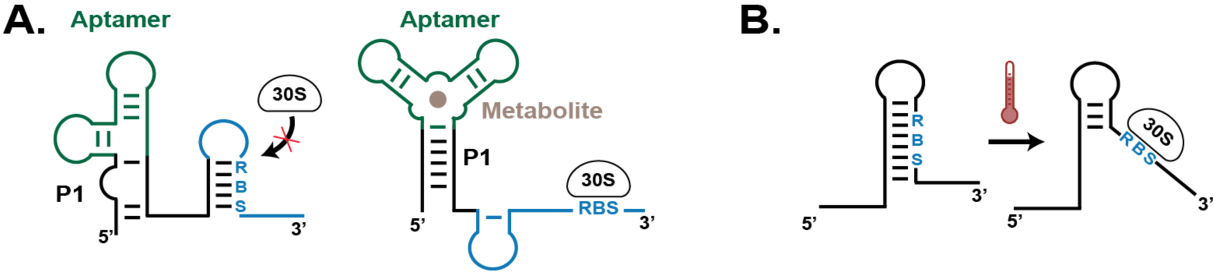


Figure 8: Simplified view of riboswitch-mediated translation initiation regulation. (A) Riboswitch-mediated positive regulation of translation initiation. At low ligand (metabolite) concentration, the ribosome binding site (RBS) is trapped into a stem structure that prevents ribosome binding. However, in the presence of a metabolite (brown circle), the riboswitch changes its conformation and releases the RBS, allowing for translation initiation to take place. **(B)** Temperature-sensitive translation initiation. Similarly to riboswitches, RBS is sequestered into stem structure that progressively opens and released the RBS while the temperature increases.

Besides small molecule sensing by aptamer, structure switching can also be triggered by physico-chemical conditions like pH and temperature. For instance, a pH responsive element (PRE) is found embedded within the 5' UTR of *alx* mRNA which encodes a putative transporter (Nechooshtan et al., 2009). Under neutral pH conditions, the RNA adopts a folding in which the RBS is sequestered in a secondary structure. Switching the pH to alkaline induces pausing of the RNA polymerase, giving time to RNA to fold into an alternative structure in which the RBS is accessible and the mRNA can be translated. RNA thermometers (RNAT) represent a second class of motifs responding to physico-chemical conditions, as they allow sensing temperature changes (Figure 8.B). These RNAs are found in the 5'UTR of genes involved in cold and heat shock response (Kortmann and Narberhaus, 2012) as well as the virulence of pathogenic bacteria by allowing them to sense their successful entry into a mammalian cell for instance (Johansson et al., 2002). On a mechanistic point of view, the RNAT is a secondary structure sequestering the RBS as long as the temperature stays below a critical threshold. However, an increase of the temperature leads a melting of the structure and the release of the RBS that is free to recruit the ribosome.

1.2.3.2.2. Translation regulatory RNAs acting in trans

Beside the *cis*-acting RNA motifs, RBS accessibility can also be modulated by the action of *trans*-acting small regulatory RNAs (sRNA) (Wagner and Romby, 2015). First, these RNAs can act by binding directly to the RBS and make it inaccessible to the ribosome (e.g. Figure 4). This action is usually complemented by the rapid degradation of the targeted RNA through the action of RNases acting on the sRNA/mRNA duplex. Alternatively, sRNAs can act by destabilizing a secondary structure element sequestering the RBS. This scenario is typically encountered in *rpos* mRNA in which the 5' UTR extremity forms hairpin structure sequestering the RBS. However, under stress conditions three sRNAs (DsrA, RprA and ArcZ) are produced and binds *rpos* 5' UTR. In addition to their role in the Rho-dependent mechanism introduced in §1.2.2.1., these RNAs also induce a conformational change that destabilize the sequestering secondary structure and release the RBS, allowing for *rpos* translation to be initiated (Battesti et al., 2011).

1.2.4. Regulation in eukaryotic cells

All the examples used above to illustrate the different strategies exploiting RNA to regulate gene expression were focused on bacterial systems for simplification reasons. However, RNA-based regulation mechanisms exploiting non-coding RNAs (ncRNAs) are also instrumental for eukaryotic systems homeostasis. In this view, microRNAs (miRNA) is probably the most important regulatory actors allowing for targeting any gene using the same machinery while just changing the sequence of the guiding miRNA (Bartel, 2009). More recently, >200 nucleotide-long ncRNA (lncRNA) have been implicated in a plethora of key biological mechanisms such as chromatin remodeling, splicing, regulate mRNA stability (Geisler and Coller, 2013). These RNAs have multiple origins (e.g. derived from introns or specific loci…) and their functions are not always fully elucidated. Their roles can also range from transcriptional to translational regulation. Some famous lncRNAs such as Xist (X inactive specific transcript) plays regulatory functions at the epigenetic level by repressing the expression of one X chromosome in mammals (Plath et al., 2002).

In compartmented eukaryotic cells, RNA localization and asymmetrical distribution are also important for many spatial gene expression events. Usually, an mRNA is transcribed in the nucleus and then exported to the cytoplasm to be translated into proteins. Directing an mRNA to a position of interest and translating it *in situ* might be more advantageous than addressing the multiple proteins produced from this RNA. Spatial localization also decentralizes the gene expression regulation allowing fastest local response to external stimuli.

The different examples introduced throughout this first chapter take place at the level of the gene or the mRNA. Of course, even though an mRNA is translated into a protein, the latter can still be targeted by several post-translational mechanisms such as phosphorylation, sequestering by another protein, tagging for the degradation… However, these mechanisms are out of the scope of this introduction, but the reader can find additional information in one of the recent reviews on the topic (Cain et al., 2014). The idea here was not to establish a comprehensive review of the current knowledges on RNA-based gene expression regulation but rather to give a rapid overview, through simple examples, of the central role of RNA in the gene expression regulation.

Being able to characterize such regulatory mechanisms requires the use of dedicated methodologies operating at the proper resolution. Indeed, up to recently, the limited sensitivity of analytical technologies imposed to use large pools of cells as starting material, assuming that every cell in this population was identical to the others. It is however established today that even though cells belong to a clonal population growing in the same environment, sometimes-significant cell-to-cell variability of gene expression can be observed.

1.3. Cell-to-cell variability

That isogenic cells can display strong cell-to-cell phenotypic heterogeneity is obvious when considering higher eukaryotes in which a single initial cell (the oocyte) divides and differentiates into a variety of cell types to form an organized organism made of specialized organs. In that case, even though cells share the same genetic information, they will differently express it in order to fit their environment and acquire their specific function. As a consequence, differentiated cells will, for instance, acquire a specific morphology, display different subsets of cell-specific receptor or even differently express genes involved in various pathways. Conversely, the extent to which gene expression may vary from one cell to the other within the same clonal population was less obvious *a priori*. However, this heterogeneity is now well established in organisms ranging from mammalian cells to bacteria (Vera et al., 2016). As before, and for simplification reason, I will focus on bacteria for which both origins and consequences of cell-to-cell phenotypic variability have been extensively studied (Veening et al., 2008a).

1.3.1. Evidence and origin of cell-to-cell phenotypic heterogeneity in clonal populations

In 1957, Novick and Weiner described the first case of phenotypic heterogeneity within an isogenic population of *E. coli* through an elegant experiment in which they studied the ß-galactosidase activity of individual cells upon various degrees of induction (Novick and Weiner, 1957)(Figure 9). More precisely, they first cultured clonal populations of *E. coli* bacteria in the absence or in the presence of low or high concentration of thiomethyl-ß-D-galactoside, an inducer that relieve lactose repressor and triggers *lac* operon transcription (see §1.2.1.1. above). Then, bacteria of each population were individualized and recultured while maintaining the inducer at the same concentration. Assaying the ß-galactosidase activity of each sub-culture then revealed two distinct phenotypes: elevated or very low of ß-galactosidase activity, leading the authors to defined it as an "all-or-none" phenotype (Figure 9). Interestingly, whereas all of the cells originally subjected to a high induction showed a ß-galactosidase positive phenotype and none of the culture inoculated with a non-induced cells gave such a phenotype, those cultures inoculated with cells subjected to a low level of induction gave a mixture of both phenotypes in which the rate of ß-galactosidase positive sub-culture was proportional to the initial induction rate (i.e. a population initially induced to produce 30% of the maximal phenotype led to 30% of single cells displaying the phenotype). This experiment is actually the first example of digital phenotypic characterization (i.e. the characterization of gene expression throughout a cell population with a single cell resolution, Figure 9) and justifies why such resolution is necessary to properly understand biological mechanisms as complex as gene expression regulation.

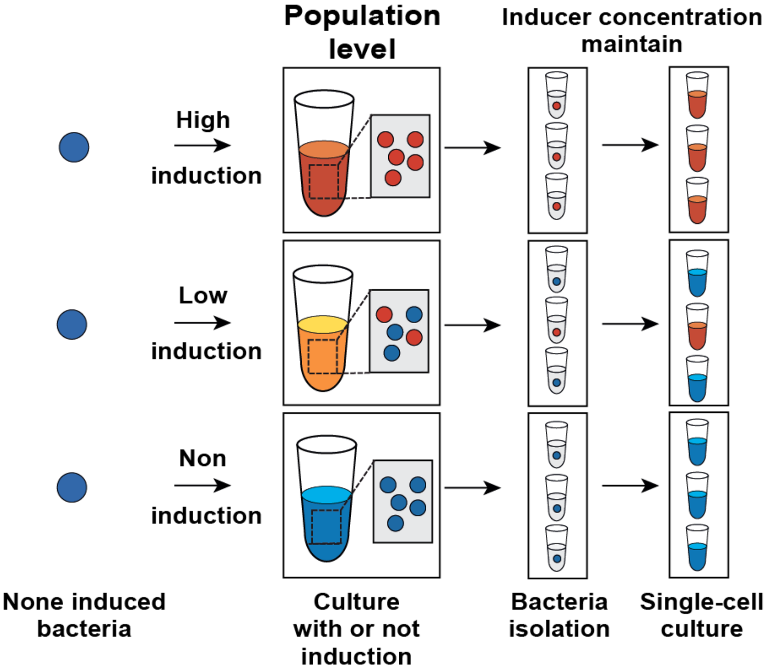


Figure 9: Phenotype diversity and single-cell culture approach. The experiment by Novick & Weiner is schematized. Uninduced bacteria are subjected to different concentrations of lactose pathway inducer. Afterwards, bacteria are individualized in tube using limiting dilution the same inducer concentration and their β-galactosidase activity assayed. This allowed observing an "all-or-none" phenomenon at low/intermediate inducer concentration where part of the bacteria displayed β-galactosidase activity and part did not.

Some 50 years later, technological advances allowed better understanding the molecular basis of the phenomenon at work using an integrated approach combining microfluidics and live-cell imaging(Cai et al., 2006; Choi et al., 2008). Indeed, monitoring either ß-galactosidase (encoded by *lacZ*, see §1.2.1.1.) activity using a fluorogenic substrate (Cai et al., 2006) or the synthesis of the lactose permease (encoded by *lacY*, see §1.2.1.1.) fused to a fluorescent protein (Choi et al., 2008) revealed that the *lac* operon gene is expressed in a bursty way. Deeper functional analysis revealed that the phenomenon primarily relies on transient dissociation of the *lac* repressor from the operator region (see §1.2.1.1. and Figure 10), which allows *lac* operon transcription to take place before the repressor re-associates to the operator (Choi et al., 2008).

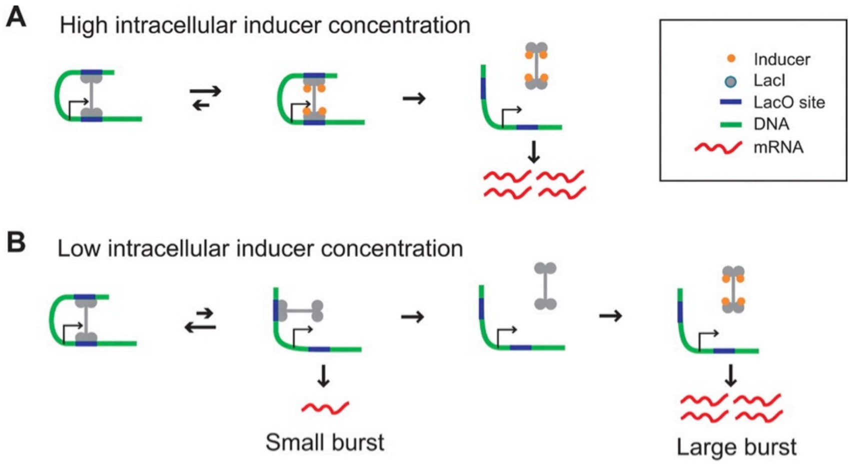

A
High intracellular inducer concentration
Inducer
LacI
LacO site
DNA
mRNA
B
Low intracellular inducer concentration
Small burst
Large burst

Figure 10: Model of Lactose operon regulation. (A) An inducer can bind to Lac^i repressor and trigger its dissociation from DNA operator (LacO), leading to the synthesis of a polycistronic mRNA. **(B)** Sometimes a partial dissociation of repressor spontaneously occurs and leads to a small burst of transcription. Moreover, more complete dissociation can also occur and lead to a larger burst of transcription. Figure adapted from Choi et al., 2008

In normal cell growth (i.e. in the absence of strong inducer) *lac* operon expression is therefore stochastic and primarily relies on stochastic nature of molecular dissociation events and diffusion. These stochastic fluctuations, also called noise, include both extrinsic and intrinsic noises (Eldar and Elowitz, 2010). The extrinsic noise relates to differences in the concentration of gene expression components (e.g. RNA polymerase or ribosome) specific to each cell and can result from unequal distribution of these components upon cell division. In a different way, intrinsic noise is defined by random molecular event occurring during gene expression, this noise becoming particularly significant in the case of poorly expressed genes (Elowitz et al., 2002). Such a pulsed gene expression has since then been identified in cell types ranging from microbes to mammalian cells and has been proposed to play key roles in several important biological mechanisms (Levine et al., 2013).

1.3.2. Consequences of cell-to-cell phenotypic heterogeneity

The direct consequence of the phenotypic heterogeneity introduced above is the quasi uniqueness of each gene expression landscape throughout the population. This palette of landscapes may predispose the population to potent environmental fluctuation and contributes to the overall robustness of the population by allowing, for instance, rapid adaptation to sudden environment change and so maximize the survival chances in case a live-threatening stress has to be faced.

Back to *E. coli* and the *lac* operon case introduced above, the expression of this operon has no real benefits while cells are growing in glucose-rich medium and may even negatively impact cell fitness by the energetic cost of synthesizing *lac* operon enzymes. Therefore, the population subset possessing an active lactose consumption machinery due to noisy gene expression has a suboptimal fitness while glucose is still the predominant carbon source in the medium. However, should the glucose concentration decrease down to a critical concentration and lactose be also present in the medium, the *lac* operon-expressing cells will then have a significant advantage over their population-mates by being able to immediately use the lactose and will therefore become the fittest cells. Such "anticipation" over future environment changes by transiently adopting a suboptimal fitness state is widely used by bacteria and has been termed "bet-hedging" (Grimbergen et al., 2015; Veening et al., 2008a).

Bet-hedging strategies usually involve bistable, or even multistable, systems (i.e. pathways leading to two, or more, mutually exclusive output phenotypes) whom one state has a negative impact on the immediate fitness but that might become highly advantageous upon environmental changes. Beside the *E. coli lac* operon case introduced above, the same phenomenon has been described in *Lactobacillus lactis* during the diauxic shift (Solopova et al., 2014). Of important note, such phenotype diversity impacts industrial process by limiting production yields. Indeed, in large-scale bioreactor microenvironments gradient (e.g. pH or nutriments) can appear and lead to heterogeneous populations made of producer and non-producer cells (Delvigne et al., 2014).

Bet-hedging is also typically exemplified by the spontaneous activation of sporulation program by *B. subtilis* (Veening et al., 2008b) and the more general spontaneous entry of a microbes in dormancy state (Harms et al., 2016). This later bet-hedging behaviour has strong medical implication as these dormant cells (also called persisters) become naturally tolerant to antibiotics due to their extremely low metabolic activity (Fisher et al., 2017). Even though failure of antibiotic treatment due to persisters was known since 1944 (Bigger, 1944), identifying the origin of the phenomenon was possible only some 60 years later with the advent of microfluidics and live-cell imaging (Balaban et al., 2004).

In light of the previous examples, it becomes incontestable that properly understanding gene expression and its regulation at a biologically relevant population-wide scale requires being able to perform characterizations with a single-cell (digital) resolution to take the cell-to-cell phenotypic heterogeneity into account. However, as highlighted in the case of the *lac* operon and the persisters, reaching such a resolution requires first the proper technology to be developed. State-of-the-art microfabrication and microfluidics as well as the discovery of naturally fluorescent proteins (see below) make possible the digital monitoring of protein-coding gene expression. However, as stated in section 1.1 and 1.2, the protein is the final output of a complex expression regulatory process in which non-coding RNAs are instrumental. Therefore, whereas monitoring protein reports on successful activation of a pathway, unfortunately key information on regulatory events occurring at the RNA level are lost. Consequently, the development of new technologies dedicated to RNA monitoring with single-cell resolution is necessary to fill the gap and access a more comprehensive picture of gene expression regulation.

2. Characterizing gene expression with single-cell resolution

Most of the biochemical methods (e.g. blots, mass spectrometry, electrophoresis-based analyses, chromatography…) routinely used to characterize gene expression operate in a so-called analogue mode in which a large number of cells are pooled together and used as starting material. As a consequence, analogue technologies return a result corresponding to the average phenotype of the population and is only partly informative (Figure 11). The main limitation that prevents these technologies to reach the required digital (single-cell) resolution is linked to their sensitivity which is far from being able to handle and analyze the few picograms to femtograms of RNA respectively released from a single mammalian or bacterial cell.

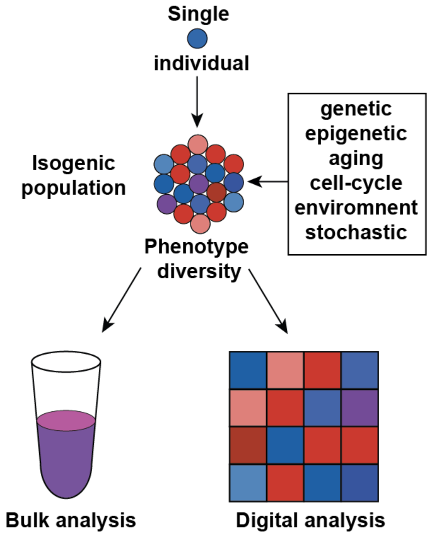


Figure 11: Biology digital concept. Phenotypic diversity of an isogenic population can be accurately profiled using single-cell resolution approaches (digital analyses) whereas analogue (bulk) approaches return a less relevant average measurement.

The advent of Next Generation Sequencing (NGS) technologies together with the development of new classes of reagents and the use of microfluidics allows, nowadays, to profile the mRNA content of eukaryotic cells population with a digital resolution (Prakadan et al., 2017). Even though they are extremely efficient, the current single-cell NGS methodologies rely on the capture of poly-adenylated RNA and are therefore limited to the analysis of eukaryotic mRNAs.

A second popular way of analyzing RNA with single-cell resolution exploits fluorescent probes and imaging technologies. Imaging offers several advantages over the other technologies as it allows both quantifying and localizing RNA molecules within the cell and live-cell compatible procedures make possible monitoring the same cell over an extended period of time, giving access to dynamic information. Such live-cell imaging can be performed by growing cells into a microfluidic device (Potvin-Trottier et al., 2018) prior to analyzing the recorded movies *via* dedicated software(Locke and Elowitz, 2009). Both imaging-dedicated microfluidics and image analysis are out of the scope of this thesis and won't be further discussed. Instead, attention will be focused on the two main types of probes used for RNA imaging: *in situ* hybridization probes and genetically encoded fluorescent probes (fluorescent proteins and fluorogenic RNA aptamers)(Tyagi, 2009).

2.1. In situ hybridization-based RNA imaging

In situ hybridization (ISH) is one of the most used visualization methods relying on labelled oligonucleotides annealing specifically to their target RNA through conventional Watson-Crick base-pairing(Cui et al., 2016). Most of the time, the experiment is performed on fixed and permeabilized cells to enable probes entry and efficient wash to be performed. This allows to strongly reducing background noise and makes the method highly sensitive. Oligonucleotide probes were initially radioactively labelled, leading to all the disadvantages associated with the use of radioactivity (exposure, waste management, impossibility to multiplex). Later on, radioactive labels were exchanged for haptens (e.g. digoxigenin) specifically recognized by enzyme-conjugated antibodies (e.g. peroxidase). Finally, advances in fluorophore chemistry now enables the synthesis of bright fluorescent probes that significantly increases the sensitivity of methods collectively called FISH (Fluorescent *In Situ* Hybridization). Not only FISH allows detecting and localizing RNAs within the cell, but the use of many probes per RNA even allow reaching single molecule resolution in smFISH (single molecule FISH) procedures (Femino et al., 1998) (Figure 12.A).

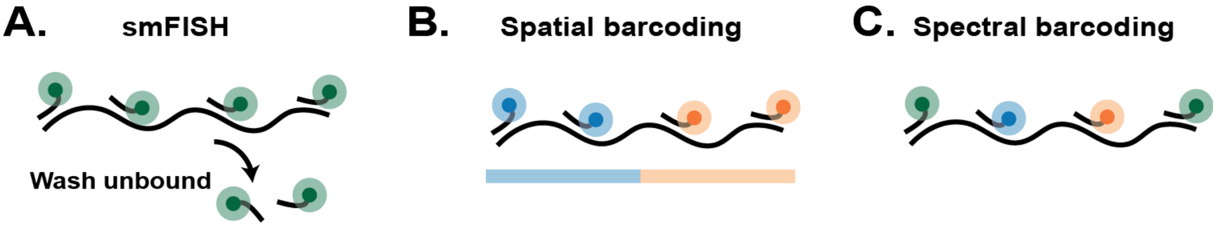


Figure 12: RNA visualization on fixed cells. (A) RNA labelling using single molecular fluorescence *in situ* hybridization (smFISH). Small fluorescently labelled anti-sense probes anneal to the target RNA into fixed cells. Excess of unbound probesis then washed away to improve the contrast **(B)** Combined use of multiple fluorescently labelled probes is used to barcode target RNA allowing for multiplex detection. **(C)** Spectral barcoding using a specific mixture of fluorophores.

Because of spectral overlap, conventional smFISH procedure allows the simultaneous monitoring of only a few (4-5) molecules at the time. However, exploiting super-resolutive microscopy and colour barcoding makes possible to drastically increase the number of co-detected RNA and nowadays hundreds to thousands distinct RNA species can be simultaneous detected (Lubeck and Cai, 2012). Such a high degree of multiplexing can be achieved *via* two main barcoding strategies: i) a spatial barcoding in which molecules are identified thanks to the order of consecutive probes labelled in different colours and ii) a spectral barcoding in which only the identity of different colours is used to encode the molecule (Figure 12.B and C)(Crosetto et al., 2015). Despite their high performances, smFISH technologies suffer from the elevated cost of the fluorescently-labelled probes and the need for cell fixation which compromises temporal monitoring and leads to the loss of information on system dynamics. Therefore, even though smFISH technologies are extremely attractive for diagnostic purposes, their interest is more limited when studying dynamic processes such as gene expression regulation.

Dynamic information can nevertheless be accessed by FISH using probes like molecular beacons (MBs). MBs are stem-loop oligonucleotides made of a large loop closed by a stem whom the extremities are coupled to a fluorophore and a quencher (Bratu et al., 2003)(Figure 13.A). In the closed conformation, the proximity between the quencher and the fluorophore prevents fluorescence emission. However, upon hybridization between the loop and the target RNA sequence, the stem structure gets opened, leading to the separation of the quencher and the fluorophore and the restoration of the fluorescence. Consequently, MBs are fluorogenic molecules (i.e. non-fluorescent molecules whom the fluorescence can be restored under defined circumstances) displaying a limited background fluorescence while becoming highly fluorescent in the presence of their specific target. A second type of fluorogenic FISH probe exploit the principle of alternative forced intercalation (Köhler et al., 2005; Kummer et al., 2012) (FIT, Figure 13.B). In FIT, the oligonucleotide is linked to a small fluorogenic dye (mostly Thiazole Orange) that is not fluorescent in aqueous environment. However, upon annealing of the probe to its target RNA, the fluorogenic dye intercalates into the newly formed helix and recovers its fluorescence capacity. Both types of probes make possible to image RNAs in living cells. However, since these polymers cannot freely cross the cell membrane, their use requires microinjecting them into cell cytoplasm, which prevents the automation of the method and seriously limits the analysis throughput. Furthermore, their use is limited to large microinjectable cells (i.e. mammalian cells). This limitation linked to the cell entry of FISH probes can however be overcome by using genetically encoded probes.

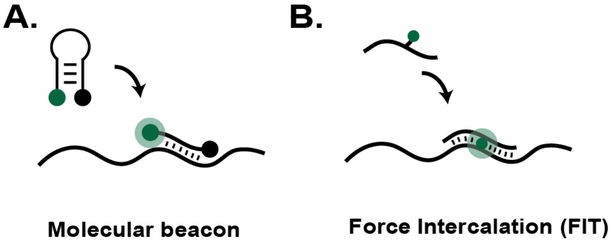


Figure 13: RNA visualization on living cells using labelled oligonucleotides. (A) Molecular beacon fold into an hairpin structure bringing together a fluorophore and a quencher reducing drastically the fluorescence of the probe. When the probe anneals to its nucleic acid target, the stem-loop opens separating the quencher from fluorophore and turning the probe fluorescent. **(B)** Force interaction (FIT) probe is a small anti-sense RNA surrogates with light-up fluorophore. The fluorescence of the latter gets activated only upon annealing of the probe to its target sequence.

2.2. *RNA imaging using genetically encoded probes*

Probes such as fluorescent proteins and fluorogenic RNAs can be genetically encoded. Indeed, inserting the coding gene into the cell (either by direct insertion into the genome or *via* plasmid transfection) enables to get the probe directly synthesized by cell transcription/translation machinery.

2.2.1. Fluorescent protein-based approaches

In 1962, Shimomura *et al* (Shimomura, 2005) described for the first time that the green fluorescence emitted by the jellyfish *Aequorea victoria* relies on a protein they later called Green Fluorescent Protein (GFP). The great capacity of this protein to support engineering allowed the development of many imaging tools, especially some dedicated to RNA imaging.

2.2.1.1. Properties and engineering of the green fluorescent protein

The GFP is organized around an 11-strand β-sheet barrel that encloses and protects the fluorophore 4'-hydroxybenzylidene-2, 3-dimethylimidazolinone (HBDI) (Figure 14). The HBDI is formed upon cyclisation of a tri-peptide (Serine, Tyrosine and Glycine) that spontaneously occurs post-translationally through an oxidation maturation step requiring molecular oxygen but no additional cell factor (Cody et al., 1993; Heim et al., 1994). This autonomous fluorescence acquisition together with the possibility to express *gfp* gene within any cell type without notable toxicity (Chalfie et al., 1994) makes the GFP an invaluable tool for characterizing gene expression among other applications.

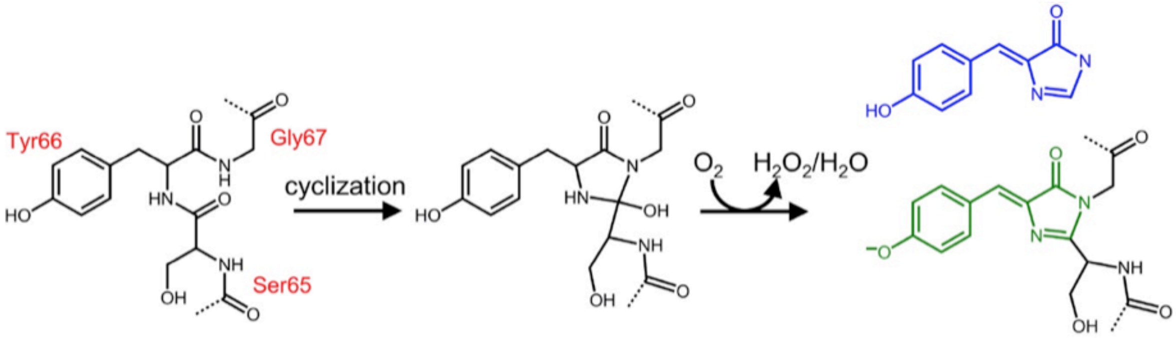


Figure 14: GFP chromophore maturation GFP fluorophore (in green), or the HDBI (in blue), readily forms *via* the cyclisation of a tri-peptide (Ser, Tyr and Gly) and followed by its oxidation. Figure adapted from Nienhaus *et al* 2016.

When excited at 395 nm, HBDI emits a green fluorescence at 508 nm with a pretty high quantum yield (QY; i.e. the number of emitted photon per exciting photon) of nearly 0.77. According to equation 1, multiplying this QY by the absorption coefficient (ε) of the fluorophore gives access to its brightness.

$$\text{Eq.1} \quad \text{Brightness} = (\text{QY} \times \varepsilon)/1000$$

Ideally, a good fluorophore should absorb as much light as possible (high ε) and re-emit as many photons as possible (high QY) to be as bright as possible. Whereas the original GFP displayed a brightness of ~ 16 $M^{-1}.cm^{-1}$, screening mutant libraries allowed isolating an improved mutant called enhanced GFP (eGFP) and characterized by an increased brightness of ~ 26 $M^{-1}.cm^{-1}$ and excitation/ emission wavelengths better fitting filter channels (ex. 489nm/em. 511nm) found on regular lab equipment (Kilgard et al., 1995). Additional work allowed identifying GFP mutants with superior folding efficiency (Pédelacq et al., 2006) that can readily be fused to other proteins. Moreover, GFP colour can easily be tuned from cyan (CFP to yellow (YFP) by inserting simple point mutations in *gfp* gene (Heim et al., 1994). Finally, proteins emitting in the orange to red region of the spectrum have also been isolated from other species and now allow the whole visible spectrum to be covered (Chudakov et al., 2010).

Placing the sequence coding for one of these fluorescent proteins (FP) under the dependency of the natural promoter of the studied gene (or fusion FP-coding sequence to that of the gene of interest) allows monitoring gene activation (Figure 15). This strategy has been typically used to decipher *lac* operon expression mode (see § 1.3.1, (Choi et al., 2008)) and is one of the most used approach.

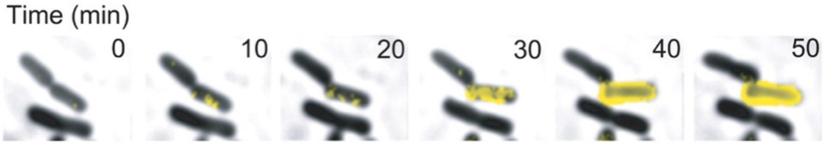


Figure 15: Monitoring phenotype diversity and single-cell resolution using a fluorescent protein. Herein, a case of phenotype switching at low concentration of inducer is visualized using LacY-YFP molecules (lactose permease conjugated with Yellow Fluorescence Protein). Figure adapted from Choi et al. 2008.

Further engineering the GFP led to the development of a split form of the protein. Each part of the protein is separately synthesized as non-fluorescent parts (Cabantous et al., 2005)that can reassemble into a full-length fluorescent molecule. This can be used for instance to monitor protein/protein interaction or assess protein aggregation state. To do so, a protein of interest is synthesized in fusion with one of the 11 GFP strands and mixed with a truncated form of the GFP containing the 10 residual strands. If the protein of interest is well folded, the 11th GFP strand is accessible and can associate with the truncated GFP to form a fluorescent complex (Figure 16.A).

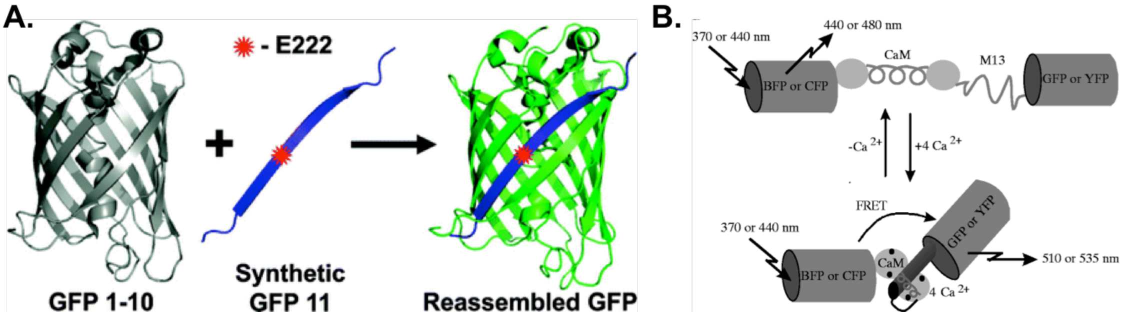


Figure 16: Green fluorescence engineering. (A) Schematic diagram of split-GFP and complementation of the 1-10 barrel by addition of eleventh strand. The labeled residue E222 is an important proton acceptor during exciting state. Figure adapted from Kent *et al*. 2008. (**B**) Schematic representation of FP-based calcium sensor. The two FPs are separated by a calmodulin domain (CAM). When calcium is present and binds to the CAM domain, the two FP get spatially closer and FRET occurs. Figure adapted from Miyaawaki *et al*. 1997

Finally, more complex FP-based probes have been developed as innovative biosensors reporting on calcium concentration (Miyawaki et al., 1997) or even intracellular cell crowding (Boersma et al., 2015) (Figure 16.B). These probes consist of a pair of FPs separated by a spacing domain (calmodulin fragment in the case of Ca^{2+} sensor) that form a FRET (Förster Resonance Energy Transfer) pair; i.e. upon excitation, the first protein emits fluorescence at a wavelength exciting the second protein, which in turn emits fluorescence. Since FRET efficiency strongly relies on the distance between the donor and the acceptor FP, measuring the FRET emitted fluorescence informs on the proximity between both molecules, and doing so gives access to the sensed parameter (calcium concentration or crowding). Even though these probes are highly efficient, their development is very tedious mainly because of the lack of efficient algorithm to predict protein folding and assist protein engineering.

2.2.1.2.Imaging RNA using fluorescent proteins

The capacity of FPs to sustain engineering allows synthesizing them in fusion with an RNA binding protein (RBP) domain (e.g. MCP, PP7 or Pumilio) (Urbanek et al., 2014). Expressing these constructs in cells together with RNAs engineered to display motifs specifically recognized by the RBP allows to specifically decorate the target RNA with FPs (Figure 17). This technology was initially pioneered by Bertrand *et al* (Bertrand et al., 1998) uses GFP and the phage-derived MS2 system in which the MS2 binding site (MBS) is specifically recognized by the membrane coat protein (MCP) with a high affinity. Co-expressing a construction in which MCP is fused to GFP therefore allows to fluorescently label the target RNA carrying MS2 stem-loop motifs. Furthermore, background fluorescence from the unbound GFP can be limited by adding a Nuclear Localization Signal (NLS) to the GFP to deplete the unbound GFP from cell cytoplasm, or by using a split form of the GFP (Wu et al., 2014).

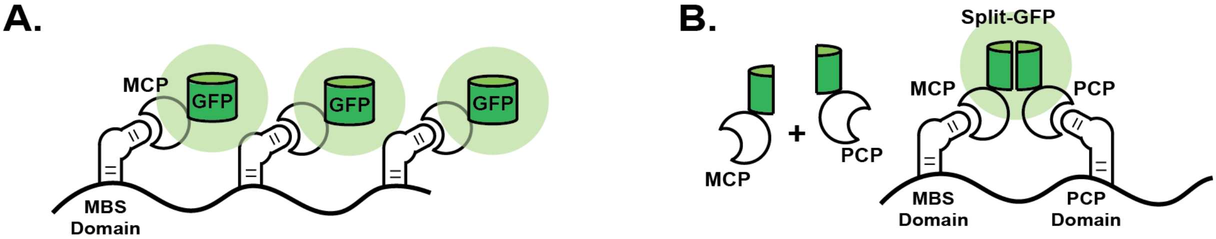


Figure 17: Fluorescent Protein-based genetically encoded RNA labelling technologies. **(A)** Schematic representation MS2-GFP RNA imaging. The MS2 binding site (MBS) made of an RNA stem loop inserted into the labelled mRNA is specifically recognized by MS2 Coat Protein (MCP) fused with GFP. **(B)** Background-reduced RNA aptamer imaging approach. Tandem of bacteriophage domains are specifically recognized by their cognate protein fused to split GFP domains.

Overall, this approach makes possible to image RNA in living cell over extended periods of time in a non-invasive and non-destructive way. However, it still suffers significant limitations. First, visualizing a single RNA may require a large number (24 to 94) of RNA motif repeats to be inserted into the target RNA (Urbanek et al., 2014), leading to significant increase of target RNA size (even without considering the presence the bulky proteins on the RNA) which can interfere with RNA function and restricts the use of this technology to large RNAs. Second, addition of an NLS to the FP fusion protein may interfere with the natural trafficking of the target RNA. Finally, even though split GFP makes possible to reduce background fluorescence, the metastability of the reassembled complex can lead to maintenance of the fluorescence signal even though the target RNA is not present anymore. These different limitations of FP-based systems can be, at least partly, overcome by the use of RNA-based fluorogenic modules

2.2.2.RNA-based fluorogenic modules

RNA-based fluorogenic modules are partly genetically encoded probes that represent an attractive alternative to the FP-based imaging methods introduced above. These modules are made of two components: i) a small fluorogenic molecule that displays a very weak fluorescence in its free state and is called a fluorogen, and ii) a light-up RNA aptamer able to specifically interact with the fluorogen and activate its fluorescence forming therefore a fluorescent complex (Figure 18).

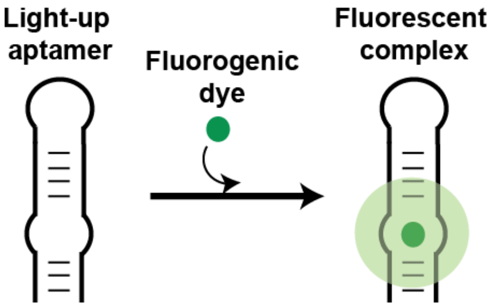


Figure 18: Working principle of an RNA-based fluorogenic module. A fluorogen poorly fluorescent in its unbound state become highly fluorescent upon specific recognition by its cognate aptamer.

This concept was originally introduced by Robert Tsien upon the discovery that the interaction between Malachite Green and its cognate aptamer leads not only to the expected production of radicals (Grate and Wilson, 1999) but also to the emission of a strong orange fluorescence (Babendure et al., 2003). Whereas the toxicity due to radical production by the complex initially limited the use of these modules for live-cell imaging purposes, the development by Samy Jaffrey of the GFP-mimicking fluorogen DFHBI together with its cognate light-up aptamer Spinach (Paige et al., 2011) contributed to repopularize this imaging strategy and paved the way to a wide set of applications.

2.2.2.1.RNA-based fluorogenic modules and their applications

RNA-based fluorogenic modules are not limited to the MG system and DFHBI/Spinach. Instead, a wide variety of fluorogens and their cognate light-up aptamers have been developed over the past decade. Furthermore, the easier engineering of RNAs in comparison to proteins made possible the development of plenty of applications. Given the importance of these systems and the fact they were central to my PhD project, I co-authored a review article, published in early 2018, that gives a comprehensive view of the field and that is reproduced hereafter (**Review Article 1**):

Review Article 1:

Light-Up RNA Aptamers and Their Cognate Fluorogens: From Their Development to Their Applications

Farah Bouhedda , Alexis Autour and Michael Ryckelynck

2018

International Journal of Molecular Sciences

International Journal of
Molecular Sciences

Review

Light-Up RNA Aptamers and Their Cognate Fluorogens: From Their Development to Their Applications

Farah Bouhedda [†], Alexis Autour [†] and Michael Ryckelynck *

Architecture et Réactivité de l'ARN, CNRS, Université de Strasbourg, UPR 9002, F-67000 Strasbourg, France; f.bouhedda@ibmc-cnrs.unistra.fr (F.B.), a.autour@ibmc-cnrs.unistra.fr (A.A.)
* Correspondence: m.ryckelynck@unistra.fr; Tel.: +33-(0)3-8841-7055
† These authors contributed equally to this work.

Received: 29 November 2017; Accepted: 15 December 2017; Published: 23 December 2017

Abstract: An RNA-based fluorogenic module consists of a light-up RNA aptamer able to specifically interact with a fluorogen to form a fluorescent complex. Over the past decade, significant efforts have been devoted to the development of such modules, which now cover the whole visible spectrum, as well as to their engineering to serve in a wide range of applications. In this review, we summarize the different strategies used to develop each partner (the fluorogen and the light-up RNA aptamer) prior to giving an overview of their applications that range from live-cell RNA imaging to the set-up of high-throughput drug screening pipelines. We then conclude with a critical discussion on the current limitations of these modules and how combining in vitro selection with screening approaches may help develop even better molecules.

Keywords: light-up aptamer; fluorogen; fluorogenic dye; fluorescence; RNA; in vitro evolution; gene expression monitoring; live-cell imaging; biosensing

1. Introduction

Quoting the famous idiom: "seeing is believing", which makes imaging and spectroscopic technologies very popular in life sciences to detect and track molecules. Among the different visualization strategies, fluorescence is the most attractive and widely used, mainly because it is safe, sensitive and it offers the possibility of analyzing multiple colors at the same time, with a long shelf life of the fluorophores when properly stored. Among the different fluorescent probes, those acquiring their fluorescence only in permissive conditions (e.g., presence of a target molecule) and hereafter defined as being fluorogenic are the most useful. These probes exhibit only a weak fluorescence in non-permissive conditions, making washing steps dispensable, and permitting the monitoring of dynamic processes in time course experiments. A plethora of such fluorogenic molecules has been developed for applications as diverse as ions and small molecules sensing, pH monitoring or environment viscosity assessment [1,2]. Moreover, significant efforts have been devoted to the development of fluorogenic tools dedicated to RNA and protein, two biological polymers particularly relevant to cellular activity. Historically, these molecules were first detected using fluorescently labeled and extracellularly supplied specific probes such as antibodies and oligonucleotides [3]. Recent developments allow these strategies to reach single-molecule sensitivity, but the requirement of fixing cells represents a major drawback that could be overcome using genetically encoded fluorescent reporters.

The discovery of the naturally fluorescing Green Fluorescent Protein (GFP) from the jellyfish *Aequorea victoria* [4] had a profound impact on protein imaging field and paved the way for major breakthroughs. Indeed, this first genetically encoded probe post-translationally acquires its fluorescence through an autocatalytic cyclization involving three amino acids (Serine, Tyrosine, and Glycine) independently of any cell factor [5,6], making it usable in any cell type. Moreover, simple point mutations can shift the excitation/emission spectra of the protein toward the blue [7] or the red [8] regions of the visible spectrum. These great properties, shared with many other fluorescent proteins (FPs), make them highly versatile (see [9] for a comprehensive review) and very attractive for biotechnological applications. However, the strict requirement of molecular oxygen for the maturation of their fluorophore also limits FPs application in some circumstances and led to the development of alternative labeling strategies in which a fluorescent dye labels the protein of interest via a peptide (e.g., tetracystein peptide labeled by the Fluorescein Arsenical Helix binder FlAsH [10]) or a whole domain (e.g., SNAP-tag labeled by dye conjugated to a benzylguanine group [11]) appended to the target protein [12].

In contrast to proteins, no naturally fluorescent RNA has been discovered yet, making the development of RNA-based genetically encoded fluorescent reporters less straightforward than their protein counterparts. A first live-cell compatible strategy pioneered by Bertrand et al. [13] consists of inserting tandem repeats of elements recognized by an RNA-binding protein (RBP) into the RNA to image. Then, co-expressing this construct with an RBP-GFP fusion protein allows the direct labeling of the target RNA with the GFP. Moreover, the use of a split form of the GFP allows converting the otherwise always fluorescent GFP into a fluorogenic system in which fluorescence is expected only upon RBP-GFP/target RNA interaction [14]. Whereas this approach proved to be efficient for live-cell monitoring of large messenger RNAs [15], it may be more challenging for smaller RNAs (e.g., regulatory RNAs). Indeed, the insertion of a large number (20–30) of RBP binding sites into a small target RNA as well as its later decoration by RBP-GFPs leads to a significant increase of RNA size that could interfere with its biological function, making a size reduction of the labels a high priority. As was the case for protein labels, significant size reduction can be obtained by exchanging the bulky GFP for a smaller fluorescent synthetic dye. Constructs in which tandem repeats of RNA aptamers (i.e., small nucleic acid folds able to specifically recognize a target molecule) specific to a dye can be inserted into the target RNA, and the labeling can be obtained by placing cells in culture medium containing a membrane permeable dye. Aptamers binding specifically to fluorescent dyes such as sulforhodamine, fluorescein [16] or modified cyanines [17] can be used in arrays [17,18]. However, since the dye emits fluorescence even in its free unbound form, such approach may suffer from significant background fluorescence, limiting its application spectrum. Nevertheless, this limitation can be overcome by exchanging the fluorescent dye for a fluorogenic one. In this view, a seminal work by Tsien's group showed that, not only an RNA aptamer can specifically interact with a target molecule, but this interaction can also strongly increase the fluorescence of compounds such as Malachite green, making such aptamer/dye couple fluorogenic [19]. This discovery was all the more astonishing that the Malachite green-binding aptamer was not originally selected to function as a light-up aptamer but rather to mediate site-specific inactivation of target RNAs [20]. Since then, a variety of fluorogenic dyes and their cognate RNAs have been developed [21] (Table 1). This short review will be primarily focused on the main design strategies of each partner (the dye and the RNA) prior to giving a rapid overview of their application scope ranging from in vivo live-cell RNA imaging to in vitro small molecule biosensing. For a more general view of the current technologies available to image RNA, the reader is redirected to other reviews [22–25].

Table 1. Main RNA-based fluorogenic modules and their properties.

Fluorogen	Light-Up Aptamer	K_D (nM)	Ex./Em. (nm)	ε [1] (M^{-1}/cm)	$\Phi^{complex}$ [2]	Brightness [3]	Relative Brightness [4]	Ref.
GFP	/	/	395/508	21,000	0.770	16.20	0.60	[26]
eGFP	/	/	490/508	39,200	0.680	26.60	1.00	[26]
OTB	DiR2s-Apt	662	380/421	73,000	0.510	37.23	1.40	[27]
Hoescht	Apt II-mini3-4 c	35	345/470	n.a.	0.260	n.a.	n.a.	[28]
DFHBI	Spinach	540	469/501	24,300	0.720	17.50	0.65	[29]
DFHBI-1T	Spinach2	560	482/505	31,000	0.940	29.10	1.10	[30]
DFHBI-1T	Broccoli	360	472/507	29,600	0.940	27.80	1.04	[31]
DFHBI-2T	Spinach2	1300	500/523	29,000	0.120	3.48	0.10	[30]
RG-DN	DNB	4480	507/534	37,350	0.320	11.90	0.44	[32]
TO-1	Mango	3	510/535	77,500	0.140	10.85	0.40	[33]
DFHO	Corn	70	505/545	29,000	0.250	7.25	0.27	[34]
CY3-BHQ1	BHQ apt (A1)	n.a.	520/565	n.a.	n.a.	n.a.	n.a.	[35]
DFHO	Red-Broccoli	206	518/582	35,000	0.340	11.90	0.44	[34]
TMR-DN	DNB	350	555/582	47,150	0.900	42.43	1.60	[32]
SR-DN	DNB	800	572/591	50,250	0.980	49.24	1.80	[32]
DIR	DIR apt	86	600/646	134,000	0.260	34.80	1.30	[36]
Mal. Green	MG aptamer	117	630/650	150,000	0.187	28.00	1.05	[19]
DIR-pro	DIR2s-Apt	252	600/658	164,000	0.330	54.12	2.00	[27]
TO-3	Mango	6–8	637/658	9300	n.a.	n.a.	n.a.	[33]
Patent Blue	SRB apt	23	n.a./665	n.a.	0.034	n.a.	n.a.	[19]

[1] Absorption coefficient (ε); [2] Quantum yield of the complex ($\Phi^{complex}$); [3] Brightness calculated as Brightness = ($\varepsilon \times \Phi^{complex}$)/1000; [4] Brightness expressed relative to eGFP. n.a.: not available. RNA modules were ordered according to their fluorescence emission wavelength.

2. Development of RNA-Based Fluorogenic Modules

RNA-based fluorogenic modules are made of two components: a fluorogenic dye (later called fluorogen) and a specific light-up RNA aptamer. In an ideal case, such module should feature five essential properties. (i) The module should be as bright as possible to ensure sensitive detection, ideally down to single molecule resolution. Therefore, the dye should have an elevated absorption coefficient (ε) and its interaction with the RNA should place it in a conformation and an environment maximizing its quantum yield (Φ^{bound}) that should be as close as possible to 1. (ii) In its free form, the fluorogen should display the lowest possible quantum yield (Φ^{free}) to minimize background fluorescence. Consequently, a good fluorogenic module is expected to have the highest possible fluorescence enhancement (calculated as the ratio of Φ^{bound} over Φ^{free}). (iii) The RNA–dye interaction should be highly specific and bio-orthogonal (i.e., cell compounds or reagents of the assay should not interfere with the interaction). (iv) Moreover, the interaction should occur with a high affinity (dissociation constant K_D in the nM range or less) to make using low concentration of the dye possible and to keep fluorescence background as low as possible, which allows getting high contrast [37]. (v) Finally, the module should be as photostable as possible to allow prolonged data acquisition. Since both components of the module affect all these parameters (with the exception of the second one that is only affected by the dye), the proper development of an RNA-based fluorogenic module is not a trivial task and both the dye and the RNA should be developed while trying to stay as close as possible to the ideal scenario depicted above.

2.1. Fluorogenic Dye Engineering

The fluorogen plays a key role in the functionality of a fluorogenic module by contributing to both its brightness and its photostability. Moreover, for live-cell applications, additional parameters such as membrane permeability and the lack of toxicity of the dye should also be considered. The fluorogenicity of a dye can be obtained in various ways encompassing but not limited to charge (or proton) transfer, conformational change, isomerization or even dye aggregation [38]. Several of these principles having been used for the development of RNA-based fluorogenic modules.

Many compounds are known to become highly fluorescent upon interaction with nucleic acids [39]. Among them, ethidium bromide [40] and Hoechst 33258 [41] are environment sensitive dyes characterized by a poor fluorescence in aqueous solution. However, upon binding to a DNA molecule (respectively by intercalation between base pairs or binding in the DNA minor groove at AT-rich region), the dye is placed into a non-polar environment that strongly increases its fluorescence (Figure 1a). The non-specific DNA binding capacity of these dyes can lead to significant unwanted background fluorescence when used in cell-based assay. However, this property can be suppressed by decorating the dye with additional chemical functions. For instance, substituting Hoechst 33258 with bulky groups allowed designing derivatives no longer capable of non-specific DNA binding, but preserving their fluorogenic capacity and becoming fluorescent only upon specific interaction with DNA and RNA aptamers [28,42] (Table 1).

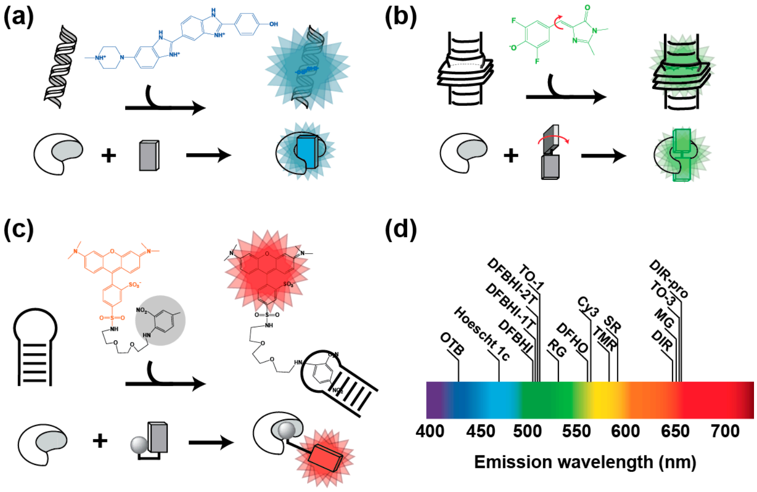


Figure 1. Main conception strategies of fluorogens. (**a**) Environment sensitive fluorogens. Molecules such as Hoescht 33258 (formula in blue) emit blue fluorescence upon association with DNA minor groove. Modifying the fluorogen with bulky groups allows aborting this non-specific DNA binding capacity, while preserving the fluorogenic capacity that can be now specifically activated by the cognate aptamer [35]. (**b**) Molecular rotor fluorogens. DFHBI (formula in green) [29] eliminates excitation energy by molecular movements (red arrows). However, upon association to the cognate aptamer, movements are restricted and fluorogen energy is eliminated by fluorescence emission. (**c**) Quenched fluorogens. Sulforhodamine B (formula in red) fluorescence is quenched by a conjugated dinitroaniline (formula in black and shaded in gray) [32]. However, the fluorescence is restored upon the specific recognition of the quencher (or the fluorophore moiety in other systems) by an aptamer. In every example, RNA aptamer is represented by the croissant-shaped object and the fluorogen by the rectangles. (**d**) Distribution of fluorogen emission wavelength along the visible spectrum.

Fluorogenicity can also result from intramolecular movements allowing for non-radiative relaxation of the dye upon excitation (Figure 1b). Such compounds, hereafter called molecular rotors, are poorly fluorescent in their unbound form in a fluid environment. However, fluorescence can be restored by restricting intramolecular movements either by strongly increasing medium viscosity, or upon specific interaction with a nucleic acid. Triphenylmethane dyes (e.g., Malachite Green (MG) and Patent Blue, Table 1) were the first class of compounds for which specific RNA light-up aptamers were identified [19,20]. Among them, MG shows an impressive ~2400-fold fluorescence enhancement when bound to its cognate RNA aptamer [19]. Nevertheless, MG light irradiation is known to elicit the generation of free radical leading to subsequent RNA cleavage [20], making this approach potentially toxic for living cells [34,43,44]. Moreover, MG was also found to be able to generate significant background fluorescence within mammalian cell [45] and bacteria [46], further limiting the use of this system. Unsymmetrical fluorogenic cyanines constitute a second attractive class of fluorogenic molecular rotors composed of two different heterocycles connected by a methine bridge subjected to twisting in fluid solution [47,48]. Thiazole Orange (TO) is a good representative of these dyes, but it is also characterized by a significant non-specific DNA binding capacity [49,50]. However, as with Hoechst 33258, modifying the dye makes it possible to strongly attenuate this adverse effect. Indeed, using a large dimethylindole heterocycle and substituting the quinolone ring with a propylsulfonate group led to Dimethyl Indole Red (DIR), a TO derivative displaying a strongly reduced non-specific interaction with nucleic acids [36]. Alternatively, adding an acetate group on the benzothiazole moiety led to TO-1, a TO derivative also displaying a reduced affinity for DNA [33]. Moreover, a specific RNA aptamer was developed for each TO derivative (Table 1). In addition to their elevated absorption coefficient, cyanines are attractive molecules from a spectral point of view since their excitation/emission spectra can be easily modulated by changing the length of the methine bridge. Indeed, simply lengthening this bridge by two carbons converted the green emitting TO-1 into a red-emitting TO-3 [33]. Despite these promising properties, one can also foresee that current cyanine-derived dyes may suffer a potential limitation regarding cell-permeability. Indeed, the charged (e.g., acetate and sulfonate [27]) and bulky polar groups (e.g., PEG-biotin [33]) used to prevent non-specific binding, may also affect the capacity of the dyes to freely cross the cell membrane. Nevertheless, transiently caging these groups (e.g., by protecting carboxylic group as acetoxymethyl (AM) esters) may improve their membrane permeability. A last class of molecular rotors is made of GFP-mimicking dyes, a set of synthetic fluorogens that mimic the fluorophore of the GFP formed upon the cyclization of the Ser-Tyr-Gly tripeptide. Many derivatives of these fluorogens have been synthesized and used to develop a variety of sensors (see [51] for a recent review). Among them, the 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI, Table 1) was developed by Jaffrey's group in 2011 together with a specific light-up aptamer [29]. Among other interesting features, DFHBI displays a ~1000 fluorescence enhancement in the presence of specific binding aptamers, it is non-toxic and has good cell permeability. Further substituting the imidazolinone cycle with a trifluoroethyl or a pentafluorophenyl group allowed obtaining DFHBI-1T [30] and DFHBI-PFP [45], two dyes with improved brightness and spectral properties. Whereas these GFP-mimicking dyes emit in the green region of the spectrum, the recent development of the DsRed-mimicking dye 3,5-difluoro-4-hydroxybenzylidene imidazolinone-2-oxime (DFHO, [34]) and others [52] also makes possible imaging nucleic acids in the orange-red region of the spectrum.

Finally, fluorogenic dyes can be obtained by appending a quenching group to a fluorescent organic dye (Figure 1c). In its free off-state, the fluorescence of the dye is quenched by a mechanism of photoinduced electron transfer (PET, [53,54]), Förster resonance energy transfer (FRET, [35]) or contact-mediated quenching [32,55]. However, the presence of an aptamer specifically recognizing either the dye [16,55] or the quencher [32,35,53] moiety allows for retrieving the fluorescence of the dye and the emission of a bright fluorescence (Table 1).

All together, these different strategies allowed designing a variety of fluorogens with fluorescence emission spanning most of the visible spectrum (Figure 1d) and brightness sometimes exceeding that of the broadly used GFP (Table 1).

2.2. Isolation of Fluorogenic RNA Aptamers

Once a fluorogen has been designed, the molecule can be used as target for the isolation and the optimization of a specific light-up RNA aptamer using two conceptually different approaches: (i) a selection strictly speaking during which an RNA pool is challenged to interact with the dye and only the best binders are recovered; or (ii) a screening approach where the light-up capacity of each molecule is analyzed and only the most fluorogenic ones are recovered.

2.2.1. Aptamers Selection Based on Binding Capacity

Historically, RNA (or DNA) aptamers are isolated using the Systematic Evolution of Ligands by EXponential enrichment (SELEX) approach [56,57], a set of technologies particularly well suited for the identification of nucleic acid specifically recognizing virtually any type of target, ranging from ions to more complex proteins [58]. Conceptually, SELEX works by performing iterative rounds of selection to exponentially enrich an RNA (or DNA) library in molecules able to bind a target molecule. Moreover, SELEX possesses the attractive capacity of handling large sequence diversity (10^{12} to 10^{15} sequences) in a single experiment.

Depending on target size and physicochemical properties, several variations of SELEX have been introduced. However, all the new light-up RNA aptamers isolated so far were obtained through the same selection scheme (Figure 2). The success of such selection relies mainly on the proper design of the starting library as well as the partition mode used during the selection process. The starting DNA gene library is usually obtained by inserting a randomized region of ~50 nucleotides or more between two constants regions. Moreover, early work on the selection of aptamers targeting small molecules showed that interrupting the randomized region with a short constant hairpin may increase the success rate of a selection process [59] and, interestingly, the use of such discontinuous randomized libraries allowed isolating some of the best fluorogenic aptamers described so far [27,29,31,36]. The selection process is then initiated by transcribing the DNA gene library into an RNA pool prior the selection step where target-binding RNAs are partitioned away from the non-binding ones. To do so, the RNA pool is usually incubated with beads (e.g., streptavidin-conjugated agarose or magnetic beads) displaying the fluorogenic dye attached via a PEG linker. After several washing steps, the retained RNAs are eluted and reverse transcribed into cDNAs. These cDNAs can be later PCR-amplified and used to prime a new round of selection. Gradually increasing the wash stringency (e.g., increase the number and/or the duration of the wash) during the selection process allows favoring the selection of aptamers with the highest affinity for the dye while the poorer binders are counter selected. Moreover, the choice of the elution mode may also significantly affect the properties of the selected molecules. For instance, immobilized RNAs can be eluted using an excess of free dye. Such a general selection scheme led to the isolation of several fluorogenic RNAs such as the Malachite Green-binding aptamer [20], the DFHBI-binding aptamer Spinach [29], the DFHO-binding aptamer Corn [34] and the Dimethylindole Red-binding aptamer [27,36]. However, even though competition-based elution may favor the isolation of aptamers having great specificity for the dye moiety, they also suffer the major drawback that variants with low dissociation constant, so very high affinity, might likely be counter selected and lost. This may partly explain why these aptamers have a moderate affinity (K_D ~ 0.1–1 μM) for their cognate dyes. This limitation could be overcome by changing the elution mode. For instance, the insertion of a disulfide bond within the linker between the dye and the bead, allows for the elution of dye/RNA complexes by adding dithiothreitol [32]. Alternatively, retained RNAs can be recovered by eluting them under denaturing conditions (e.g., heating, addition of urea, formamide, NaOH, etc.) as used for the selection of Hoechst-binding aptamer [28] and TO-1-binding Mango RNA aptamer [33]. Combined with drastic washing conditions (ionic strength reduction and addition of free competitive dye in the washing buffer), denaturing elution allowed the isolation of Mango RNA, a fluorogenic aptamer forming a fluorescent complex with the thiazole orange derivative TO-1-biotin characterized by a K_D in the nM range, the highest affinity described so far for a fluorogenic module (Table 1). Nevertheless, one should note that RNAs isolated by non-competitive elution might also display affinity not only for the dye moiety but also accessory elements. This is well exemplified by Mango RNA that recognizes the TO-1-biotin not only via its thiazole orange moiety, but also displays some affinity for the PEG linker and the terminal biotin [60].

Upon isolation, fluorogenic aptamers can be further optimized by truncation and rational design. For instance, rationally inserting a few mutations in the Spinach aptamer resulted in Spinach2, a DFHBI-binding aptamer with improved folding properties [61]. However, some beneficial mutations may be difficult to predict and improving these aptamers may require individually testing many mutants. Moreover, an intrinsic limitation of SELEX is linked to the way selections are performed. Indeed, in this format, molecules are selected for their capacity to bind a fluorogen rather than for their capacity to trigger its fluorescence. Therefore, screening approaches, in which molecules are directly analyzed for their fluorogenic capacity, represent an attractive complement to SELEX.

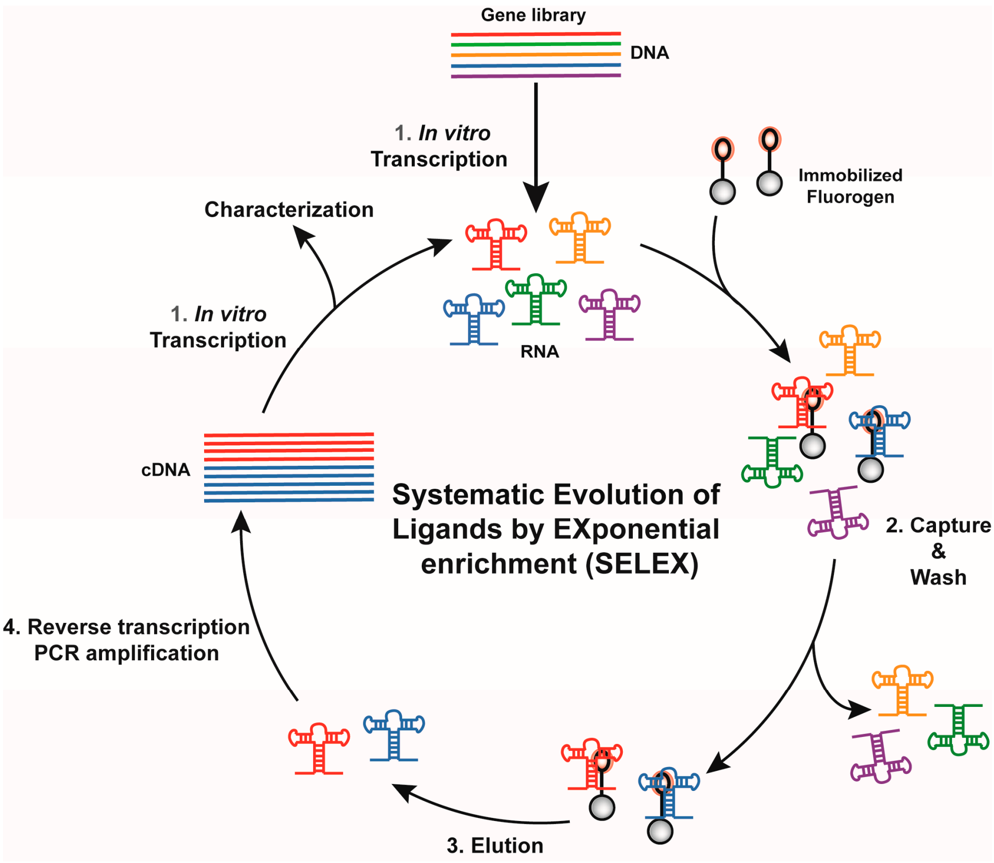

Gene library
DNA
1. *In vitro* Transcription
Immobilized Fluorogen
Characterization
1. *In vitro* Transcription
RNA
cDNA
Systematic Evolution of Ligands by EXponential enrichment (SELEX)
2. Capture & Wash
4. Reverse transcription PCR amplification
3. Elution

Figure 2. General overview of an in vitro selection using Systematic Evolution of Ligands by EXponential enrichment (SELEX). Each round proceeds in four main steps. A gene library is in vitro transcribed (Step 1) and mixed with beads displaying the fluorogen. Poor binders are eliminated by wash of variable stringency (selection pressure, Step 2) prior to recovering binding RNAs during the elution step (Step 3). Finally, RNAs are converted into cDNAs (Step 4), later used to prime a new round of selection.

2.2.2. Aptamers Isolation Based on Their Fluorogenic Capacity

Screening a light-up aptamer gene library requires isolating each gene prior to its expression into RNA followed by the assessment of its light-up capacity. In a first scheme, the library can be inserted into an expression plasmid later used to transform bacterial cells that will express light-up aptamers [31,62]. Bacteria can then be screened in two ways. First, they can be plated onto a solid medium supplemented with the fluorogen and allowed to grow until forming colonies [62]. The plates are then observed under fluorogen exciting light and the most fluorescent clones are recovered (Figure 3a). Even though it is simple and direct, such approach suffers from a limited throughput (a few thousand clones per day) and exploring large libraries can rapidly become time consuming, tedious and expensive. Alternatively, upon gene expression the cells can be incubated with the fluorogen prior to being screened using a Fluorescence-Activated Cell Sorter (FACS) (Figure 3b) [31], which substantially increases the analytical throughput (up to several million genes analyzed per day). However, even though it is more accurate, these screening approaches explore a much smaller fraction of the sequence space than SELEX does. Therefore, preferred methods combine library pre-enrichment by SELEX followed by ultrahigh-throughput screening using FACS. Such a tandem procedure was already used to obtain different light-up aptamers such as Broccoli [31], a new DFHBI-binding molecules [63] and Corn [34]. These in vivo approaches offer the great advantage of directly assaying the performance of the aptamers in the cellular context, and in doing so to select those RNAs with optimal in vivo performances. However, they also suffer two major drawbacks: (i) the absolute need of using a cell-permeable fluorogen (a property dispensable for in vitro applications); and (ii) the applicable selection pressures are limited to those compatible with living cells (e.g., presence of potassium, physiological pH and temperature). Moreover, in the case of FACS, since the analysis is performed at single cell level, the screening may also face cell-to-cell expression variability and requires the co-expression of an internal fluorescent standard (e.g., GFP and mCherry).

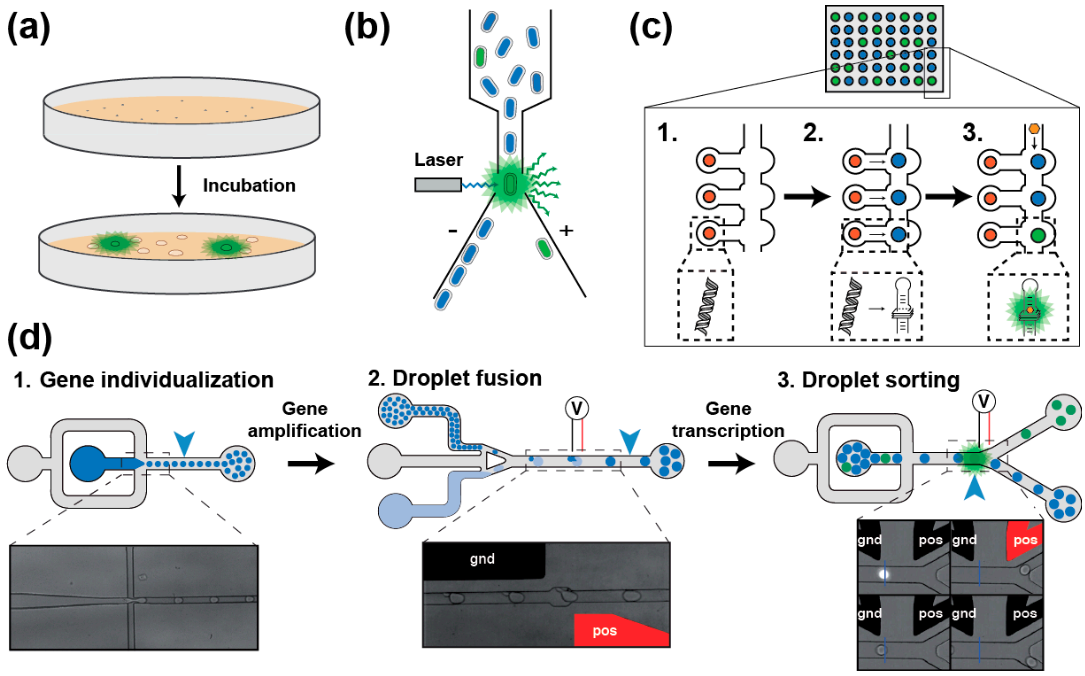


Figure 3. Main screening strategies for isolating light-up RNA aptamers. (**a**) Colony screening [31,62]. SELEX-enriched gene library is cloned and expressed in bacteria plated on a solid medium supplemented with the fluorogen. Upon incubation, colonies expressing light-up aptamers are identified by fluorescence emission (green shadow) when illuminated with fluorogen excitation wavelength. (**b**) FACS-based bacteria screening [31]. SELEX-enriched gene library is cloned and expressed in bacteria incubated in the presence of the fluorogen. Bacteria fluorescence is then analyzed on a FACS using a laser exciting the fluorogen. Fluorescence emitting bacteria (green shadow) are then deflected and sorted from the rest of the population. (**c**) Miniaturized in vitro screening using large integration scale microfluidic devices [64]. DNA coding for each variant is first spotted (red dots, Step 1) onto a surface prior to assembling the microfluidic chip. Then, each DNA cluster is transcribed in RNA later captured on a second spot of the chip (blue spot, Step 2). Finally, flowing the fluorogen into the microfluidic channels (Step 3) allows the detection of light-up aptamers (green spot). Quantifying the fluorescence emitted by each construct at various concentrations of fluorogen allows the determination of parameters such as the brightness and the dissociation constant (K_D). (**d**) Droplet-based microfluidic in vitro screening workflow [65]. A gene library is diluted into a PCR mixture (in dark blue) prior to individualizing the molecules into picoliter-sized water-in-oil droplets carried by an oil phase (in gray, Step 1). Upon off-chip PCR amplification, small and amplified DNA-containing droplets (dark blue) are synchronized and fused with larger droplets containing an in vitro transcription mixture supplemented with the fluorogen (light blue, Step 2). Upon an incubation step allowing for in vitro transcription to take place, the fluorescence of each droplet is analyzed and those droplets containing a light-up aptamer (in green) are sorted from the rest of the population (in blue). Both fusion and sorting events are triggered by the application of an electric field to built-in positive (pos, shown in red) and ground (gnd, shown in black) electrodes.

Most of the cell-based screening limitations can be overcome by switching to an in vitro expression approach. Performing such screening in microtiter plate can rapidly become very expensive, making the miniaturization of the process using microfluidics highly desirable [66]. A first level of miniaturization can be achieved by using a large-scale integration microfluidic device made of 640 independent one-nanoliter chambers, each containing several copies of the DNA coding for the variant to assay (Figure 3c) [64]. Then, genes are in vitro transcribed into RNA, which are later individually assayed for fluorogenic capacity. Moreover, the possibility of varying ligand concentration during the experiment allows the extraction of thermodynamic parameters of each variant in an automated way. However, the low throughput of the method (640 chambers per run) mainly restricts its use for refinement purposes. Further substantial throughput increase and cost reduction can be achieved by transposing the screening to droplet-based microfluidics [65,67] (Figure 3d). In this format, DNA molecules of a library are diluted into a PCR mixture and individualized at very high throughputs (several thousands of droplets generated per second) in small two-picoliter water-in-oil droplets. Upon thermocycling, each small droplet is fused to a larger 18-picoliter droplet containing an in vitro transcription mixture supplemented with the fluorogen. Finally, upon a last incubation, the fluorescence of each droplet is profiled and the most fluorescent ones (so those containing the best light-up aptamers) are sorted. Using this approach, we recently isolated iSpinach, a variant of the Spinach light-up aptamer optimized for in vitro applications (far less salt-sensitive, brighter and more thermo-stable) [65]. The isolation of this variant was made possible by the great control over the reaction conditions and the possibility of applying strong selection pressures difficult to apply in cell-based screening (e.g., warming and complete replacement of potassium ions by sodium). Therefore, droplet-based microfluidics screening is a viable and efficient way of isolating optimized light-up aptamers for in vitro application. Finally, further automation of the process could be obtained by integrating a Next Generation Sequencing analysis at the end of the process.

2.3. Features of the Best Characterized RNA-Based Fluorogenic Modules

Among the variety of RNA-based fluorogenic modules developed so far (Table 1), those involving MG, fluorescent protein mimicking dyes (e.g., DFHBI and DFHO) and TO-1 are the best characterized, especially from a structural point of view. Interestingly, even though each of these RNAs adopts a distinct folding, they all possess a fluorogen-binding pocket comprising an extended planar platform made of at least one base quadruple onto which the fluorogen is accommodated in a near planar conformation competent for efficient fluorescence emission (see [68] for a recent critical review on the topic). Whereas the MG-binding aptamer platform is made of mixed base quadruple [69], the platform of all the other light-up aptamers is made from a stack of two or more G-quartets stabilized by potassium ions [60,70–73], indicating that such platform might be a consensus solution to fluorogens binding. However, beside this consensus platform, the rest of the fluorogen-binding pocket as well as the recognition strategy vary from one aptamer to the other. Indeed, for instance, MG-binding pocket acquires its structure only upon dye binding. On the other hand, in Spinach and its derivatives, the DFHBI intercalates between the platform and a base triple whilst additional contacts are established with lateral bases [70–72]. This pocket is formed prior to dye binding and was found to accommodate other dyes [74] limiting the possibility of using these aptamers concomitantly to other ones in multiplex experiments. Moreover, the pocket does not strongly constrain the dye that is free to rapidly isomerize upon illumination followed by a rapid exchange with a non-isomerized fluorogen to restore the fluorescent module [75]. DFHBI-based fluorogenic modules are therefore poorly photostable under continuous-wave illumination and are better imaged using a pulsed mode illumination scheme. This behavior challenges the accurate monitoring of some biological phenomena (e.g., RNA movement in the cell). Moreover, even though no crystal structure has been solved yet for Broccoli (another DFHBI-binding aptamer), the strong sequence homologies as well as photophysical behavior shared with Spinach suggest that both aptamers are very likely to possess an identical fluorogen-binding pocket [76]. Recently, the structure of Corn RNA in complex with DFHO has been resolved [73] and revealed that, surprisingly, the fluorogen-binding pocket does not lie into a single RNA molecule but instead is formed by the interface of two interacting protomers of identical sequence, each made of four base quadruples and six adenine residues asymmetrically oriented. This preformed binding pocket can specifically accommodate a single DFHO and not only triggers its fluorescence but it also constitutes a highly protective environment preserving the dye from photobleaching over a long period of time [34]. Consequently, even though Corn/DFHO module is dimmer than the other systems, its strong photostability makes it well suited for the prolonged imaging of RNAs in live cells. Finally, the crystal structure of RNA Mango in complex with TO-1 helped explain in part the origin of the elevated affinity in the complex. Indeed, like the other aptamers, Mango possesses a planar platform made of G-quartets and interacts not only with the cyanine moiety but also with the PEG linker and the biotin group of the fluorogen [60]. Interestingly, both heterocycles of the cynanine are not coplanar but instead they make a 45 °C angle, suggesting that brighter mutants of the aptamer that would accommodate the cyanine more coplanarly could, in principle, be isolated by functional screening (a work currently in progress in our lab).

Up to now, RNA-based fluorogenic modules involving quenched fluorogen have been less characterized. However, even though these modules display a lower affinity than the aforementioned ones (Table 1), they have the great advantage of using very bright and photostable organic dyes as fluorogenic moiety (i.e., symmetrical cynanines [35,77], Sulforhodamines [55] or Rhodamine Green [32]), making them attractive for future developments.

3. Applications of RNA-Based Fluorogenic Modules

The development of RNA-based fluorogenic modules was initially motivated by the need for genetically encoded tools to monitor gene expression and regulation at the RNA level. However, the great flexibility and the ease to engineer these RNA molecules rapidly led to the development of new sets of fluorogenic reporters with applications both in vivo and in vitro (Figure 4).

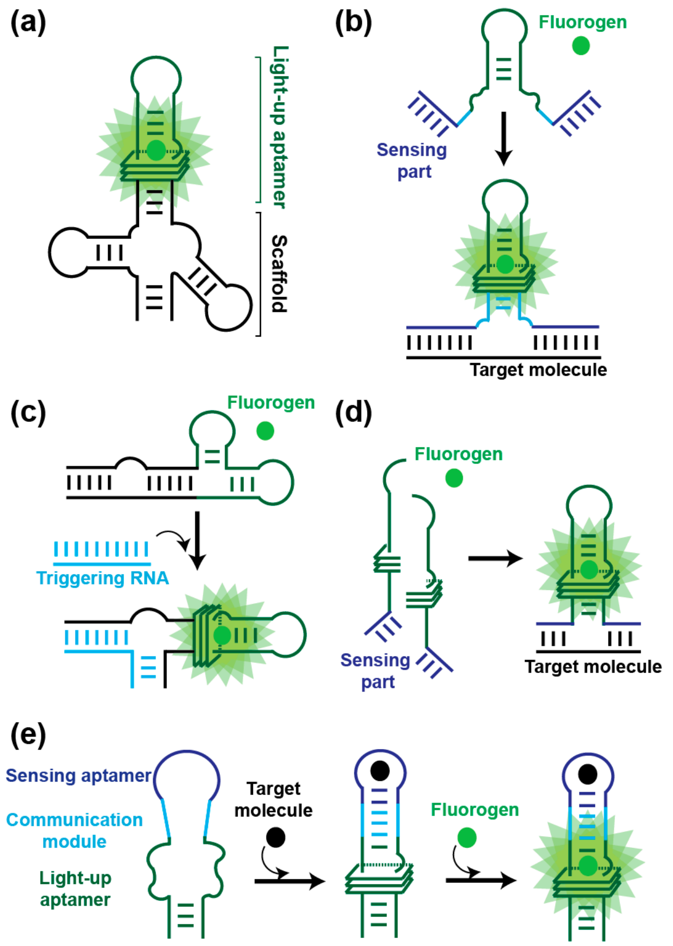


Figure 4. Light-up aptamers engineering and uses thereof. (**a**) Light-up RNA aptamers can be embedded into a scaffold RNA (here a tRNA) to improve folding efficiency and resistance to RNases [29,61]. The construct can either be directly expressed from an independent promoter or inserted into the sequence of a target mRNA. (**b**) Light-up aptamer sensing a target RNA in trans [77,78]. A key helix (light blue) of the light-up aptamer is engineered to weaken the structure of the RNA and destabilized the fluorogen-binding site. Furthermore, sequences complementary to the target nucleic acid (acting as a sensing part) are appended to both ends of the molecule (dark blue). Upon binding to the target molecule (in black), the structure of the light-up moiety is stabilized and its fluorogen-binding capacity is restored, leading to fluorescence emission (green shadow). (**c**) Light-up aptamer sensing a target RNA in trans via strand displacement [79–81]. Sequences are appended to both ends of the light-up aptamer and induce an alternative folding preventing the interaction of the aptamer with the fluorogen. However, the binding of the target RNA to the construct induces a conformational change restoring the structure of the light-up aptamer moiety and its fluorogen-binding capacity, leading to fluorescence emission (green shadow). (**d**) Sensing target nucleic acids using split light-up aptamers [82,83]. The aptamer is split into two halves, each possessing a sequence complementary to the target nucleic acid (acting as a sensing part, in dark blue). Whereas both halves cannot form a functional aptamer in the absence of the target nucleic acid, the presence of the latter drives the productive association of both molecules into a functional light-up aptamer able to emit fluorescence. (**e**) Allosteric light-up RNA aptamer [84,85]. The light-up aptamer is engineered to weaken the structure of the RNA and destabilize the fluorogen-binding site. Moreover, the sequence of a sensing aptamer (in dark blue) is inserted into that of the light-up aptamer via a communication module (in light blue). Upon the interaction of the sensing aptamer with a target molecule (e.g., a metabolite or a protein), the structure of the light-up aptamer is stabilized and its fluorogen-binding capacity is restored. Note that all the examples are shown with the Spinach aptamer, but these concepts can be extended to any other light-up aptamer.

3.1. Live-Cell Imaging of Biomolecules

3.1.1. Live-Cell RNA Imaging

Direct expression of light-up RNA aptamers in cells can be challenging due to their degradation by cell nucleases. Therefore, the molecules are usually expressed inserted into a second RNA such as a transfer RNA (tRNA) [29,86] or a three-way junction [87] acting as a scaffold to both protect the aptamer from rapid degradation and assist its folding (Figure 4a). Expressing such constructs from a strong promoter (e.g., T7 RNA polymerase promoter in *E. coli*, 5S promoter in mammalian cells) allows the expression of abundant non-coding RNAs to be monitored in prokaryotes [29,32,55] and eukaryotes [29,31,34]. Moreover, recent publications reported on the possibility to insert Spinach-derived aptamers into the variable region of bacterial tRNAs [88], within an apical loop of the bacterial 16S ribosomal RNA [89] or in metazoan tRNA introns [90] while preserving the natural functionality of the carrier molecules. Being able to image these non-coding RNAs not only informs on their own expression but it can also be used to diagnose the whole associated pathway (e.g., Pol III transcription, translation, etc.). Spinach was also found suited for imaging messenger RNAs in organisms as diverse as bacteria [91–93], yeast [94], mammalian cells [61], viruses [95–97] and algae [98]. The limited brightness and poor photostability of Spinach and its derivatives initially limited its use to the imaging of mRNAs either aggregated [61] or labeled with tandem repeats of up to 64 copies of the aptamer [93]. However, an improved understanding the photophysics of DFHBI inactivation allowed devising pulsed excitation-based imaging procedures making it possible to monitor Spinach-labeled RNAs over a much longer period of time [75] and enabling imaging much less abundant mRNAs [93]. Furthermore, using spinning disk confocal microscopy and post-acquisition image analysis algorithms it was possible to track the nuclear export of yeast mRNA labeled with a single copy of Spinach [94]. Interestingly, these studies also revealed that inserting the light-up aptamer into the 3$'$ untranslated region of the mRNA did not affect the fate (i.e., localization, translation or stability) of the labeled molecule.

Direct insertion of light-up aptamers into target RNAs is mainly limited by the laborious genetics required for modifying the locus of each target gene. This limitation can nevertheless be overcome by using in trans-acting fluorogenic modules that are destabilized forms of the light-up aptamer [77,78] (Figure 4b). In this approach, the aptamer is engineered at two levels. First, a helix of the aptamer is truncated to weaken the structure of the RNA and abolish its capacity to interact with the fluorogen. Second, sequences complementary to the target RNA are appended at both extremities of the helix. Upon binding to the target RNA, the structure of the aptamer is stabilized and its capacity to bind and activate the fluorogen is restored. Provided an unstructured region is accessible in the target RNA, this approach can be used to image both mRNA [77,78] and small non-coding RNA [99]. Moreover, embedding the aptamer into a scaffold RNA and triggering its activation by a strand-displacement mechanism respectively improve the stability of the fluorogenic module and increase signal amplitude [80,81] (Figure 4c). Further improvement in detection specificity and signal amplitude can be obtained by using split aptamers [82,83] (Figure 4d). Recently, a split version of the Broccoli light-up aptamer was also generated and embedded into an AND logic gate for sensing the simultaneous presence of two target RNAs in vivo [100].

3.1.2. Live-Cell Imaging of Metabolites and Proteins

Its great structural flexibility enables RNA to fold into aptamers able to specifically recognize virtually any type of molecule. Moreover, if the interaction between the aptamer and its target leads to a structural remodeling or even the stabilization of RNA structure, then such aptamer can be used for the conception of an allosteric biosensor in which a sensing module is connected to a reporting module via a communication module (Figure 4e). Briefly, the structure of the reporting module (a ribozyme or a light-up aptamer) is destabilized which keeps its function (e.g., RNA cleavage or fluorescence emission) in an off state. However, the presence of the target analyte induces a structural change (or stabilization) of the sensing aptamer that is transmitted to the reporting module via the communication module. The structure of the reporting RNA is in turn stabilized and its function restored. This strategy was originally pioneered using ribozymes as reporting modules [101] and early work with MG aptamer allowed establishing the proof-of-concept experiment using light-up aptamers as reporter [84]. Since then, Spinach RNA-based fluorogenic biosensors have been developed to specifically report on the presence of metabolites such as FMN [84], cyclic-di-GMP [85,102–104], cyclic-di-AMP [105], cyclic AMP-GMP [106,107], cyclic-AMP [108], S-adenosylmethionine (SAM) [109], S-adenosyl-L-homocysteine (SAH) [110], Thiamine Pyrophosphate (TPP) [111] and neurotransmitter precursors [112]. The majority of these biosensors are built using riboswitch-derived aptamers. Indeed, since these molecules have naturally evolved to switch RNA structure in cellular environment, these aptamers possess the required structural flexibility and are able to efficiently discriminate their cognate target from closely related analogues contained in the cell. Finally, even though they were less used, allosteric biosensors can also be developed for protein detection [113]. With such biosensor in hands it is then possible to precisely detect and quantify a target molecule in a dynamic way and with single cell resolution. This allows the dynamics of a biological pathway to be monitored, but also to characterize the enzymes involved in this pathway, paving the way for drug discovery applications [105–107].

Whereas both sensing and reporting modules have key roles in the proper function of allosteric fluorogenic biosensors, the communication module is probably the most critical element, as it drives the information from one aptamer to the other. Usually, a communication module is rationally designed based on thermodynamic considerations since such module would be extremely challenging to isolate using a SELEX approach. However, the recent advances in high-throughput screening technologies introduced above offer a new way of rapidly identifying optimal communication modules. For instance, a large-scale integration microfluidic device was used to systematically screen a library of 94 different Spinach-based glycine biosensor containing different communication modules while measuring their response to various glycine concentrations [114]. Therefore, a single experiment allows accessing both the fluorescence amplitude of many biosensors as well as their affinity for the glycine and their dynamic range. Consequently, assisting the development of new biosensors by high-throughput screening technologies should both accelerate their development and improve their performances. Finally, one should note that a communication might be dispensable if one simply wants to develop an aptamer-based protein-targeting fluorescent probe. In that case, the binding aptamer can be directly connected to a light-up aptamer as recently exemplified by the use of a fluorogenic cyanine-binding aptamer fused with a VEGF-binding aptamer [27].

3.2. *In Vitro Applications*

The possibility of detecting RNA via fluorogenic assays opened a whole bunch of applications in vitro. First, inserting the sequence of a light-up aptamer into that of another RNA allows monitoring in real-time the in vitro transcription of the construct [35,115] and, doing so, to compute its transcription rate while decorrelating it from the translation rate of an encoded protein. This is of particular importance in experiments aiming at engineering genetic circuits, especially for generating artificial cells [116–118]. Being able to monitor transcription activity in an automatable way makes also possible to set-up high-throughput microtiter plate-based screening of drugs targeting RNA polymerase activity [115]. Drug screening actually represents a second important set of applications of light-up aptamers in vitro. For instance, the fluorogenic function of the RNA can be transiently inactivated by modifying (e.g., methylating) a key nucleotide to convert the aptamer into a fluorogenic substrate of a modification-removing enzyme such as demethylases, a class of enzymes involved in several diseases [119]. Monitoring reaction product apparition by the mean of the fluorogenic allosteric biosensors introduced above [110,120,121] represents an alternative way of assaying enzyme activity. In any case, these different strategies can easily be transposed to microtiter plate format to set-up a high-throughput screening workflow [115,119,120] and even an ultrahigh-throughput one by using droplet-based microfluidics [121].

The great ability to predict to some extent the folding of an RNA sequence makes possible the design of molecular circuits composed of two or several RNA (or DNA) elements trapped by folding to keep the circuit silent [122,123]. However, in the presence of one or several target molecules, sensing elements undergo structural rearrangements activating the circuit to compute the information and eventually reports the presence of the targets via engineered light-up aptamer [79,83,124]. Whereas simple circuit made of a split light-up aptamer directly sensing the target molecule would lead to a signal stoichiometric to the input [124], the implementation of an enzyme-free amplification loop [122,123,125] is expected to significantly increase the sensitivity of these assays.

Split and full-length light-up aptamers can also be used to devise in vitro procedures aimed at assisting catalytic RNAs engineering [126,127], tracking RNA-based molecular complexes as well as monitoring the formation of supramolecular assemblies such as those used in RNA nanotechnology [128–131] directly through fluorescence emission measurements. Moreover, the proper size and integrity of light-up-labeled RNAs can also be verified by gel electrophoresis followed by specific staining of the target RNAs by the fluorogen directly in the gel [87,131]. Finally, light-up aptamers inserted into a target RNA can serve as handles to specifically fish-out the labeled RNA and its associated partners in native conditions by incubating a reaction mixture (e.g., a cell lysate) with a biotinylated fluorogen and purifying the resulting complex on a streptavidin-conjugated resin [132].

4. Conclusions

Over the past decade, significant efforts were devoted to the development of new RNA-based fluorogenic modules leading to a toolbox of modules covering the whole visible spectrum (Table 1). Many of these modules are significantly brighter than the widely used GFP and, conversely to their protein counterparts, they do not require oxygen since no maturation step is needed. Therefore, labeling target RNAs with light-up aptamers is not only subjected to a low fluorescent background, but RNAs can also be monitored in conditions in which fluorescent proteins could not be used (e.g., anaerobiosis [104]). Besides direct RNA imaging in vivo, RNA-based fluorogenic modules also found a wide range of applications as briefly reviewed in this article and it is very likely that many new ones shall be developed in a near future (see below).

Interestingly, even though current modules are already pretty efficient and permitted multiple questions to be answered, none of them completely fulfills all the criteria enounced in Section 2 of this review, suggesting that there is still room for improvement in terms of sensitivity and robustness. Indeed, most of the current light-up aptamers were obtained by SELEX using moderate selection pressures and have dissociation constant (so affinity) for the fluorogen of tens of nM or more. Thus far, Mango is the only RNA displaying a very high affinity for its fluorogen (K_D of a few nM), likely resulting from the use of very stringent wash and elution conditions. Therefore, applying the same concept to other systems could yield modules with much higher affinity. This would reduce the amount of dye in the assays and lead to a gain in sensitivity by improving the contrast of the experiment [37]. Moreover, since SELEX-derived RNAs were selected for their capacity to bind the fluorogen rather than their capacity to generate fluorescent complexes, it is likely that the fluorescence of most of the current modules may still be sub-optimal. Indeed, our own experience with Spinach [65] and Mango (on-going work in our lab) RNAs shows that revisiting these aptamers by a screening approach may significantly increase their light-up capacity. Such improvement could result from an increased brightness but also from a higher photostability. Indeed, both properties rely not only on the fluorogen itself but also on capacity of the RNA to properly accommodate it and protect it from unwanted photoisomerization and photobleaching. This last point was recently exemplified by Corn, an RNA isolated by SELEX in tandem with FACS screenings, that forms with DFHO a complex displaying an extraordinary high photostability [34]. This photostability likely results from a caging of the fluorogen preventing its rapid isomerization [73]. This property is clearly contributed by the RNA since lighting-up DFHO with Broccoli did not yield a significant photostability [34]. Therefore, we anticipate that future selection schemes combining conventional SELEX with an (ultra)high-throughput functional screening (e.g., FACS or microfluidic-based screening) together with next-generation sequencing could lead to the efficient discovery of light-up aptamers with superior turn-on and photostabilizing properties. This new generation of molecules should further enlarge the already wide application range of RNA-based fluorogenic modules. For example, they would allow labeling target RNAs with a limited number of aptamer repeats (ideally a single one) to make possible single molecule resolution, thus enabling the tracking of low abundant RNA molecules, for instance using super-resolution microscopy [133]. New generations of fluorogenic modules would not only allow highly sensitive gene expression monitoring, but they could also serve to set-up new drug screening pipelines as well as a variety of analytical platforms such as microarrays developed to sense target molecules with size ranging from ions (recently exemplified by the capacity of Spinach to specifically sense lead [134]) to more complex protein using allosteric biosensors such as those introduce in this review. Finally, whereas this review was focused on RNA aptamers, DNA light-up aptamers were also reported [28,52] and one should consider the possibility of applying the concepts exposed throughout this review to the development of new DNA-based sensors.

Acknowledgments: We thank Redmond Smyth for his help in manuscript proofreading. This work has been published under the framework of the LabExNetRNA (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 also received the financial support of the Agence Nationale de la Recherche (ANR-16-CE11-0010-01) and was supported by the Université de Strasbourg and the Centre National de la Recherche Scientifique.

Author Contributions: All authors analyzed the literature and wrote the paper. Farah Bouhedda and Alexis Autour equally contributed to this work.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Cy3-BHQ1	Cyanine 3-Black Hole Quencher 1
DIR	Dimethyl Indole Red
DFHBI	3,5-difluoro-4-hydroxybenzylidene imidazolinone
DFHBI-1(2)T	3,5-difluoro-4-hydroxybenzylidene imidazolinone 1 (2) trifluoroethyl
DFHO	3,5-difluoro-4-hydroxybenzylidene imidazolinone-2-oxime
FACS	Fluorescence Activated Cell Sorter
GFP	Green Fluorescent Protein
MG	Malachite Green
OTB	Oxazole Thiazole Blue
RG-DN	Rhodamine Green-DiNitroaniline
SELEX	Systematic Evolution of Ligands by EXponential enrichment
SR-DN	Sulforhodamine-DiNitroaniline
TMR-DN	Tertamethyl rhodamine-DiNitroaniline
TO-1 (3)	Thiazole Orange 1 (3)

References

1. Schaferling, M. The art of fluorescence imaging with chemical sensors. *Angew. Chem.* **2012**, *51*, 3532–3554. [CrossRef] [PubMed]
2. Wong, J.K.; Todd, M.H.; Rutledge, P.J. Recent advances in macrocyclic fluorescent probes for ion sensing. *Molecules* **2017**, *22*, 200. [CrossRef] [PubMed]
3. Cui, C.; Shu, W.; Li, P. Fluorescence in situ hybridization: Cell-based genetic diagnostic and research applications. *Front. Cell Dev. Biol.* **2016**, *4*, 89. [CrossRef] [PubMed]
4. Shimomura, O.; Johnson, F.H.; Saiga, Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J. Cell. Comp. Physiol.* **1962**, *59*, 223–239. [CrossRef] [PubMed]

5. Cody, C.W.; Prasher, D.C.; Westler, W.M.; Prendergast, F.G.; Ward, W.W. Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein. *Biochemistry* **1993**, *32*, 1212–1218. [CrossRef] [PubMed]
6. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W.W.; Prasher, D.C. Green fluorescent protein as a marker for gene expression. *Science* **1994**, *263*, 802–805. [CrossRef] [PubMed]
7. Heim, R.; Prasher, D.C.; Tsien, R.Y. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 12501–12504. [CrossRef] [PubMed]
8. Ormo, M.; Cubitt, A.B.; Kallio, K.; Gross, L.A.; Tsien, R.Y.; Remington, S.J. Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **1996**, *273*, 1392–1395. [CrossRef] [PubMed]
9. Chudakov, D.M.; Matz, M.V.; Lukyanov, S.; Lukyanov, K.A. Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol. Rev.* **2010**, *90*, 1103–1163. [CrossRef] [PubMed]
10. Griffin, B.A.; Adams, S.R.; Tsien, R.Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **1998**, *281*, 269–272. [CrossRef] [PubMed]
11. Correa, I.R., Jr. Live-cell reporters for fluorescence imaging. *Curr. Opin. Chem. Biol.* **2014**, *20*, 36–45. [CrossRef] [PubMed]
12. Li, C.; Tebo, A.G.; Gautier, A. Fluorogenic labeling strategies for biological imaging. *Int. J. Mol. Sci.* **2017**, *18*, 1473. [CrossRef] [PubMed]
13. Bertrand, E.; Chartrand, P.; Schaefer, M.; Shenoy, S.M.; Singer, R.H.; Long, R.M. Localization of *ASH1* mRNA particles in living yeast. *Mol. Cell* **1998**, *2*, 437–445. [CrossRef]
14. Wu, B.; Chen, J.; Singer, R.H. Background free imaging of single mRNAs in live cells using split fluorescent proteins. *Sci. Rep.* **2014**, *4*, 3615. [CrossRef] [PubMed]
15. Buxbaum, A.R.; Haimovich, G.; Singer, R.H. In the right place at the right time: Visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 95–109. [CrossRef] [PubMed]
16. Holeman, L.A.; Robinson, S.L.; Szostak, J.W.; Wilson, C. Isolation and characterization of fluorophore-binding RNA aptamers. *Fold. Des.* **1998**, *3*, 423–431. [CrossRef]
17. Shin, I.; Ray, J.; Gupta, V.; Ilgu, M.; Beasley, J.; Bendickson, L.; Mehanovic, S.; Kraus, G.A.; Nilsen-Hamilton, M. Live-cell imaging of pol II promoter activity to monitor gene expression with RNA IMAGEtag reporters. *Nucleic Acids Res.* **2014**, *42*, e90. [CrossRef] [PubMed]
18. Eydeler, K.; Magbanua, E.; Werner, A.; Ziegelmuller, P.; Hahn, U. Fluorophore binding aptamers as a tool for RNA visualization. *Biophys. J.* **2009**, *96*, 3703–3707. [CrossRef] [PubMed]
19. Babendure, J.R.; Adams, S.R.; Tsien, R.Y. Aptamers switch on fluorescence of triphenylmethane dyes. *J. Am. Chem. Soc.* **2003**, *125*, 14716–14717. [CrossRef] [PubMed]
20. Grate, D.; Wilson, C. Laser-mediated, site-specific inactivation of RNA transcripts. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6131–6136. [CrossRef] [PubMed]
21. Ouellet, J. RNA fluorescence with light-up aptamers. *Front. Chem.* **2016**, *4*, 29. [CrossRef] [PubMed]
22. Rath, A.K.; Rentmeister, A. Genetically encoded tools for RNA imaging in living cells. *Curr. Opin. Biotechnol.* **2015**, *31*, 42–49. [CrossRef] [PubMed]
23. Pauff, S.; Withers, J.M.; McKean, I.J.; Mackay, S.P.; Burley, G.A. Synthetic biological approaches for RNA labelling and imaging: Design principles and future opportunities. *Curr. Opin. Biotechnol.* **2017**, *48*, 153–158. [CrossRef] [PubMed]
24. Schneider, A.F.; Hackenberger, C.P. Fluorescent labelling in living cells. *Curr. Opin. Biotechnol.* **2017**, *48*, 61–68. [CrossRef] [PubMed]
25. Van Gijtenbeek, L.A.; Kok, J. Illuminating messengers: An update and outlook on RNA visualization in bacteria. *Front. Microbiol.* **2017**, *8*, 1161. [CrossRef] [PubMed]
26. Heim, R.; Cubitt, A.B.; Tsien, R.Y. Improved green fluorescence. *Nature* **1995**, *373*, 663–664. [CrossRef] [PubMed]
27. Tan, X.; Constantin, T.P.; Sloane, K.L.; Waggoner, A.S.; Bruchez, M.P.; Armitage, B.A. Fluoromodules consisting of a promiscuous RNA aptamer and red or blue fluorogenic cyanine dyes: Selection, characterization, and bioimaging. *J. Am. Chem. Soc.* **2017**, *139*, 9001–9009. [CrossRef] [PubMed]
28. Sando, S.; Narita, A.; Hayami, M.; Aoyama, Y. Transcription monitoring using fused RNA with a dye-binding light-up aptamer as a tag: A blue fluorescent RNA. *Chem. Commun.* **2008**, *44*, 3858–3860. [CrossRef] [PubMed]
29. Paige, J.S.; Wu, K.Y.; Jaffrey, S.R. RNA mimics of green fluorescent protein. *Science* **2011**, *333*, 642–646. [CrossRef] [PubMed]

30. Song, W.; Strack, R.L.; Svensen, N.; Jaffrey, S.R. Plug-and-play fluorophores extend the spectral properties of Spinach. *J. Am. Chem. Soc.* **2014**, *136*, 1198–1201. [CrossRef] [PubMed]
31. Filonov, G.S.; Moon, J.D.; Svensen, N.; Jaffrey, S.R. Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc.* **2014**, *136*, 16299–16308. [CrossRef] [PubMed]
32. Arora, A.; Sunbul, M.; Jaschke, A. Dual-colour imaging of RNAs using quencher- and fluorophore-binding aptamers. *Nucleic Acids Res.* **2015**, *43*, e144. [CrossRef] [PubMed]
33. Dolgosheina, E.V.; Jeng, S.C.; Panchapakesan, S.S.; Cojocaru, R.; Chen, P.S.; Wilson, P.D.; Hawkins, N.; Wiggins, P.A.; Unrau, P.J. RNA Mango aptamer-fluorophore: A bright, high-affinity complex for RNA labeling and tracking. *ACS Chem. Biol.* **2014**, *9*, 2412–2420. [CrossRef] [PubMed]
34. Song, W.; Filonov, G.S.; Kim, H.; Hirsch, M.; Li, X.; Moon, J.D.; Jaffrey, S.R. Imaging RNA polymerase III transcription using a photostable RNA-fluorophore complex. *Nat. Chem. Biol.* **2017**, *13*, 1187–1194. [CrossRef] [PubMed]
35. Murata, A.; Sato, S.; Kawazoe, Y.; Uesugi, M. Small-molecule fluorescent probes for specific RNA targets. *Chem. Commun.* **2011**, *47*, 4712–4714. [CrossRef] [PubMed]
36. Constantin, T.P.; Silva, G.L.; Robertson, K.L.; Hamilton, T.P.; Fague, K.; Waggoner, A.S.; Armitage, B.A. Synthesis of new fluorogenic cyanine dyes and incorporation into RNA fluoromodules. *Org. Lett.* **2008**, *10*, 1561–1564. [CrossRef] [PubMed]
37. Dolgosheina, E.V.; Unrau, P.J. Fluorophore-binding RNA aptamers and their applications. *Wiley Interdiscip. Rev. RNA* **2016**, *7*, 843–851. [CrossRef] [PubMed]
38. Klymchenko, A.S. Solvatochromic and fluorogenic dyes as environment-sensitive probes: Design and biological applications. *Acc. Chem. Res.* **2017**, *50*, 366–375. [CrossRef] [PubMed]
39. Lo̊ber, G. The fluorescence of dye-nucleic acid complexes. *J. Lumin.* **1981**, *22*, 221–265. [CrossRef]
40. Olmsted, J., 3rd; Kearns, D.R. Mechanism of ethidium bromide fluorescence enhancement on binding to nucleic acids. *Biochemistry* **1977**, *16*, 3647–3654. [CrossRef] [PubMed]
41. Jin, R.; Breslauer, K.J. Characterization of the minor groove environment in a drug-DNA complex: Bisbenzimide bound to the poly[d(AT)].Poly[d(AT)]duplex. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 8939–8942. [CrossRef] [PubMed]
42. Sando, S.; Narita, A.; Aoyama, Y. Light-up hoechst-DNA aptamer pair: Generation of an aptamer-selective fluorophore from a conventional DNA-staining dye. *Chembiochem* **2007**, *8*, 1795–1803. [CrossRef] [PubMed]
43. Kraus, G.A.; Jeon, I.; Nilsen-Hamilton, M.; Awad, A.M.; Banerjee, J.; Parvin, B. Fluorinated analogs of malachite green: Synthesis and toxicity. *Molecules* **2008**, *13*, 986–994. [CrossRef] [PubMed]
44. Lux, J.; Pena, E.J.; Bolze, F.; Heinlein, M.; Nicoud, J.F. Malachite green derivatives for two-photon RNA detection. *Chembiochem* **2012**, *13*, 1206–1213. [CrossRef] [PubMed]
45. Ilgu, M.; Ray, J.; Bendickson, L.; Wang, T.; Geraskin, I.M.; Kraus, G.A.; Nilsen-Hamilton, M. Light-up and FRET aptamer reporters; evaluating their applications for imaging transcription in eukaryotic cells. *Methods* **2016**, *98*, 26–33. [CrossRef] [PubMed]
46. Saurabh, S.; Perez, A.M.; Comerci, C.J.; Shapiro, L.; Moerner, W.E. Super-resolution imaging of live bacteria cells using a genetically directed, highly photostable fluoromodule. *J. Am. Chem. Soc.* **2016**, *138*, 10398–10401. [CrossRef] [PubMed]
47. Silva, G.L.; Ediz, V.; Yaron, D.; Armitage, B.A. Experimental and computational investigation of unsymmetrical cyanine dyes: Understanding torsionally responsive fluorogenic dyes. *J. Am. Chem. Soc.* **2007**, *129*, 5710–5718. [CrossRef] [PubMed]
48. Armitage, B.A. Cyanine dye–nucleic acid interactions. In *Heterocyclic Polymethine Dyes: Synthesis, Properties and Applications*; Strekowski, L., Ed.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 11–29.
49. Lee, L.G.; Chen, C.H.; Chiu, L.A. Thiazole orange: A new dye for reticulocyte analysis. *Cytometry* **1986**, *7*, 508–517. [CrossRef] [PubMed]
50. Nygren, J.; Svanvik, N.; Kubista, M. The interactions between the fluorescent dye thiazole orange and DNA. *Biopolymers* **1998**, *46*, 39–51. [CrossRef]
51. Walker, C.L.; Lukyanov, K.A.; Yampolsky, I.V.; Mishin, A.S.; Bommarius, A.S.; Duraj-Thatte, A.M.; Azizi, B.; Tolbert, L.M.; Solntsev, K.M. Fluorescence imaging using synthetic GFP chromophores. *Curr. Opin. Chem. Biol.* **2015**, *27*, 64–74. [CrossRef] [PubMed]

52. Feng, G.; Luo, C.; Yi, H.; Yuan, L.; Lin, B.; Luo, X.; Hu, X.; Wang, H.; Lei, C.; Nie, Z.; et al. DNA mimics of red fluorescent proteins (RFP) based on G-quadruplex-confined synthetic RFP chromophores. *Nucleic Acids Res.* **2017**, *45*, 10380–10392. [CrossRef] [PubMed]
53. Sparano, B.A.; Koide, K. A strategy for the development of small-molecule-based sensors that strongly fluoresce when bound to a specific RNA. *J. Am. Chem. Soc.* **2005**, *127*, 14954–14955. [CrossRef] [PubMed]
54. Sparano, B.A.; Koide, K. Fluorescent sensors for specific RNA: A general paradigm using chemistry and combinatorial biology. *J. Am. Chem. Soc.* **2007**, *129*, 4785–4794. [CrossRef] [PubMed]
55. Sunbul, M.; Jaschke, A. Contact-mediated quenching for RNA imaging in bacteria with a fluorophore-binding aptamer. *Angew. Chem.* **2013**, *52*, 13401–13404. [CrossRef] [PubMed]
56. Ellington, A.D.; Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346*, 818–822. [CrossRef] [PubMed]
57. Tuerk, C.; Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990**, *249*, 505–510. [CrossRef] [PubMed]
58. Stoltenburg, R.; Reinemann, C.; Strehlitz, B. Selex—A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol. Eng.* **2007**, *24*, 381–403. [CrossRef] [PubMed]
59. Davis, J.H.; Szostak, J.W. Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11616–11621. [CrossRef] [PubMed]
60. Trachman, R.J., 3rd; Demeshkina, N.A.; Lau, M.W.L.; Panchapakesan, S.S.S.; Jeng, S.C.Y.; Unrau, P.J.; Ferre-D'Amare, A.R. Structural basis for high-affinity fluorophore binding and activation by RNA Mango. *Nat. Chem. Biol.* **2017**, *13*, 807–813. [CrossRef] [PubMed]
61. Strack, R.L.; Disney, M.D.; Jaffrey, S.R. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat. Methods* **2013**, *10*, 1219–1224. [CrossRef] [PubMed]
62. Lee, J.; Lee, K.H.; Jeon, J.; Dragulescu-Andrasi, A.; Xiao, F.; Rao, J. Combining SELEX screening and rational design to develop light-up fluorophore-RNA aptamer pairs for RNA tagging. *ACS Chem. Biol.* **2010**, *5*, 1065–1074. [CrossRef] [PubMed]
63. Zou, J.; Huang, X.; Wu, L.; Chen, G.; Dong, J.; Cui, X.; Tang, Z. Selection of intracellularly functional RNA mimics of green fluorescent protein using fluorescence-activated cell sorting. *J. Mol. Evol.* **2015**, *81*, 172–178. [CrossRef] [PubMed]
64. Ketterer, S.; Fuchs, D.; Weber, W.; Meier, M. Systematic reconstruction of binding and stability landscapes of the fluorogenic aptamer Spinach. *Nucleic Acids Res.* **2015**, *43*, 9564–9572. [CrossRef] [PubMed]
65. Autour, A.; Westhof, E.; Ryckelynck, M. iSpinach: A fluorogenic RNA aptamer optimized for in vitro applications. *Nucleic Acids Res.* **2016**, *44*, 2491–2500. [CrossRef] [PubMed]
66. Autour, A.; Ryckelynck, M. Ultrahigh-throughput improvement and discovery of enzymes using droplet-based microfluidic screening. *Micromachines* **2017**, *8*, 128. [CrossRef]
67. Ryckelynck, M.; Baudrey, S.; Rick, C.; Marin, A.; Coldren, F.; Westhof, E.; Griffiths, A.D. Using droplet-based microfluidics to improve the catalytic properties of RNA under multiple-turnover conditions. *RNA* **2015**, *21*, 458–469. [CrossRef] [PubMed]
68. Trachman, R.J., 3rd; Truong, L.; Ferre-D'Amare, A.R. Structural principles of fluorescent RNA aptamers. *Trends Pharmacol.Sci.* **2017**, *38*, 928–939. [CrossRef] [PubMed]
69. Baugh, C.; Grate, D.; Wilson, C. 2.8 a crystal structure of the malachite green aptamer. *J. Mol. Biol.* **2000**, *301*, 117–128. [CrossRef] [PubMed]
70. Huang, H.; Suslov, N.B.; Li, N.S.; Shelke, S.A.; Evans, M.E.; Koldobskaya, Y.; Rice, P.A.; Piccirilli, J.A. A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.* **2014**, *10*, 686–691. [CrossRef] [PubMed]
71. Warner, K.D.; Chen, M.C.; Song, W.; Strack, R.L.; Thorn, A.; Jaffrey, S.R.; Ferre-D'Amare, A.R. Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat. Struct. Mol. Biol.* **2014**, *21*, 658–663. [CrossRef] [PubMed]
72. Fernandez-Millan, P.; Autour, A.; Ennifar, E.; Westhof, E.; Ryckelynck, M. Crystal structure and fluorescence properties of the iSpinach aptamer in complex with dfhbi. *RNA* **2017**, *23*, 1788–1795. [CrossRef] [PubMed]
73. Warner, K.D.; Sjekloca, L.; Song, W.; Filonov, G.S.; Jaffrey, S.R.; Ferre-D'Amare, A.R. A homodimer interface without base pairs in an RNA mimic of red fluorescent protein. *Nat. Chem. Biol.* **2017**, *13*, 1195–1201. [CrossRef] [PubMed]

74. Jeng, S.C.; Chan, H.H.; Booy, E.P.; McKenna, S.A.; Unrau, P.J. Fluorophore ligand binding and complex stabilization of the RNA Mango and RNA Spinach aptamers. *RNA* **2016**, *22*, 1884–1892. [CrossRef] [PubMed]
75. Han, K.Y.; Leslie, B.J.; Fei, J.; Zhang, J.; Ha, T. Understanding the photophysics of the Spinach-DFHBI RNA aptamer-fluorogen complex to improve live-cell RNA imaging. *J. Am. Chem. Soc.* **2013**, *135*, 19033–19038. [CrossRef] [PubMed]
76. Ageely, E.A.; Kartje, Z.J.; Rohilla, K.J.; Barkau, C.L.; Gagnon, K.T. Quadruplex-flanking stem structures modulate the stability and metal ion preferences of RNA mimics of GFP. *ACS Chem. Biol.* **2016**, *11*, 2398–2406. [CrossRef] [PubMed]
77. Sato, S.; Watanabe, M.; Katsuda, Y.; Murata, A.; Wang, D.O.; Uesugi, M. Live-cell imaging of endogenous mRNAs with a small molecule. *Angew. Chem.* **2015**, *54*, 1855–1858. [CrossRef] [PubMed]
78. Ong, W.Q.; Citron, Y.R.; Sekine, S.; Huang, B. Live cell imaging of endogenous mRNA using RNA-based fluorescence "turn-on" probe. *ACS Chem. Biol.* **2017**, *12*, 200–205. [CrossRef] [PubMed]
79. Bhadra, S.; Ellington, A.D. A Spinach molecular beacon triggered by strand displacement. *RNA* **2014**, *20*, 1183–1194. [CrossRef] [PubMed]
80. Huang, K.; Doyle, F.; Wurz, Z.E.; Tenenbaum, S.A.; Hammond, R.K.; Caplan, J.L.; Meyers, B.C. Fastmir: An RNA-based sensor for in vitro quantification and live-cell localization of small RNAs. *Nucleic Acids Res.* **2017**, *45*, e130. [CrossRef] [PubMed]
81. Ying, Z.M.; Wu, Z.; Tu, B.; Tan, W.; Jiang, J.H. Genetically encoded fluorescent RNA sensor for ratiometric imaging of microRNA in living tumor cells. *J. Am. Chem. Soc.* **2017**, *139*, 9779–9782. [CrossRef] [PubMed]
82. Kolpashchikov, D.M. Binary malachite green aptamer for fluorescent detection of nucleic acids. *J. Am. Chem. Soc.* **2005**, *127*, 12442–12443. [CrossRef] [PubMed]
83. Kikuchi, N.; Kolpashchikov, D.M. A universal split Spinach aptamer (USSA) for nucleic acid analysis and DNA computation. *Chem. Commun.* **2017**, *53*, 4977–4980. [CrossRef] [PubMed]
84. Stojanovic, M.N.; Kolpashchikov, D.M. Modular aptameric sensors. *J. Am. Chem. Soc.* **2004**, *126*, 9266–9270. [CrossRef] [PubMed]
85. Kellenberger, C.A.; Wilson, S.C.; Sales-Lee, J.; Hammond, M.C. RNA-based fluorescent biosensors for live cell imaging of second messengers cyclic di-GMP and cyclic AMP-GMP. *J. Am. Chem. Soc.* **2013**, *135*, 4906–4909. [CrossRef] [PubMed]
86. Ponchon, L.; Dardel, F. Recombinant RNA technology: The tRNA scaffold. *Nat. Methods* **2007**, *4*, 571–576. [CrossRef] [PubMed]
87. Filonov, G.S.; Kam, C.W.; Song, W.; Jaffrey, S.R. In-gel imaging of RNA processing using Broccoli reveals optimal aptamer expression strategies. *Chem. Biol.* **2015**, *22*, 649–660. [CrossRef] [PubMed]
88. Masuda, I.; Igarashi, T.; Sakaguchi, R.; Nitharwal, R.G.; Takase, R.; Han, K.Y.; Leslie, B.J.; Liu, C.; Gamper, H.; Ha, T.; et al. A genetically encoded fluorescent tRNA is active in live-cell protein synthesis. *Nucleic Acids Res.* **2017**, *45*, 4081–4093. [CrossRef] [PubMed]
89. Okuda, M.; Fourmy, D.; Yoshizawa, S. Use of baby Spinach and Broccoli for imaging of structured cellular RNAs. *Nucleic Acids Res.* **2017**, *45*, 1404–1415. [CrossRef] [PubMed]
90. Lu, Z.; Filonov, G.S.; Noto, J.J.; Schmidt, C.A.; Hatkevich, T.L.; Wen, Y.; Jaffrey, S.R.; Matera, A.G. Metazoan tRNA introns generate stable circular RNAs in vivo. *RNA* **2015**, *21*, 1554–1565. [CrossRef] [PubMed]
91. Pothoulakis, G.; Ceroni, F.; Reeve, B.; Ellis, T. The Spinach RNA aptamer as a characterization tool for synthetic biology. *ACS Synth. Biol.* **2013**, *3*, 182–187. [CrossRef] [PubMed]
92. Ellefson, J.W.; Meyer, A.J.; Hughes, R.A.; Cannon, J.R.; Brodbelt, J.S.; Ellington, A.D. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat. Biotechnol.* **2014**, *32*, 97–101. [CrossRef] [PubMed]
93. Zhang, J.; Fei, J.; Leslie, B.J.; Han, K.Y.; Kuhlman, T.E.; Ha, T. Tandem Spinach array for mRNA imaging in living bacterial cells. *Sci. Rep.* **2015**, *5*, 17295. [CrossRef] [PubMed]
94. Guet, D.; Burns, L.T.; Maji, S.; Boulanger, J.; Hersen, P.; Wente, S.R.; Salamero, J.; Dargemont, C. Combining Spinach-tagged RNA and gene localization to image gene expression in live yeast. *Nat. Commun.* **2015**, *6*, 8882. [CrossRef] [PubMed]
95. Tsvetkova, I.B.; Yi, G.; Yi, Y.; Kao, C.C.; Dragnea, B.G. Segmented GFP-like aptamer probes for functional imaging of viral genome trafficking. *Virus Res.* **2015**, *210*, 291–297. [CrossRef] [PubMed]
96. Burch, B.D.; Garrido, C.; Margolis, D.M. Detection of human immunodeficiency virus RNAs in living cells using Spinach RNA aptamers. *Virus Res.* **2017**, *228*, 141–146. [CrossRef] [PubMed]

97. Nilaratanakul, V.; Hauer, D.A.; Griffin, D.E. Development and characterization of sindbis virus with encoded fluorescent RNA aptamer Spinach2 for imaging of replication and immune-mediated changes in intracellular viral RNA. *J. Gen. Virol.* **2017**, *98*, 992–1003. [CrossRef] [PubMed]
98. Guzman-Zapata, D.; Dominguez-Anaya, Y.; Macedo-Osorio, K.S.; Tovar-Aguilar, A.; Castrejon-Flores, J.L.; Duran-Figueroa, N.V.; Badillo-Corona, J.A. mRNA imaging in the chloroplast of chlamydomonas reinhardtii using the light-up aptamer Spinach. *J. Biotechnol.* **2017**, *251*, 186–188. [CrossRef] [PubMed]
99. Aw, S.S.; Tang, M.X.; Teo, Y.N.; Cohen, S.M. A conformation-induced fluorescence method for microRNA detection. *Nucleic Acids Res.* **2016**, *44*, e92. [CrossRef] [PubMed]
100. Alam, K.K.; Tawiah, K.D.; Lichte, M.F.; Porciani, D.; Burke, D.H. A fluorescent split aptamer for visualizing RNA-RNA assembly in vivo. *ACS Synth. Biol.* **2017**, *6*, 1710–1721. [CrossRef] [PubMed]
101. Breaker, R.R. Engineered allosteric ribozymes as biosensor components. *Curr. Opin. Biotechnol.* **2002**, *13*, 31–39. [CrossRef]
102. Nakayama, S.; Luo, Y.; Zhou, J.; Dayie, T.K.; Sintim, H.O. Nanomolar fluorescent detection of c-di-GMP using a modular aptamer strategy. *Chem. Commun.* **2012**, *48*, 9059–9061. [CrossRef] [PubMed]
103. Inuzuka, S.; Matsumura, S.; Ikawa, Y. Optimization of RNA-based c-di-GMP fluorescent sensors through tuning their structural modules. *J. Biosci. Bioeng.* **2016**, *122*, 183–187. [CrossRef] [PubMed]
104. Wang, X.C.; Wilson, S.C.; Hammond, M.C. Next-generation RNA-based fluorescent biosensors enable anaerobic detection of cyclic di-GMP. *Nucleic Acids Res.* **2016**, *44*, e139. [CrossRef] [PubMed]
105. Kellenberger, C.A.; Chen, C.; Whiteley, A.T.; Portnoy, D.A.; Hammond, M.C. RNA-based fluorescent biosensors for live cell imaging of second messenger cyclic di-AMP. *J. Am. Chem. Soc.* **2015**, *137*, 6432–6435. [CrossRef] [PubMed]
106. Kellenberger, C.A.; Wilson, S.C.; Hickey, S.F.; Gonzalez, T.L.; Su, Y.; Hallberg, Z.F.; Brewer, T.F.; Iavarone, A.T.; Carlson, H.K.; Hsieh, Y.F.; et al. Gemm-I riboswitches from geobacter sense the bacterial second messenger cyclic AMP-GMP. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5383–5388. [CrossRef] [PubMed]
107. Hallberg, Z.F.; Wang, X.C.; Wright, T.A.; Nan, B.; Ad, O.; Yeo, J.; Hammond, M.C. Hybrid promiscuous (hypr) ggdef enzymes produce cyclic AMP-GMP ($3'$, $3'$-cgamp). *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 1790–1795. [CrossRef] [PubMed]
108. Sharma, S.; Zaveri, A.; Visweswariah, S.S.; Krishnan, Y. A fluorescent nucleic acid nanodevice quantitatively images elevated cyclic adenosine monophosphate in membrane-bound compartments. *Small* **2014**, *10*, 4276–4280. [CrossRef] [PubMed]
109. Paige, J.S.; Nguyen-Duc, T.; Song, W.; Jaffrey, S.R. Fluorescence imaging of cellular metabolites with RNA. *Science* **2012**, *335*, 1194. [CrossRef] [PubMed]
110. Su, Y.; Hickey, S.F.; Keyser, S.G.; Hammond, M.C. In vitro and in vivo enzyme activity screening via RNA-based fluorescent biosensors for *S*-adenosyl-L-homocysteine (SAH). *J. Am. Chem. Soc.* **2016**, *138*, 7040–7047. [CrossRef] [PubMed]
111. You, M.; Litke, J.L.; Jaffrey, S.R. Imaging metabolite dynamics in living cells using a Spinach-based riboswitch. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E2756–E2765. [CrossRef] [PubMed]
112. Porter, E.B.; Polaski, J.T.; Morck, M.M.; Batey, R.T. Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nat. Chem. Biol.* **2017**, *13*, 295–301. [CrossRef] [PubMed]
113. Song, W.; Strack, R.L.; Jaffrey, S.R. Imaging bacterial protein expression using genetically encoded RNA sensors. *Nat. Methods* **2013**, *10*, 873–875. [CrossRef] [PubMed]
114. Ketterer, S.; Gladis, L.; Kozica, A.; Meier, M. Engineering and characterization of fluorogenic glycine riboswitches. *Nucleic Acids Res.* **2016**, *44*, 5983–5992. [CrossRef] [PubMed]
115. Hofer, K.; Langejurgen, L.V.; Jaschke, A. Universal aptamer-based real-time monitoring of enzymatic RNA synthesis. *J. Am. Chem. Soc.* **2013**, *135*, 13692–13694. [CrossRef] [PubMed]
116. Van Nies, P.; Nourian, Z.; Kok, M.; van Wijk, R.; Moeskops, J.; Westerlaken, I.; Poolman, J.M.; Eelkema, R.; van Esch, J.H.; Kuruma, Y.; et al. Unbiased tracking of the progression of mRNA and protein synthesis in bulk and in liposome-confined reactions. *Chembiochem* **2013**, *14*, 1963–1966. [CrossRef] [PubMed]
117. Chizzolini, F.; Forlin, M.; Cecchi, D.; Mansy, S.S. Gene position more strongly influences cell-free protein expression from operons than T7 transcriptional promoter strength. *ACS Synth. Biol.* **2014**, *3*, 363–371. [CrossRef] [PubMed]
118. Chizzolini, F.; Forlin, M.; Yeh Martin, N.; Berloffa, G.; Cecchi, D.; Mansy, S.S. Cell-free translation is more variable than transcription. *ACS Synth. Biol.* **2017**, *6*, 638–647. [CrossRef] [PubMed]

119. Svensen, N.; Jaffrey, S.R. Fluorescent RNA aptamers as a tool to study RNA-modifying enzymes. *Cell Chem. Biol.* **2016**, *23*, 415–425. [CrossRef] [PubMed]
120. Bose, D.; Su, Y.; Marcus, A.; Raulet, D.H.; Hammond, M.C. An RNA-based fluorescent biosensor for high-throughput analysis of the cGAS-cGAMP-sting pathway. *Cell Chem. Biol.* **2016**, *23*, 1539–1549. [CrossRef] [PubMed]
121. Abatemarco, J.; Sarhan, M.F.; Wagner, J.M.; Lin, J.L.; Liu, L.; Hassouneh, W.; Yuan, S.F.; Alper, H.S.; Abate, A.R. RNA-aptamers-in-droplets (RAPID) high-throughput screening for secretory phenotypes. *Nat. Commun.* **2017**, *8*, 332. [CrossRef] [PubMed]
122. Li, B.; Ellington, A.D.; Chen, X. Rational, modular adaptation of enzyme-free DNA circuits to multiple detection methods. *Nucleic Acids Res.* **2011**, *39*, e110. [CrossRef] [PubMed]
123. Bhadra, S.; Ellington, A.D. Design and application of cotranscriptional non-enzymatic RNA circuits and signal transducers. *Nucleic Acids Res.* **2014**, *42*, e58. [CrossRef] [PubMed]
124. Rogers, T.A.; Andrews, G.E.; Jaeger, L.; Grabow, W.W. Fluorescent monitoring of RNA assembly and processing using the split-Spinach aptamer. *ACS Synth. Biol.* **2015**, *4*, 162–166. [CrossRef] [PubMed]
125. Akter, F.; Yokobayashi, Y. RNA signal amplifier circuit with integrated fluorescence output. *ACS Synth. Biol.* **2015**, *4*, 655–658. [CrossRef] [PubMed]
126. Auslander, S.; Fuchs, D.; Hurlemann, S.; Auslander, D.; Fussenegger, M. Engineering a ribozyme cleavage-induced split fluorescent aptamer complementation assay. *Nucleic Acids Res.* **2016**, *44*, e94. [CrossRef] [PubMed]
127. Furukawa, A.; Tanaka, T.; Furuta, H.; Matsumura, S.; Ikawa, Y. Use of a fluorescent aptamer RNA as an exonic sequence to analyze self-splicing ability of agroup I intron from structured RNAs. *Biology* **2016**, *5*, 43. [CrossRef] [PubMed]
128. Afonin, K.A.; Danilov, E.O.; Novikova, I.V.; Leontis, N.B. TokenRNA: A new type of sequence-specific, label-free fluorescent biosensor for folded RNA molecules. *Chembiochem* **2008**, *9*, 1902–1905. [CrossRef] [PubMed]
129. Shu, D.; Shu, Y.; Haque, F.; Abdelmawla, S.; Guo, P. Thermodynamically stable RNA three-way junction for constructing multifunctional nanoparticles for delivery of therapeutics. *Nat. Nanotechnol.* **2011**, *6*, 658–667. [CrossRef] [PubMed]
130. Reif, R.; Haque, F.; Guo, P. Fluorogenic RNA nanoparticles for monitoring RNA folding and degradation in real time in living cells. *Nucleic Acid Ther.* **2012**, *22*, 428–437. [PubMed]
131. Shu, D.; Khisamutdinov, E.F.; Zhang, L.; Guo, P. Programmable folding of fusion RNA in vivo and in vitro driven by pRNA 3WJ motif of phi29 DNA packaging motor. *Nucleic Acids Res.* **2014**, *42*, e10. [CrossRef] [PubMed]
132. Panchapakesan, S.S.S.; Ferguson, M.L.; Hayden, E.J.; Chen, X.; Hoskins, A.A.; Unrau, P.J. Ribonucleoprotein purification and characterization using RNA Mango. *RNA* **2017**, *23*, 1592–1599. [CrossRef] [PubMed]
133. You, M.; Jaffrey, S.R. Structure and mechanism of RNA mimics of green fluorescent protein. *Annu. Rev. Biophys.* **2015**, *44*, 187–206. [CrossRef] [PubMed]
134. DasGupta, S.; Shelke, S.A.; Li, N.S.; Piccirilli, J.A. Spinach RNA aptamer detects lead(II) with high selectivity. *Chem. Commun.* **2015**, *51*, 9034–9037. [CrossRef] [PubMed]

© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Since the publication of this review, new RNA-based fluorogenic modules have been described in the literature. First of all, the collection of fluorogens activated by SRB-2 aptamer has been greatly enlarged upon the discovery that the pyronin core is the minimal structure recognized by the aptamer(Sunbul and Jäschke, 2018). Then, a collection of pyronin derivatives was synthesized conjugated with a dinitroaniline quenching group. Among the new fluorogens, TMR-DN revealed to be particularly interesting by displaying a low background fluorescence while being able to form with SRB-2 a complex as bright as the eGFP and of a pretty high affinity (K_D ~ 30nM, Table 1). These properties allowed TMR-DN/SRB2 complex to visualize mRNAs in living cells by inserting as few as 6 to 15 repeats of the aptamer in the target RNA. Whereas other pyronin derivatives were able to generate even brighter complexes, they suffered either form a low affinity or from a strong background fluorescence of the fluorogen in its free state, limiting imaging contrast, therefore the applicability of these fluorogens. This also underlines the importance of properly designing the fluorogen.

The second new module was developed on the basis of the cobalamin (Cbl) riboswitch (Braselmann et al., 2018). The authors discovered that the covalent attachment of the Cbl to a dye quenches the latter and converts it into a fluorogen. Moreover, incubating this fluorogen with Cbl riboswitch leads to the formation of a complex in which the Cbl is "neutralized" and the fluorescence of the dye is restored. However, Cbl-based fluorogen cannot freely diffuse through cell membrane limiting their practical application. Moreover, being based on vitamin naturally found in the cell, these dyes could face competition issues with the endogenous Cbl.

Table 1: New fluorophores/aptamers modules described after the publication of the Review Article 1. The new modules are shaded in blue (Braselmann et al., 2018; Sunbul and Jäschke, 2018).

Dyes	**Aptamers**	**Kd (nM)**	**Ex/Em (nm)**	**ε**	**Complex Φ**	**Brightness**	**Relative Brightness**
eGFP	n.r	n.r	490/508	39,200	0.68	26.6	1.00
TO-1	Mango	3.2	510/535	77,500	0.14	10.85	0.4
DFHBI 1-T	Broccoli	360	472/507	29,600	0.94	27.8	1.04
Atto 495-DN	SRB-2	358	509 /526	87,400	0.98	85 .6	3.2
TMR-DN	DNB	350	555/582	47,150	0.90	42.43	1.6
TMR-DN	SRB-2	35	564/587	90,500	0.33	29.86	1.1
Pyronin-B-DN	SRB-2	251	559/576	110,400	0.90	99.3	3.7
Cy5-Cbl	Cbl-ribo	34	646/662	270,000	0.2	54.0	2.0

Finally, the application scope of the RNA-based fluorogenic modules was further enlarged by adapting to live cells two concepts that were previously described only *in vitro*: the accurate detection of target RNAs using split aptamers (Alam et al., 2017; Wang et al., 2018) and the increase of RNA detection sensitivity using enzyme-free RNA-based amplification circuits (Karunanayake Mudiyanselage et al., 2018). Furthermore, as this was described above with FPs, RNA-based fluorogenic modules can also be used as FRET pairs. Following this idea, YOYO/Mango and DFHBI/Spinach were used as orthogonal fluorogenic FRET pair to characterize folding changes of RNA origamis both *in vitro* and *in vivo* in response to triggering RNA or metabolites (Jepsen et al., 2018).

From what precedes, one can see that numerous RNA-based fluorogenic modules have been developed over the past decade opening the way to a myriad of potential applications. Nevertheless, and as discussed in the review, significant gain in efficiency (especially at the level of the brightness and folding efficiency) could be expected by changing the way light-up RNA aptamers are developed.

2.2.2.2. Development strategies of light-up RNA aptamers

As stated in the **review article 1** (Bouhedda et al., 2017), all the known light-up aptamers (with the exception of Broccoli) have been isolated using SELEX (Systematic Evolution of Ligands by Exponential Enrichment), a methodology introduced in early 90's (Ellington and Szostak, 1990; Tuerk et al., 1990) and that aims at isolating RNAs for their ability to recognize a ligand (Darmostuk et al., 2014; Stoltenburg et al., 2007). Briefly, large starting libraries (~10^{15} variants) are prepared either by chemical synthesis with targeted randomization, or error-prone PCR using nucleotide analogues or even recombination methods (Vorobyeva et al., 2018). Molecules contained in theses libraries are then challenged to interact with a target molecule immobilized on beads, displayed on cell surface (e.g. cell-SELEX) or free in solution and later separated by capillary electrophoresis (e.g. CE-SELEX). Then, extensive washes allow for discarding poor binders while enriching the library in RNA (or DNA) endowed with good affinity for the target. Enriched pools are then subjected to several additional cycles while increasing wash stringency. At the end of the process, libraries are cloned and screened for the best variants, a labor-intensive and time-consuming step. Coupling SELEX with next generation sequencing technologies (Quang et al., 2016) allowed substantial gain in efficiency by making possible to monitor the whole evolution process and enabling the early detection of promising molecules.

Despite their efficiency at isolating high affinity binders, in some cases SELEX methods are limited by the fact that they primarily isolate RNA for their binding capacity rather than for their final function. As a first consequence, selected molecules are often high affinity binders organized as rigid scaffold. This structural rigidity makes extremely complex the later conversion of these aptamers into flexible RNA modules like those found in riboswitches and used to build-up biosensors. This

limitation can nevertheless be overcome by using dedicated SELEX approaches such as Capture-SELEX ((Stoltenburg et al., 2012), see §5.4.1.2). A second consequence of selecting RNAs only for their binding capacity is obvious in the case of light-up aptamers: a good binder is not synonymous with good lighting-up variant. Indeed, strongly interacting with a compound does not mean that this compound is placed in a configuration maximizing fluorescence emission and reducing unwanted photoinactivation
(two parameters the SELEX cannot easily select for). In support to this idea, it was estimated that less than 1% of the RNA variants isolated by SELEX were actually efficient at lighting-up the fluorogen they were able to interact with (Filonov et al., 2014).

Considering what precedes, it sounds reasonable to assume that an approach in which the lighting-up capacity of the aptamer would be used as a selection criterium should be far more performant than SELEX at isolating efficient light-up aptamers. On the first hand, performing such a screening in the gold standard microtiter plates would be cost-prohibitive even though it makes possible to work *in vitro* by applying any type of selection pressure. On the other hand, cloning the library into expression plasmids later transformed into bacteria makes possible to functionally screen the library in ultrahigh-throughput regimes using a Fluorescence-Activated Cell Sorter (FACS). However, this approach is time-consuming and the range of applicable selection pressures is limited to those cell-compatible while excluding non-permeable dyes such as Cy5-Cbl (Braselmann et al., 2018) and DIR (Tan et al., 2017).

To overcome these limitations, one goal of this PhD thesis was to explore a third screening strategy exploiting droplet-based microfluidics, a technology combining the advantages of the two other aforementioned technologies. Indeed, confining biological reaction within picoliter-sized water-in-oil droplets allows preforming *in vitro* functional screening in an ultrahigh-throughput, quantitative and cost-effective way.

3. Droplet-based microfluidics: principle and applications in biology

Over the past decade, microfluidics (i.e. the discipline in which microfabricated micrometer scale systems are used to handle and manipulate liquids) allowed several breakthroughs to occur in biology. Indeed, as rapidly introduced below, many laboratory equipments are now built around a microfluidic component. Beside these state-of-the-art equipments, a new set of methodologies exploiting droplet-based microfluidics has rapidly expended these last years due to their throughput, their scale and the control they offer over reaction conditions.

3.1. Impact of microfluidics on biology

The most remarkable impact of microfluidics on biology is exemplified by the advent of high-throughput sequencing (also called Next Generation Sequencing or NGS) technologies (Metzker, 2010). Indeed, whatever the NGS technology (e.g. 4.5.4, Illumina, PacBio…), they all involve a micrometer-scale flow-cell (Figure 19) in which reagents circulate to allow the sequencing reaction to proceed (Metzker, 2010). Additionally, as highlighted above (Chapter 2), microfluidics also had an important contribution in the characterization of cell-to-cell gene expression heterogeneity by making possible to keep cells alive and under perfusion of culture medium while being imaged in constrained geometries (Potvin-Trottier et al., 2018).

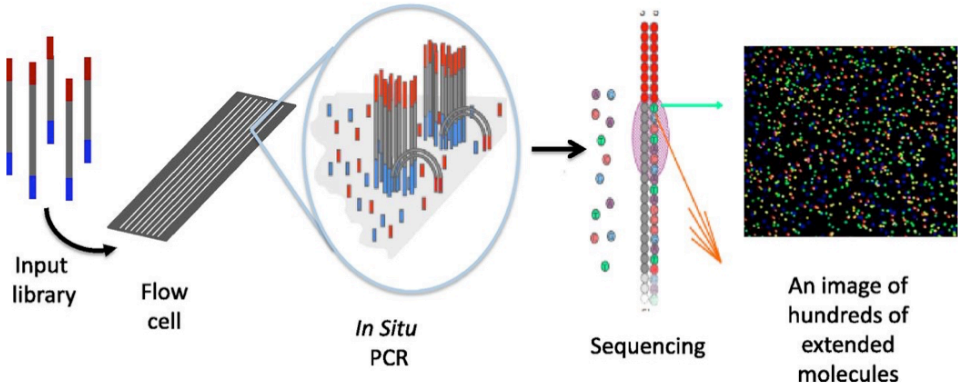


Figure 19: Working principle of an Illumina flow cell. DNA library displaying adaptors is introduced into the flow-cell. Upon hybridization and bridge-PCR, the sequencing reaction is performed by flowing fluorescent nucleotides into flow-cell. Sequencing reaction is revealed by reading fluorescent spots. Figure adapted from Johnsen, 2013.

In addition to exploiting channels geometry, microsystems functionality can be further expanded by functionalizing channels surface with biomolecules (e.g. to set-up biosensors (Zhou et al., 2010)) or even by patterning electrodes underneath the chip to manipulate liquids in a reconfigurable way like this is done in digital microfluidics (Choi et al., 2012). Such devices may have a wide range of application, especially in personalized medicine, when they are incorporated into point-of-care portable devices. Last but not least, incorporating valves into the device makes possible to create microchambers on-demand and, this way, to produce Large Scale Integration (LSI) devices in which hundreds of independent reactions can be performed in parallel (Figure 20). Such devices allowed first miniaturized screening and single-cell genomics to be performed (Marcus et al., 2006). However, the pretty large volume of reaction chambers (nanoliters) and their finite low number (at most a few hundred per chip) reduces their application scope. However, such limitations can easily by overcome by exchanging the rigid polymer-based chambers for soft water-in-oil droplets.

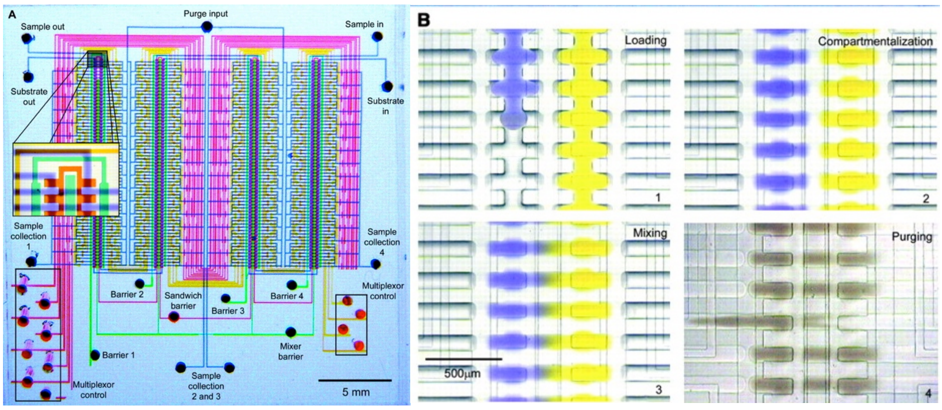

A
Sample out
Purge input
Sample in
Substrate out
Substrate in
Sample collection 1
Sample collection 4
Barrier 2
Sandwich barrier
Barrier 3
Barrier 4
Multiplexor control
Barrier 1
Mixer barrier
Sample collection 2 and 3
Multiplexor control
5 mm
B
Loading
1
Compartmentalization
2
Mixing
500μm
3
Purging
4

Figure 20: Large Scale Integration (LSI) microfluidic device. Photographs of microfluidic device in which micro-valves are used to segment channels into micro-chambers **(A)** or to mix reagents (**B**). Figure adapted from Thorsen, 2002.

3.2. Droplet-based microfluidics

Whereas the microfluidic devices introduced above handle liquids as continuous phases, combining aqueous and oil streams allows for segmenting the former into water-in-oil droplets.

3.2.1.Droplet generation and manipulation

When confined and forced to circulate into a microfluidic channel, liquids streams are characterized by a low Reynolds number, indicating that they adopt a laminar flow where liquids circulate side by side with a poor mixing-efficiency. This behavior allows for precisely controlling liquids. For instance, forcing an aqueous and oil phase to cross each other into a microfluidic device allows for producing water-in-oil droplets of highly reproducible volumes (Figure 21.A); such emulsion being said monodisperse. Moreover, droplet stability is ensured by the addition of a surface-active agent (surfactant) in the oil phase. This compounds partitions at the water/oil interface and, doing so, stabilizes the droplets by reducing the surface tension (Baret, 2012).

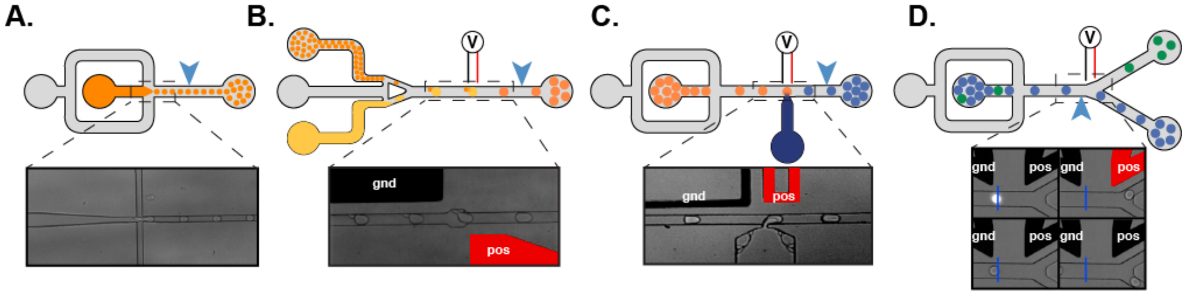


Figure 21: Schematic representation of the main droplet-based microfluidic devices. (A) Droplet generator disperses aqueous PCR mixture (in orange) into droplets carry by a fluoronitated oil phase (in grey). The blue arrow indicates the measurement point where the proper monodispersity of the emulsion is monitored by fluorescence measurement. (B) Representation of the fusion module, where PCR emulsion is reinjected onto the top while a second emulsion is generated into the chip (in yellow). Both emulsions are synchronized and fused together when passing between two electrodes (labelled gnd and pos). (C) Schematic of a pico-injector module. Droplets are reinjected and spaced by oil (in grey) and receive a controlled aliquot of aqueous (in dark blue) when passing in front of an injection channel located in between two electrodes. (D) Fluorescence activated droplet sorting (FADS). The fluorescence of each droplet is measured at the level of the blue arrow. When a droplet of interest is detected a pair of electrodes is energized, an electric gradient is created that pulls the droplet toward the sort channel. Else, droplets flow toward the waste channel. Figure adapted from Ryckelynck et al., 2015.

Besides generating droplet of controlled volume, using dedicated microfluidic devices also makes possible to manipulate these droplets (see **Review Article 2** for a more comprehensive presentation). For instance, droplet content can be modified *via* droplet fusion. In this scheme, two sets of droplets (the first made of preformed reinjected droplets and the second being generated on-chip) are synchronized on the device to form pairs (Figure 21.B). These pairs are then fused while passing in between a pair of electrodes subjected to an AC current (Mazutis et al., 2009). Alternatively, large droplets can be reinjected into a device in which a pico-injector (Abate et al., 2010) allows for delivering a controlled amount of reagent to each droplet in a precise way (Figure 21.C, see **Review Article 2** for a deeper explanation). Moreover, using a dedicated optical set-up, the fluorescence content of each droplet can be monitored and even used to trigger further action. For instance, coupling fluorescence detection with droplet deflection allowed to set-up a microfluidic Fluorescent-Activated Droplet Sorting (FADS (Baret et al., 2009)) that makes possible to analyze and sort droplets in high-throughput regime (Figure 21.D).

Using these different microfluidic devices, alone or in combination, enables setting-up new applications characterized by high sensitivity and ultrahigh-throughput capacity.

3.2.2. Analysing nucleic acids with droplet-based microfluidics

Nucleic acids (DNA and RNA) form a very attractive class of molecules on an analytical point of view. Indeed, their biological relevance highlighted in Chapter 1 makes them good markers of gene expression, infection or even pathologic state. Moreover, nucleic acids have the advantage of being easily amplified and identified by reading their sequence. Therefore, first commercial equipment exploiting droplet-based have been focussed on the detection and preparation of nucleic acids.

3.2.2.1.Digital droplet PCR

Nowadays, nucleic acids are usually quantified using quantitative PCR (qPCR), an analogue quantification method in which the concentration of the target molecule is determined by measuring the number of PCR cycles required to reach a detection threshold (Ct) and comparing this value to that obtained with a reference molecule. Whereas qPCR offers a great detection sensitivity, the method also strongly relies on the detection efficiency of both the target and the reference. Moreover, despite its high sensitivity, qPCR becomes more limited when searching for rare variants a gene (e.g. point mutant) contained within a large population of wild type molecule. For instance, detecting *KRAS* gene mutants (marker of colorectal cancer) by qPCR requires the mutant to represent at least 10 % of the total *KRAS* pool to achieve reliable detection, making such qPCR-based detection poorly efficient(Pekin et al., 2011).

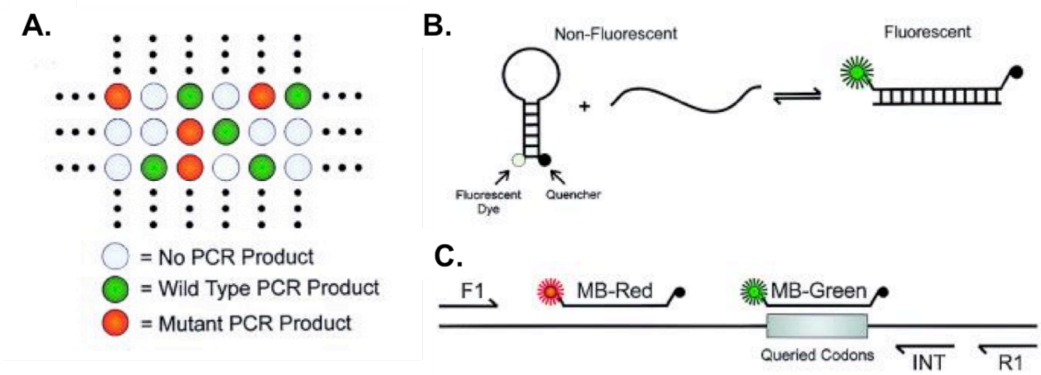


Figure 22: Principe of the digital PCR (dPCR). (A) DNA samples are diluted into a PCR mixture supplemented with fluorogenic probes allowing for discriminating a mutant sequence from the wild-type (B&C). Figure adapted from Vogelstein 1999.

Substantial gain in sensitivity can be obtained by digitalizing the PCR (Figure 22), a concept pioneered in 1999 by Vogelstein *et al* (Vogelstein and Kinzler, 1999). In a typical digital PCR (dPCR) procedure, a DNA mixture is diluted into a qPCR mixture (i.e. containing a fluorogenic probe such as SybrGreen or TaqMan probes) prior to being distributed into compartments. Initial DNA solution is diluted such that, on average, a compartment contains less than one DNA molecule. Upon thermocycling, only those compartments that initially contained a template DNA molecule display a fluorescent signal and counting the fraction of fluorescent compartments gives direct access to the occupancy. Since, DNA molecules distribute into compartments according to Poisson statistics, DNA concentration of the sample can then be calculated back using equation 2 where λ corresponds to the average number of DNA molecule per compartment of known volume.

$$\text{Eq. 2} \quad \lambda = 1 - \ln(\text{Occupancy})$$

The original experiment by Vogelstein was performed in 96-well plates, a format in which such experiments become rapidly cost-prohibitive and the throughput is seriously limited. Consequently, several groups and companies have developed microfluidic devices to miniaturize dPCR and doing so reduce its cost while increasing its throughput (Quan et al., 2018). Among these technologies, the digital droplet PCR (ddPCR) is particularly efficient and dedicated equipment are now broadly sold (e.g. QX-200 platform from Bio-Rad). In ddPCR, DNA-containing qPCR mixture is distributed into water-in-oil droplets. Upon production, droplets are thermocycled prior to analyzing their fluorescence on a dedicated analysis module. Using such an approach, *KRAS* mutants can now be detected with a sensitive as high 1 mutant in 200,000 molecules, opening new perspectives in personalized medicine (Pekin et al., 2011).

3.2.2.2. Sequencing-based gene expression analysis with single-cell resolution

The advent of next-generation sequencing (NGS) made possible genome-wide quantitative analysis of the transcriptome to be performed. However, as stated above (section 1.3), significant cell-to-cell heterogeneity occurs within a cell population. In an effort to get NGS analysis working at the required resolution, Tang and co-workers introduced the single-cell RNA sequencing (scRNA-seq). To do so, they manually individualized single cells prior to isolating mRNA using poly-T capture primer and sequencing the recovered mRNAs (converted into cDNAs) to establish transcriptome profiles (Tang et al., 2009). However, collecting statistically relevant information requires analyzing a large number of single cells in a single run of NGS.

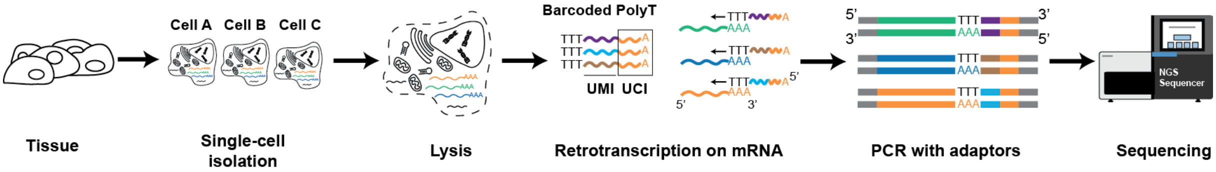


Figure 23: Schematic workflow of a single-cell RNA-Seq experiment. Upon cell individualization from biological tissue, single-cells are lyzed and mixed with oligo-dT primers containing two randomized regions and defined as being unique cellule identifier (UDI) and unique molecular identifier (UMI). Upon RNA reverse transcription, cDNAs carry UDI and UMI. Prior to sequencing the library on a NGS device, adaptors are added to barcoded cDNAs. Then cDNAs are pooled, analyzed in a single NGS run prior to being identified and quantified by bioinformatics.

Preserving single-cell resolution while pooling many cells in a single multiplexed run of NGS is possible by appending a DNA barcode to each cDNA. A barcode is a DNA sequence unique to each cell but common to every cDNA (obtained upon mRNA reverse transcription) originating from the same cell (Figure 23). To do so, the cells to analyze are first individualized into wells prior to extracting their mRNA. Next, mRNAs are reverse transcribed into cDNAs using primers containing a different barcode for each cell. Finally, cDNA samples are pooled and analyzed in a single run of NGS. This concept was validated by Islam *et al* who managed to pool and analyze ~100 cells in a single experiment (Islam et al., 2011). In order to get enough material for the next steps of library preparation, cDNAs must be amplified; a step during which significant amplification bias can take place and challenge the quantitativity of the method. This limitation can nevertheless be overcome by introducing a second barcode known as unique molecular identifier (UMI) into the reverse transcription primer (Figure 23). The UMI is a random sequence of 5 to 20 nucleotides that will be different for each molecule. The likelihood of having twice the same UMI being very low, simply counting the number of different UMIs for each mRNA in each cell directly gives access to the copy number of each mRNA per cell (Islam et al., 2014). Whereas this double barcoding revealed to be very efficient, the throughput of the method is limited by the need of handling cells and performing reverse transcription manually.

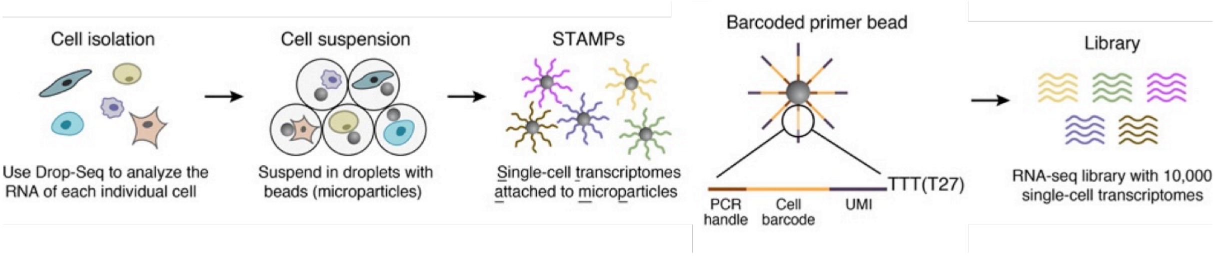

Cell isolation
Use Drop-Seq to analyze the RNA of each individual cell
Cell suspension
Suspend in droplets with beads (microparticles)
STAMPs
Single-cell transcriptomes attached to microparticles
Barcoded primer bead
TTT(T27)
PCR handle
Cell barcode
UMI
Library
RNA-seq library with 10,000 single-cell transcriptomes

Figure 24: Schematic of cDNA barcoding workflow using beads and microfluidics. Cells are encapsulated with beads into water-in-oil droplets. Beads carry oligo-dT reverse transcription primers labelled by barcodes identifying the cell (UCI) and the molecule (UMI). Upon pooling and sequencing, those barcodes allow to retrieve and identify the cell each originates from as well as establishing molecules copy number. Figure adapted from Macosko *et al* 2015.

As this was this case above with dPCR, adapting the barcoding process to a droplet-based microfluidics format allowed automating the process while significantly increasing the throughput to ~ 10.000 cells analyzed per run. In this scheme (Figure 24), dually barcoded reverse transcription primers are directly synthesized on beads, each one displaying several millions of oligo-dT reverse transcription primers, each labelled with a different UMI while all the primer of the same bead also share the same unique cell barcode (Macosko et al., 2015). This technology is now exploited by the Chromium platform distributed by 10x Genomics.

3.2.3.Droplet-based microfluidics screening applications

In addition to analytical applications such as those exemplified above, the ultrahigh-throughput capacity offered by droplet-based microfluidics also makes the technology perfectly suited for screening applications. In this view, many platforms have been developed for drug and cell screening (Shembekar et al., 2016). However, since these applications are out of the scope of this thesis, I won’t further discuss them, and I will instead focus on the use of droplet-based microfluidics for screening gene libraries.

3.2.3.1. *From IVC to μIVC*

In 1998, Tawfik and Griffiths introduced the concept of *in vitro* compartmentalization (IVC) in which biological reactions are individualized into small (femtoliter size) water-in-oil droplets (Tawfik and Griffiths, 1998). In the original methodology, a gene library is diluted into an *in vitro* expression mixture (*in vitro* transcription or coupled *in vitro* transcription/translation) just prior to being emulsified by dispersing an aqueous phase into an oil phase using a homogenizer (Miller et al., 2006). Here, single genes are isolated and expressed into the droplets (Figure 25). IVC allowed several proteins and RNA to be improved (Aharoni et al., 2005; Ghadessy et al., 2001; Levy et al., 2005; Mastrobattista et al., 2005; Zaher and Unrau, 2007). However, the mechanical emulsification process leads to a strongly polydisperse emulsions limiting the method on a quantitative point of view. Moreover, once these micro-droplets are generated it is difficult to modify their content. However, both limitations can be overcome by preparing IVC emulsion using a microfluidic device, an approach that will be called hereafter μIVC (microfluidic-assisted IVC).

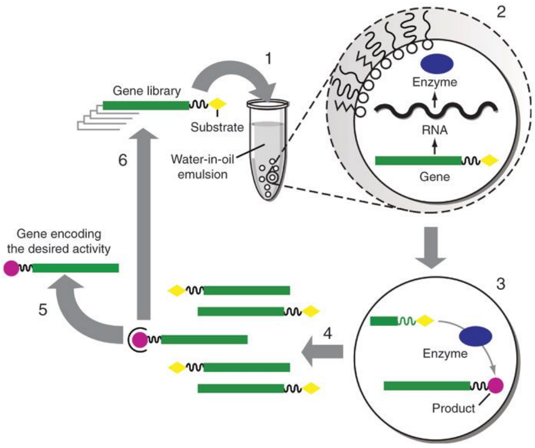


Figure 25: General principle of a IVC directed evolution process. Genes are linked to the substrate and individualized into water-in-oil droplets containing an *in vitro* transcription and translation mixture. Upon incubation, the protein with an activity will have converted the substrate into product. The emulsion is broken and gene coding for an active enzyme, so displaying the reaction product will be selectively enriched before starting a new round. Figure adapted from Miller *et al* 2006.

As mentioned above, microfluidics not only allows to generate highly homogeneous droplets (so monodisperse emulsion), but dedicated device can also be used to modify their content on-demand, as well as to analyze and sort them. Therefore, using the proper enchainment of microfluidic device allows to form, modify and sort droplets based on their content, making possible to develop a whole bunch of new molecules in a rapid and highly quantitative way.

3.2.3.2. *Searching for new or improved catalysts using μIVC*

μIVC has been extensively used to screen gene libraries to search for new or improved enzyme activities. This is mainly due to the fact that enzymes are important biological molecules on a fundamental and industrial point of view. The important literature on the topic together with the overall goal of my thesis (see section 4.) encouraged us to write a second review article (**Review Article 2**) that was published in early 2017 and can be found hereafter.

Review Article 2:

Ultrahigh-Throughput Improvement and Discovery of Enzymes Using Droplet-Based Microfluidic Screening

Alexis Autour and Michael Ryckelynck

2017

Micromachines

micromachines

Review

Ultrahigh-Throughput Improvement and Discovery of Enzymes Using Droplet-Based Microfluidic Screening

Alexis Autour and Michael Ryckelynck *

Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, F-67000 Strasbourg, France; a.autour@ibmc-cnrs.unistra.fr
* Correspondence: m.ryckelynck@unistra.fr

Academic Editors: Andrew J. deMello and Xavier Casadevall i Solvas
Received: 21 January 2017; Accepted: 13 April 2017; Published: 18 April 2017

Abstract: Enzymes are extremely valuable tools for industrial, environmental, and biotechnological applications and there is a constant need for improving existing biological catalysts and for discovering new ones. Screening microbe or gene libraries is an efficient way of identifying new enzymes. In this view, droplet-based microfluidics appears to be one of the most powerful approaches as it allows inexpensive screenings in well-controlled conditions and an ultrahigh-throughput regime. This review aims to introduce the main microfluidic devices and concepts to be considered for such screening before presenting and discussing the latest successful applications of the technology for enzyme discovery.

Keywords: droplet-based microfluidics; high-throughput screening; enzyme improvement; directed evolution; single-cell

1. Introduction

Enzymes constitute a fascinating class of biological polymers (proteins or nucleic acids) able to efficiently catalyze virtually any chemical reaction by placing a substrate molecule in optimal configuration and environment to stabilize the transition state, and doing so promotes its transformation (e.g., cleavage, addition, or modification) in an enantioselective manner [1]. These properties make enzymes extremely attractive for industrial applications (e.g., green chemistry, bioethanol production), bioremediation, biotechnologies, bioengineering, and synthetic biology. However, even though a plethora of natural enzymes have been identified, improvement and/or adaptation of existing molecules, or even discovery of new ones, may still be necessary for several reasons [2]. First, the wild-type molecule has naturally evolved to display an optimal activity in physicochemical conditions (e.g., crowded medium, temperature, ionic strength …) that may significantly differ from those of the application one may wish to use it for [3]. Second, the activity of interest may be a secondary weak activity of the molecule. This scenario is typically encountered with promiscuous enzymes, a set of catalysts with relaxed specificity that are able to catalyze other reactions and/or transforming other substrates than those they originally evolved for [4]. Promiscuity has been proposed to be an important evolutionary engine predisposing the cell with a basal level of activity allowing for gene speciation and rapid adaptation to environmental changes [5]. However, the prediction of such activity, as well as the identity of the residues to mutate to improve this activity, is extremely challenging. Finally, the catalyst may have been lost during the evolution as it is the case of many catalytic RNAs that one may want to resurrect, for instance, when studying the origin of life [6].

Micromachines **2017**, *8*, 128; doi:10.3390/mi8040128

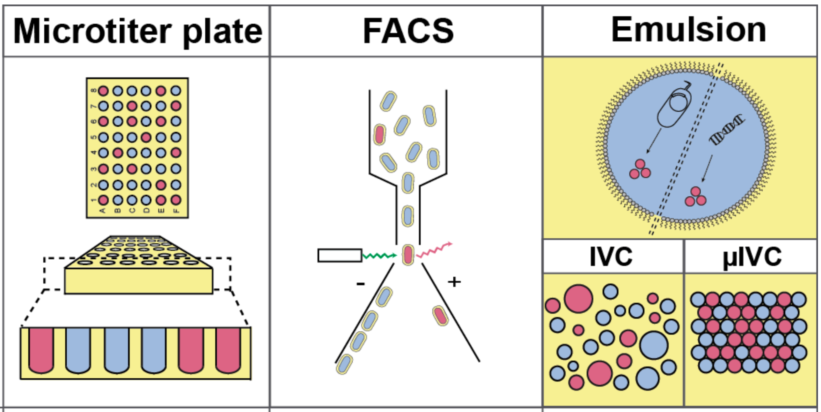
In any of the aforementioned cases, the catalyst can be improved, or generated *de novo*, using directed evolution, a set of laboratory methodologies allowing for rapid evolution of biological molecules (proteins or nucleic acids) by iterative rounds of (i) mutagenesis to generate mutant libraries (genetic diversity) and (ii) selection aiming at enriching the libraries in molecules with the desired properties [7–14]. Owing to the 20-letter alphabet of proteins (4-letter for nucleic acids), the size of gene libraries rapidly grows with increasing the number of residues to be mutated. Therefore, the selection process has to be able to handle such a large diversity (>10^6 mutants). Affinity panning technologies can easily manipulate such diversity [6,15] but, whereas they are very efficient at identifying good binders, these approaches are less suited for enzyme evolution since a single turnover is often sufficient to make a gene selectable, which preclude the enrichment of molecules with elevated turnover in the library. Consequently, a high-throughput screening approach where each variant of a library needs to perform several turnovers to become selected has a better chance of successfully identifying efficient enzymes.

In this review, we present the latest developments and use of droplet-based microfluidics for the discovery of new enzymes by ultrahigh-throughput screening of either mutant libraries or environmental samples. This is however only one of the various applications of the technology and the reader is referred to other recent reviews for a more exhaustive view of its scope of application [16–19].

2. (Ultra) High-Throughput Screening Strategies and Their Limitations

Nowadays, most high-throughput screenings aimed at improving/discovering enzymes are performed by specialized facilities using microtiter plate (MTP) format and requiring sizeable, costly, and sophisticated colony pickers and liquid-handling robots able to perform up to 10^5 assays per day (Figure 1) [20]. Therefore, screening a million variants with such platform would take at least 10 days. Increasing MTP well density from 96 (100–200 μL per well) to 384 (30–100 μL per well) and even 1536 (2.5–10 μL per well) wells per plate can increase analytical throughput while decreasing the cost. However, further reduction is prevented by capillarity and fast evaporation that would dominate at sub-microliter volumes. Beside MTPs, the fluorescence-activated cell sorter (FACS) is common bench-top laboratory equipment able to analyze and sort up to 10,000–15,000 cells per second, an ultrahigh-throughput regime (more than 10^5 analyses per day) allowing a million variants to be screened in less than 2 min. However, whereas in MTP the well boundary physically confines the genotype (catalyst-coding gene) with the phenotype (reaction product), the use of the FACS requires this linkage to be performed at the level of the cell [21]. Therefore, FACS-based screenings are restricted to only two scenarios: either the produced enzyme stays in the cytoplasm and the fluorogenic substrate has to be able to cross the cell membrane while the fluorescent product should stay trapped in the cell; or the enzyme is displayed at cell surface and the fluorescent product should be captured at cell surface as well.

Emulsion-based screening represents a third strategy in which a biological (or chemical) reaction is dispersed into an oil phase to form water-in-oil (w/o) droplets, whose boundary ensures the genotype/phenotype linkage [22]. As originally conceptualized by Tawfik and Griffiths [23], in vitro compartmentalization (IVC) was limited to DNA modifying enzymes or required complex chemistry and/or selection procedure to be set up [24,25]. However, the possibility of re-emulsifying w/o droplets into water-in-oil-in-water (w/o/w) double emulsions made it possible to screen droplets using fluorogenic assays and FACS [26]. By combining the advantages of MTP and FACS, IVC allowed improving the activity of many enzymes [27–38], a list that may lengthen in the future with the development of new strategies [39]. Finally, the small droplet volume (femtoliters to picoliters) permits screening directly at the level of single cells, and even single genes. Despite the great perspectives opened by the IVC, the technology still suffered limitations challenging its flexibility and that could ultimately preclude the identification of efficient catalysts [24,25]. Indeed, high polydispersity within emulsions significantly limits IVC quantitatively, a phenomenon further amplified when a second emulsification step is used as it may lead to the co-encapsulation of several w/o droplets into a single w/o/w droplet, increasing the false positive rate. Finally, the difficulty of modifying the content of a droplet once it has been formed strongly reduces IVC flexibility. Nevertheless, all these limitations can be overcome by transposing IVC to a microfluidic format, summarized hereafter as µIVC; standing for microfluidic-assisted In Vitro Compartmentalization.



	Microtiter plate	FACS	Emulsion
Throughput	High (<10^5 per day)	Ultrahigh (>10^5 per day)	Ultrahigh (>10^5 per day)
Genotype/phenotype link	Well boundary	Cell membrane	Droplet boundary
Assay type*	Versatile	Cell only	Versatile
Single-cell compatible	No	Yes	Yes
Cost per reaction	High	Low	Low

Figure 1. Principal (ultra)high-throughput screening strategies for enzyme improvement/discovery. The different methods (Microtiter plate, fluorescence-activated cell sorting (FACS) and emulsion) are schematized with the boundary (well, cell membrane, or oil) confining the phenotype with the genotype shaded in yellow. High-throughput (up to 10^5 analyses par day) is distinguished from ultrahigh-throughput (more than 10^5 analyses per day) regimes. Note that the emulsion-based approach is represented its two format: (i) in vitro compartmentalization (IVC) manipulating polydisperse emulsions and (ii) microfluidic-assisted IVC (µIVC) manipulating highly monodisperse emulsions. Compartments containing the target phenotype (red) are discriminated from those that do not possess it (blue). *Assay type can be cell-based, cell lysate-based, cell-free, or all of them (versatile).

3. Droplet-Based Microfluidics

Setting up a µIVC screening pipeline requires generating monodisperse water-in-oil droplets but it may also necessitate to modify their content and, at the end of the process, the enzymatic activity contained in each droplet should be quantified and used to sort them accordingly.

3.1. Droplet Production

Water-in-oil (w/o) emulsions are obtained by dispersing an aqueous phase into an immiscible carrier oil phase, usually supplemented with a surfactant molecule to stabilize the emulsion (see Section 4). However, because of the laminar flow adopted by liquids circulating within a microfluidic device, simply flowing the continuous (oil) and the dispersed (aqueous) phases together in a channel is not enough to generate droplets of controlled volume. Instead, the production of highly monodisperse emulsions (less than 5% of variation in droplet volumes) requires the use of well-defined geometries, eventually producing fL to nL droplets. Many geometries and strategies have been developed over the last years. Whereas an exhaustive introduction of these systems can be accessed elsewhere [40], the main strategies are summarized on Figure 2.

A first set of droplet generators for producing droplets of reliable and controlled size used geometries that apply a shear stress to the dispersed phase. In early 2000s, Umbanhowar et al. first introduced the co-flow strategy (Figure 2a) where the dispersed phase is released from a capillary into a carrier oil phase flowing in a parallel direction outside the capillary [41]. Nevertheless, the complexity of integrating capillaries with conventional microfabrication processes has precluded the wide use of such devices. Nowadays, the most widely used droplet generators have planar geometries consisting of intersecting channels bringing together an aqueous and an oil flow (Figure 2b–d). Among them, T-junction [42,43] is frequently used for the production of large droplets while hydrodynamic flow-focusing [44] is preferred for generating droplets of small size and/or at very high production frequency. Moreover, the pinned-jet flow focusing [45] is an attractive variation of the T-junction, allowing for producing droplets at a production frequency matching the operating frequency of a downstream module (e.g., droplet fuser or droplet sorter).

Beside shearing-based strategies, droplets can also be produced through a sharp change in the capillary pressure as this occurs during step emulsification. On such a device (Figure 2e), the aqueous and continuous phases are first combined in a shallow region acting as a Hele–Shaw cell that prevents droplet formation by suppressing the interfacial instability. However, when the liquids reach the much deeper region of the device, the abrupt relaxation of the confinement leads to droplet formation [46,47]. Interestingly, an abrupt step is not an absolute requirement, since the effect can be obtained by using a progressive gradient of confinement [48]. Step emulsification is well suited for producing monodisperse emulsions even in extreme conditions of viscosity [46] as well as for the production of monodisperse emulsions made of small femtoliter droplets [49].

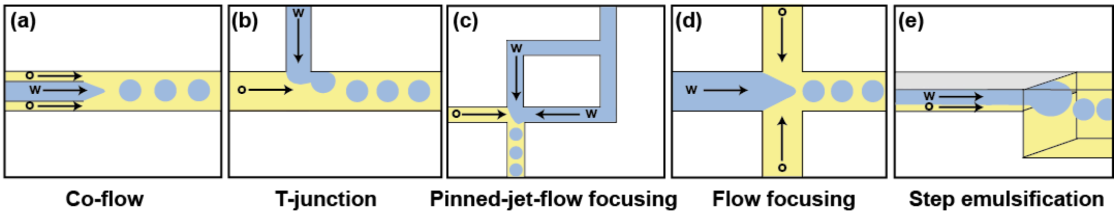


Figure 2. Main geometries used for droplet generation. (**a**) Droplet generation by co-flowing an aqueous phase (in blue) arriving from a capillary immersed in an oil stream (yellow). (**b**–**d**) Droplet generators with planar geometries. Channels in which oil (shaded in yellow) and aqueous phase (shaded in blue) flow are respectively labeled "o" and "w". Arrows indicate flow directions. (**e**) Schematic of a step emulsification process.

Upon droplet production, their contents start to mix because of recirculating streamlines inside the droplet [50], but faster mixing (less than 2 ms) can be achieved by circulating the droplets into a zigzag-shaped channel inducing chaotic advection into the droplet [51]. Finally, droplets can be incubated on-chip, either by storing them in dedicated chambers [52,53] and microarrays [54]; or by circulating them into various geometries of channels [55,56]. Alternatively, droplets can be collected into a variety of off-chip reservoirs (e.g., tube closed by a polydimethylsiloxane (PDMS) plug [57], a capillary [58], a Pasteur pipette [59], or a syringe [60]) and incubated for longer times.

3.2. Droplet Content Modification

A great advantage of droplet-based microfluidics is the possibility of modifying droplet composition on demand and in a controlled manner. For instance, combining different aqueous streams just upstream the droplet production nozzle makes it possible to continuously produce droplets while modulating their content by changing the composition of one of the streams. For instance, plugs of compounds (e.g., enzyme inhibitors, substrates . . .) can be prepared from microtiter plate [61,62] or tubes [63–65] and sequentially injected into the droplet generator (Figure 3a). Alternatively, compounds can be loaded onto beads prior to individualizing them into droplets and release their payload by UV illumination [66]. Droplets of variable but controlled composition can also be produced by 'printing' them using valve-based droplet generators [67,68]. Moreover, the concentration of the reagents added to the droplets can be varied using one of the several reported microfluidic gradient generator [69,70], a concentrator module [71], or an ultra-performance liquid chromatography [72].

Droplet content can also be modified after the droplet has been formed and stabilized by a surfactant, which is especially important when implementing an experimental workflow in which two (or more) mutually incompatible steps must be decoupled (e.g., cell growth and activity assay [73–75], or protein synthesis and activity assay in harsh conditions [76]) as this may occur in some µIVC procedures. A first approach consists in fusing each droplet with a second one containing the reagents to be delivered [77] (Figure 3b). Pairs of droplets can be formed and synchronized using dedicated structures such as pillar arrays [78] or by exploiting the Poiseuille parabolic profile adopted by a liquid circulating in a microfluidic channel, which causes a small droplet to circulate faster than a larger one, eventually leading the former to catch up the latter [79]. If one of the droplets is deprived of surfactant [80] or only partially stabilized [81], the pairs of droplets can be passively fused upon a decompression, likely as the result of a local depletion of surfactant molecules [82]. However, this approach may be sensitive to surfactant batch-to-batch variations. Therefore, active strategies in which an external event triggers droplet fusion are more attractive due to their robustness. Laser-assisted local warming of droplet–droplet interface was shown to efficiently promote droplet fusion [83] but is not frequently used as it can damage the biological sample. Instead, most of the active droplet fusion devices use either charged droplets [84] and more frequently electrocoalescence where an AC field is used to destabilize droplets' interface and promote their fusion (Figure 3b) [85]. Electrodes can be directly inserted into the channel [86], built under it [79] or, more frequently, made of a solder [87] or a salt solution [88] injected in channels surrounding the channel of interest.

Finally, addition of reagents to surfactant-stabilized droplets can also be performed by the direct injection of the solution into circulating droplets [89] (Figure 3c). In a droplet picoinjector, a solution is delivered to the droplets while they pass in front a pressurized injection channel subjected to an electric field that transiently destabilizes the droplet/aqueous solution interface. Even though this module can be prone to hydrodynamic instabilities, the implementation of an upstream pressure stabilizer was found to greatly improve its reliability [90].

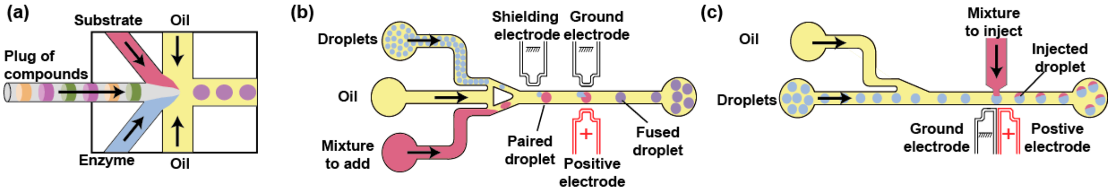


Figure 3. Main strategies for modifying droplet content. (**a**) Modification before droplet production. Plugs of compounds spaced by a carrier liquid (in gray) are injected into a droplet generator and combined with two other aqueous phases (containing enzyme and substrate) just prior to being emulsified. (**b**) Droplet fusion device. Schematic of an active droplet fusion device where droplets are synchronized and fused when passing between a pair of electrodes. Here, small droplets (blue) are reinjected, spaced, and synchronized with larger droplets (red) generated on-chip. (**c**) Droplet picoinjection device. Droplets are reinjected and spaced by an oil stream. When passing in between an orthogonal channel containing the fluid to inject and a pair of electrodes that destabilize droplet/liquid interface, the droplets receive a controlled volume of fluid. Arrows indicate flow directions.

3.3. Droplet Analysis and Sorting

The last step of a µIVC screening consists in measuring the phenotype (i.e., the enzymatic activity) of each droplet and using it to trigger the specific recovery of droplets containing the corresponding genotype.

Over the past decade, a large set of analytical techniques has been adapted to droplet-based microfluidics as reviewed in [91]. These techniques include label-free methods such Raman spectroscopy and mass spectrometry, but also electrochemical detection of reaction products and, more frequently, optical detections exploiting the plethora of fluorogenic and chromogenic substrates commercially available. Light-induced fluorescence (LIF) is currently the most-used detection approach as it offers high sensitivity, makes possible to track several parameters (e.g., enzyme activity, droplet size . . .) in parallel by using different colors, and is compatible with high throughput measurements on droplets circulating with a short residence time in the analytical area. Indeed, since the fluorescence lifetime (ns) is much lower than the time a droplet spends in the analysis window (µs-ms), the analysis time is not a limiting factor, making it possible to analyze the fluorescence of several tens of thousands of droplets per second. Most of the LIF-based assays are performed using continuous illumination at a single measurement point giving access only to a snapshot of the reaction. However, recent work showed that, even though at a lower throughput (150 droplets per second), reaction kinetics could be accessed by combining wide field microscopy and stroboscopic illumination [92]. Moreover, despite the small optical path length of the droplets, recent studies have demonstrated the possibility of monitoring droplet absorbance [93,94], which opens exciting perspectives by expanding the range of assays achievable in droplets.

Phenotypes can be used to trigger the specific recovery of the droplets of interest by deflecting them into a dedicated channel. Typical sorting devices consist of a re-injection module where the droplets are injected and spaced by an oil stream before arriving at a junction where the main channel splits into at least two exhaust channels: the "waste" channel into which the droplets flow per default and the "collection" channel the droplets of interest are deflected in (Figure 4a). As recently reviewed elsewhere [95], there is a large variety of geometries and active strategies that can be used to sort droplets including laser-induced thermophoresis [83], mechanical sorting using piezoelectric elements [96] or single-layer built-in membrane valves [97], acoustic waves [98,99], as well as electrical approaches in which droplets are charged on-chip prior to being deflected into one [84] or several channels by dedicated sets of electrodes [100]. One of the most used deflection methods in µIVC is based on dielectrophoresis (DEP). In DEP, the application of an AC field to a pair of electrodes creates a non-uniform electric field polarizing the droplet and pulling it toward the most intense part of the field, which ultimately leads to its deflection toward the "collection" channel [101]. Depending on the biological assay, integrating a fluorescence or absorbance read-out with DEP sorting produces FADS (Fluorescence-Activated Droplet Sorting) [59] and AADS (Absorbance-Activated Droplet Sorting) [94] microfluidic platforms, respectively. Recently, an optimized FADS geometry allowed sorting up to 30,000 droplets per second (30 kHz) [102]. While the continuous phase circulating in the sort channel usually consists of the fluorinated oil spacing the reinjected droplets, the use of a H-shaped geometry allows filling the collection channel with an aqueous stream so that, upon a sort signal, a droplet is both deflected and its content (a bead [103] or a gene [104,105]) is extracted into an aqueous phase (Figure 4b).

Aside from the microfluidic droplet sorting devices, droplets can also be handled on a conventional FACS (Figure 4c), provided the continuous oil phase is exchanged for an aqueous phase. A first strategy consists of generating a double emulsion by re-emulsifying water-in-oil droplets into water-in-oil-in-water droplets carried by a sheath fluid [106]. Alternatively, droplets can be converted into hydrogel beads [107]. For instance, droplets can initially contain a polymer (e.g., agarose) kept in a liquid form and gelified only after the droplet has been formed. Upon droplet breaking, the beads form a physical boundary that confines compartment content and can be further tightened using additives such as charged polymers [108]. The beads can then be suspended into conventional sheath fluid prior to being loaded, analyzed, and sorted on a FACS machine.

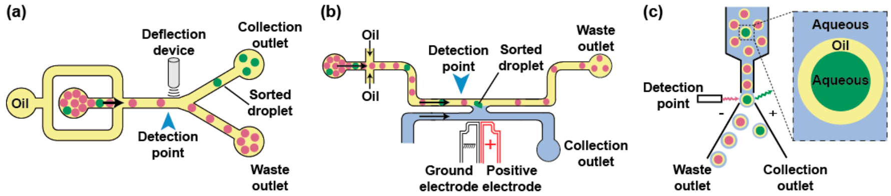


Figure 4. Main droplet sorting strategies. (**a**) Microfluidic droplet sorting using droplet deflection. Droplets are reinjected into a sorting device and spaced by an oil flow. Droplet content (e.g., fluorescence) is analyzed at a detection point (blue arrow) and a deflection device (e.g., pairs of electrodes, surface acoustic wave generator, piezoelectric component . . .) is activated accordingly to deflect the droplets of interest (green) into the collection channel. (**b**) Droplet sorting based on droplet fusion with an aqueous phase. Droplets are reinjected into the sorting device and spaced by an oil flow. Droplet content (e.g., fluorescence) is analyzed at a detection point (blue arrow) and a pair of electrodes is activated to fuse the droplet of interest with a continuous aqueous flow (blue) accordingly. (**c**) FACS-based droplet sorting. Water-in-oil-in-water droplets are carried by a sheath fluid (blue), analyzed and sorted by FACS.

4. Genotype/Phenotype Confinement Maintenance

The success of µIVC screening is primarily conditioned by the proper maintenance of genotype/phenotype confinement. However, this linkage can be challenged by two main scenarios: (i) an uncontrolled coalescence resulting from a lack of droplet stability and (ii) droplet–droplet exchange of reaction products, especially fluorescent dyes, leading to an uniformization of droplet phenotypes. In both cases, various strategies have been proposed to limit these adverse effects.

4.1. Droplet Stabilization

Dispersing an aqueous medium into an immiscible oil phase allows producing droplets whose interface is subjected to a high surface tension. As a consequence, droplets tend to coalesce to minimize oil/water surface area, a phenomenon that disrupts genotype/phenotype linkage. Nevertheless, droplet stability can be preserved by supplementing the oil phase with a "surface active agent" (or surfactant). Surfactants are amphiphilic molecules composed of different groups having affinity for each phase (oil and aqueous) and that are able to reduce the surface tension by partitioning at the oil/aqueous interface (Figure 5a). Once at the interface, the surfactant was proposed to prevent droplet coalescence by steric hindrance as well as by the establishment of a Marangoni flow, counteracting the oil drainage between droplets contacting each other [109]. A variety of oils and surfactants can be used as reviewed elsewhere [109,110] and, whereas IVC protocols usually employ non-fluorinated oils [111], nowadays droplet-based microfluidic approaches mainly use fluorinated carrier oils (e.g., FC40 and Novec 7500) since they both limit droplet–droplet molecular exchanges, as organic molecules are poorly soluble in these oils [112], and they are well compatible with PDMS (avoiding PDMS swelling [113] as mineral oil would [52]). In addition, fluorinated oils efficiently dissolve respiratory gases [114], a particularly interesting feature when cells need to be grown in the droplets. Most of the fluorinated surfactants are based on the perfluorinated polyether (PFPE) Krytox™ (Figure 5b) that possesses a carboxylic head group. However, the ammonium salt of the molecule has poor biocompatibility [115], a key parameter that can be efficiently assessed using cell-based [115] or in vitro gene expression [116] assays. The toxicity of Krytox™ was likely due to the nature of the counter cation used, since a recent study showed that the addition of free polyetherdiamine in the aqueous phase (Figure 5c) restores the biocompatibility of droplets stabilized by Krytox™ [117]. However, most of the surfactants used nowadays consist of PFPE tails covalently conjugated with polar heads like dimorpholinophosphate [115], polyethylene glycols [118], or polyetherdiamine [71]. Several of these block co-polymers are now commercially available (e.g., from RAN biotechnologies (Beverly, MA, USA) or Sphere Fluidics (Cambridge, UK)) which eases the access to the technology.

Furthermore, since surfactant toxicity may be due to non-specific adhesion to surface, additives such as Pluronic®, can be added to the aqueous phase to limit adsorption of cells or molecules [119].

Besides its function in droplet stabilization, the surfactant may also fulfill additional roles. For instance, using a surfactant affording a polar head functionalized with a nitrilotriactetate group was used to both stabilize the droplets and capture his-tagged green fluorescent protein at their surface [120]. Similarly, gold-conjugated surfactant was used to immobilize peptides for capturing cells at the droplet surface [121]. Finally, inner droplet surface could also be functionalized using surfactant affording hydroxyl groups such as the fluorinated polyglycerols [122] or *via* bio-orthogonal click chemistry using azide surfactant [123], both approaches opening up exciting perspectives in future applications exploiting the large inner droplet surface.

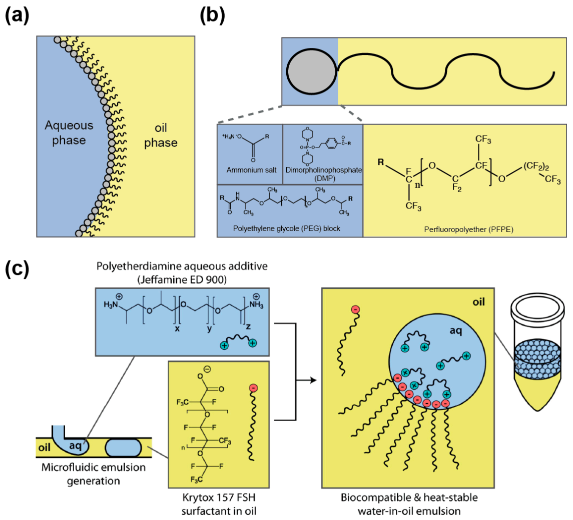

(a)
(b)
Aqueous phase
oil phase
Ammonium salt
Dimorpholinophosphate (DMP)
Polyethylene glycole (PEG) block
Perfluoropolyether (PFPE)
(c)
Polyetherdiamine aqueous additive (Jeffamine ED 900)
oil
aq
Microfluidic emulsion generation
Krytox 157 FSH surfactant in oil
Biocompatible & heat-stable water-in-oil emulsion

Figure 5. Chemical structure of fluorinated surfactants. (**a**) Surfactants are amphiphilic molecules that stabilize droplets by partitioning at water/oil interface. (**b**) Structure of the fluorinated surfactants. Perfluoropolyether (PFPE) fluorinated tail (shaded in yellow) is covalently attached to different polar head groups (shaded in blue). (**c**) Droplet stabilization by bipartite surfactant. Krytox™ 157 FSH fluorinated tail (shaded in yellow) stabilizes the droplet by partitioning at the interface and interacts with Jeffamine ED 900 polar head group (shaded in blue) by ionic interactions. Reprinted with permission [117]. Copyright 2013 American Chemical Society.

4.2. Limiting Droplet–Droplet Exchange

Preserving the genotype/phenotype link also requires preventing information exchange between the droplets. Bulky charged genetic polymers (RNA and DNA) are unlikely to easily pass through the droplet/oil interface. However, smaller molecules like fluorescent products (phenotype) are more prone to be released from and/or exchanged between the droplets. Such a leakage can occur in two different ways: (i) partition of the compound into the oil phase or (ii) micellar transport. Whereas direct partition of the dyes into the oil phase has been described for droplets carried by hydrocarbon oils [124,125], the phenomenon is less common in fluorinated oils even though some dyes like coumarin were found to efficiently diffuse into Novec 7500 oil [126]. Molecular retention is however strongly challenged by micellar transport, a dynamic process that has been modeled [127,128] and in which the free surfactant contained into the oil self-organizes into micelles that can act as cargo, transporting molecules from one droplet to the other. Micelles formation can be limited by working at surfactant concentrations below the critical micellar concentration (CMC) at which they start to be significant, but this would also challenge droplets stability. Therefore, several strategies have been explored to limit micellar transport while working at optimal surfactant concentrations. Simple actions such as adjusting surfactant concentration and maintaining spacing in between droplets were shown to significantly decrease exchange kinetics [129]. Nevertheless, this may not always be applicable, especially when extended incubation times are required, as it is the case during most of the biological screenings, and for which droplets are incubated as dense and compact emulsions.

The degree to which a dye is loaded into micelles differs from one molecule to other. Indeed, while some dyes, like fluorescein (Figure 6a), have a retention time in droplets of several days, others like rhodamine 6G [129], coumarin [126], and resorufin [127] tend to have very fast leakage kinetics with retention times in the order of seconds to minutes. Several works have correlated the propensity of a dye to be exchanged by micellar transport with its distribution coefficient (LogD) [126,130], suggesting that molecules of higher hydrophobicity are more likely to be efficiently loaded into micelles. Therefore, increasing the hydrophilicity of the dye by adding polar groups was expected to reduce micellar transport of the molecule. This hypothesis was first validated by the observation that substituting a coumarin with a sulfonate group increased its retention time in the droplets from a few minutes to hours and even days [126,131]. Addition of hydrocarbon (dodecyl) groups was also found to significantly increase the retention of resorufin by a mechanism that still needs to be deciphered [132]. Beside the chemical modification of the fluorophore, Janiesch et al. showed that tailoring buffer composition, as well as surfactant geometry and concentration, to the properties of the dye may also be used to modulate its retention [130]. Reaction mixtures can also be supplemented with additives such as sugars [133] and bovine serum albumin (BSA, [134]) that were both found to significantly limit fluorescent dye leakage. While the exact mechanism of action underlying sugars is unclear, BSA was suggested to act by increasing the solubility of the fluorophore into the droplets [127].

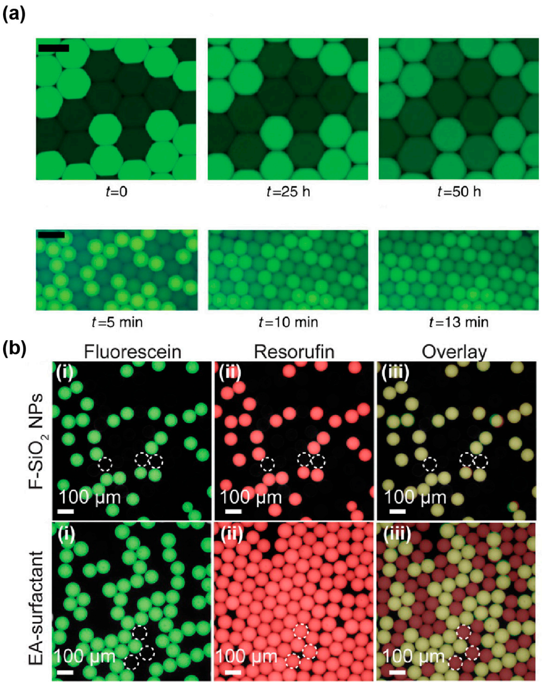


Figure 6. Retention of different fluorescent dyes into droplets. (**a**) Retention of fluorescent dyes into droplets stabilized by a surfactant. Whereas fluorescein has an extended retention time (top panel) in surfactant-stabilized droplets, the rhodamine 6G is rapidly exchanged (bottom panel). Adapted from [129] under a Creative Commons License. (**b**) Retention of fluorescent dyes into droplets stabilized by a nanoparticle. Whereas both fluorescein and resorufin are efficiently confined into nanoparticle-stabilized droplets (top panel), the latter is rapidly exchanged between surfactant-stabilized droplets (bottom panel). Adapted with permission from [135]. Copyright 2014 American Chemical Society.

Finally, a new solution to both stabilize droplets and efficiently prevent droplet–droplet exchanges may come from the use of nanoparticules (NPs) instead of conventional surfactants. Indeed, a recent report showed that fluorinated silica NPs efficiently stabilize droplets carried in a surfactant-free oil phase, while abolishing resorufin leakage (Figure 6b) [135] and being biocompatible [136]. For µIVC screening applications, these attractive properties might however be balanced by the rather high stability of the interface. Indeed, the authors pointed out that NPs were likely to irreversibly adsorb at the interface, raising questions about the possibility of modifying droplet content after they have been formed, a limitation shared with IVC. In addition, the possibility of recovering the content from droplets of interest still needs to be addressed.

5. Discovery and Improvement of Biological Catalysts

Combining several of the microfluidic devices introduced above, together with the use of proper reagents, allowed devising µIVC screening workflows for the isolation of efficient catalysts, either from mutant libraries by directed evolution of natural or synthetic molecules, or through the bioprospecting of new activities from environmental samples. In any case, the low cost and the ultrahigh-throughput of the µIVC are two key factors enabling rapid and efficient screening of large libraries for rare events that would be extremely difficult to identify by lower throughput approaches such as MTPs.

5.1. Cell-Based Directed Evolution

The first directed evolution performed with µIVC was achieved by Agresti et al. who developed an integrated microfluidic platform composed of a droplet generator directly connected to a FADS module by a delay line, which was used to improve the catalytic properties of the horseradish peroxidase (HRP) [137]. HRP mutant gene libraries were prepared by error-prone PCR prior to being expressed in yeast as surface-displayed proteins. Since cell encapsulation follows a Poisson distribution [115,138], adjusting the dilution of cell suspension allowed encapsulating single cells in the presence of a fluorogenic substrate of the enzyme, therefore confining a single genotype with its phenotype. Upon short on-chip incubation, droplets presenting the highest fluorescence signal were sorted and the cells were recovered. Performing a few rounds of screening identified improved variants of the already efficient HRP that were able to catalyze the reaction (10 times higher catalytic constant k_{cat}) with a near diffusion-limited efficiency. A second major outcome of this work was the demonstration of the significant saving of time (1000 times faster) and cost (1 million times cheaper) allowed by such µIVC workflow with respect to a conventional robotic-based screening.

While yeast display allows protein synthesis in a eukaryotic context, gene libraries can also be expressed in prokaryotes, especially in the bacterium *Escherichia coli*, with the protein of interest accumulating either in the periplasm or in the cytoplasm. The former allows the substrate to directly access the enzyme as shown by a model experiment in which *Escherichia coli* cells expressing *Bacillus subtilis* CotA laccase were subjected to a µIVC procedure to profile the activity of a mutant library [73]. Nevertheless, most of the evolutions performed so far used cytoplasmic expression of the protein, requiring the cell to be lysed into the droplet to release the enzyme [60,94,105,108,139–141]. In these approaches (Figure 7a), the cell suspension is first combined on-chip with an aqueous stream containing a lysis agent (usually a detergent, but the combined use of lysozyme with an electric field [141] and heat-lysis [139] were also demonstrated) and a fluorogenic substrate prior to generating water-in-oil droplets [60,142]. Droplets are then incubated either on-chip [140] or off-chip [60,94,105,108,139], prior to being analyzed and sorted by FADS [60,105,140], FACS [108,139] or AADS [94]. This strategy was used to screen mutants of the promiscuous *Pseudomonas aeruginosa* arylsulfatase for their capacity to transform a phosphonate molecule, a secondary substrate of the enzyme [60]. This led to the isolation of mutants with an overall increased expression level and a mutant six-fold more efficient. On another hand, the improvement of the detoxifying phosphotriesterase from *Pseudomonas diminuta* was of particular interest [108]. Indeed, whereas the µIVC procedure could identify mutants with a 19-fold improved activity toward the fluorogenic substrate analog used for the screening, the activity with the native substrate (the paraoxon pesticide) was increased only 8-fold, highlighting a potent limitation of using a fluorogenic substrate analog as the enzyme can evolve to recognize it better than the native substrate. Such a limitation can nevertheless be overcome by using the natural substrate of the enzyme and by detecting the reaction product with a chain of enzymatic reactions that convert the product into a fluorescent signal. This strategy was demonstrated for the detection of cellulase activity [143] as well as xylose consumption and lactate production [75]. Beside the natural enzymes, cell-based μIVC was also recently used to improve 30-fold the catalytic properties of an artificial computationally designed aldolase [140], opening new perspectives in the conception of artificial catalysts. Another μIVC strategy called DrOPS, recently allowed isolating a mutant of the 9n-GLK DNA polymerase (an engineered version of the DNA polymerase from *Thermococcus* sp. 9 °N) able to efficiently copy a DNA template into α-L-threofuranosyl xeno nucleic acid with a high degree of fidelity [139].

Whereas in most of the applications presented above the bacteria were lysed immediately upon their encapsulation, several works showed that integrating a picoinjection [73,74] or a fusion [75] step in the μIVC workflow further increases the flexibility of the process by successively performing steps with different timescale. This allowed, for instance, to first grow cells individualized into droplets for several hours (or days) prior to adding activity assay reagents and incubate the mixture for a few minutes before analyzing and sorting the droplets.

Mutational scanning, a strategy combining μIVC screening and next generation sequencing (NGS) represents an alternative strategy to improve enzymes in which the active mutants of a library are first sorted by μIVC before the sequences contained in both selected and unselected libraries are analyzed by NGS [105]. While this approach primarily allows for the discrimination of key residues (invariant) from those supporting variations, the application of a selection pressure and the insertion of cherry-picked mutations into the wild-type molecule allows the engineering of optimized enzymes, as demonstrated with the Bgl3 β-gluconidase from *Streptomyces* sp. whose thermostability was increased by more than 5 °C by a single point mutation identified by this approach.

All the screenings introduced above were performed with host cells expressing a library of mutant genes carried by a plasmid. However, another approach consists of introducing mutations at the level of the whole organism and selecting those survivals improved for the target activity. This was recently shown with the fungus *Aspergillus niger* and allowed the isolation of individuals displaying a significantly improved amylase activity among the rare survivals of a UV-irradiated population of spores [144].

Altogether, cell-based μIVC strategies are extremely attractive since they are cheap and they allow the isolation of catalysts not only improved for the target activity but also optimized in terms of expression level [75] and solubility [94], two important features when evolving molecules expected to be later produced at an industrial scale. Besides the direct improvement of the catalyst, cell-based μIVC can also be used to screen libraries of microbes prepared by UV irradiation and to search for host microorganisms with improved secretion properties [145,146].

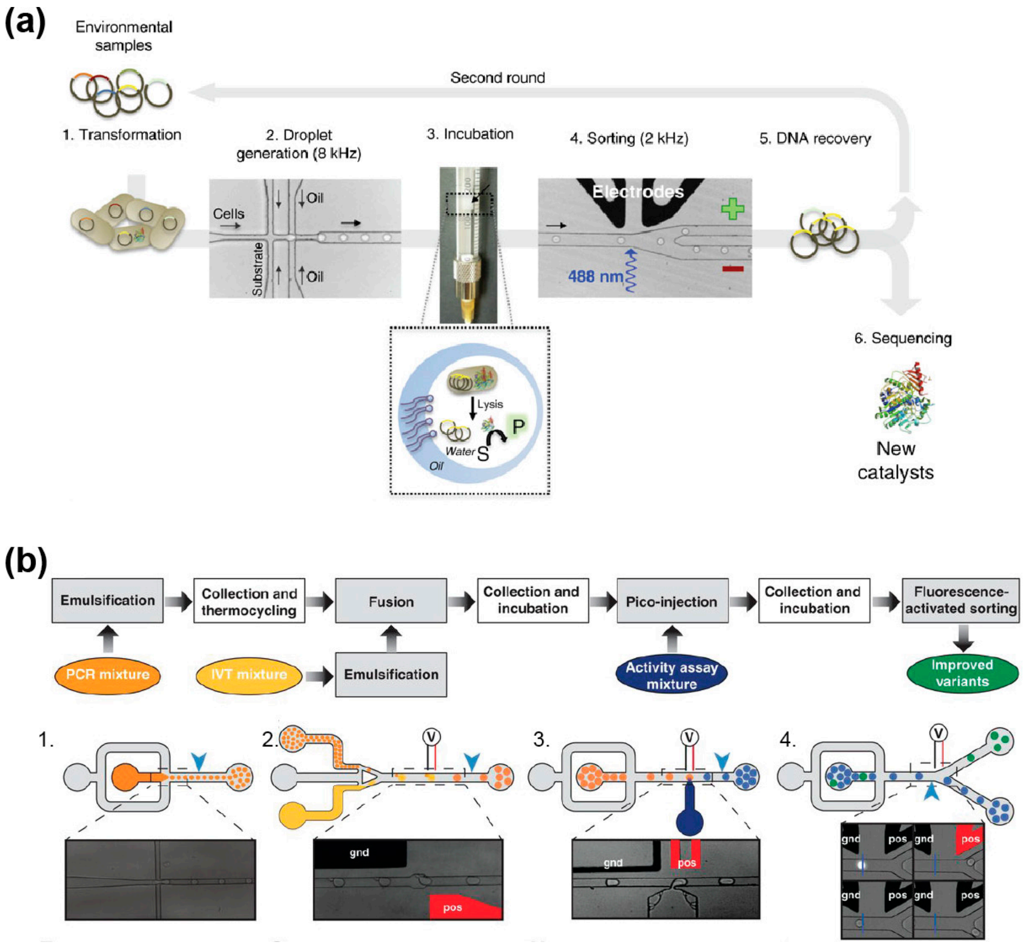

(a)
Environmental samples
Second round
1. Transformation
2. Droplet generation (8 kHz)
3. Incubation
4. Sorting (2 kHz)
5. DNA recovery
Cells
Oil
Substrate
Oil
Electrodes
+
488 nm
−
Lysis
P
Water
S
Oil
6. Sequencing
New catalysts
(b)
Emulsification
Collection and thermocycling
Fusion
Collection and incubation
Pico-injection
Collection and incubation
Fluorescence-activated sorting
PCR mixture
IVT mixture
Emulsification
Activity assay mixture
Improved variants
1.
2.
3.
4.
V
gnd
pos
gnd
pos
gnd
pos
gnd
pos
gnd
pos
gnd
pos

Figure 7. Microfluidic workflow for directed evolution of biological catalysts. (**a**) Cell-based μIVC screening workflow. Bacteria were transformed with a metagenomic library from an environmental sample (1) and individualized into droplets together with a fluorogenic substrate and a lysis agent (2). Upon collection and incubation in a syringe (3), the emulsion was reinjected into a microfluidic droplet sorter (4), the positive droplets were sorted and the DNA recovered (5). The variants were either analyzed or used for a new round of screening (6). Adapted from [147] under a Creative Commons License. (**b**) Cell-free μIVC screening workflow. Genes of a library were individualized into PCR droplets (1) and, upon DNA amplification by thermocycling, each droplet was synchronized and fused with a droplet containing an in vitro transcription mixture (2). Once the transcription occurred, a substrate solution was delivered to each droplet by picoinjection (3) prior to a short incubation followed by droplet fluorescence analysis and sort (4). Adapted from [148] under a Creative Commons License.

5.2. Cell-Free Directed Evolution

μIVC-assisted directed evolution can also be performed in a cell-free manner using in vitro gene expression mixtures either made of purified components (in vitro transcription and Protein synthesis Using Recombinant Elements (PURE) [149] in vitro translation system) or based on cell extracts. This strategy ensures that all the droplets of an emulsion have identical composition, further increasing the accuracy of the method. In addition, using in vitro expression format allows for the expression of proteins toxic for a cell or even the introduction of non-natural amino acids. The first cell-free μIVCs were performed by mixing a cell extract-based in vitro expression mixture with a DNA solution on-chip just prior to producing the droplets [52,150]. Since DNA molecules distribute into the droplets following Poisson statistics [58,151], limiting dilution produces droplets containing at most one DNA molecule. The small volume of the droplets (femtoliters to picoliters) concentrates expression products to a detectable level [52,150,152], but working at the single molecule level may also lead to a significant variance resulting from gene-to-gene expression variation. Nevertheless, this variance can be mostly suppressed by increasing the copy number of gene molecules prior to expressing them [152]. Clonal population of a DNA molecule can easily be generated in droplets using isothermal [58] or PCR amplification [151,153]. However, since DNA amplification and gene expression are mutually exclusive events, a droplet fusion step is required upon DNA amplification to feed each droplet with the in vitro expression mixture [58]. Therefore, a typical cell-free µIVC screening of a mutant gene library occurs in three main steps: (i) DNA individualization and amplification, (ii) gene expression, and (iii) droplets analysis and sorting [104]. Whereas such a pipeline has not been applied to protein evolution yet, it has been used to improve the catalytic properties of RNA molecules [148] (Figure 7b), a process that would have been extremely challenging to perform using cell-based µIVC because of the strong cellular RNase background activity. Furthermore, the same procedure was used to improve the folding properties of the fluorogenic RNA aptamer Spinach by using selection pressures that would have been impossible to apply in a cell-based format (i.e., complete absence of potassium ions) [154]. Therefore, even though cell-free µIVC workflows are more complex than their cell-based counterparts, they are extremely valuable for applications requiring high control over reaction conditions and/or that are sensitive to cellular background.

5.3. Discovery of New Catalysts from Environmental Samples

Given the limited fraction of environmental microbes that are culturable [155], only a small fraction of the existing catalysts has been accessed so far, whilst the remaining microbiological dark matter has been explored mainly at a metagenomic level making possible only the identification of catalysts sharing homologies with the known enzymes [156]. Therefore, the ultrahigh-throughput regime of µIVC and its capacity to handle single cells make it an extremely attractive option for the functional screening of environmental samples during bioprospecting programs. This was recently demonstrated by the successful identification of cellulolytic microbes collected from a wheat stubble field [157]. Indeed, using a µIVC procedure, more than 100,000 microbes were screened in less than 20 min without the need of any pre-culturing step leading to the identification of several taxons with increased cellulobiohydrolase activity. Even more excitedly, another recent study used a cell-based µIVC approach to functionally screen a pool of metagenomic libraries from diverse sources cloned and expressed into *Escherichia coli* [147]. 1,250,000 variants were screened for promiscuous sulfatase and phosphotriesterase activities. Importantly, despite the fact that all these new enzymes shared a common activity, sequence-similarity networks showed that they did not strongly cluster together, but were distributed among various unrelated super-families. This finding is of great importance as it demonstrates that the sequence is insufficient for predicting a function and that functional screening by µIVC is a robust way to discover new catalysts.

6. Conclusions and Outlook

In this review, we presented the main microfluidic modules that can be combined to devise µIVC screening procedures for performing highly quantitative measurements while giving great flexibility to the experimental workflows. The possibility of lysing cells on demand, either upon their encapsulation or after a defined culture time while preserving genotype/phenotype confinement has made this technology extremely attractive for cell-based screening performed at the single-cell level. While today only a handful of works has been reported, it is very likely that, in a near future, a plethora of new catalysts will be discovered and improved. The possibility of exploring the microbial dark matter should have a major impact on these new developments. The capacity of µIVC to identify the rare variants of interest contained in very large populations and at a very low cost makes a three-step scenario that could be routinely adopted in a near future highly plausible. In this scenario, µIVC would be first used to enrich environmental samples from various origins (e.g., agricultural soil, contaminated soils from industrial plant ...) in organisms displaying the highest levels of the target activity (e.g., [157]). Then, metagenomic DNA libraries could be prepared from these microbes and a second μIVC screening could be used to identify the genes encoding for the target activity (e.g., [75,147]). Finally, the properties of the newly identified catalysts would be improved and tailored for their final application using one of the directed evolution workflows presented in this review. There is no doubt that the increasing number of companies selling microfluidic equipment and reagents will greatly boost new developments and offer a wider use of μIVC.

Acknowledgments: We warmly thank Redmond Smyth for critical reading of the manuscript. This work has been published under the framework of the LABEX: ANR-10-LABX-0036_NETRNA and benefits from funding from the state-managed 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.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; W.H. Freeman: New York, NY, USA, 1999; p. xxi.
2. Jacques, P.; Bechet, M.; Bigan, M.; Caly, D.; Chataigne, G.; Coutte, F.; Flahaut, C.; Heuson, E.; Leclere, V.; Lecouturier, D.; et al. High-throughput strategies for the discovery and engineering of enzymes for biocatalysis. *Bioprocess Biosyst. Eng.* **2016**, *40*, 161–180. [CrossRef] [PubMed]
3. Molina-Espeja, P.; Vina-Gonzalez, J.; Gomez-Fernandez, B.J.; Martin-Diaz, J.; Garcia-Ruiz, E.; Alcalde, M. Beyond the outer limits of nature by directed evolution. *Biotechnol. Adv.* **2016**, *34*, 754–767. [CrossRef] [PubMed]
4. Khersonsky, O.; Tawfik, D.S. Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **2010**, *79*, 471–505. [PubMed]
5. Pandya, C.; Farelli, J.D.; Dunaway-Mariano, D.; Allen, K.N. Enzyme promiscuity: Engine of evolutionary innovation. *J. Biol. Chem.* **2014**, *289*, 30229–30236. [CrossRef] [PubMed]
6. Joyce, G.F. Forty years of in vitro evolution. *Angew. Chem.* **2007**, *46*, 6420–6436. [CrossRef] [PubMed]
7. Romero, P.A.; Arnold, F.H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 866–876. [CrossRef] [PubMed]
8. Goldsmith, M.; Tawfik, D.S. Directed enzyme evolution: Beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.* **2012**, *22*, 406–412. [CrossRef] [PubMed]
9. Denard, C.A.; Ren, H.; Zhao, H. Improving and repurposing biocatalysts via directed evolution. *Curr. Opin. Struct. Biol.* **2015**, *25*, 55–64. [CrossRef] [PubMed]
10. Zhao, H.; Chockalingam, K.; Chen, Z. Directed evolution of enzymes and pathways for industrial biocatalysis. *Curr. Opin. Biotechnol.* **2002**, *13*, 104–110. [CrossRef]
11. Wang, M.; Si, T.; Zhao, H. Biocatalyst development by directed evolution. *Bioresour. Technol.* **2012**, *115*, 117–125. [CrossRef] [PubMed]
12. Xiao, H.; Bao, Z.; Zhao, H. High throughput screening and selection methods for directed enzyme evolution. *Ind. Eng. Chem. Res.* **2015**, *54*, 4011–4020. [CrossRef] [PubMed]
13. Porter, J.L.; Rusli, R.A.; Ollis, D.L. Directed evolution of enzymes for industrial biocatalysis. *Chembiochem Eur. J. Chem. Biol.* **2016**, *17*, 197–203. [CrossRef] [PubMed]
14. Kumar, A.; Singh, S. Directed evolution: Tailoring biocatalysts for industrial applications. *Crit. Rev. Biotechnol.* **2013**, *33*, 365–378. [CrossRef] [PubMed]
15. Fernandez-Gacio, A.; Uguen, M.; Fastrez, J. Phage display as a tool for the directed evolution of enzymes. *Trends Biotechnol.* **2003**, *21*, 408–414. [CrossRef]
16. Guo, M.T.; Rotem, A.; Heyman, J.A.; Weitz, D.A. Droplet microfluidics for high-throughput biological assays. *Lab Chip* **2012**, *12*, 2146–2155. [CrossRef] [PubMed]
17. Price, A.K.; Paegel, B.M. Discovery in droplets. *Anal. Chem.* **2016**, *88*, 339–353. [CrossRef] [PubMed]
18. Rakszewska, A.; Tel, J.; Chokkalingam, V.; Huck, W.T.S. One drop at a time: Toward droplet microfluidics as a versatile tool for single-cell analysis. *NPG Asia Mater.* **2014**, *6*, e133. [CrossRef]
19. Shembekar, N.; Chaipan, C.; Utharala, R.; Merten, C.A. Droplet-based microfluidics in drug discovery, transcriptomics and high-throughput molecular genetics. *Lab Chip* **2016**, *16*, 1314–1331. [CrossRef] [PubMed]
20. Mayr, L.M.; Bojanic, D. Novel trends in high-throughput screening. *Curr. Opin. Pharmacol.* **2009**, *9*, 580–588. [CrossRef] [PubMed]

21. Yang, G.; Withers, S.G. Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *Chembiochem Eur. J. Chem. Biol.* **2009**, *10*, 2704–2715. [CrossRef] [PubMed]
22. Colin, P.Y.; Zinchenko, A.; Hollfelder, F. Enzyme engineering in biomimetic compartments. *Curr. Opin. Struct. Biol.* **2015**, *33*, 42–51. [CrossRef] [PubMed]
23. Tawfik, D.S.; Griffiths, A.D. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* **1998**, *16*, 652–656. [CrossRef] [PubMed]
24. Griffiths, A.D.; Tawfik, D.S. Miniaturising the laboratory in emulsion droplets. *Trends Biotechnol.* **2006**, *24*, 395–402. [CrossRef] [PubMed]
25. Kelly, B.T.; Baret, J.C.; Taly, V.; Griffiths, A.D. Miniaturizing chemistry and biology in microdroplets. *Chem. Commun.* **2007**, *18*, 1773–1788. [CrossRef] [PubMed]
26. Bernath, K.; Hai, M.; Mastrobattista, E.; Griffiths, A.D.; Magdassi, S.; Tawfik, D.S. In vitro compartmentalization by double emulsions: Sorting and gene enrichment by fluorescence activated cell sorting. *Anal. Biochem.* **2004**, *325*, 151–157. [CrossRef] [PubMed]
27. Aharoni, A.; Amitai, G.; Bernath, K.; Magdassi, S.; Tawfik, D.S. High-throughput screening of enzyme libraries: Thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chem. Biol.* **2005**, *12*, 1281–1289. [CrossRef] [PubMed]
28. Gupta, R.D.; Goldsmith, M.; Ashani, Y.; Simo, Y.; Mullokandov, G.; Bar, H.; Ben-David, M.; Leader, H.; Margalit, R.; Silman, I.; et al. Directed evolution of hydrolases for prevention of g-type nerve agent intoxication. *Nat. Chem. Biol.* **2011**, *7*, 120–125. [CrossRef] [PubMed]
29. Hardiman, E.; Gibbs, M.; Reeves, R.; Bergquist, P. Directed evolution of a thermophilic beta-glucosidase for cellulosic bioethanol production. *Appl. Biochem. Biotechnol.* **2010**, *161*, 301–312. [CrossRef] [PubMed]
30. Ma, F.; Xie, Y.; Huang, C.; Feng, Y.; Yang, G. An improved single cell ultrahigh throughput screening method based on in vitro compartmentalization. *PLoS ONE* **2014**, *9*, e89785. [CrossRef] [PubMed]
31. Mastrobattista, E.; Taly, V.; Chanudet, E.; Treacy, P.; Kelly, B.T.; Griffiths, A.D. High-throughput screening of enzyme libraries: In vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. *Chem. Biol.* **2005**, *12*, 1291–1300. [CrossRef] [PubMed]
32. Ostafe, R.; Prodanovic, R.; Commandeur, U.; Fischer, R. Flow cytometry-based ultra-high-throughput screening assay for cellulase activity. *Anal. Biochem.* **2013**, *435*, 93–98. [CrossRef] [PubMed]
33. Ostafe, R.; Prodanovic, R.; Nazor, J.; Fischer, R. Ultra-high-throughput screening method for the directed evolution of glucose oxidase. *Chem. Biol.* **2014**, *21*, 414–421. [CrossRef] [PubMed]
34. Tu, R.; Martinez, R.; Prodanovic, R.; Klein, M.; Schwaneberg, U. A flow cytometry-based screening system for directed evolution of proteases. *J. Biomol. Screen.* **2011**, *16*, 285–294. [CrossRef] [PubMed]
35. Ghadessy, F.J.; Ong, J.L.; Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4552–4557. [CrossRef] [PubMed]
36. Zaher, H.S.; Unrau, P.J. Selection of an improved rna polymerase ribozyme with superior extension and fidelity. *RNA* **2007**, *13*, 1017–1026. [CrossRef] [PubMed]
37. Levy, M.; Griswold, K.E.; Ellington, A.D. Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. *RNA* **2005**, *11*, 1555–1562. [CrossRef] [PubMed]
38. Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **2011**, *332*, 209–212. [CrossRef] [PubMed]
39. Ma, F.; Fischer, M.; Han, Y.; Withers, S.G.; Feng, Y.; Yang, G.Y. Substrate engineering enabling fluorescence droplet entrapment for IVC-FACS-based ultrahigh-throughput screening. *Anal. Chem.* **2016**, *88*, 8587–8595. [CrossRef] [PubMed]
40. Zhu, P.; Wang, L. Passive and active droplet generation with microfluidics: A review. *Lab Chip* **2016**, *17*, 34–75. [CrossRef] [PubMed]
41. Umbanhowar, P.B.; Prasad, V.; Weitz, D.A. Monodisperse emulsion generation via drop break off in a coflowing stream. *Langmuir* **2000**, *16*, 347–351. [CrossRef]
42. Thorsen, T.; Roberts, R.W.; Arnold, F.H.; Quake, S.R. Dynamic pattern formation in a vesicle-generating microfluidic device. *Phys. Rev. Lett.* **2001**, *86*, 4163–4166. [CrossRef] [PubMed]
43. Garstecki, P.; Fuerstman, M.J.; Stone, H.A.; Whitesides, G.M. Formation of droplets and bubbles in a microfluidic T-junction—Scaling and mechanism of break-up. *Lab Chip* **2006**, *6*, 437–446. [CrossRef] [PubMed]
44. Anna, S.L.; Bontoux, N.; Stone, H.A. Formation of dispersions using "flow focusing" in microchannels. *Appl. Phys. Lett.* **2003**, *82*, 364–366. [CrossRef]

45. Abate, A.R.; Poitzsch, A.; Hwang, Y.; Lee, J.; Czerwinska, J.; Weitz, D.A. Impact of inlet channel geometry on microfluidic drop formation. *Phys. Rev. E* **2009**, *80*, 026310. [CrossRef] [PubMed]
46. Chan, E.M.; Alivisatos, A.P.; Mathies, R.A. High-temperature microfluidic synthesis of CdSe nanocrystals in nanoliter droplets. *J. Am. Chem. Soc.* **2005**, *127*, 13854–13861. [CrossRef] [PubMed]
47. Priest, C.; Herminghaus, S.; Seemann, R. Generation of monodisperse gel emulsions in a microfluidic device. *Appl. Phys. Lett.* **2006**, *88*, 024106. [CrossRef]
48. Dangla, R.; Kayi, S.C.; Baroud, C.N. Droplet microfluidics driven by gradients of confinement. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 853–858. [CrossRef] [PubMed]
49. Malloggi, F.; Pannacci, N.; Attia, R.; Monti, F.; Mary, P.; Willaime, H.; Tabeling, P.; Cabane, B.; Poncet, P. Monodisperse colloids synthesized with nanofluidic technology. *Langmuir* **2010**, *26*, 2369–2373. [CrossRef] [PubMed]
50. Rhee, M.; Burns, M.A. Drop mixing in a microchannel for lab-on-a-chip platforms. *Langmuir* **2008**, *24*, 590–601. [CrossRef] [PubMed]
51. Song, H.; Tice, J.D.; Ismagilov, R.F. A microfluidic system for controlling reaction networks in time. *Angew. Chem. Int. Ed.* **2003**, *42*, 768–772. [CrossRef] [PubMed]
52. Courtois, F.; Olguin, L.F.; Whyte, G.; Bratton, D.; Huck, W.T.; Abell, C.; Hollfelder, F. An integrated device for monitoring time-dependent in vitro expression from single genes in picolitre droplets. *Chembiochem Eur. J. Chem. Biol.* **2008**, *9*, 439–446. [CrossRef] [PubMed]
53. Khorshidi, M.A.; Rajeswari, P.K.P.; Wahlby, C.; Joensson, H.N.; Svahn, H.A. Automated analysis of dynamic behavior of single cells in picoliter droplets. *Lab Chip* **2014**, *14*, 931–937. [CrossRef] [PubMed]
54. Huebner, A.; Bratton, D.; Whyte, G.; Yang, M.; Demello, A.J.; Abell, C.; Hollfelder, F. Static microdroplet arrays: A microfluidic device for droplet trapping, incubation and release for enzymatic and cell-based assays. *Lab Chip* **2009**, *9*, 692–698. [CrossRef] [PubMed]
55. Song, H.; Ismagilov, R.F. Millisecond kinetics on a microfluidic chip using nanoliters of reagents. *J. Am. Chem. Soc.* **2003**, *125*, 14613–14619. [CrossRef] [PubMed]
56. Frenz, L.; Blank, K.; Brouzes, E.; Griffiths, A.D. Reliable microfluidic on-chip incubation of droplets in delay-lines. *Lab Chip* **2009**, *9*, 1344–1348. [CrossRef] [PubMed]
57. Pekin, D.; Skhiri, Y.; Baret, J.C.; Le Corre, D.; Mazutis, L.; Salem, C.B.; Millot, F.; El Harrak, A.; Hutchison, J.B.; Larson, J.W.; et al. Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. *Lab Chip* **2011**, *11*, 2156–2166. [CrossRef] [PubMed]
58. Mazutis, L.; Araghi, A.F.; Miller, O.J.; Baret, J.C.; Frenz, L.; Janoshazi, A.; Taly, V.; Miller, B.J.; Hutchison, J.B.; Link, D.; et al. Droplet-based microfluidic systems for high-throughput single DNA molecule isothermal amplification and analysis. *Anal. Chem.* **2009**, *81*, 4813–4821. [CrossRef] [PubMed]
59. Baret, J.C.; Miller, O.J.; Taly, V.; Ryckelynck, M.; El-Harrak, A.; Frenz, L.; Rick, C.; Samuels, M.L.; Hutchison, J.B.; Agresti, J.J.; et al. Fluorescence-activated droplet sorting (FADS): Efficient microfluidic cell sorting based on enzymatic activity. *Lab Chip* **2009**, *9*, 1850–1858. [CrossRef] [PubMed]
60. Kintses, B.; Hein, C.; Mohamed, M.F.; Fischlechner, M.; Courtois, F.; Laine, C.; Hollfelder, F. Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution. *Chem. Biol.* **2012**, *19*, 1001–1009. [CrossRef] [PubMed]
61. Pei, J.; Li, Q.; Kennedy, R.T. Rapid and label-free screening of enzyme inhibitors using segmented flow electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1107–1113. [CrossRef] [PubMed]
62. Sun, S.; Kennedy, R.T. Droplet electrospray ionization mass spectrometry for high throughput screening for enzyme inhibitors. *Anal. Chem.* **2014**, *86*, 9309–9314. [CrossRef] [PubMed]
63. Sun, M.; Fang, Q. High-throughput sample introduction for droplet-based screening with an on-chip integrated sampling probe and slotted-vial array. *Lab Chip* **2010**, *10*, 2864–2868. [CrossRef] [PubMed]
64. Cai, L.F.; Zhu, Y.; Du, G.S.; Fang, Q. Droplet-based microfluidic flow injection system with large-scale concentration gradient by a single nanoliter-scale injection for enzyme inhibition assay. *Anal. Chem.* **2012**, *84*, 446–452. [CrossRef] [PubMed]
65. Gielen, F.; van Vliet, L.; Koprowski, B.T.; Devenish, S.R.; Fischlechner, M.; Edel, J.B.; Niu, X.; deMello, A.J.; Hollfelder, F. A fully unsupervised compartment-on-demand platform for precise nanoliter assays of time-dependent steady-state enzyme kinetics and inhibition. *Anal. Chem.* **2013**, *85*, 4761–4769. [CrossRef] [PubMed]

66. Price, A.K.; MacConnell, A.B.; Paegel, B.M. Hnusabr: Photochemical dose-response bead screening in droplets. *Anal. Chem.* **2016**, *88*, 2904–2911. [CrossRef] [PubMed]
67. Jambovane, S.; Kim, D.J.; Duin, E.C.; Kim, S.K.; Hong, J.W. Creation of stepwise concentration gradient in picoliter droplets for parallel reactions of matrix metalloproteinase ii and ix. *Anal. Chem.* **2011**, *83*, 3358–3364. [CrossRef] [PubMed]
68. Rane, T.D.; Zec, H.C.; Wang, T.H. A barcode-free combinatorial screening platform for matrix metalloproteinase screening. *Anal. Chem.* **2015**, *87*, 1950–1956. [CrossRef] [PubMed]
69. Toh, A.G.G.; Wang, Z.P.; Yang, C.; Nguyen, N.T. Engineering microfluidic concentration gradient generators for biological applications. *Microfluid. Nanofluid.* **2014**, *16*, 1–18. [CrossRef]
70. Bui, M.P.N.; Li, C.A.; Han, K.N.; Choo, J.; Lee, E.K.; Seong, G.H. Enzyme kinetic measurements using a droplet-based microfluidic system with a concentration gradient. *Anal. Chem.* **2011**, *83*, 1603–1608. [CrossRef] [PubMed]
71. Chen, C.H.; Sarkar, A.; Song, Y.A.; Miller, M.A.; Kim, S.J.; Griffith, L.G.; Lauffenburger, D.A.; Han, J. Enhancing protease activity assay in droplet-based microfluidics using a biomolecule concentrator. *J. Am. Chem. Soc.* **2011**, *133*, 10368–10371. [CrossRef] [PubMed]
72. Theberge, A.B.; Whyte, G.; Huck, W.T. Generation of picoliter droplets with defined contents and concentration gradients from the separation of chemical mixtures. *Anal. Chem.* **2010**, *82*, 3449–3453. [CrossRef] [PubMed]
73. Beneyton, T.; Coldren, F.; Baret, J.C.; Griffiths, A.D.; Taly, V. Cota laccase: High-throughput manipulation and analysis of recombinant enzyme libraries expressed in *E. coli* using droplet-based microfluidics. *Analyst* **2014**, *139*, 3314–3323. [CrossRef] [PubMed]
74. Hammar, P.; Angermayr, S.A.; Sjostrom, S.L.; van der Meer, J.; Hellingwerf, K.J.; Hudson, E.P.; Joensson, H.N. Single-cell screening of photosynthetic growth and lactate production by cyanobacteria. *Biotechnol. Biofuels* **2015**, *8*, 193. [CrossRef] [PubMed]
75. Wang, B.L.; Ghaderi, A.; Zhou, H.; Agresti, J.; Weitz, D.A.; Fink, G.R.; Stephanopoulos, G. Microfluidic high-throughput culturing of single cells for selection based on extracellular metabolite production or consumption. *Nat. Biotechnol.* **2014**, *32*, 473–478. [CrossRef] [PubMed]
76. Mazutis, L.; Baret, J.C.; Treacy, P.; Skhiri, Y.; Araghi, A.F.; Ryckelynck, M.; Taly, V.; Griffiths, A.D. Multi-step microfluidic droplet processing: Kinetic analysis of an in vitro translated enzyme. *Lab Chip* **2009**, *9*, 2902–2908. [CrossRef] [PubMed]
77. Baroud, C.N.; Gallaire, F.; Dangla, R. Dynamics of microfluidic droplets. *Lab Chip* **2010**, *10*, 2032–2045. [CrossRef] [PubMed]
78. Niu, X.; Gulati, S.; Edel, J.B.; deMello, A.J. Pillar-induced droplet merging in microfluidic circuits. *Lab Chip* **2008**, *8*, 1837–1841. [CrossRef] [PubMed]
79. Ahn, K.; Agresti, J.; Chong, H.; Marquez, M.; Weitz, D.A. Electrocoalescence of drops synchronized by size-dependent flow in microfluidic channels. *Appl. Phys. Lett.* **2006**, *88*, 264105. [CrossRef]
80. El Debs, B.; Utharala, R.; Balyasnikova, I.V.; Griffiths, A.D.; Merten, C.A. Functional single-cell hybridoma screening using droplet-based microfluidics. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11570–11575. [CrossRef] [PubMed]
81. Mazutis, L.; Griffiths, A.D. Selective droplet coalescence using microfluidic systems. *Lab Chip* **2012**, *12*, 1800–1806. [CrossRef] [PubMed]
82. Bremond, N.; Thiam, A.R.; Bibette, J. Decompressing emulsion droplets favors coalescence. *Phys. Rev. lett.* **2008**, *100*, 024501. [CrossRef] [PubMed]
83. Baroud, C.N.; de Saint Vincent, M.R.; Delville, J.P. An optical toolbox for total control of droplet microfluidics. *Lab Chip* **2007**, *7*, 1029–1033. [CrossRef] [PubMed]
84. Link, D.R.; Grasland-Mongrain, E.; Duri, A.; Sarrazin, F.; Cheng, Z.; Cristobal, G.; Marquez, M.; Weitz, D.A. Electric control of droplets in microfluidic devices. *Angew. Chem. Int. Ed.* **2006**, *45*, 2556–2560. [CrossRef] [PubMed]
85. Chabert, M.; Dorfman, K.D.; Viovy, J.L. Droplet fusion by alternating current (AC) field electrocoalescence in microchannels. *Electrophoresis* **2005**, *26*, 3706–3715. [CrossRef] [PubMed]
86. Priest, C.; Herminghaus, S.; Seemann, R. Controlled electrocoalescence in microfluidics: Targeting a single lamella. *Appl. Phys. Lett.* **2006**, *89*, 134101. [CrossRef]

87. Frenz, L.; El Harrak, A.; Pauly, M.; Begin-Colin, S.; Griffiths, A.D.; Baret, J.C. Droplet-based microreactors for the synthesis of magnetic iron oxide nanoparticles. *Angew. Chem. Int. Ed.* **2008**, *47*, 6817–6820. [CrossRef] [PubMed]
88. Sciambi, A.; Abate, A.R. Generating electric fields in PDMS microfluidic devices with salt water electrodes. *Lab Chip* **2014**, *14*, 2605–2609. [CrossRef] [PubMed]
89. Abate, A.R.; Hung, T.; Mary, P.; Agresti, J.J.; Weitz, D.A. High-throughput injection with microfluidics using picoinjectors. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 19163–19166. [CrossRef] [PubMed]
90. Rhee, M.; Light, Y.K.; Yilmaz, S.; Adams, P.D.; Saxena, D.; Meagher, R.J.; Singh, A.K. Pressure stabilizer for reproducible picoinjection in droplet microfluidic systems. *Lab Chip* **2014**, *14*, 4533–4539. [CrossRef] [PubMed]
91. Zhu, Y.; Fang, Q. Analytical detection techniques for droplet microfluidics—A review. *Anal. Chim. Acta* **2013**, *787*, 24–35. [CrossRef] [PubMed]
92. Hess, D.; Rane, A.; deMello, A.J.; Stavrakis, S. High-throughput, quantitative enzyme kinetic analysis in microdroplets using stroboscopic epifluorescence imaging. *Anal. Chem.* **2015**, *87*, 4965–4972. [CrossRef] [PubMed]
93. Deal, K.S.; Easley, C.J. Self-regulated, droplet-based sample chopper for microfluidic absorbance detection. *Anal. Chem.* **2012**, *84*, 1510–1516. [CrossRef] [PubMed]
94. Gielen, F.; Hours, R.; Emond, S.; Fischlechner, M.; Schell, U.; Hollfelder, F. Ultrahigh-throughput-directed enzyme evolution by absorbance-activated droplet sorting (AADS). *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E7383–E7389. [CrossRef] [PubMed]
95. Xi, H.D.; Zheng, H.; Guo, W.; Ganan-Calvo, A.M.; Ai, Y.; Tsao, C.W.; Zhou, J.; Li, W.; Huang, Y.; Nguyen, N.T.; et al. Active droplet sorting in microfluidics: A review. *Lab Chip* **2017**, *17*, 751–771. [CrossRef] [PubMed]
96. Shemesh, J.; Bransky, A.; Khoury, M.; Levenberg, S. Advanced microfluidic droplet manipulation based on piezoelectric actuation. *Biomed. Microdevices* **2010**, *12*, 907–914. [CrossRef] [PubMed]
97. Abate, A.R.; Agresti, J.J.; Weitz, D.A. Microfluidic sorting with high-speed single-layer membrane valves. *Appl. Phys. Lett.* **2010**, *96*, 203509. [CrossRef]
98. Franke, T.; Abate, A.R.; Weitz, D.A.; Wixforth, A. Surface acoustic wave (SAW) directed droplet flow in microfluidics for PDMS devices. *Lab Chip* **2009**, *9*, 2625–2627. [CrossRef] [PubMed]
99. Schmid, L.; Weitz, D.A.; Franke, T. Sorting drops and cells with acoustics: Acoustic microfluidic fluorescence-activated cell sorter. *Lab Chip* **2014**, *14*, 3710–3718. [CrossRef] [PubMed]
100. Ahn, B.; Lee, K.; Panchapakesan, R.; Oh, K.W. On-demand electrostatic droplet charging and sorting. *Biomicrofluidics* **2011**, *5*, 024113. [CrossRef] [PubMed]
101. Ahn, K.; Kerbage, C.; Hunt, T.P.; Westervelt, R.M.; Link, D.R.; Weitz, D.A. Dielectrophoretic manipulation of drops for high-speed microfluidic sorting devices. *Appl. Phys. Lett.* **2006**, *88*, 024104. [CrossRef]
102. Sciambi, A.; Abate, A.R. Accurate microfluidic sorting of droplets at 30 kHz. *Lab Chip* **2015**, *15*, 47–51. [CrossRef] [PubMed]
103. Fidalgo, L.M.; Whyte, G.; Bratton, D.; Kaminski, C.F.; Abell, C.; Huck, W.T. From microdroplets to microfluidics: Selective emulsion separation in microfluidic devices. *Angew. Chem. Int. Ed.* **2008**, *47*, 2042–2045. [CrossRef] [PubMed]
104. Fallah-Araghi, A.; Baret, J.C.; Ryckelynck, M.; Griffiths, A.D. A completely in vitro ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution. *Lab Chip* **2012**, *12*, 882–891. [CrossRef] [PubMed]
105. Romero, P.A.; Tran, T.M.; Abate, A.R. Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 7159–7164. [CrossRef] [PubMed]
106. Zinchenko, A.; Devenish, S.R.; Kintses, B.; Colin, P.Y.; Fischlechner, M.; Hollfelder, F. One in a million: Flow cytometric sorting of single cell-lysate assays in monodisperse picolitre double emulsion droplets for directed evolution. *Anal. Chem.* **2014**, *86*, 2526–2533. [CrossRef] [PubMed]
107. Zhu, Z.; Yang, C.J. Hydrogel droplet microfluidics for high-throughput single molecule/cell analysis. *Acc. Chem. Res.* **2016**, *50*, 22–31. [CrossRef] [PubMed]
108. Fischlechner, M.; Schaerli, Y.; Mohamed, M.F.; Patil, S.; Abell, C.; Hollfelder, F. Evolution of enzyme catalysts caged in biomimetic gel-shell beads. *Nat. Chem.* **2014**, *6*, 791–796. [CrossRef] [PubMed]
109. Baret, J.C. Surfactants in droplet-based microfluidics. *Lab Chip* **2012**, *12*, 422–433. [CrossRef] [PubMed]

110. Gruner, P.; Riechers, B.; Orellana, L.A.C.; Brosseau, Q.; Maes, F.; Beneyton, T.; Pekin, D.; Baret, J.C. Stabilisers for water-in-fluorinated-oil dispersions: Key properties for microfluidic applications. *Curr. Opin. Colloid Interface Sci.* **2015**, *20*, 183–191. [CrossRef]
111. Miller, O.J.; Bernath, K.; Agresti, J.J.; Amitai, G.; Kelly, B.T.; Mastrobattista, E.; Taly, V.; Magdassi, S.; Tawfik, D.S.; Griffiths, A.D. Directed evolution by in vitro compartmentalization. *Nat. Methods* **2006**, *3*, 561–570. [CrossRef] [PubMed]
112. Studer, A.; Hadida, S.; Ferritto, R.; Kim, S.Y.; Jeger, P.; Wipf, P.; Curran, D.P. Fluorous synthesis: A fluorous-phase strategy for improving separation efficiency in organic synthesis. *Science* **1997**, *275*, 823–826. [CrossRef] [PubMed]
113. Lee, J.N.; Park, C.; Whitesides, G.M. Solvent compatibility of poly(dimethylsiloxane)-based microfluidic devices. *Anal. Chem.* **2003**, *75*, 6544–6554. [CrossRef] [PubMed]
114. Lowe, K.C.; Anthony, P.; Wardrop, J.; Davey, M.R.; Power, J.B. Perfluorochemicals and cell biotechnology. *Artif. Cell Blood Substit. Immobil. Biotechnol.* **1997**, *25*, 261–274. [CrossRef]
115. Clausell-Tormos, J.; Lieber, D.; Baret, J.C.; El-Harrak, A.; Miller, O.J.; Frenz, L.; Blouwolff, J.; Humphry, K.J.; Koster, S.; Duan, H.; et al. Droplet-based microfluidic platforms for the encapsulation and screening of mammalian cells and multicellular organisms. *Chem. Biol.* **2008**, *15*, 427–437. [CrossRef] [PubMed]
116. Kaltenbach, M.; Devenish, S.R.; Hollfelder, F. A simple method to evaluate the biochemical compatibility of oil/surfactant mixtures for experiments in microdroplets. *Lab Chip* **2012**, *12*, 4185–4192. [CrossRef] [PubMed]
117. DeJournette, C.J.; Kim, J.; Medlen, H.; Li, X.; Vincent, L.J.; Easley, C.J. Creating biocompatible oil-water interfaces without synthesis: Direct interactions between primary amines and carboxylated perfluorocarbon surfactants. *Anal. Chem.* **2013**, *85*, 10556–10564. [CrossRef] [PubMed]
118. Holtze, C.; Rowat, A.C.; Agresti, J.J.; Hutchison, J.B.; Angile, F.E.; Schmitz, C.H.J.; Koster, S.; Duan, H.; Humphry, K.J.; Scanga, R.A.; et al. Biocompatible surfactants for water-in-fluorocarbon emulsions. *Lab Chip* **2008**, *8*, 1632–1639. [CrossRef] [PubMed]
119. Brouzes, E.; Medkova, M.; Savenelli, N.; Marran, D.; Twardowski, M.; Hutchison, J.B.; Rothberg, J.M.; Link, D.R.; Perrimon, N.; Samuels, M.L. Droplet microfluidic technology for single-cell high-throughput screening. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 14195–14200. [CrossRef] [PubMed]
120. Kreutz, J.E.; Li, L.; Roach, L.S.; Hatakeyama, T.; Ismagilov, R.F. Laterally mobile, functionalized self-assembled monolayers at the fluorous-aqueous interface in a plug-based microfluidic system: Characterization and testing with membrane protein crystallization. *J. Am. Chem. Soc.* **2009**, *131*, 6042–6043. [CrossRef] [PubMed]
121. Platzman, I.; Janiesch, J.W.; Spatz, J.P. Synthesis of nanostructured and biofunctionalized water-in-oil droplets as tools for homing T cells. *J. Am. Chem. Soc.* **2013**, *135*, 3339–3342. [CrossRef] [PubMed]
122. Wagner, O.; Thiele, J.; Weinhart, M.; Mazutis, L.; Weitz, D.A.; Huck, W.T.; Haag, R. Biocompatible fluorinated polyglycerols for droplet microfluidics as an alternative to peg-based copolymer surfactants. *Lab Chip* **2016**, *16*, 65–69. [CrossRef] [PubMed]
123. Ursuegui, S.; Mosser, M.; Wagner, A. Copper-free click chemistry for microdroplet's w/o interface engineering. *RSC Adv.* **2016**, *6*, 94942–94948. [CrossRef]
124. Pays, K.; Giermanska-Kahn, J.; Pouligny, B.; Bibette, J.; Leal-Calderon, F. Double emulsions: How does release occur? *J. Control. Release* **2002**, *79*, 193–205. [CrossRef]
125. Hai, M.; Magdassi, S. Investigation on the release of fluorescent markers from w/o/w emulsions by fluorescence-activated cell sorter. *J. Control. Release* **2004**, *96*, 393–402. [CrossRef] [PubMed]
126. Woronoff, G.; El Harrak, A.; Mayot, E.; Schicke, O.; Miller, O.J.; Soumillion, P.; Griffiths, A.D.; Ryckelynck, M. New generation of amino coumarin methyl sulfonate-based fluorogenic substrates for amidase assays in droplet-based microfluidic applications. *Anal. Chem.* **2011**, *83*, 2852–2857. [CrossRef] [PubMed]
127. Skhiri, Y.; Gruner, P.; Semin, B.; Brosseau, Q.; Pekin, D.; Mazutis, L.; Goust, V.; Kleinschmidt, F.; El Harrak, A.; Hutchison, J.B.; et al. Dynamics of molecular transport by surfactants in emulsions. *Soft Matter* **2012**, *8*, 10618–10627. [CrossRef]
128. Chen, Y.; Wijaya Gani, A.; Tang, S.K. Characterization of sensitivity and specificity in leaky droplet-based assays. *Lab Chip* **2012**, *12*, 5093–5103. [CrossRef] [PubMed]
129. Gruner, P.; Riechers, B.; Semin, B.; Lim, J.; Johnston, A.; Short, K.; Baret, J.C. Controlling molecular transport in minimal emulsions. *Nat. Commun.* **2016**, *7*, 10392. [CrossRef] [PubMed]

130. Janiesch, J.W.; Weiss, M.; Kannenberg, G.; Hannabuss, J.; Surrey, T.; Platzman, I.; Spatz, J.P. Key factors for stable retention of fluorophores and labeled biomolecules in droplet-based microfluidics. *Anal. Chem.* **2015**, *87*, 2063–2067. [CrossRef] [PubMed]
131. Najah, M.; Mayot, E.; Mahendra-Wijaya, I.P.; Griffiths, A.D.; Ladame, S.; Drevelle, A. New glycosidase substrates for droplet-based microfluidic screening. *Anal. Chem.* **2013**, *85*, 9807–9814. [CrossRef] [PubMed]
132. Scheler, O.; Kaminski, T.S.; Ruszczak, A.; Garstecki, P. Dodecylresorufin (C12R) outperforms resorufin in microdroplet bacterial assays. *ACS Appl. Mater. Interfaces* **2016**, *8*, 11318–11325. [CrossRef] [PubMed]
133. Sandoz, P.A.; Chung, A.J.; Weaver, W.M.; Di Carlo, D. Sugar additives improve signal fidelity for implementing two-phase resorufin-based enzyme immunoassays. *Langmuir* **2014**, *30*, 6637–6643. [CrossRef] [PubMed]
134. Courtois, F.; Olguin, L.F.; Whyte, G.; Theberge, A.B.; Huck, W.T.; Hollfelder, F.; Abell, C. Controlling the retention of small molecules in emulsion microdroplets for use in cell-based assays. *Anal. Chem.* **2009**, *81*, 3008–3016. [CrossRef] [PubMed]
135. Pan, M.; Rosenfeld, L.; Kim, M.; Xu, M.; Lin, E.; Derda, R.; Tang, S.K. Fluorinated pickering emulsions impede interfacial transport and form rigid interface for the growth of anchorage-dependent cells. *ACS Appl. Mater. Interfaces* **2014**, *6*, 21446–21453. [CrossRef] [PubMed]
136. Pan, M.; Lyu, F.; Tang, S.K. Fluorinated pickering emulsions with nonadsorbing interfaces for droplet-based enzymatic assays. *Anal. Chem.* **2015**, *87*, 7938–7943. [CrossRef] [PubMed]
137. Agresti, J.J.; Antipov, E.; Abate, A.R.; Ahn, K.; Rowat, A.C.; Baret, J.C.; Marquez, M.; Klibanov, A.M.; Griffiths, A.D.; Weitz, D.A. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4004–4009. [CrossRef] [PubMed]
138. Huebner, A.; Srisa-Art, M.; Holt, D.; Abell, C.; Hollfelder, F.; deMello, A.J.; Edel, J.B. Quantitative detection of protein expression in single cells using droplet microfluidics. *Chem. Commun.* **2007**, 1218–1220. [CrossRef] [PubMed]
139. Larsen, A.C.; Dunn, M.R.; Hatch, A.; Sau, S.P.; Youngbull, C.; Chaput, J.C. A general strategy for expanding polymerase function by droplet microfluidics. *Nat. Commun.* **2016**, *7*, 11235. [CrossRef] [PubMed]
140. Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P.R.; Baker, D.; Griffiths, A.D.; Hilvert, D. Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **2017**, *9*, 50–56. [CrossRef] [PubMed]
141. De Lange, N.; Tran, T.M.; Abate, A.R. Electrical lysis of cells for detergent-free droplet assays. *Biomicrofluidics* **2016**, *10*, 024114. [CrossRef] [PubMed]
142. Najah, M.; Griffiths, A.D.; Ryckelynck, M. Teaching single-cell digital analysis using droplet-based microfluidics. *Anal. Chem.* **2012**, *84*, 1202–1209. [CrossRef] [PubMed]
143. Ostafe, R.; Prodanovic, R.; Lloyd Ung, W.; Weitz, D.A.; Fischer, R. A high-throughput cellulase screening system based on droplet microfluidics. *Biomicrofluidics* **2014**, *8*, 041102. [CrossRef] [PubMed]
144. Beneyton, T.; Wijaya, I.P.; Postros, P.; Najah, M.; Leblond, P.; Couvent, A.; Mayot, E.; Griffiths, A.D.; Drevelle, A. High-throughput screening of filamentous fungi using nanoliter-range droplet-based microfluidics. *Sci. Rep.* **2016**, *6*, 27223. [CrossRef] [PubMed]
145. Sjostrom, S.L.; Bai, Y.; Huang, M.; Liu, Z.; Nielsen, J.; Joensson, H.N.; Andersson Svahn, H. High-throughput screening for industrial enzyme production hosts by droplet microfluidics. *Lab Chip* **2014**, *14*, 806–813. [CrossRef] [PubMed]
146. Huang, M.; Bai, Y.; Sjostrom, S.L.; Hallstrom, B.M.; Liu, Z.; Petranovic, D.; Uhlen, M.; Joensson, H.N.; Andersson-Svahn, H.; Nielsen, J. Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E4689–E4696. [CrossRef] [PubMed]
147. Colin, P.Y.; Kintses, B.; Gielen, F.; Miton, C.M.; Fischer, G.; Mohamed, M.F.; Hyvonen, M.; Morgavi, D.P.; Janssen, D.B.; Hollfelder, F. Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat. Commun.* **2015**, *6*, 10008. [CrossRef] [PubMed]
148. Ryckelynck, M.; Baudrey, S.; Rick, C.; Marin, A.; Coldren, F.; Westhof, E.; Griffiths, A.D. Using droplet-based microfluidics to improve the catalytic properties of rna under multiple-turnover conditions. *RNA* **2015**, *21*, 458–469. [CrossRef] [PubMed]
149. Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. Cell-free translation reconstituted with purified components. *Nat Biotechnol.* **2001**, *19*, 751–755. [CrossRef] [PubMed]

150. Dittrich, P.S.; Jahnz, M.; Schwille, P. A new embedded process for compartmentalized cell-free protein expression and on-line detection in microfluidic devices. *Chembiochem Eur. J. Chem. Biol.* **2005**, *6*, 811–814. [CrossRef] [PubMed]
151. Kiss, M.M.; Ortoleva-Donnelly, L.; Beer, N.R.; Warner, J.; Bailey, C.G.; Colston, B.W.; Rothberg, J.M.; Link, D.R.; Leamon, J.H. High-throughput quantitative polymerase chain reaction in picoliter droplets. *Anal. Chem.* **2008**, *80*, 8975–8981. [CrossRef] [PubMed]
152. Woronoff, G.; Ryckelynck, M.; Wessel, J.; Schicke, O.; Griffiths, A.D.; Soumillion, P. Activity-fed translation (AFT) assay: A new high-throughput screening strategy for enzymes in droplets. *Chembiochem Eur. J. Chem. Biol.* **2015**, *16*, 1343–1349. [CrossRef] [PubMed]
153. Schaerli, Y.; Wootton, R.C.; Robinson, T.; Stein, V.; Dunsby, C.; Neil, M.A.; French, P.M.; Demello, A.J.; Abell, C.; Hollfelder, F. Continuous-flow polymerase chain reaction of single-copy DNA in microfluidic microdroplets. *Anal. Chem.* **2009**, *81*, 302–306. [CrossRef] [PubMed]
154. Autour, A.; Westhof, E.; Ryckelynck, M. Ispinach: A fluorogenic RNA aptamer optimized for in vitro applications. *Nucleic Acids Res.* **2016**, *44*, 2491–2500. [CrossRef] [PubMed]
155. Vartoukian, S.R.; Palmer, R.M.; Wade, W.G. Strategies for culture of 'unculturable' bacteria. *FEMS Microbiol. Lett.* **2010**, *309*, 1–7. [CrossRef] [PubMed]
156. Ekkers, D.M.; Cretoiu, M.S.; Kielak, A.M.; Elsas, J.D. The great screen anomaly—A new frontier in product discovery through functional metagenomics. *Appl. Microbiol. Biotechnol.* **2012**, *93*, 1005–1020. [CrossRef] [PubMed]
157. Najah, M.; Calbrix, R.; Mahendra-Wijaya, I.P.; Beneyton, T.; Griffiths, A.D.; Drevelle, A. Droplet-based microfluidics platform for ultra-high-throughput bioprospecting of cellulolytic microorganisms. *Chem. Biol.* **2014**, *21*, 1722–1732. [CrossRef] [PubMed]

© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

3.2.3.3. Evolving RNA using µIVC

As stated in the **Review Article 2**, µIVC has mainly been applied to proteins. However, RNA can also be evolved using similar procedure and our laboratory is actually specialized in such activity. Conversely to protein enzyme, RNA cannot be conveniently expressed into bacteria because of the strong bacterial RNase activity. Instead, the ease of expressing RNA-coding genes *in vitro* makes *in vitro* transcription-based approach more appealing for RNA development.

The first RNA evolved in the lab was the catalytic RNA X-motif. X-motif ribozyme was initially isolated by SELEX due to its very efficient self-cleaving activity. It is therefore very efficient at catalyzing a single turn-over reaction. However, when converted into a bimolecular trans-cleaving ribozyme, the molecule revealed far less efficient, mainly because of a strong product inhibition (Ryckelynck et al., 2015). Trans-cleaving X-motif represented therefore a good model for a proof-of-concept experiment aiming at demonstrating the efficiency of µIVC to evolve RNAs.

After having developed a fluorogenic assay of X-motif activity and prepared mutant libraries of the ribozyme, our group set-up a 4-step µIVC screening workflow (Figure 21) consisting of: i) the individualization and amplification of mutant genes (ddPCR), ii) the addition of *in vitro* transcription (IVT) mixture to each PCR droplet by droplet fusion to transcribe X-motif mutant genes into RNA, iii) the addition of X-motif fluorogenic substrate by pico-injection and iv) finally the fluorescence-activated droplet sorting of droplets of interest. Iterative repetitions of the process with gradual increase of the selection pressure as well as additional mutagenesis steps finally allowed isolating new improved variants of the ribozyme with an almost 30-fold higher catalytic activity and that no longer suffered from the original product inhibition (Ryckelynck et al., 2015). This work demonstrated for the first time how efficient µIVC can be at evolving RNAs and paved the way for the thesis project I then developed.

4. Thesis objectives

As stated in this introduction (Chapter 1), RNA is a key molecule in gene expression and its regulation. Therefore, being able to monitor RNA by live-cell imaging would represent an important step toward a better understanding of gene expression regulation by making possible to monitor regulatory events involving non-coding RNA in a dynamic way and with single-cell resolution. As stated in Chapter 2, RNA-based fluorogenic modules are extremely promising tools to reach this goal. At the time this thesis project was started, two light-up aptamers (Spinach and Mango) displayed attractive properties but suffered from a limited brightness. Since previous work of the group demonstrated the possibility to evolve RNA using μIVC by selecting droplets (so genes) based on their fluorogenic capacity, this technology appeared to be well suited to evolve fluorogenic RNA as well.

The first goal of this thesis was therefore to adapt the μIVC procedure to the selection of efficient light-up RNA aptamers. Next, I used the technology to improve the properties (especially the brightness) of both Spinach and Mango aptamers to develop a new generation of brighter RNA-based fluorogenic modules. Finally, coupling μIVC with NGS I further developed the technology into a more integrated approach in which RNAs are selected in ultrahigh-throughput regime by μIVC prior to identifying the best variants by an automated bioinformatic process following NGS analysis. Using this approach, I was able to identify new efficient biosensor of metabolites.

To summarize the main objective of this thesis was to establish robust μIVC screening workflows for the discovery of novel efficient light-up RNA aptamers as well as metabolites sensors.

Results & discussion:

5. Improvement, characterization and application of Spinach-derived aptamers

6. Improvement, characterization and application of Mango RNA aptamers

II. Results & discussion

RNA-based fluorogenic modules are promising RNA imaging tools for characterizing gene expression regulation (especially those mechanisms involving non-coding RNAs) with single live cell resolution. Among the different fluorogen/light-up RNA aptamer couples described in the literature (Bouhedda et al., 2017), DFHBI/Spinach and TO1/Mango modules are particularly attractive since i) both fluorogens are cell-permeable and non-toxic, ii) they displays a very low background fluorescence in their free state and their fluorescence is not triggered by cell compounds, iii) aptamers are of small size and recognize their cognate fluorogen respectively with high (K_D ~ 500 nM) and very high affinity (K_D ~ 5 nM). Nevertheless, both original modules are also characterized by a sub-optimal brightness. Consequently, they both represented perfect targets to be improved using a μIVC procedure similar to that previously used by the group to improve ribozyme activities (Ryckelynck et al., 2015).

In the work presented hereafter, I will show how, adapting the μIVC screening pipeline to the selection of light-up aptamers, I successfully converted the original Spinach molecule into an improved Spinach (iSpinach) variant that we later structurally characterized prior to further engineering it into an RNA-based allosteric biosensor of a metabolite. Next, using a similar overall strategy, I also identified new variants of the Mango RNA aptamer also endowed with improved brightness and folding properties. These new variants revealed to be highly efficient in monitoring expression of small non-coding RNAs.

5. Improvement, characterization and application of Spinach-derived aptamers

As stated above, the first part of my PhD thesis work was devoted to the development of an improved version of the Spinach RNA aptamer. After a quick reminder on the "story" of the original Spinach aptamer, I will present how I improved the folding properties and the stability of the molecule prior to presenting its crystal structure in complex with its cognate fluorogen (DFHBI). Finally, I will conclude this chapter by introducing a new advance in μIVC screening I developed by coupling μIVC and NGS and that allows identifying optimal communication modules in a semi-automated manner.

5.1. A brief story of Spinach RNA aptamer

In 2011, the team of Samy Jaffrey reported on the synthesis of the fluorogenic compound 3,5-difluoro-4-hydroxybenzylideneimidazolinone (DFHBI), a mimic of the HBI (the fluorophore found in the GFP, see §2.2.1.1.) and that was found to freely diffuse through cell membrane while being harmless for the cell (Paige et al., 2011). In parallel, using SELEX they also isolated Spinach, an RNA aptamer able to specifically recognize the DFHBI and activate its fluorescence by forming a complex displaying the same spectral properties as the eGFP (ex: 482 nm, em: 505 nm) and therefore compatible with usual optical lab equipment. Moreover, this DFHBI/Spinach module was found to be functional in living cells. This capacity to work in living cells together with the commercial availability of dye encouraged many groups worldwide to use Spinach as starting point in the development of new assays and functional biomolecules (Jaffrey, 2017). Nevertheless, the module still suffered serious limitations, mainly on the aptamer side. Indeed, both the folding efficiency and thermal stability of the DFHBI/Spinach complex were limited. For instance, the system kept less than 50% of its fluorescence at 37°C (Strack et al., 2013). Rationally redesigning the molecule led to Spinach2, a substantially improved aptamer displaying a ~ 1.5-fold fluorescence increase but still requiring elevated Mg^{2+} and K^+ concentrations to display optimal performances. This need for K^+ was later explained upon the resolution of the crystal structure of the aptamer in complex with its cognate fluorogen(Huang et al., 2014; Warner et al., 2014). Indeed, these studies showed that Spinach (and its derivatives) is organized around a G-quadruplex; i.e. a stable structure consisting of planar guanine quartet connected to each other *via* Hoogsteen interactions and stacked on top of each other (Figure 26). Such a structure is further stabilized by accommodating a metal ion (preferentially a potassium ion) that intercalates between two layers of G-quartets.

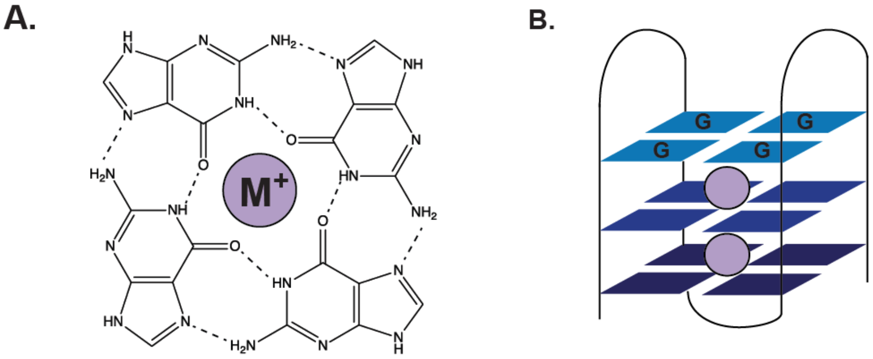


Figure 26: G-quadruplex structural organization. **(A)** G-quartet structure. Four guanines form planar structure link to each other by hydrogen bonding on their Hoogsteen face and are stabilized by metallic ion, preferentially potassium ion. **(B)** Schematic folding of G-quadruplex made of a stacking of three G-quartets. Purple circles represent metal ions.

Spinach was isolated upon 10 rounds of SELEX (Paige et al., 2011). However, as stated in the introduction, the SELEX is primarily a method aiming at isolating ligands and good ligand is not necessarily synonymous with good lighting-up capacity. This was later confirmed by the selection of Broccoli, another DFHBI-binding aptamer sharing strong sequence homologies with Spinach(Filonov et al., 2014). Broccoli was isolated using a mixt SELEX/screening approach, since the starting library was first pre-enriched in DFHBI-binding aptamers by performing 5 rounds of SELEX prior to cloning and expressing it into *E. coli* cells. FACS sorting these cells based on their fluorescence then allowed to isolated Broccoli, which is notably characterized by a low Mg^{2+} dependency (Filonov et al., 2014) but that I found later being poorly efficient in the absence of the tRNA scaffold into which it was originally identified. Moreover, and similarly to other Spinach derivatives, proper RNA folding and module fluorogenic capacity are quite dependent on re-folding protocol used (Ageely et al., 2016; Alam et al., 2017; Filonov et al., 2014; Okuda et al., 2017), suggesting that the folding of all known DFHBI-binding aptamers was still significantly sub-optimal.

5.2. Development of an improved Spinach (iSpinach) RNA aptamer

Considering the apparent sub-optimal folding capacity of Spinach and the other DFHBI-binding aptamers, I designed a μIVC screening strategy aiming at selecting molecules endowed with better fluorogenic capacity while being challenged in their folding efficiency. To achieve this, I replaced potassium ion by sodium to challenge G-quartet stability. Moreover, I warmed the sorting device at 45°C to favor the isolation of the most thermostable complexes. Therefore, selecting the most fluorescent droplets containing sodium as the only monovalent cation and warming the sorting device at 45°C was expected to make possible isolating RNAs with superior folding capacity and thermal stability, with the net effect of forming brighter complexes.

5.2.1. iSpinach development

Applying the screening strategy proposed above to Spinach mutant libraries allowed me to isolate iSpinach, a molecule endowed with the expected superior folding capacity and thermal stability surpassing that of all the previously described DFHBI-binding aptamers. The originality of the screening strategy together with the superior performances of iSpinach enabled this work to get published in Nucleic Acids Research in 2016 as an article (**Original Article 1**) reproduced hereafter.

Original Article 1:

iSpinach: a fluorogenic RNA aptamer optimized for in vitro applications.

Alexis Autour Eric Westhof Michael Ryckelynck

2016

Nucleic Acids Research

Published online 1 March 2016 *Nucleic Acids Research, 2016, Vol. 44, No. 6* ***2491–2500***
doi: 10.1093/nar/gkw083

iSpinach: a fluorogenic RNA aptamer optimized for *in vitro* applications

Alexis Autour, Eric Westhof and Michael Ryckelynck*

Architecture et Réactivité de l'ARN, Université de Strasbourg, Institut de biologie moléculaire et cellulaire du CNRS, 15 rue René Descartes, 67084, Strasbourg, France

Received December 08, 2015; Revised January 29, 2016; Accepted February 01, 2016

ABSTRACT

Using random mutagenesis and high throughput screening by microfluidic-assisted *In Vitro* Compartmentalization, we report the isolation of an order of magnitude times brighter mutants of the light-up RNA aptamers Spinach that are far less salt-sensitive and with a much higher thermal stability than the parent molecule. Further engineering gave iSpinach, a molecule with folding and fluorescence properties surpassing those of all currently known aptamer based on the fluorogenic co-factor 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI). We illustrate the potential of iSpinach in a new sensitive and high throughput-compatible fluorogenic assay that measures co-transcriptionally the catalytic constant (k_{cat}) of a model ribozyme.

INTRODUCTION

The recent development of new fluorogenic dyes and the concomitant isolation of specific RNA aptamers (1–10) offer the possibility to monitor in real-time the synthesis of a specific RNA both *in vitro* and *in vivo* with applications as diverse as live-cell imaging (11–13), biosensing (14,15) or screening (16). Among this new generation of dyes, the 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), a commercially available dye mimicking the natural fluorophore of the green fluorescent protein (17), proved to be particularly well suited for live-cell imaging as it is non toxic, cell membrane-permeable, does not interact with cell components and has a low fluorescence in its free state (18). These attractive properties led to the isolation of a first DFHBI-binding aptamer termed Spinach (6). While Spinach was able to enhance DFHBI fluorescence more than ≈2000 times upon binding, it however suffered from several limitations such as a limited folding efficiency and thermal instability. These limitations were partly overcome in a second version of the aptamer (Spinach2) obtained by rational design (19). Structural studies shed light on the recognition mechanism between the dye and both RNA (20,21) by revealing that both aptamers were built around a G-quadruplex structure and that DFHBI was recognized in a pocket encompassing the top G-quartet, a base triple and a unpaired G residue. The stabilization of this structure requires potassium ions, making the fluorescence of the complex sensitive to the nature of the cation present in the reaction mixture (20,21).

Like every artificial RNA aptamer discovered to date, Spinach was initially isolated by SELEX, a powerful approach to identify specific and high affinity RNA ligand starting from large libraries (22–24) but that does not allow to select for fluorescence phenotypes. Consequently, light-up aptamers are usually identified across a second labor-intensive and low throughput post-selection screening step that allows exploring only a limited fraction of the selected variants. Therefore, as confirmed by a recent study (25), Spinach and its derivatives were likely to be sub-optimal fluorogenic aptamers. This was further supported by a recent work in which FACS-based screening of SELEX-enriched libraries led to the isolation of Broccoli, a DFHBI-binding aptamer optimized for *in vivo* applications with superior stability and fluorescence properties (4).

Beside live-cell imaging applications, several new *in vitro* fluorogenic assays were developed using either the full-length (16,26) or a split version of DFHBI-binding RNA aptamers (27) and Spinach was even used as extracellular metabolite fluorogenic sensor (28). However, despite its great potential, little was known about how well Spinach and its derivatives were adapted to *in vitro* conditions.

In the work presented herein, we show that none of the DFHBI-binding aptamers currently known was optimal for *in vitro* applications since the fluorescence of the complex they generate with DFHBI had a limited thermal stability and strongly depended on the nature of the salt present in the reaction mixture. Using the microfluidic-assisted *In Vitro* Compartmentalization (μIVC) procedure we recently introduced (29) allowed us to screen Spinach gene libraries while applying selection pressures favoring the isolation of aptamers able to work at high temperature and in a potassium free environment. Identified beneficial mutations were combined and the molecule further engineered to afford iSpinach, an aptamer with properties surpassing those

---

*To whom correspondence should be addressed. Tel: +33 3 88 41 70 55; Fax: +33 3 88 60 22 18; Email: m.ryckelynck@ibmc-cnrs.unistra.fr

© The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

of all DFHBI-binding aptamers described so far. Finally, we demonstrate the potent value of this new aptamer by setting-up a sensitive and high throughput-compatible fluorogenic assay able to assess co-transcriptionally the catalytic constant (k_{cat}) of a model ribozyme.

MATERIALS AND METHODS

Gene library generation

The sequence coding for Spinach 1.1 (19) was flanked with constant regions at 5′ (GGAAGACGTAGCAAG) and 3′ ends (CAGAGAGATGACAGAGAACACA) to afford SpiSel. Mutant libraries were generated by error-prone polymerase chain reaction (PCR) by subjecting 0.04 ng of DNA to 4 amplification cycles in the presence of Rev (TGTGTTCTCTGTCATCTCTCTG) and Fwd (GTAAAACGACGGCCAGTGCCAAGCTTGCATGC-TAATACGACTCACTATAGGAAGACGTAGCAAG) primers as well as nucleotide analogues (JBS dNTP-Mutagenesis Kit, Jena Bioscience) as described before (29). 1 ng of PCR products was amplified in a second PCR mixture containing 10 pmol of each primer (Fwd and Rev), 0.2 mM of each dNTPs, 5 U of DreamTaq™ and the corresponding buffer (Fermentas). The mixture was thermocycled starting with an initial step of denaturation of 2 min at 92°C followed by 4 cycles of: 1 min at 92°C, 1 min at 55°C and 5 min at 72°C, and PCR products were purified. The starting library was TA-cloned and the sequence of 26 clones revealed an average mutation rate of 4.23 mutations per variant.

TA-cloning and sequencing

Genes contained in mutant libraries were diluted in a mixture containing 10 pmol of each primer (Fwd and Rev), 0.2 mM of each dNTPs, 5 U of DreamTaq™ and the corresponding buffer (Fermentas). The mixture was then subjected to an initial denaturation step of 3 min at 95°C, followed by 25 cycles of: 30 s at 95°C, 30 s at 60°C and a final extension step of 10 min at 72°C. PCR products were inserted in pTZ57R/T vector following manufacturer's instructions (InsTAclone PCR cloning Kit, Thermo-Scientific). Ligation products were recovered by phenol/chloroform extraction and 100 ng of DNA used to transform Electro-10 blue bacteria (Agilent) placed in 2 mm electroporation (MicroPulser, Bio-Rad). After an hour of recovery at 37°C under agitation, bacteria were plated on Luria broth (LB)-Amplicillin agar plate and incubated overnight at 37°C. The colonies were picked, used to inoculate liquid LB and grown at 37°C until saturation. Plasmids DNA were extracted using 'GenJet Plasmid Miniprep kit' (Thermo-Scientific), and sequences determined by Sanger approach (GATC Biotech).

Microfluidic chip fabrication and manipulation

The description of the microfluidic devices, their fabrication as well as the microfluidic workstation used in this work can be found in (29).

Microfluidic-assisted *In vitro* Compartmentalization screening

i/ Digital droplet PCR. DNA mutant libraries were diluted in 200 μg/ml yeast total RNA solution (Ambion) down to ≈8 template DNA molecules per picoliter. 1 μl of this dilution was then introduced in 100 μl of PCR mixture containing 20 pmol of each primer (Fwd and Rev), 0.2 mM of each dNTPs, 0.67 mg/ml Dextran-Texas Red 70 kDa (Molecular Probes), 0.1% Pluronic F68, 1x EvaGreen (Biotium), 5 U of DreamTaq™ and the corresponding buffer (Fermentas). The mixture was loaded in a length of PTFE tubing and infused into droplet generator microfluidic chip where it was dispersed in 2.5 pl droplets (production rate of ≈12 000 droplets/s) carried by HFE 7500 fluorinated oil (3M) supplemented with 3% of a fluorosurfactant (29). Droplet production frequency was monitored and used to determined droplet volume, and pumps flow rates (MFCS, Fluigent) adjusted to generate 2.5 pl droplets. Emulsions were collected in 0.2 μl tubes as described before (29) and subjected to an initial denaturation step of 1 min at 95°C followed by 30 cycles of: 1 min at 95°C, 1 min at 55°C, 2 min at 72°C. Droplets were then reinjected into a droplet fusion microfluidic device.

ii/ Droplet fusion. PCR droplets were reinjected and spaced into a fusion device at a rate of ≈1500 droplets/s. Measuring the fraction of droplets with an increased EvaGreen fluorescence allowed assessing droplet occupancy. Each PCR droplet was then synchronized with a 16 pl *in vitro* transcription (IVT) droplet containing 2.2 mM of each NTP (Larova), 24 mM $MgCl_2$, 44 mM Tris-HCl pH 8.0, 5 mM DTT, 1 mM Spermidine, 35 μg/ml of Dextran-Texas Red 70 kDa (Molecular Probes), 0.1% Pluronic F68, 20 μg/ml T7 RNA polymerase, 200 μM DFHBI (Lucerna), 1 μg inorganic pyrophosphatase (Roche) and 50 mM NaCl or KCl. IVT mixture was loaded in a length of PTFE tubing kept on ice during all experiment. PCR droplets were spaced and IVT droplets produced using a single stream HFE 7500 fluorinated oil (3M) supplemented with 2% (w/w) of fluorinated. Flow-rates (MFCS, Fluigent) were adjusted to generate 16 pl IVT droplets and maximize synchronization of 1 PCR droplet with 1 IVT droplet. Pairs of droplets were then fused with an AC field (350 V at 30 kHz) and the resulting emulsion collected off-chip and incubated for 2 h at 37°C.

iii/ Droplet analysis and sort. The emulsion was finally reinjected into of analysis and sorting microfluidic device mounted on Thermo plate (Tokai Hit) holding the temperature at 45°C. Prior to starting droplet fluorescence analysis and sort, the temperature of the microfluidic device was allowed to equilibrate at 45°C for 10 min. The proper warming of the chip was controlled with a thermocouple. The micrometric depth of the channels together with the low droplets re-injection flow rate used were expected to allow equilibration of droplets temperature to that of the device prior to analyzing their fluorescence. Droplets were reinjected at a frequency of ≈200 droplets/s and spaced with a stream of surfactant-free HFE 7500 fluorinated oil (3M). The green fluorescence (DFHBI complexed with the aptamer) of each droplet was analyzed and the 1–2% most green fluorescence droplets were sorted. The gated droplets were deflected into collecting channel by applying a 1 ms AC fields (1200 V 30 kHz) and collected into a 1.5 ml tube. Residual droplets were recovered from the collection tubing by flushing 200 μl of HFE fluorinated oil (3M). 100 μl of 1H, 1H, 2H, 2H-perfluoro-1-octanol (Sigma-Aldrich) and 200 μl of 200 μg/ml yeast total RNA solution (Ambion) were then added and droplets broken by vortexing the mixture. DNA-containing aqueous phase was then recovered.

Enrichment measurement and colony screening

Genes contained in 2 μl of DNA-containing aqueous phase from unsorted or sorted fractions were placed in 100 μl of PCR mixture identical to that used for TA-cloning and thermocycled starting with an initial cycle of 30 s at 95°C, followed by 25 cycles of: 5 s at 95°C and 30 s at 60°C. Then 2.5 μl of this PCR product were incubated with 16 μl of IVT mixture identical to that used for the μIVC screening containing 200 μM DFHBI but deprived of dextran-Texas Red and pluronic. The transcription mixture was then incubated at 37°C in a real-time thermocycler (Mx 3005P, Agilent) and the green fluorescence (ex: 492 nm/em: 516 nm) measured every minute for 2 h.

Enriched libraries from the last round of screening in the presence of sodium or potassium were TA-cloned. 90 clones were randomly picked, genes coding for the DFHBI-binding aptamers were PCR-amplified and *in vitro* transcribed as for enrichment measurement. Following an hour of transcription real-time monitoring at 37°C, the mutants with the highest apparent transcription rate were sequenced.

RNA synthesis and purification

Genes coding for aptamers of interest were PCR amplified with the same procedure used for enrichment measurements. 20 μl of PCR products were then *in vitro* transcribed in 230 μl of mixture containing 2 mM of each NTP, 25 mM $MgCl_2$, 40 mM Tris-HCl pH 8.0, 5 mM DTT, 1 mM Spermidine and 70 μg/ml T7 RNA polymerase. After 4 h of incubation at 37°C, 10 units of Baseline-Zero^{TM} DNase (Epicentre) and the corresponding buffer were added to the mixture and a second incubation of 1 h at 37°C performed. RNAs was recovered by phenol extraction followed by an ethanol precipitation in the presence of 300 mM sodium acetate pH 5.5 (Sigma-Aldrich). After centrifugation and a wash in 70% ethanol, the pellets were dissolved in denaturing buffer (0.05% bromophenol blue, 0.05% xylene cyanol, 20% glycerol, TBE 1x, 8M urea) and the solution loaded on a 12% denaturing 8 M urea acrylamide/bisacrylamide gel. The piece of gel containing RNA was identified by UV shadowing, sliced from the gel and transferred in a dialyze tube (MWCO = 3500, Spectrum Lab) filled with TBE. RNA was electro-eluted by placing the montage in TBE for 1h00 at 100 V. Eluted RNA were filtered in centrifuge tube 0.45 μm (VWR) and ethanol precipitated in the presence of 300 mM sodium acetate pH 5.5. After centrifugation and a wash in 70% ethanol, the pellets were dissolved in DEPC-Treated water and quantified with Nanodrop (Thermo Scientific).

Fluorescence measurement of purified RNA

A 2 μM solution of purified RNA was heated for 2 min at 85°C and cooled at 25°C for 5 min. The solution was next mixed with 1 volume of a mixture containing 80 mM Tris-HCl pH 7.5, 2 mM $MgCl_2$, 20 μM DFHBI, 200 mM of salt (KCl, NaCl, LiCl or CsCl). The mixture was then incubated at 25°C for 15 min prior to fluorescence measurement. 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). For T_m^{app} measurement, the temperature was gradually increased from 25 to 60°C with a step of 5°C and a 5 min fluorescence monitoring was performed at each temperature.

Folding efficiency

Folding Efficiency was determined using procedure described in (19). The green fluorescence was measured (ex: 492 nm/em: 516 nm) at three temperatures (25°C, 37°C, 45°C) on a real-time thermocycler (Mx 3005P, Agilent).

Affinity measurements

To measure K_d, the concentration of DFHBI was progressively increased from 100 nM to 200 μM with 100 nM of renatured RNA in 40 mM Tris-HCL pH 7.5, 1 mM $MgCl_2$, 100 mM NaCl or KCl and 2.5% DMSO). The fluorescence was measured (ex: 492 nm/em: 516 nm) at three temperatures (25°C, 37°C, 45°C) on a real-time thermocycler (Mx 3005P, Agilent).

Co-transcriptional k_{cat}^{app} measurement

0.6 pmol of DNA bearing T7 promoter were introduced in 60 μl of 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 and 65 U/μl T7 RNA polymerase (NEB). 200 pmol of DFHBI or 20 pmol of S21-Atto (29) were then added to 20 μl aliquots and the green fluorescence (ex: 492 nm/em: 516 nm) was monitored for 1h00 at 37°C in a real-time thermocycler (Mx 3005P, Agilent). Fluorescence values were converted into pmol of generated product or synthesized RNA using experimentally measured molar fluorescence of S21-Atto and DFHBI, respectively.

Spectral analysis

Absorbance was measured on a UV-Vis spectrophotometer (Agilent Technology) from 350 nm to 500 nm. Emission (from 410 nm to 60 nm) and excitation (from 350 to 490 nm) was recorded by FluoroMax-4 spectrofluorometer (Horiba Scientific). This measurement was realized using an excess of RNA.

RESULTS AND DISCUSSION

Evaluation of DFHBI-binding aptamers performances *in vitro*

We started our study by characterizing the performances of the three DFHBI-binding aptamers described so far:

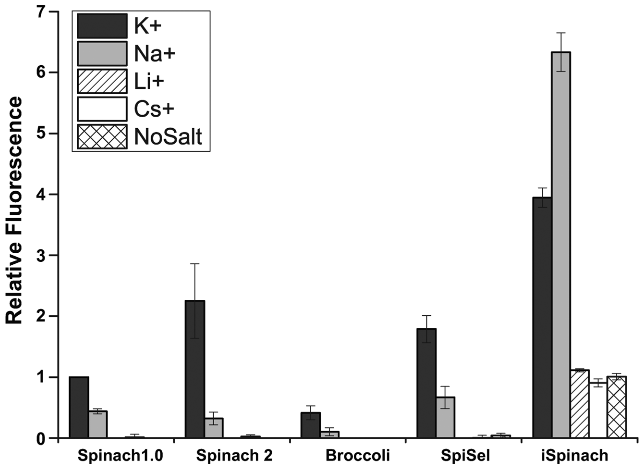


Figure 1. Effect of monovalent cations on the fluorescence of complexes formed between DFHBI and different DFHBI-binding RNA aptamers. 1 μM of RNA was incubated with 10 μM DFHBI in 40 mM Tris-HCl, 1 mM $MgCl_2$ and either 100 mM of salt (K, Na, Li or Cs) or no monovalent cation (no salt). The fluorescence of the complex was measured at 25°C. Values are the mean of three independent experiments and error bars correspond to ± 1 standard error.

Spinach1.0, Spinach2 and Broccoli (4,6,19). To assess unambiguously the intrinsic fluorescence and the folding properties of these molecules, aptamers were used in their minimal version instead of being inserted into a tRNA scaffold (30) as in previously reported (4,19). Coding-genes were *in vitro* transcribed, RNA were gel-purified and renaturated just prior to being used.

First, 10 pmol of aptamer (1 μM) were incubated with 1 nmol of DFHBI in a buffer mimicking *in vivo* conditions (40 mM Tris-HCl pH 7.5, 1 mM MgCl2, 100 mM KCl) and the fluorescence of DFHBI/RNA complexes was measured. As expected from a previous report (19) Spinach2 was twice more fluorescent than Spinach 1.0 in our conditions (Figure 1). Surprisingly, Broccoli, a variant with superior properties *in vivo* (4), generated the dimmest complex with DFHBI *in vitro* (Figure 1). This bad performance was partly explained by a low folding efficiency (F.E. < 18% at 25°C, Supplementary Table S1) of the aptamer in our conditions. A lower F.E. of Broccoli compared to Spinach2 was already suggested in a previous work in which such a strong difference was probably minimized as the aptamer was inserted into a tRNA scaffold that assisted aptamer folding (4). Consequently to this low fluorescence *in vitro*, Broccoli was no longer considered in the rest of the study.

We next investigated the effect of exchanging potassium for another monovalent cation on the fluorescence of DFHBI in complex with Spinach1.0 or 2. Indeed, whereas potassium is the main monovalent cation found *in cellulo* (31), it might not be dominant ion in the extracellular environment and even be absent from *in vitro* reaction mixtures. Using 100 mM NaCl instead of KCl in the reaction mixture led to a significant drop of the fluorescence of both Spinach variants (Figure 1) confirming data previously obtained with Spinach (20). Whereas a residual fluorescence was still observed with sodium, the use of lithium, cesium or the lack of monovalent cation completely abolished the fluorescence of Spinach aptamers (Figure 1). The differential effect of the cations confirmed that proper stabilization of the G-quadruplex motif of the aptamer was necessary for efficient recognition of DFHBI and that the fluorescence of DFHBI/aptamer complex strongly depends on the nature of the monovalent cation present.

Beside the strong dependence of Spinach1.0 and 2 on potassium, both molecules were also reported to have a limited thermal stability as the complexes they generate with DFHBI have melting temperature of ≈34°C and ≈38°C, respectively (19). Accordingly, we found that, in the presence of potassium, the fraction of Spinach2 aptamer properly folded decreased from 61% to 27% at 25°C and 37°C, respectively (Supplementary Table S1).

Overall, these different observations show that none of the DFHBI-binding aptamer described so far was optimized for *in vitro* application (e.g. at 37°C in a reaction mixture not supplemented with potassium). The folding efficiency of the RNA was the main limiting parameter, suggesting that re-evolving the molecule *in vitro* by applying selection pressures favoring the isolation of a thermostable fluorogenic aptamer in a potassium-free environment should lead to the identification of a mutant with properties surpassing those of the rationally designed Spinach2. The use of such conditions that destabilize the molecular folding has already proved to be an efficient strategy as it allowed, for instance, the identification of the Superfolder GFP (32).

Improvement of Spinach properties by *in vitro* evolution

Spinach1.1 was used as starting point of the evolutionary process as its sequence differed from the best variant isolated from SELEX only by a single stabilizing mutation in the basal stem (19). Constant regions were appended at the 5′ of the molecule to ensure a homogeneous transcription initiation and at its 3′ end to enable PCR amplification of the gene. The resulting molecule (SpiSel) was found to have a fluorescence slightly improved (Figure 1). Mutant libraries were then generated by error-prone PCR using a mutation rate of ≈4 mutations per gene and subjected to several rounds of high throughput screening using a procedure of microfluidic-assisted *In Vitro* Compartmentalization (μIVC) we recently developed (29). Even though, this method has a lower throughput than conventional SELEX, it has the great advantage of directly screening libraries for mutants able to generate a fluorescent complex with DFHBI.

Each round of μIVC selection comprised 3 main steps (Figure 2). First, the DNA library was diluted into a PCR mixture down to a concentration of 1 molecule/12.5 pl and the aqueous phase was dispersed into highly monodisperse 2.5 pl droplets carried by a fluorinated oil phase using a first microfluidic device (Figure 2B). The used dilution allowed having, on average, ≈0.2 molecule of DNA per droplet to maximize droplet occupancy (≈18%, Supplementary Table S2) while limiting multiple encapsulation events (<10% of the occupied droplets contained more than 1 gene). Droplets production monitoring was possible by supplementing the PCR mixture with a high concentration of an orange fluorescent dye (670 μg/ml Dextran-Texas

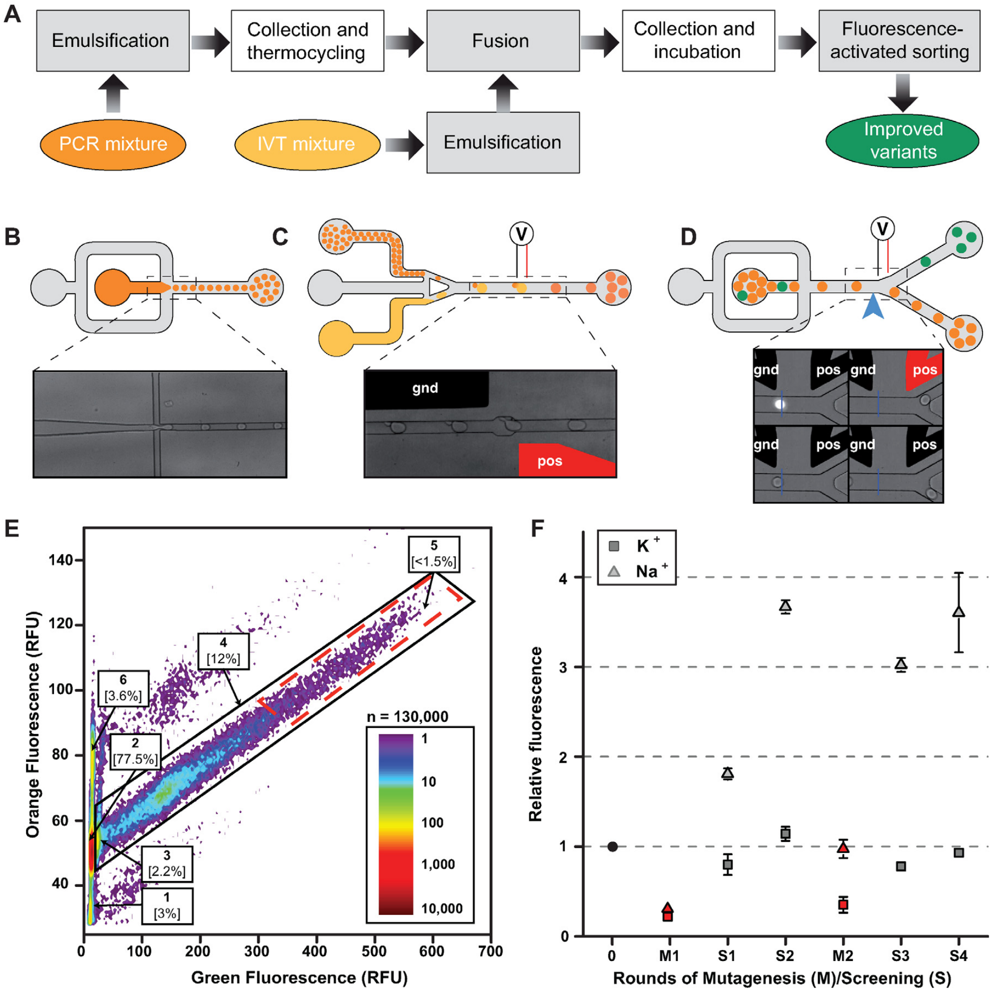


Figure 2. Microfluidic-assisted screening. (**A**) Experimental workflow. Steps performed on-chip (gray boxes) were distinguished from those performed off-chip (white boxes). (**B**) PCR droplets production. Aqueous phase supplemented with a high concentration of a orange fluorescent dye was injected into droplet generator device and 2.5 pl droplets were generated by focusing aqueous (dark orange) and oil (gray) flows. Emulsions were collected and thermocycled. (**C**) Droplets fusion. Small PCR droplets were reinjected into droplet-fusion device and spaced by a stream of oil. 16 pl droplets containing *In Vitro* Transcription (IVT, light orange) mixture supplemented with DFHBI were concomitantly produced and synchronized with PCR droplets. Pairs of droplets were then fused when passing in between a ground-connected electrode (gnd, in black) and an electrode to which tension (pos, in red) was applied. (**D**) Droplets sorting. After incubation, emulsions were reinjected into a Fluorescence Activated Droplet Sorting device and the fluorescence of each droplet read at a detection point (blue arrow and blue line on the micrograph). Based on the fluorescence signal, droplets of interest (green) were deflected into sort channel by applying tension to one of the electrode (pos, in red) whereas non-fluorescent droplets (orange) flowed into the waste channel. (**E**) Typical fluorescence profile of screened emulsion. The analysis of DFHBI green fluorescence and Texas-Red orange fluorescence allowed identifying the different populations composing the emulsion. Indeed, using orange fluorescence signal, IVT droplets fused to single PCR droplets (populations 2–5) were easily discriminated from unfused (population 1) and double fused (population 6) IVT droplets. Green fluorescence resulting from EvaGreen intercalation allowed discriminating droplets containing amplified DNA (population 4) from droplet resulting from fusion with an initially empty PCR droplet (population 2). Finally, the stronger DFHBI green fluorescence allowed discriminating droplets containing non-fluorogenic (population 2) and highly fluorogenic aptamers (population 5, red dashed boxed). Population 5 was gated and corresponding droplets sorted. (**F**) Evolution profiles of SpiSel-derived mutants. Gene libraries obtained after mutagenesis (in red) or screening steps (in gray) were transcribed, the RNAs purified and their fluorogenic properties assayed in the presence of the salt used for the selection. Fluorescence values were normalized to that of SpiSel (black circle) in the same conditions. Screenings performed in potassium (triangles) were distinguished from those performed in sodium (squares). Values are the mean of two independent experiments and error bars correspond to ± 1 standard error.

Red). The emulsion was collected and thermocycled to amplify the DNA prior to reinjecting the droplets into a second microfluidic chip (Figure 2C) where each small droplet was synchronized and fused with an on-chip generated 16 pl droplet containing an *In Vitro* Transcription (IVT) mixture supplemented with 200 μM DFHBI, 50 mM NaCl and a low concentration of the orange fluorescent dye (35 μg/ml Dextran-Texas Red). The new emulsion was collected and incubated. The different concentrations of orange fluorescent dye in PCR and IVT droplets allowed discriminating unfused IVT droplets (low orange fluorescent population 1 on Figure 2E) from IVT droplets fused with 1 (populations 2–5 on Figure 2E) or 2 PCR droplets (population 6 on Figure 2E) and to adjust the flow parameters to maximize the fusion of 1 IVT droplet to 1 PCR droplet (70–92%, Supplementary Table S2). On another hand, the use of sodium was expected to weaken the G-quadruplex structure while maintaining a residual fluorescence allowing for detecting positive droplets. This was used as a selection pressure favoring the isolation of mutants with improved folding properties. Finally, the emulsion was reinjected into a droplet sorting microfluidic device (Figure 2D) warmed at 45°C, a selection pressure expected to favor the isolation of RNA with improved thermal stability. The orange and green fluorescence of each droplet was measured and the 1% most green fluorescent droplets of the single fused population (population 5 on Figure 2E) were recovered at the end of each round (Supplementary Table S2). To summarize, three selection pressures (use of sodium, sorting device warming and selection of the DFHBI/aptamer complexes based on their brightness) were simultaneously applied to favor the isolation of DFHBI-binding aptamers able to fold efficiently and to form a thermostable fluorescent complex with DFHBI.

A total of two pairs of rounds of selection, interspersed by a round of mutagenesis to maintain genetic diversity, were performed and allowed increasing the average molar fluorescence of the population ≈3.5 times (gray triangles on Figure 2F) with respect to the starting molecule (SpiSel). Interestingly, performing the same *in vitro* evolution experiment using KCl instead of NaCl did not significantly improved the fluorescence of the population after 4 rounds of selection (gray squares on Figure 2F), confirming that sodium was an efficient selection pressure. The enriched library from the fourth round of selection in sodium was then cloned and the fluorescence of 85 individual mutants was analyzed (Supplementary Figure S1). More than 95% of the tested clones were more fluorescent than SpiSel. Whereas the majority of these clones were 2–6 times more fluorescent than SpiSel, the variant 4–68 was surprisingly ≈10 times more fluorescent (Supplementary Figure S1). Sequence analysis (Supplementary Table S3 and Figure S2) revealed that 4–68 differed from SpiSel by 8 mutations (Figure 3A).

Miniaturization and engineering of the improved Spinach

Prior to individually evaluating the contribution of each mutation, we defined the minimal size of 4–68. Indeed, in a previous study Spinach was successfully shortened down to a 51 nucleotides long molecule (BabySpinach) that retained fluorogenic properties (21).

Sequence alignment of the most fluorescent variants revealed that a significant variability (especially deletions) was found in J1/2 (Supplementary Table S3) region indicating a dispensable role of this domain. Accordingly, 4–68-J1/2, a variant deleted of J1/2 (Figure 3B), conserved intact fluorogenic properties (Figure 3E). Conversely, the G-quadruplex domain was strongly conserved and, within the sampled population, only one mutation (A42U) was tolerated (mutant 49 in Supplementary Table S3). Transplanting this mutation into 4–68-J1/2 afforded the variant 4–68/49 (Figure 3C) that conserved identical fluorogenic properties in potassium-containing buffer but had a fluorescence increased ≈3 times in sodium-containing buffer (Figure 3E). This improvement was likely to result from a higher stabilization of the G-quadruplexes in sodium and at 37°C when U was present at this position. Interestingly, a recent thermodynamic-based scanning of Spinach, also identified this mutation as the only one tolerated in this region of the molecule, further supporting our results (25).

The aptamer miniaturization process was pursued by removing 6 base pairs from P3 region to generate a shorter and continuous apical helix (mutant min68). In addition, the first 3 or 6 base pairs of P1 were deleted to generate respectively variants min68/49–3 and min68/49–6 (Figure 3D). While the fluorescence of both constructs remained identical to that of 4–68/49 in the absence of constant regions (Figure 3F), their presence drastically reduced the fluorescence of min68/49–6. Conversely, the fluorescence of the longer variant min68/49–3 was unaffected by the constant regions. Even though more experiments are necessary, it is tempting to propose that the long P1/2 helix of min68/49–3 could act as an insulation module, similar to a three-way junction recently engineered (33), that protects the aptamer from the surrounding sequences.

In addition to A17U (originally A42U), our miniaturized variant (min68/49–3) still contained 6 of the 8 original mutations. U8C and U39C were likely to act by stabilizing RNA structure by converting G.U in stronger G-C base pairs. Mutations A14G and A56G were always found associated together and were the only common feature shared by the two most active mutants (clone 68 and 34 in Supplementary Table S3). To test the contribution of this pair of mutations, we produced a double mutant G14A/G56A and measured its capacity to form a fluorescent complex with DFHBI. Whereas reverting these mutations had only a small effect in the presence of potassium, they drastically reduced the fluorescence in other salts pointing a key role of these mutations in salt resistance of the molecule (Figure 3G). Their location may suggest that these mutations could form a third G-quadruplex with surrounding Gs and, doing so, increase the overall stability of the aptamer. To shed light on the contribution of the two last mutations (C31U and U32C) found in the apical loop of the aptamer, we generated a second double mutant (U31C/C32U) and tested its capacity to form fluorescent complex with DFHBI. Surprisingly, reverting these two positions was found to increase the fluorescence of the complex by ≈25% whereas its salt resistance stayed unchanged (Figure 3G). We considered this double mutant, (U31C/C32U) of min68/49–3, to be the optimal form of the aptamer that could be isolated by our approach and we named it iSpinach.

Performances and applications of iSpinach

We next compared the performances of iSpinach and Spinach2 (Figure 4A). We found that, in addition to being 1.3 times smaller then Spinach2, iSpinach was 1.4 times more fluorescent in potassium (Figure 1). This improvement was likely a combined effect of slightly better folding ability of iSpinach (F.E. ≈70%), a twice-higher affinity for the DFHBI and a slightly increased molar fluorescence of the DFHBI in complex with iSpinach (Figure 4B). We noticed that K_d values obtained with Spinach2 were slightly higher than those from the literature. We attributed this difference to the temperature (25°C) used in our measurements. Indeed, we found that K_d values increased with the temperature (Supplementary Figure S3). However, since

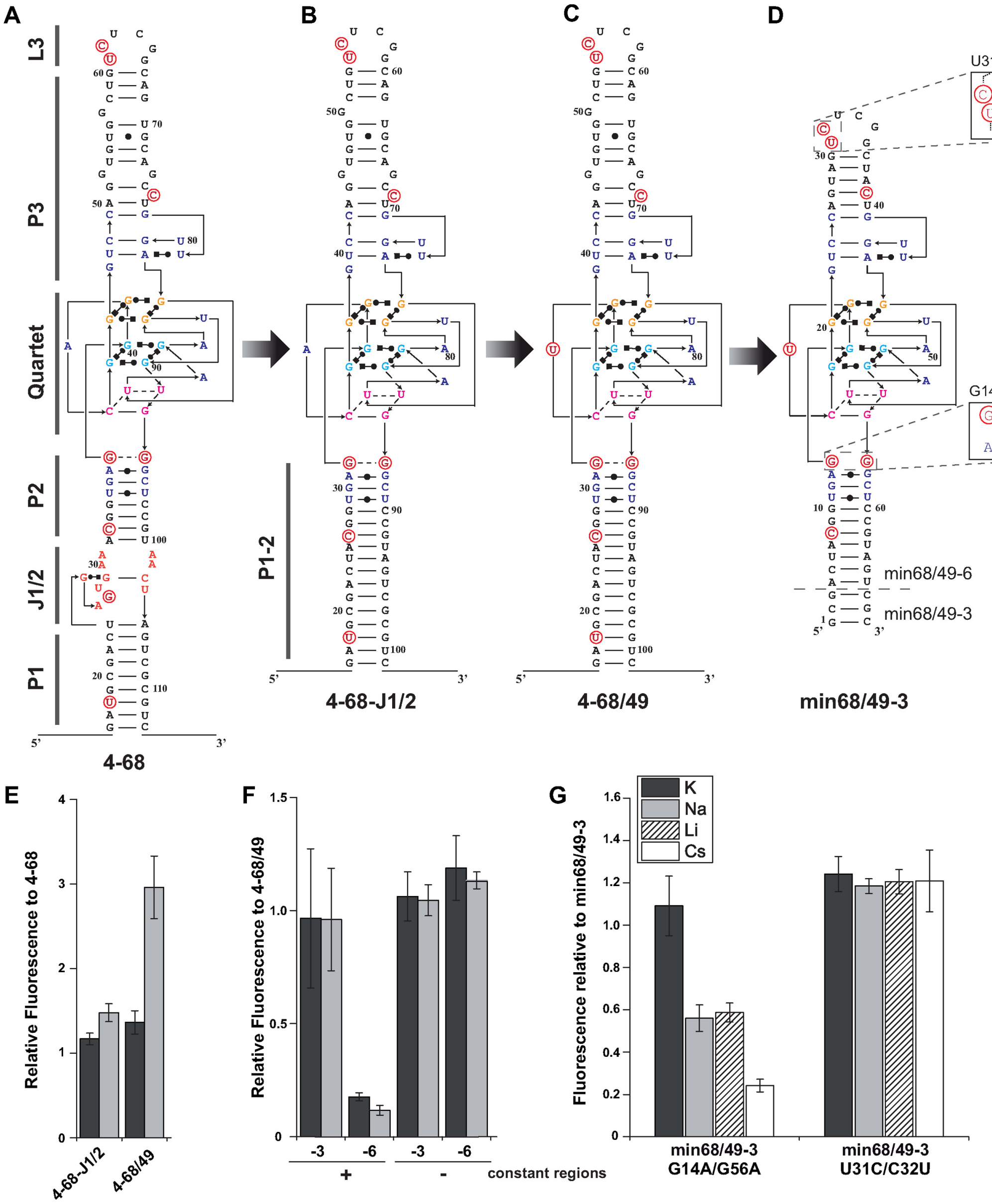
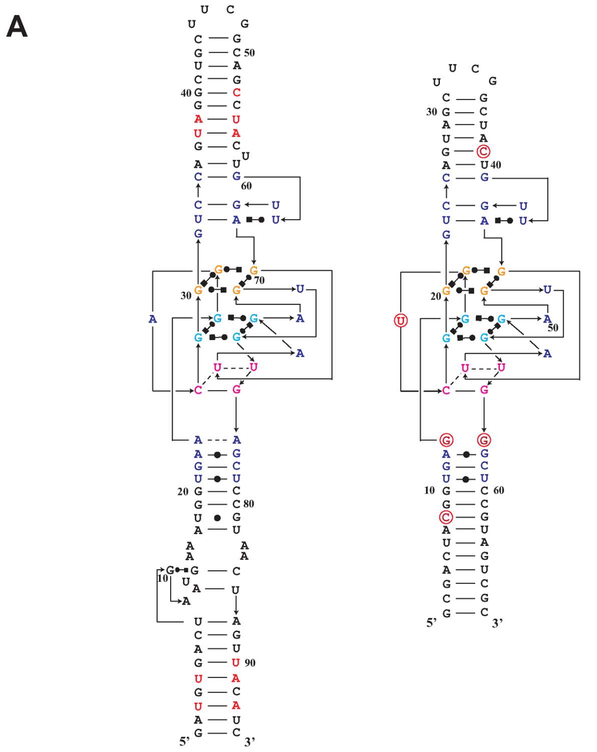


Figure 3. Engineering of improved aptamers. (**A**–**D**) Secondary structure models of aptamers. Molecules are represented according to the model introduced in(21). Mutations are shown in red and numbering adapted to molecule length. (**E**) Effect of J1/2 deletion and A42U mutation on the fluorescence of aptamer 4–68. (**F**) Effect of min68/49 miniaturization on complex fluorescence in the presence or absence of constant regions. Mutant min68/49 was shortened by 3 (−3) or 6 (−6) base pairs. The fluorescence of constructs was tested in the presence and absence of constant regions. (**G**) Effect of double mutations on min68/49–3 fluorescence. The nature of the monovalent cation used is indicated by the color code shown in the inset. Assays were performed with 1 μM RNA and 10 μM DFHBI at 37°C. Values are the mean of three independent experiments and error bars correspond to ± 1 standard error.

none of the previous studies indicated the temperature used to measure this constant, more accurate comparison could not be performed. Finally, iSpinach also formed a more thermostable complex with DFHBI displaying a melting temperature $\left(T_m^{cplx}\right)$ increased by more than 8°C (Figure 4B). iSpinach also outperformed Spinach2 in the presence of lithium and cesium, but we were more surprised to find that the iSpinach/DFHBI complex was not only stable in the presence of sodium but even twice more fluorescent than with potassium (Figure 1). This impressive improvement of the molecule resulted from a better capacity of the aptamer to fold (F.E. ≈ 60%) as compared with Spinach2 leading to an affinity for DFHBI increased by at least an order of magnitude, as well as a doubling of the molar fluorescence of DFHBI (Figure 4B). At this stage, only speculations can be proposed to explain how the nature of the cation affects the brightness of the complex.

The improved folding and fluorescence properties of iSpinach in a wider range of conditions should make it appealing for the design of novel fluorogenic probes as well as of sensitive fluorogenic assays. To demonstrate the advantages that iSpinach may offer, we set-up an *in vitro* experi-



B

	Spinach 2.0		iSpinach	
	K^+	Na^+	K^+	Na^+
F.E. (%)	61.0 ±12.2	n.d	69.0 ±4.0	59.6 ±8.1
T_m^{Cplx} (°C)	36.2 ±0.2	30.6 ±0.3	44.6 ±0.2	36.3 ±0.10
K_d (μM)	1.45 ± 0.2	>53.4 ± 7	0.92 ±0.1	4.46 ±0.2
Max ex. (nm)	445	n.d	442	458
Max em. (nm)	501	n.d	503	507
Rel. Molar Fluo. of DFHBI	1	n.d	1.4 ±0.1	2.1 ± 0.2

Figure 4. Comparison of Spinach2 and iSpinach. (**A**) Secondary structure models. RNAs were represented according to the model introduced in (21). Mutations are shown in red and numbering adapted to molecule length. (**B**) Aptamers properties. Main parameters including Folding Efficiency (F.E.), melting temperature of the complex $\left(T_m^{cplx}\right)$ and dissociation constant (K_d) of each RNA as well as the Relative Molar Fluorescence of DFHBI were determined at 25°C. Values are the mean of three independent experiments and error bars correspond to ± 1 standard error. Excitation/emission spectra are shown on Supplementary Figure S4.

ment aiming at measuring the apparent catalytic constant (k_{cat}^{app}) of a ribozyme directly from the *in vitro* transcription mixture. As model ribozyme, we choose iXm1, a variant of X-motif we recently isolated and which transforms its substrate through a simple first order kinetics (29). We fused the DNA coding either for iSpinach, or for Spinach2, to the 3′ end of iXm1 coding region and placed the construct under the control of T7 RNA polymerase promoter sequence (Figure 5A). The resulting gene was then *in vitro* transcribed in a mixture containing a low concentration of T7 RNA polymerase (50 times less than for expression in droplets) to limit the amount of ribozyme generated in the reaction mixture and properly assess k_{cat}^{app} values. In addition, since X-motif was shown not to require monovalent cations (34), the reaction mixture was not supplemented with extra monovalent cations. Upon DNA addition, the reaction mixture was immediately split and mixed with either DFHBI or S21-Atto (a fluorogenic substrate of iXm1(29)) and the fluorescence monitored at 37°C. Ribozyme activity was readily measurable and was not affected by the fusion of the catalyst with fluorogenic aptamers (Figure 5B). Conversely, DFHBI fluorescence was observed only when iSpinach was used alone or fused with iXm1, whereas no fluorescence was observed in the presence of Spinach2 (Figure 5C). The lack of fluorescence of Spinach2-based constructs was attributed to the poor folding performance of the Spinach2 in the experimental conditions used (i.e. potassium-free reaction mixture incubated at 37°C). Spinach2 folding inefficiency could be partly overcome by increasing the potassium concentration up to 100 mM (Figure 1) (20), a concentration of monovalent cations also shown to strongly inhibit T7 RNA polymerase activity (35). Therefore, in addition of being brighter, iSpinach allows to perform experiments in conditions optimal for the activity of the polymerase. The co-transcriptional kinetics of iXm1 activity could be approximated by Equation (1):

$$\frac{dP}{dt} = k_{cat}^{app}\left(\sigma^{rbz} - \delta^{rbz}\right) + k_{uncat} \qquad (1)$$

with σ^{rbz} the synthesis rate of the ribozyme, δ^{rbz} its degradation rate and k_{uncat} the substrate auto-hydrolysis rate resulting form T7 RNA polymerase side activity(29). Considering the high efficiency of T7 RNA polymerase, δ^{rbz} was neglected, whereas σ^{rbz} and k_{uncat} were experimentally measured to be 2.1 pmol/min and 0.17 pmol/min respectively and allowed extracting a k_{cat}^{app} value of 0.36 min^{-1}, a value that is close to the 0.48 min^{-1} previously measured with gel-purified RNA (29). This fluorogenic approach can be used to screen ribozyme mutant libraries in high throughput regimes for identifying molecules with high k_{cat}^{app}, but it also demonstrate how the performances of iSpinach is advantageous. In addition, changing the dye labeling the fluorogenic substrate of X-motif for a red emitting one would allow for simultaneously monitoring RNA synthesis (*via* iSpinach fluorescence) and ribozyme activity in a single well.

The present work allowed us to develop iSpinach, a DFHBI-binding aptamer with properties (thermal stability and brightness) surpassing those of Spinach2 and all other DFHBI-binding aptamers in the conditions tested. Isolating this new optimized aptamer was possible by using selection pressures difficult, if not impossible, to apply on living cells (i.e. high temperature, potassium-free environment) but easy to implement *in vitro* using μIVC. In addition, evolving fluorogenic aptamers *in vivo* faces cell-size

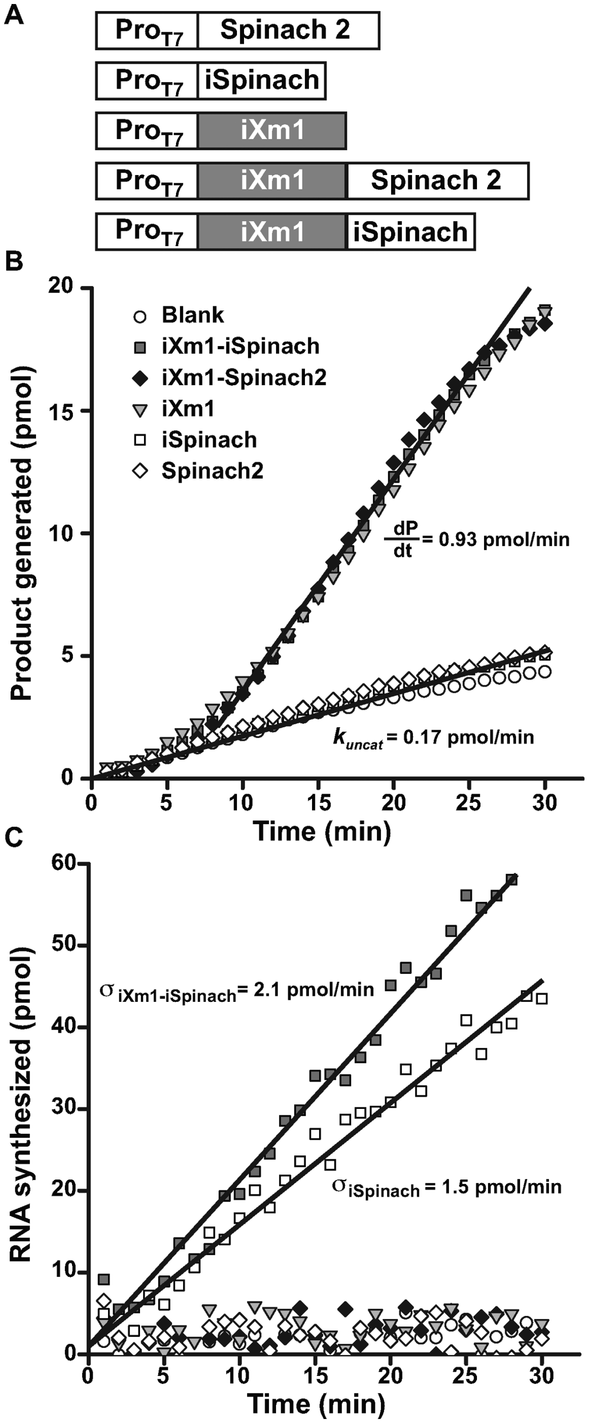


Figure 5. Real-time monitoring of RNA synthesis and ribozyme activity. (**A**) Constructs used. DNA fragments coding for iSpinach, Spinach2 or iXm1 ribozyme were combined in different ways and placed under the control of T7 RNA polymerase promoter sequence (Pro_{T7}). (**B**) Real-time ribozyme activity monitoring. The different constructs were *in vitro* transcribed in the presence of ribozyme fluorogenic substrate (S21-Atto) and the fluorescence monitored at 37°C. Product generation rate $\left(\frac{dP}{dt}\right)$ and uncatalyzed reaction rate (k_{uncat}) were determined as the slope of the linear phase of reaction respectively in the presence and absence of ribozyme. (**C**) Real-time transcription monitoring. The *in vitro* transcription mixture used in B. was supplemented with DFHBI (instead of S21-Atto) and the fluorescence monitored at 37°C. The synthesis rate (σ) was determined as the slope of the linear phase.

variations issue making necessary the use of a second fluorescent reporter as internal reference (4), whereas the strong monodispersity (polydispersity < 3% (29)) and homogeneity of the droplets used in μIVC makes this dispensable.

iSpinach acquired several mutations that might have been difficult to predict, especially in the absence of a crystal structure model. Our work shows how powerful *in vitro* evolution may be, provided proper selection pressures and selection format are used. Indeed, whereas Spinach2 was obtained by rational design, the tRNA scaffold-free molecule remained sub-optimal in term of fluorescence and folding stability. In addition, one should also note that none of Spinach2 mutations was found in iSpinach. Recently, an alternative thermodynamic-based approach using large-scale integrated microfluidic chips was introduced to screen libraries of point mutants of Spinach (25). Whereas this technology is extremely powerful at screening small libraries of mutants carrying single mutations, its limited throughput (<1000 mutants per round of screening) makes it less suited at screening more complex libraries (≥2 mutations per molecule). Consequently, even though part of iSpinach mutations were revealed by this methodology, none of them was considered and the improved version of the aptamer was both less thermostable and less fluorescent than iSpinach. The improved properties of iSpinach make it appealing for the development of new *in vitro* applications such as biosensing or high throughput screening. The wider salt-tolerance of iSpinach also makes it a good candidate for the development of new fluorogenic probes working in cellular ionic microenvironments (e.g. endosomes and organelles) to sense their composition (e.g. metabolites, proteins) as this was already described with DNA-based nanodevices (36).

The main limitation of iSpinach and of every other DFHBI-binding aptamers is their low affinity for the DFHBI (K_d > 400 nM) compared to what could be expected. However, Mango RNA was recently isolated and shown to have an affinity 2 order of magnitude higher for its fluorogenic co-factor (K_d ≈3 nM) (10). One can therefore predict that performing SELEX in more stringent selection conditions that favor the isolation of high affinity binders in tandem with μIVC screening should make possible to isolate aptamers able to form bright and high affinity complexes. Finally, Spinach/DFHBI complex was also shown to be prone to light-induced dissociation (37). Since our screening is based on light-induced fluorescence measurement, we anticipate that, in the future, using a stronger light source as well as increasing the illumination time may offer the possibility to isolate new mutants with improved photostability.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Stéphanie Baudrey, Ketty Pernod, Christian Rick and Duygu Yilmaz for technical assistance.

FUNDING

Funding for open access charge: This work has 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.

Conflict of interest statement. None declared.

REFERENCES

1. Arora,A., Sunbul,M. and Jaschke,A. (2015) Dual-colour imaging of RNAs using quencher- and fluorophore-binding aptamers. *Nucleic Acids Res.*, **43**, e144.
2. Babendure,J.R., Adams,S.R. and Tsien,R.Y. (2003) Aptamers switch on fluorescence of triphenylmethane dyes. *J. Am. Chem. Soc.*, **125**, 14716–14717.
3. Constantin,T.P., Silva,G.L., Robertson,K.L., Hamilton,T.P., Fague,K., Waggoner,A.S. and Armitage,B.A. (2008) Synthesis of new fluorogenic cyanine dyes and incorporation into RNA fluoromodules. *Org. Lett.*, **10**, 1561–1564.
4. Filonov,G.S., Moon,J.D., Svensen,N. and Jaffrey,S.R. (2014) Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc.*, **136**, 16299–16308.
5. Lee,J., Lee,K.H., Jeon,J., Dragulescu-Andrasi,A., Xiao,F. and Rao,J. (2010) Combining SELEX screening and rational design to develop light-up fluorophore-RNA aptamer pairs for RNA tagging. *ACS Chem. Biol.*, **5**, 1065–1074.
6. Paige,J.S., Wu,K.Y. and Jaffrey,S.R. (2011) RNA mimics of green fluorescent protein. *Science*, **333**, 642–646.
7. Pei,R., Rothman,J., Xie,Y. and Stojanovic,M.N. (2009) Light-up properties of complexes between thiazole orange-small molecule conjugates and aptamers. *Nucleic Acids Res.*, **37**, e59.
8. Sando,S., Narita,A., Hayami,M. and Aoyama,Y. (2008) Transcription monitoring using fused RNA with a dye-binding light-up aptamer as a tag: a blue fluorescent RNA. *Chem. Commun.*, **44**, 3858–3860.
9. Sparano,B.A. and Koide,K. (2005) A strategy for the development of small-molecule-based sensors that strongly fluoresce when bound to a specific RNA. *J. Am. Chem. Soc.*, **127**, 14954–14955.
10. Dolgosheina,E.V., Jeng,S.C., Panchapakesan,S.S., Cojocaru,R., Chen,P.S., Wilson,P.D., Hawkins,N., Wiggins,P.A. and Unrau,P.J. (2014) RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking. *ACS Chem. Biol.*, **9**, 2412–2420.
11. Strack,R.L. and Jaffrey,S.R. (2015) Live-cell imaging of mammalian RNAs with Spinach2. *Methods Enzymol.*, **550**, 129–146.
12. Sunbul,M. and Jaschke,A. (2013) Contact-mediated quenching for RNA imaging in bacteria with a fluorophore-binding aptamer. *Angew. Chem.*, **52**, 13401–13404.
13. Sato,S., Watanabe,M., Katsuda,Y., Murata,A., Wang,D.O. and Uesugi,M. (2015) Live-cell imaging of endogenous mRNAs with a small molecule. *Angew. Chem.*, **54**, 1855–1858.
14. Paige,J.S., Nguyen-Duc,T., Song,W. and Jaffrey,S.R. (2012) Fluorescence imaging of cellular metabolites with RNA. *Science*, **335**, 1194.
15. You,M., Litke,J.L. and Jaffrey,S.R. (2015) Imaging metabolite dynamics in living cells using a Spinach-based riboswitch. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E2756–E2765.
16. Hofer,K., Langejurgen,L.V. and Jaschke,A. (2013) Universal aptamer-based real-time monitoring of enzymatic RNA synthesis. *J. Am. Chem. Soc.*, **135**, 13692–13694.
17. Chudakov,D.M., Matz,M.V., Lukyanov,S. and Lukyanov,K.A. (2010) Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol. Rev.*, **90**, 1103–1163.
18. You,M. and Jaffrey,S.R. (2015) Structure and Mechanism of RNA Mimics of Green Fluorescent Protein. *Annu. Rev. Biophys.*, **44**, 187–206.
19. Strack,R.L., Disney,M.D. and Jaffrey,S.R. (2013) A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat. Methods*, **10**, 1219–1224.
20. Huang,H., Suslov,N.B., Li,N.S., Shelke,S.A., Evans,M.E., Koldobskaya,Y., Rice,P.A. and Piccirilli,J.A. (2014) A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.*, **10**, 686–691.
21. Warner,K.D., Chen,M.C., Song,W., Strack,R.L., Thorn,A., Jaffrey,S.R. and Ferre-D'Amare,A.R. (2014) Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat. Struct. Mol. Biol.*, **21**, 658–663.
22. Ellington,A.D. and Szostak,J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
23. Stoltenburg,R., Reinemann,C. and Strehlitz,B. (2007) SELEX–a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol. Eng.*, **24**, 381–403.
24. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
25. Ketterer,S., Fuchs,D., Weber,W. and Meier,M. (2015) Systematic reconstruction of binding and stability landscapes of the fluorogenic aptamer spinach. *Nucleic Acids Res.*, **43**, 9564–9572.
26. Bhadra,S. and Ellington,A.D. (2014) A Spinach molecular beacon triggered by strand displacement. *RNA*, **20**, 1183–1194.
27. Rogers,T.A., Andrews,G.E., Jaeger,L. and Grabow,W.W. (2015) Fluorescent monitoring of RNA assembly and processing using the split-spinach aptamer. *ACS Synthetic Biol.*, **4**, 162–166.
28. Sharma,S., Zaveri,A., Visweswariah,S.S. and Krishnan,Y. (2014) A fluorescent nucleic acid nanodevice quantitatively images elevated cyclic adenosine monophosphate in membrane-bound compartments. *Small*, **10**, 4276–4280.
29. Ryckelynck,M., Baudrey,S., Rick,C., Marin,A., Coldren,F., Westhof,E. and Griffiths,A.D. (2015) Using droplet-based microfluidics to improve the catalytic properties of RNA under multiple-turnover conditions. *RNA*, **21**, 458–469.
30. Ponchon,L. and Dardel,F. (2007) Recombinant RNA technology: the tRNA scaffold. *Nat. Methods*, **4**, 571–576.
31. Shabala,L., Bowman,J., Brown,J., Ross,T., McMeekin,T. and Shabala,S. (2009) Ion transport and osmotic adjustment in Escherichia coli in response to ionic and non-ionic osmotica. *Environ. Microbiol.*, **11**, 137–148.
32. Pedelacq,J.D., Cabantous,S., Tran,T., Terwilliger,T.C. and Waldo,G.S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.*, **24**, 79–88.
33. Filonov,G.S., Kam,C.W., Song,W. and Jaffrey,S.R. (2015) In-gel imaging of RNA processing using broccoli reveals optimal aptamer expression strategies. *Chem. Biol.*, **22**, 649–660.
34. Lazarev,D., Puskarz,I. and Breaker,R.R. (2003) Substrate specificity and reaction kinetics of an X-motif ribozyme. *RNA*, **9**, 688–697.
35. Chamberlin,M. and Ring,J. (1973) Characterization of T7-specific ribonucleic acid polymerase. II. Inhibitors of the enzyme and their application to the study of the enzymatic reaction. *J. Biol. Chem.*, **248**, 2245–2250.
36. Saha,S., Prakash,V., Halder,S., Chakraborty,K. and Krishnan,Y. (2015) A pH-independent DNA nanodevice for quantifying chloride transport in organelles of living cells. *Nat. Nanotechnol.*, **10**, 645–651.
37. Han,K.Y., Leslie,B.J., Fei,J., Zhang,J. and Ha,T. (2013) Understanding the photophysics of the spinach-DFHBI RNA aptamer-fluorogen complex to improve live-cell RNA imaging. *J. Am. Chem. Soc.*, **135**, 19033–19038.

5.2.2.*Additional data and discussion*

After the publication of the **Original Article 1**, several additional experiments were performed to further characterize the molecule and its performances.

5.2.2.1.*iSpinach versatility in several buffers*

iSpinach displays improved folding efficiency and substantially decreased salt dependency. Interestingly, even though less impressive, the increased fluorescence of the complex in the presence of sodium was also noticed with the wild-type form of Spinach shortly after publication of the **Original Article 1** (Ausländer et al., 2016). Observing proper fluorescence enhancement with the wild type form of Spinach requires providing a minimal monovalent cation concentration of 25 mM and 140 mM of potassium or sodium respectively (Ausländer et al., 2016). On the contrary, iSpinach does not have an absolute requirement for potassium, works better in the presence of sodium and conserves even a substantial fraction of its properties in the absence of monovalent cation. This makes the aptamer highly versatile and able to work in various *in vitro* and extra-cellular conditions in which other DFHBI-binding aptamers would poorly perform. In agreement with this idea, iSpinach works extremely well in a modified minimal hemolymph-like solution (HL3 buffer) mimicking extra-cellular environment, which contains mainly sodium ion (110 mM of sodium and only 5 mM of potassium), conditions in which Spinach2 is almost an order of magnitude dimmer (Figure27). Therefore, iSpinach could serve as starting point in the development of fluorogenic biosensors expected to work in the extracellular space, as this was recently proposed for screening secreting activity of microbes in droplet-based microfluidic screening(Abatemarco et al., 2017).

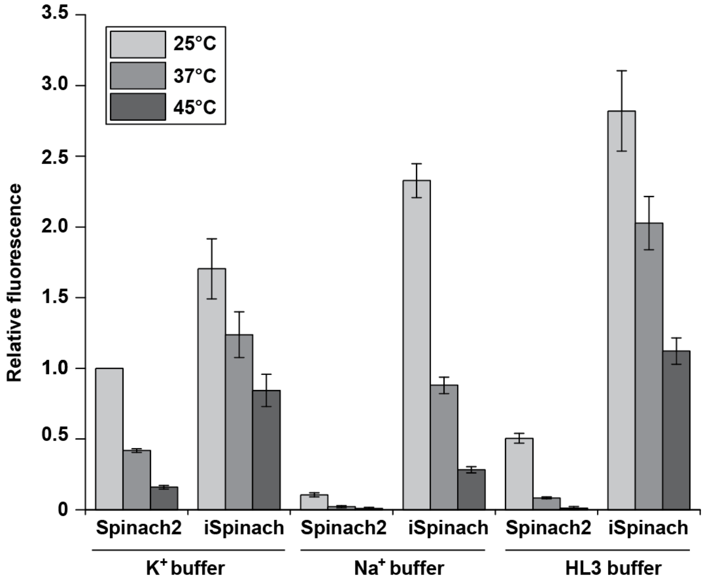


Figure 27: iSpinach and Spinach2 fluorescence in various buffers and temperatures. RNA was renatured in water, then incubated with dedicated buffer for 15 min at 25 °C. Green fluorescence was recorded every minute during 5 minutes at three different temperatures. Fluorescence were relativized to that of Spinach2 in potassium buffer at 25°C. Concerning to potassium or sodium buffer, each of them contain 40 mM Tris-HCl pH 7.5, 1 mM $MgCl_2$ 10 μM DFHBI, and 100 mM of dedicated monovalent cation. HL-3 buffer is more elaborate with 110 mM sodium, 5 mM potassium, 10 mM $NaHCO_3$, 5 mM HEPES pH 7.5, 30 mM sucrose, 5 mM Threalose, and 10 mM $MgCl_2$. Errors bars represent standard deviation of 3 different experiments.

5.2.2.2.*Comparative analysis of DFHBI-binding aptamer sequences*

Even though a pretty high mutation rate (~4 mutations per variants) was used during the error-prone PCR step, the G-quartet domain and the DHFBI-binding pocket were largely preserved and only a single point mutation (A17U according to iSpinach numbering, Figure 28.E) was selected. This general intolerance to mutations of these regions agrees with another study in which all the possible single point mutants of Spinach were individually tested and only the A17U mutation was found tolerated in the G-quartet domain and the DHFBI-binding pocket (Ketterer et al., 2015). Interestingly, this mutation is also shared with Broccoli (A8U according to Broccoli numbering), a DFHBI-binding aptamer also identified *via* a fluorescence-based screening (see above). In addition to this mutation, Broccoli also displays U32C mutation nearby the DFHBI-binding pocket, a mutation that is unique to Broccoli and shared with none of the other DFHBI-binding aptamers. However, a study by Ageely and co-worker (Ageely et al., 2016), indicates that this mutation is likely to be sub-optimal, explaining probably why our screening did not selected it.

An additional difference is found between iSpinach and Broccoli at the level the closing base pair immediately underneath the G-quadruplex. Indeed, whereas the original AoA pair found in Spinach was exchanged for a GoG in iSpinach, it is a canonical C=G pair in Broccoli (Figure 28.D). This base-pair was proposed to be involved in the magnesium dependency of Spinach family, whereas Boccoli was largely magnesium-independent (Ageely et al., 2016). However, to what extend the presence of this canonical base pair may be involved in the limited fluorogenicity we observed with the scaffold-free version of Broccoli is left to be tested. Furthermore, Broccoli displays a second mutation (U2A according to Broccoli numbering) in the basal stem instead of the UoU pair observed in the Spinach family (Figure 28). Interestingly, turning this UoU pair into a canonical A-U was also found to be beneficial in a study performed by Ketterer and co-coworker (Ketterer et al., 2015). Even though, the positive role of this mutation in the context of iSpinach is still left to be tested, the absence of this mutation in our selected pool can be explained by the mutagenesis kit we used to prepare the libraries. Indeed, this kit is based on the use of nucleotide analogues that mediate only poorly T to A transversions (Zaccolo et al., 1996).

To conclude on this sequence analysis, should the screening be repeated, a potential improvement of the method could be the use an additional step consisting in pico-injecting some EDTA together with the DFHBI into the droplets (similarly to the strategy used by the team during X-motif evolution) after RNA synthesis and prior to measuring droplets fluorescence and sorting them. This could be expected to add a new selection pressure favoring the isolation of molecules having a low magnesium-dependency and elevated k_{on} for the DFHBI while still being salt-independent (see also General Discussion Chapter), bright and thermostable. Such experiment would allow seeing to what extend sequence changes other than the AoA to C=G substitution allow reducing the magnesium-dependency of DFHBI-binding aptamers.

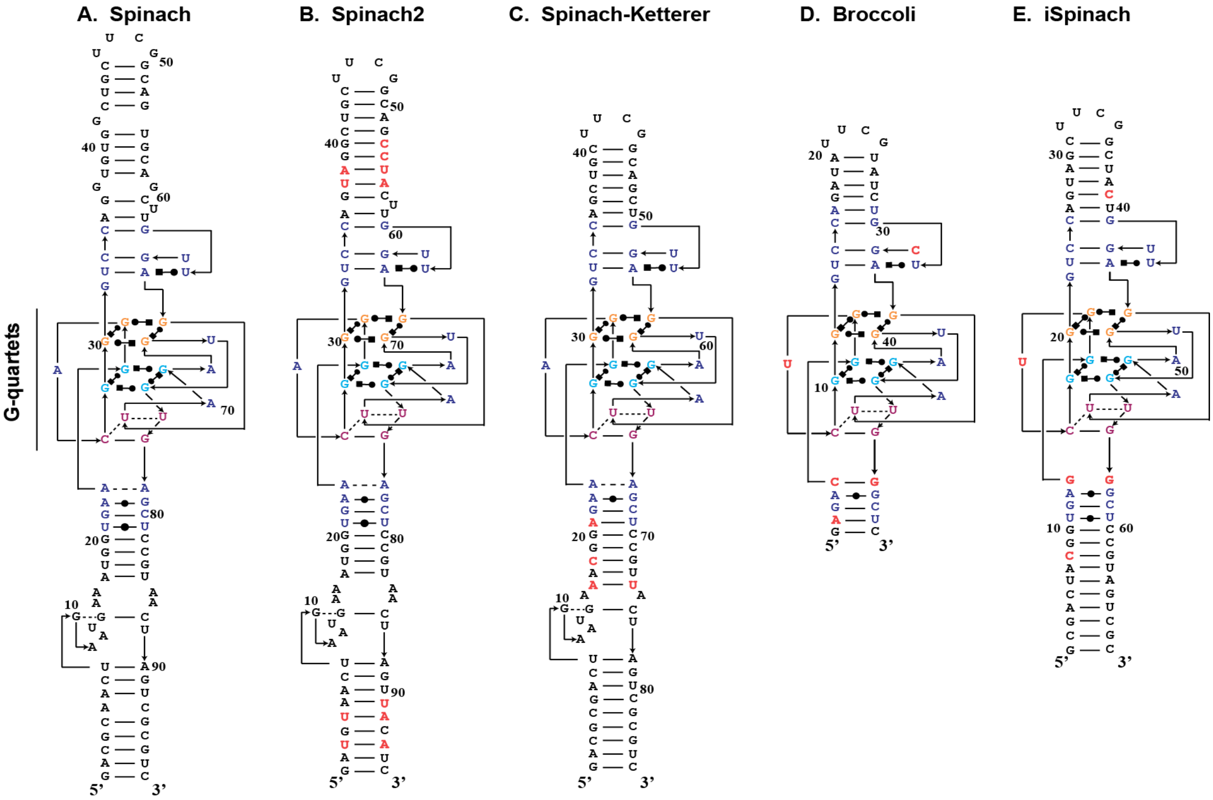


Figure 28: DFHBI light-up aptamers. (A) Original Spinach structure (Warner et al., 2014)isolated from SELEX procedure (Paige et al., 2011). **(B)** Spinach2 obtained with rational design to improve fluorescence and folding (Strack et al., 2013). **(C)** Spinach variant engineered on the basis of a thermodynamic-based scanning(Ketterer et al., 2015). **(D)** Broccoli was isolated using a new selection scheme coupling SELEX and FACS (Filonov et al., 2014). Despite that 3D structure of Broccoli has not been determined, this representation rely on presumed model due to high similarity to Spinach sequence and 2D model(Ageely et al., 2016; Svensen and Jaffrey, 2016). **(E)** iSpinach isolated after droplet-based microfluidics screening (Autour et al., 2016). All mutations for the Spinach variants are labelled in red (B-E).

5.2.2.3.iSpinach imaging in prokaryotes

Even though iSpinach was isolated for *in vitro* purposes, I managed to get it working in bacteria as well (Figure 29). Indeed, whereas the overexpression of a gene coding for the non-related SRB-2 light-up aptamer (Sunbul and Jäschke, 2013)did not yield any fluorescence in the presence of the DFHBI-1T (a brighter analogue of DFHBI (Song et al., 2014), expressing a gene in which iSpinach was embedded either within a tRNA scaffold, or within the highly structured RNAIII (a non-coding RNA found in *Staphylococcus aureus*) allowed observing high level of green fluorescence, demonstrating that iSpinach could also be used for intracellular imaging purposes.

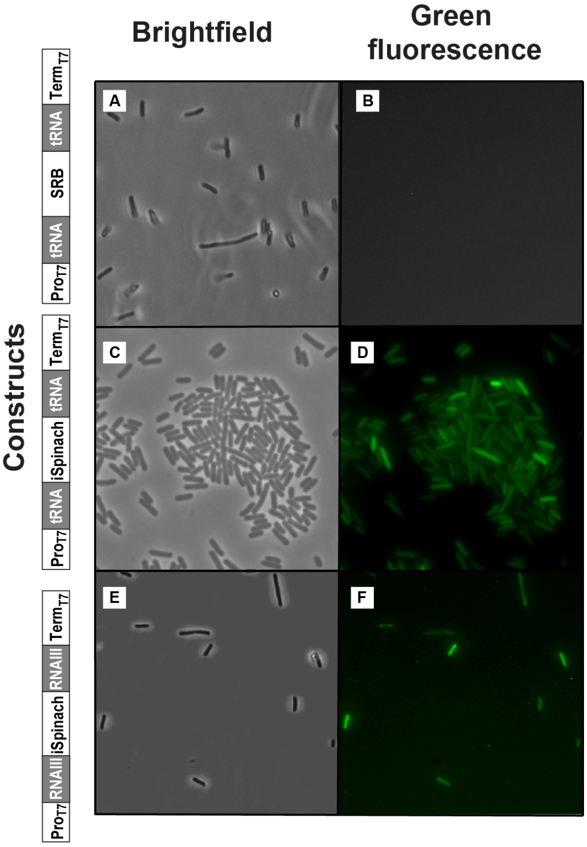

Brightfield
Green fluorescence
Constructs
Pro$_{T7}$ tRNA SRB tRNA Term$_{T7}$
Pro$_{T7}$ tRNA iSpinach tRNA Term$_{T7}$
Pro$_{T7}$ RNAIII iSpinach RNAIII Term$_{T7}$
A
B
C
D
E
F

Figure 29 : *In vivo* imaging of iSpinach using DFBHI-1T. (A-F) All the constructs were placed under control of T7 promotor and terminator and were cloned into pET28a plasmid. BL21(DE3) were transformed with the plasmid and plated onto a 2YT medium supplemented with kanamycin. A colony was picked and grown in 2YT liquid media supplemented with kanamycin at 37°C. Upon reaching an O.D of ~ 0.5, RNA expression was induced for 2 hours by addition of IPTG. Then bacteria were incubated in PBS supplemented with 20 μM of DFHBI-1T prior to being mounted between a glass slide and cover slip. SRB-2 aptamer **(A-B)** or iSpinach **(C-D)** inserted into tRNA scaffold were observed in brightfield **(A&C)** or in epi-fluorescence (ex./em. 488 nm/525 nm) **(B&D)**. **(E-F)** Herein, iSpinach aptamer was inserted within the stem 8 of RNA III.

As an additional control, total RNA was extracted from both induced and non-induced bacteria and controlled amount of RNA was loaded onto a polyacrylamide denaturing gel (Figure 30.A). Upon electrophoresis and renaturation protocol as described elsewhere(Filonov et al., 2015), gels were stained with DFHBI prior to being restained with Sybr-Gold. Upon DFHBI staining, a major band corresponding to the tRNA-iSpinach construct was observed only with induced cells confirming that the aptamer is properly expressed and stable. One can observed additional bands of lower molecular weight on the gel. These bands were already observed in a previous study (Filonov et al., 2015)and are due to a cleavage occurring between the tRNA scaffold and iSpinach, likely to be catalyzed by tRNA maturing enzymes naturally present in the bacterium. Such unwanted cleavage could however be reduced in the future by using alternative scaffold such as the F30 three-way junction recently described (Filonov et al., 2015) instead of the tRNA. Interestingly, revealing the total RNA fraction using Sybr-Gold staining (Figure 30.B) also showed that iSpinach represents a substantial fraction of the transcriptome indicating that imaging less abundant but biologically more relevant RNAs might reveal more challenging even though iSpinach is significantly more efficient the other DFHBI-binding aptamers.

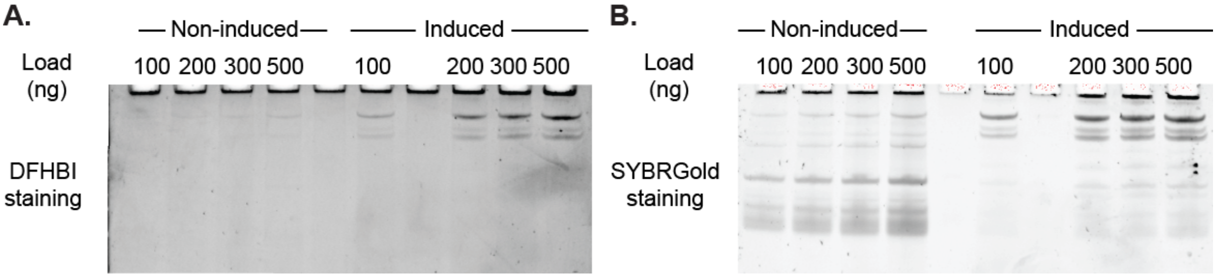


Figure 30: In gel imaging. Total RNAs was retrieved after Trizol extraction and loaded onto 12 % denaturing polyacrylamide gel. Total RNA was extracted from bacteria transformed with pET28 plasmid containing iSpinach inserted into tRNA scaffold and either induced by IPTG or kept non-induced. (A) Total RNA was stained with DFHBI according to (Filonov et al., 2015). (B) Prior SYBR-Gold staining, the same gel was washed by 3 water baths.

Finally, a recent publication by Wang *et al*, also showed that iSpinach and its split-version can be used as imaging tools to monitor RNAs in living mammalian cells (Wang et al., 2018). Interestingly, and as anticipated from our own data, it seems that iSpinach could even behave better than Broccoli. However, additional experiments should be performed prior to firmly conclude on this later point.

To conclude this first part, the results I obtained demonstrate how powerful μIVC can be in identifying light-up aptamers endowed with improved folding efficiency, stability and brightness, provided proper selection pressures are used. Besides the engineering used to shorten the molecule, a total of 5 point mutations were identified to be responsible for the improved performances of the aptamer. However, properly understanding the contribution of each mutation required the crystal structure of the complex to be solved.

5.3. iSpinach structural characterization

To get insight into the origin of iSpinach superior properties, we crystallized iSpinach in complex with DFHBI and solved the X-ray structure of the complex. X-ray crystallography was chosen because iSpinach was slightly too large to be analyzed by NMR and definitely too small to be studied by CryoEM. Moreover, crystallogenesis conditions were already known for Spinach (Huang et al., 2014; Warner et al., 2014), which made it easier to obtain crystals in a short period of time.

Crystals were indeed obtained for the complex crystallized in the presence of sodium or potassium. However, whereas those crystals obtained in the presence of potassium readily diffracted, none of the crystals obtained in the sodium allowed us to solve the structure of the complex (a possible explanation is discussed below). Despite the partial success of our crystallization campaign, the data collected were sufficient to explain the effect of 4 out of the 5 mutations and allowed me to co-author a publication (**Original Article 2**) that is been published in RNA journal in 2017 and is presented hereafter.

Original Article 2:

Crystal structure and fluorescence properties of the iSpinach aptamer in complex with DFHBI

Pablo Fernandez-Millan, Alexis Autour, Eric Ennifar, Eric Westhof and Michael Ryckelynck

2017

RNA

REPORT

Crystal structure and fluorescence properties of the iSpinach aptamer in complex with DFHBI

PABLO FERNANDEZ-MILLAN, ALEXIS AUTOUR, ERIC ENNIFAR, ERIC WESTHOF, and MICHAEL RYCKELYNCK
Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, F-67000 Strasbourg, France

ABSTRACT

Fluorogenic RNA aptamers are short nucleic acids able to specifically interact with small molecules and strongly enhance their fluorescence upon complex formation. Among the different systems recently introduced, Spinach, an aptamer forming a fluorescent complex with the 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), is one of the most promising. Using random mutagenesis and ultrahigh-throughput screening, we recently developed iSpinach, an improved version of the aptamer, endowed with an increased folding efficiency and thermal stability. iSpinach is a shorter version of Spinach, comprising five mutations for which the exact role has not yet been deciphered. In this work, we cocrystallized a reengineered version of iSpinach in complex with the DFHBI and solved the X-ray structure of the complex at 2 Å resolution. Only a few mutations were required to optimize iSpinach production and crystallization, underlying the good folding capacity of the molecule. The measured fluorescence half-lives in the crystal were 60% higher than in solution. Comparisons with structures previously reported for Spinach sheds some light on the possible function of the different beneficial mutations carried by iSpinach.

Keywords: fluorogenic RNA aptamer; DFHBI; crystal structure; fluorescence

INTRODUCTION

The discovery of naturally fluorescent proteins had a profound impact on biology by making it possible to track proteins in the cell, but also monitoring gene expression as well as devising new innovative fluorogenic sensors (Enterina et al. 2015; Mishin et al. 2015). Even though no naturally fluorescent RNA has been discovered yet, several artificial fluorogenic RNAs have been developed. These RNAs are short nucleic acids able to specifically interact with small molecules that become fluorescent upon complex formation. This capacity of an RNA molecule to light up a fluorogenic dye was first demonstrated with the aptamer selected to specifically interact with malachite green (Grate and Wilson 1999), which was later found to be able to efficiently activate the fluorescence of triphenylmethane dyes (Babendure et al. 2003). Since then, a variety of alternative fluorogenic dye/RNA aptamer couples have been developed (Dolgosheina and Unrau 2016; Ouellet 2016; Pauff et al. 2017). Similarly to fluorescent proteins, fluorescent RNAs have a wide application spectrum (e.g., live-cell gene expression imaging [Guet et al. 2015; Zhang et al. 2015], biosensing of proteins [Song et al. 2013], nucleic acids [Sato et al. 2015; Aw et al. 2016; Huang et al. 2017; Ong et al. 2017], or even metabolites [Nakayama et al. 2012; Paige et al. 2012; Kellenberger et al. 2013, 2015a,b; Sharma et al. 2014; You et al. 2015; Bose et al. 2016; Hallberg et al. 2016; Su et al. 2016; Wang et al. 2016]), and one can anticipate their significant impact on RNA biology.

Among the variety of fluorogenic aptamers currently known, Spinach is the best characterized and the most widely used. This RNA was originally identified by in vitro selection of aptamers for their capacity to interact with the 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), a GFP-mimicking dye that is nontoxic, bio-orthogonal, and cell permeable (Paige et al. 2011). Further engineering of the original aptamer allowed the development of Spinach 2, a molecule with increased folding properties and thermal stability (Strack et al. 2013). Structural characterization of both DFHBI-binding aptamers revealed that they are organized around a three-tier G-quadruplex core forming the bottom layer of the DFHBI-binding pocket that accommodates the dye in planar conformation (Huang et al. 2014; Warner et al. 2014). The remainder of the DFHBI-binding pocket consists of a top layer made of a base-triple while the side of the dye establishes interactions with nucleobases. Despite the tight accommodation of the dye and the improved properties of Spinach 2, we recently found that the folding properties of the aptamer were still suboptimal (Autour et al. 2016). To further optimize Spinach RNA, we

---

Corresponding author: m.ryckelynck@unistra.fr
Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.063008.117. Freely available online through the *RNA* Open Access option.

© 2017 Fernandez-Millan et al. This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

RNA 23:1788–1795; Published by Cold Spring Harbor Laboratory Press for the RNA Society

generated mutant libraries and screened them for brighter and better folding mutants using our recently introduced microfluidic-assisted in vitro compartmentalization pipeline (Ryckelynck et al. 2015). During this process we used selection pressures (i.e., warming, complete absence of K^+ and use of Na^+ instead) aimed at challenging RNA structural stability, especially that of the G-quadruplex core (Autour et al. 2016). Sequence analysis of the most fluorescent mutants allowed us to identify several beneficial mutations, of which the five displaying the highest effect were combined into the same aptamer. Further engineering the molecule led to an improved Spinach molecule (iSpinach), a DFHBI-binding aptamer forming a fluorescent complex that is characterized by a high thermal stability, a reduced salt-sensitivity, and an increased brightness. Interestingly, none of the selected mutations occurred in the DFHBI-binding pocket, suggesting that they may primarily have an effect on aptamer folding.

In the present work, we determined the crystal structure of iSpinach in complex with the DFHBI to gain insight into the role of the different mutations introduced in the aptamer. Our data indicate that, as anticipated, the mutations act at the structural level by compacting and stabilizing the structure of the aptamer. More specifically, we found that the replacement of an AoA base pair interacting via a single hydrogen bond and connecting the G-quadruplex to the helix distal from the DFHBI-binding site by a GoG base pair forming a Watson–Crick/Hoogsteen interaction allows a better coordination of potassium by the G-quadruplex through the formation of a water-mediated network.

RESULTS AND DISCUSSION

Molecular engineering and crystallogenesis

The original iSpinach aptamer was first reengineered to optimize its production and to promote intermolecular interactions during the crystallization process. First, the double mutation C2G/C5G was introduced at the 5′ end of the sequence of all the constructs to increase the in vitro transcription reaction yield (Supplemental Fig. S1) by providing an optimal internal transcription sequence to the enzyme; the consequence was only the inversion of the Watson–Crick base pair C=G by G=C in the basal stem (Fig. 1). Second,

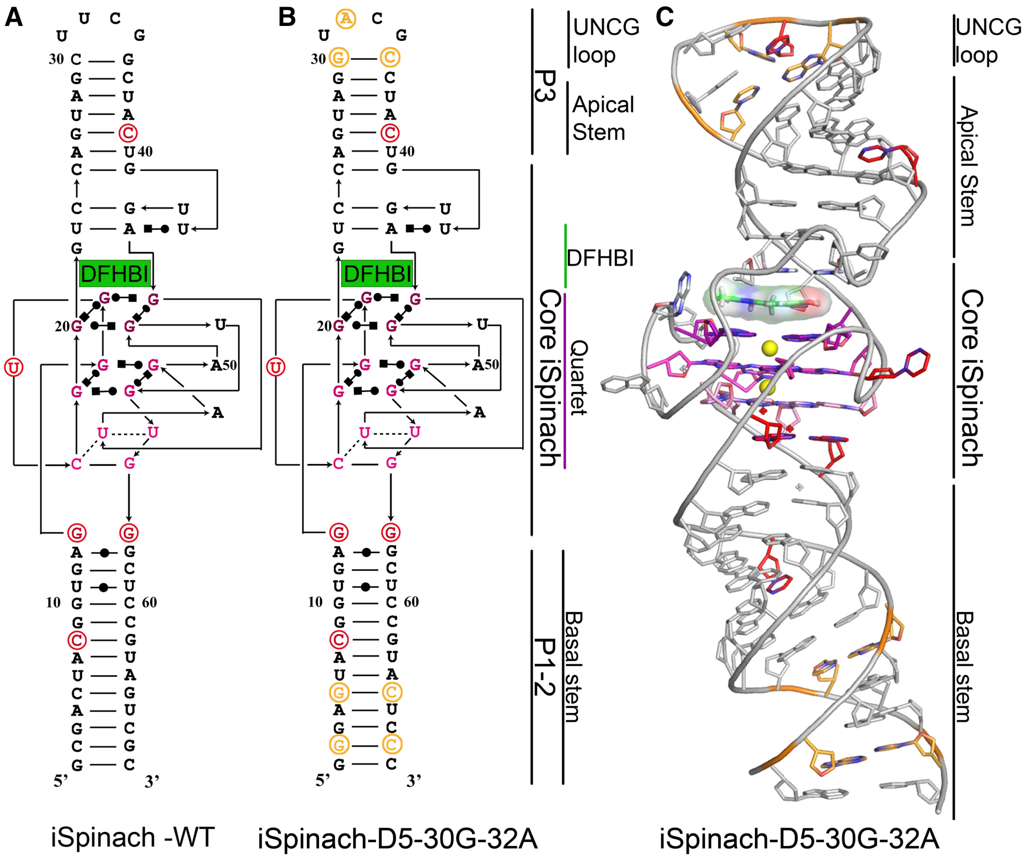


FIGURE 1. Structural organization of iSpinach and its crystallized mutant. (*A*) Secondary structure of iSpinach. (*B*) Secondary structure of iSpinach-D5-G30–A32. On both models, the beneficial mutations combined to obtain iSpinach are shown and circled in red, while those introduced to increase the transcription rate of the molecule and promote its crystallization are shown in orange. (*C*) Crystal structure of iSpinach-D5-G30–A32. The residues are shown using the same color code used on secondary structure models. Potassium ions are shown as yellow spheres.

the second position of the UUCG tetraloop capping the apical hairpin was varied (UNCG motif) in order to promote intermolecular contacts in the crystal and favor the crystallization success. Finally, it is reported that the two base pairs proximal to the UNCG may affect its folding and stability (Molinaro and Tinoco 1995; Davis and Znosko 2007). Therefore, we assayed both the original base pair C=G and a G=C mutant. All tested constructs conserved intact fluorogenic properties and several of them cocrystallized in the presence of DFHBI (Supplemental Tables S1, S2), especially iSpinach-D5-G30–A32, which yielded crystals diffracting at a resolution of 2–2.4 Å, a resolution exceeding that obtained in previous studies. Remarkably, iSpinach readily crystallized without the assistance of a protein as reported previously (Huang et al. 2014), underlying the improved folding capacity of the aptamer. This makes iSpinach the first DFHBI-binding aptamer crystallized in its native state with minor sequence changes and without the need for sequence insertion (Huang et al. 2014) or cleavage (Warner et al. 2014).

As a quality control, we first verified the fluorescence of the crystallized iSpinach-D5-G30–A32/DFHBI complex by fishing one crystal from each crystallization condition and, after washing, imaging the crystal by epi-fluorescence microscopy (Fig. 2A). We then evaluated the photostability of crystal and soluble complexes by placing them within a capillary under continuous illumination at DFHBI excitation wavelength (i.e., 475 nm) while monitoring fluorescence decay (Fig. 2B). To avoid artifactual fluorescence recovery of the soluble complex through the diffusion of fluorescent complexes from a nonilluminated area of the capillary, the solution of RNA/DFHBI complex was encapsulated into 100 pL water-in-oil droplets using a microfluidic droplet generator (Anna et al. 2003) and the resulting emulsion was loaded into the capillary (Fig. 2A). As expected from the difference in local concentration, the fluorescence intensity of the crystallized complex largely exceeded that of the soluble complex. Nevertheless, normalizing the fluorescence to that of the starting exposure time revealed that both types of complexes displayed similar fluorescence half-life, with the crystallized form having a fluorescence half-life only twice higher than the soluble one (Fig. 2B). Moreover, we observed that crystals melted and cracked during the experiment, suggesting conformational changes consistent with a DFHBI photo-inactivation model in which the dye isomerizes upon illumination and is rapidly ejected from the aptamer (Han et al. 2013). Altogether, these results suggest that the crystallized form of the complex is representative of its native state in solution.

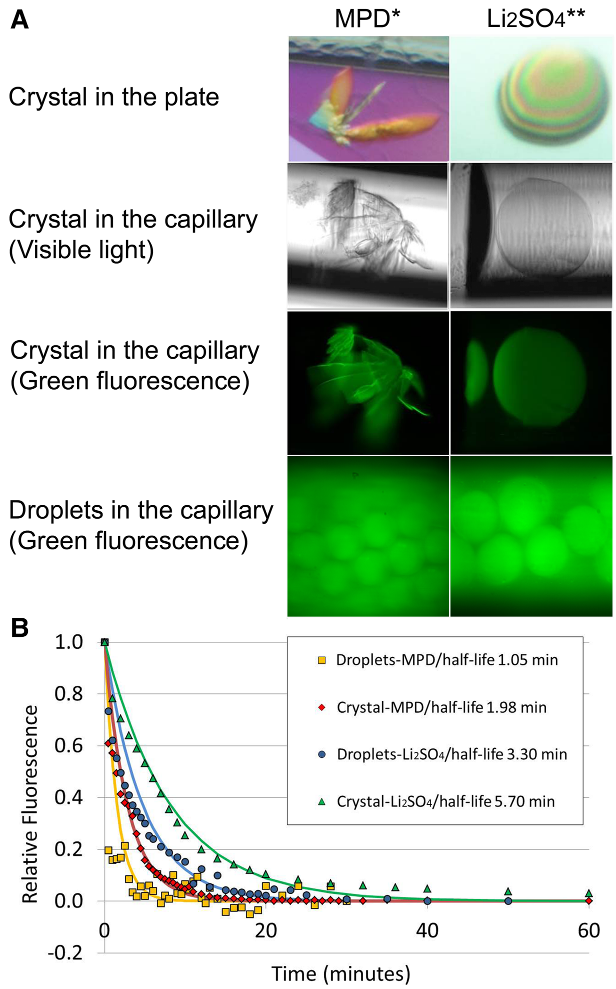


FIGURE 2. Fluorescence properties of the iSpinach-D5-G30–A32/DFHBI crystallized complex. (*A*) Micrographs of crystals and droplets. Micrographs of the crystals were taken while they were in the crystallization plate or loaded into the capillary and observed under visible light or in epi-fluorescence mode. The conditions in which the crystal was grown or the droplets prepared (MPD or Li_2SO_4) are indicated *under* the corresponding column. (*) MPD crystallization conditions were 32% MPD, 40 mM NaCl, 40 mM sodium cacodylate pH 7.0 and 12 mM spermine. (**) Li_2SO_4 crystallization conditions were 1.9 M Li_2SO_4, 50 mM MES pH 5.6, 5 mM $MgCl_2$. (*B*) Fluorescence decay monitoring. The fluorescence of the crystals or the droplets was monitored along the time. The values were relativized to that of the first measurement point, and an exponential decay was fit to the data to compute the fluorescence half-life (values given in the legend of the plot).

Crystal structure of the complex

The iSpinach-D5-G30–A32 molecule folds into an 80.5 Å long double-stranded hairpin, containing in its center a G-quadruplex surrounded by two sets of coaxially stacked helical stems. The hairpin is capped by a tetraloop composed of UACG, from U31 to G34, that adopts a conventional fold for UNCG loops where G34 interacts with the ribose edge of U31 with an extra H-bond to the A32 ribose (Ennifar et al. 2000). Moreover, C33 (N4) forms an H-bond to an anionic A32 oxygen atom. Therefore, as initially expected, modulating the sequence of this tetraloop promoted crystal packing since the exposed A32 forms a base-pair kissing contact with a neighboring iSpinach molecule (Supplemental Fig. S2D). Besides A32, U43 and A48 are also exposed to the solvent and involved in the crystal packing. Both U43 and A48 from two different neighboring molecules and related by crystallographic symmetry are interacting together by a double set of interactions: first through stacking interaction and second by H-bonds between hydroxyl A48(O2′) and U43 (O4). At the same time, U43 and A48 have contacts with a third molecule: U43(O2) interacts with G19(N2), A48(N1) with G55(O2′), and A48 (N1) with G55(N3) by H-bonds (Supplemental Fig. S2E). Finally, the first base pair of the basal stem, G1=C69, interacts through stacking interaction with the homologous base pair of another RNA molecule (Supplemental Fig. S2F).

The functional core of iSpinach-D5-G30–A32 is located at the center of the molecule and defines the DFHBI binding site. The core is composed of two structural groups: an ensemble of three quadruplex layers and one layer of triple base pairs, with the DFHBI being sandwiched in between these two structural groups.

The quadruplexes are composed of two regular G-quartet layers and a third mixed layer (Fig. 3). In the double G-quartet layer, the top layer (layer 1) is defined by the interaction of G20 and G16 with an *anti* glycosyl conformation, while G46 and G51 are in the *syn* glycosyl conformation. The second quartet (layer 2) is composed of G49 in the *anti* glycosyl conformation with G15, G19, and G53 in the *syn* glycosyl conformation. The third layer, below layer 2, is a mixed sequence tetrad (layer 3) in which C18 and G55 bind as a Watson–Crick pair with, in addition, U47 pairing with C18 in the *anti* glycosyl conformation and U54 in the *syn* glycosyl conformation binding G55. Using anomalous data collected at 1.8 Å wavelength ($f' \sim 1.5\ \bar{e}$), the nature and position of potassium ions were unambiguously confirmed by peaks at 6 − σ level in anomalous difference maps. A first potassium ion (K1) with octahedron coordination is found in between the first two layers, in agreement with the previous structural studies (Huang et al. 2014; Warner et al. 2014). Moreover, we identified a second potassium ion (K2) located between "layers 2 and 3." This ion also has an octahedron coordination through the participation of a water molecule (W1) fixed by the bases of "layer 3" with the participation of G14 and G56 (Fig. 3), two mutations specific of iSpinach (see below).

The second structural motif forming the DFHBI-binding site is a triple base pair interaction in planar orientation (*Lid layer*) where U22 and A45 interact through canonical Watson–Crick pair while U42 and A45 form a Watson–Crick/Hoogsteen *cis* base pair (Fig. 3). This triple forms the upper part of the DFHBI-binding pocket into which the DFHBI intercalates in a planar conformation and makes extensive aromatic stacking interactions with the first G-quartet layer. In addition to stacking interaction with layer 1 of the G-quartet, the phenol group of the DFHBI makes an interaction involving its hydroxyl and the ribose hydroxyl of G16. Moreover, the imidazole ring of the dye makes direct contacts with A50 and G21. The high resolution of our structure allows the unambiguous fit of the active *cis* isomer of the DFHBI in the density map (Supplemental Fig. S3), further confirming that the complex is in its native form.

Finally, the center core of the molecule is flanked by two helical regions composed of Watson–Crick interactions involving nucleotides 1–14 and 56–69 in the basal stem and nucleotides 24–30 in 35–41 in the apical stem, the latter being capped by a tetraloop.

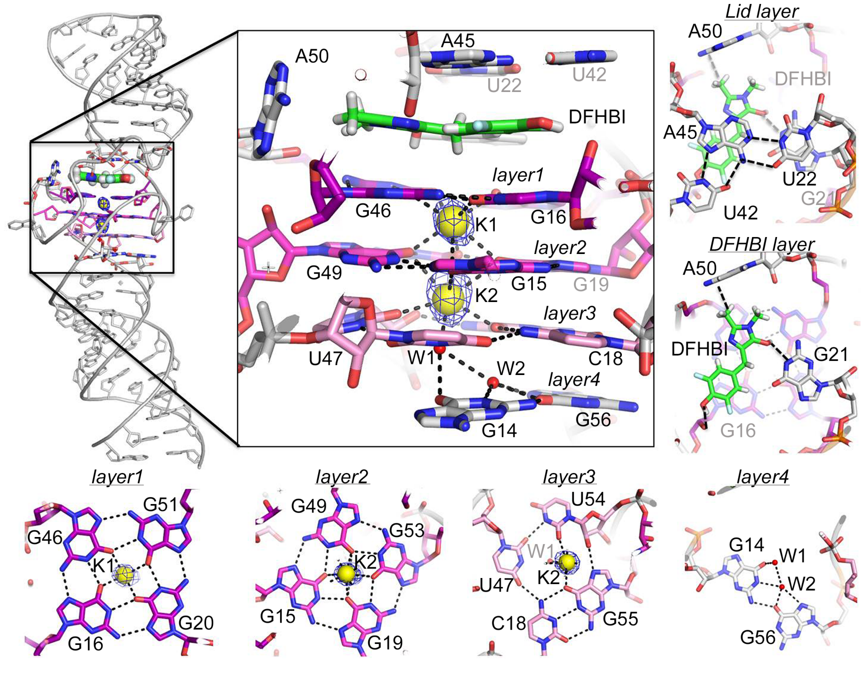


FIGURE 3. iSpinach crystallographic structure. The crystal structure of iSpinach-D5-G30–A32 is shown in light gray (*top left*) together with a zoom on the core region of the molecule involved in DFHBI binding. The ligand DFHBI is colored in green. The quartet motif is in magenta from dark magenta in layer 1, magenta layer 2, and light magenta layer 3. Potassium ions are represented by yellow spheres and the anomalous differential electronic density maps (Fo–Fc at 6.0 σ level) are shown as a blue mesh *around* the spheres. The interactions are represented by black dashed lines. The waters are represented by red dots. All the zoomed layers are shown *perpendicular* to the helical axis as well.

Structural comparative analysis of iSpinach and Spinach aptamers

The overall structure of iSpinach/DFHBI complex is very similar to that reported for Spinach 1 and 2 aptamers, suggesting that the beneficial mutations contained in iSpinach may have subtle yet important effects. To shed light on the contribution of these mutations, we compared the structure obtained in the present study with those previously reported with different RNA constructs (Huang et al. 2014; Warner et al. 2014). To this end, we selected from each study the structure of the aptamer in complex with DFHBI having the highest resolution (i.e., PDB code 4KZD.pdb [Huang et al. 2014] and 4TS0.pdb [Warner et al. 2014]). Hereafter, these structures will be called, respectively, Spinach-P and Spinach-FdA for ease of reading (Supplemental Fig. S4).

The iSpinach core region spans from the G14–G56 base pair to the triple base pairs that define the lid layer of the DFHBI-binding pocket (Fig. 1). The superposition of the three structures (Fig. 4) indicates that they are very similar in their core region as highlighted by the very low RMSD (Supplemental Table S4). Consistently, the position and the relative orientation of the DFHBI inside the core of the RNA aptamer do not show differences between the three structures (Supplemental Fig. S5). Yet, interesting differences were observed. First, U17 in iSpinach (originally A24 in Spinach-P and A27 in Spinach-FdA) is exposed to the solvent (Fig. 4, bottom left), making a clear-cut interpretation of the effect of interchanging a purine for a pyrimidine challenging. Nevertheless, the distance between the OP1 phosphate of U17 and the O2′ of U42 ribose is 3.6 Å (U42 is involved in Hoogsteen triple interaction in the Lid layer of DHFBI), a distance significantly shorter than in Spinach-FdA (4.3 Å) and in Spinach-P (4.5 Å) and compatible with the formation of an electrostatic contact. Therefore, even though additional experiments will be required to conclude on the exact contribution of U17, our data suggest that this residue could act by increasing the folding stability of the DFHBI-binding pocket. This mutation was also found to be the only one tolerated in the G-quadruplex region and its introduction in the original Spinach aptamer was also found to increase the fluorescence of the complex (Ketterer et al. 2015). Moreover, Broccoli, another DFHBI-binding aptamer (Filonov et al. 2014) proposed to share the same core structure as Spinach (Ageely et al. 2016), also possesses a U at the same position, further supporting the key role of this residue. Second, of the three structures, the present one is the only one in which both potassium ions of the G-quartet core are seen with the proper coordination state. Indeed, in Spinach-P only the first potassium ion (K1) was observed between layers 1 and 2, while a water molecule occupies the K2 position. In addition, even though both K1 and K2 ions are seen in Spinach-FdA structure, the coordination of K2 is only seven. On the contrary, in the present structure a water molecule (W1) coordinating K2 is clearly observed on the density map (Fig. 3). W1 is held in place by an interaction network involving a second water molecule (W2) and the residues G14 and G56. Interestingly, in Spinach these residues are two adenines (A14 and A56) interacting by a single H-bond between N1 and N6 in both structures (Fig. 4, top right). Conversely, in iSpinach, there are bifurcated H-bonds between G14 (N1) and G56 (O6) with G56 (N2), while W2 mediates four H-bonds between this base pair and layer 3 (Fig. 3). Regarding the impact of these mutations, 4KZD and 4KZE PDB structures (Huang et al. 2014) provide important information. The comparison of the DFHBI-bound Spinach-P

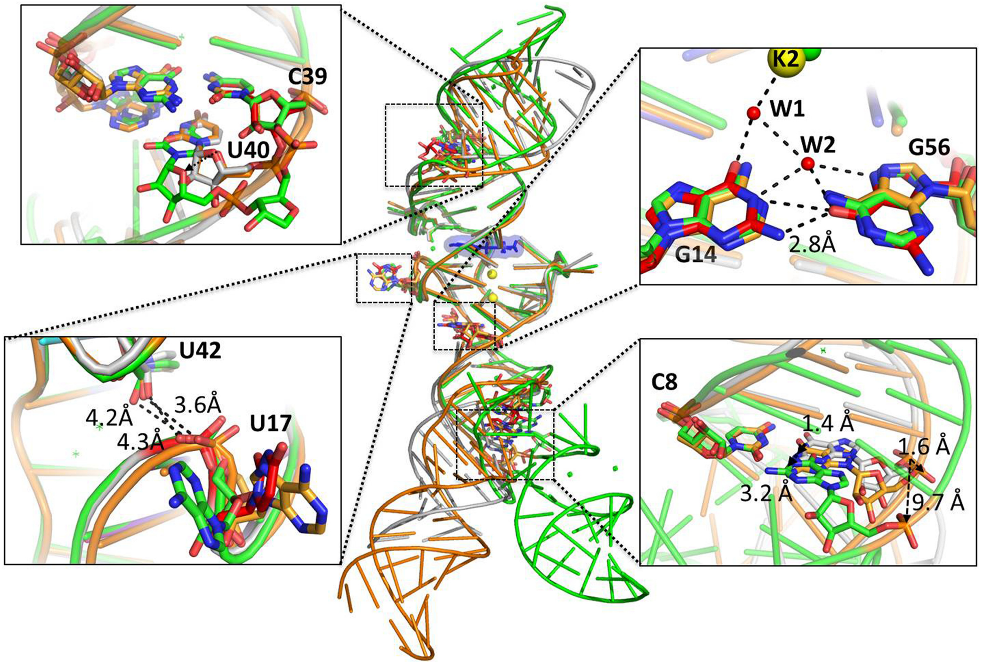


FIGURE 4. Superposition of different Spinach/DFHBI complex crystal structures. The structure obtained in the present study is shown in gray and is superimposed to that of Spinach-FdA (shown in green) and that of Spinach-P (shown in orange). The regions containing the beneficial mutations are zoomed-in and the mutated residues are shown in red, whereas the wild-type ones are shown according to the color code used before. Potassium ions are represented by yellow spheres and water molecules by red dots.

(4KZD) with the DFHBI-free Spinach-P (4KZE) shows the substitution of U54 by A48 in layer 3. It is possible that a movement consequent to DFHBI binding promotes the loss of K2 coordination. Therefore, we suggest that GoG has a positive impact during this transition by holding the K2 in position via the water W1 and helping to stabilize the G-quartet core. This observation is consistent with the fact that, in our original screen, the two G's were always found together (Autour et al. 2016). This G pair was also recently found to increase the fluorescence of Broccoli, but at the expense of an increased dependency on magnesium (Ageely et al. 2016). The possibility that W2 is an ion (magnesium or sodium) was taken into account during the refinement process; however, this option was rejected after an exhaustive analysis because a correct coordination geometry was absent (Supplemental Fig. S6; Zheng et al. 2009, 2017).

Extending the superimposition of the three structures beyond the core significantly increased the RMSD values as a consequence of the variations in the stems of the RNAs. Indeed, the basal stems have different orientations with respect to the G-quadruplex domains, which are especially accentuated in the Spinach-FdA structure (in green on Fig. 4). Furthermore, the apical stem is even more challenging to compare because the designs of the different aptamers are completely different. Indeed, while in iSpinach the apical stem is closed by a UNCG tetraloop, in Spinach-P the loop is replaced by an RNA element bound to an antibody, and the Spinach-FdA structure results from the assembly of two single strands (Supplemental Fig. S4). Nevertheless, local superimpositions of the three structures allowed us to further decipher the contribution of the two left mutations (C39 and C8). As expected for the conversion of a weak GoU into a more stable G=C pair, no structural difference was observed at the level of the C39 mutation (Fig. 4, top left). This confirms our initial suggestion (Autour et al. 2016) that the evolutionary process selected the mutation U39C only to increase the stability of the helix without modifying its structure. Since iSpinach and Spinach-P shared the same sequence from G28–C39 to the first layer of G-quartet, both molecules superposed well over this region. Conversely, the equivalent region of Spinach-FdA (G36–C57/U58/A35–U59) displays a bulged U residue that slightly disturbs the continuity of the helix. Finally, the U to C substitution at position 8 produces a better interaction with the G of the complementary strand. The GoU pair promotes the displacement of the complementary base by 1.4 Å and 3.2 Å (measured from N1 of the base) and a displacement of 1.6 Å and 9.7 Å in Spinach-P and Spinach-FdA structures, respectively (Fig. 4, bottom right). Overall, the major groove is wider in the Spinach-FdA structure due to the presence of a 2-nt bulge (Supplemental Fig. S7).

In conclusion, the cocrystal structure of iSpinach in complex with the DFHBI helps in the understanding of the contribution of the different mutations combined in the new aptamer. Indeed, whereas C8 and C39 were found to have the anticipated stabilization action by changing GoU pairs into more stable G=C's, the effect of U17, G14, and G56 were less obvious a priori. Nevertheless, our model now allows us to propose that these mutations could, at least in part, act by forming a new electrostatic interaction (U17) as well as ensuring a tighter coordination of the second potassium ion of the G-quartet via a water molecule (G14 and G56). The net effect of these different mutations is to compact and stabilize the structure of the aptamer, especially that of the G-quadruplex core and of the DFHBI binding pocket, making the iSpinach/DFHBI complex far less salt sensitive and more thermostable than the parental molecule as previously observed (Autour et al. 2016). This improved folding capacity is also supported by the fact that we could obtain diffracting crystals with minimal modification of the sequence. Indeed, only two base-pair inversions were introduced in the basal stem to optimize RNA synthesis, and a single point mutation was introduced in the apical loop to promote crystal-packing interactions. This clearly contrasts with previous work in which the RNA was formed from two half-molecules annealed together (Warner et al. 2014), or where the apical loop was substituted by a Fab-binding domain (Huang et al. 2014). This capacity of iSpinach to crystallize in its near-native state suggests a high structural homogeneity inside the population of molecules underlining the good folding capacity of iSpinach. Nevertheless, despite this improved folding capacity, iSpinach still shares the complex structural organization of the Spinach family (i.e., a mixed-base quadruplex core surrounded by two flanking A-form helices), characterized by a complex connectivity. Whereas building the dye-binding pocket around a base quadruple is common to several fluorogenic RNA aptamers (Trachman et al. 2017b), the use of more conventional G-quadruplex folding with short propeller arms may enable the connectivity of these molecules to be simplified by using, for instance, a simple GNRA-derived motif as recently described for Mango RNA (Trachman et al. 2017a). Therefore, future work may lead to the isolation of fluorogenic aptamers with folding properties even surpassing that of iSpinach. Finally, the present structure will also be of great interest for future experiments aimed at engineering the aptamer, for instance to convert iSpinach into an allosteric fluorogenic RNA.

MATERIALS AND METHODS

iSpinach RNA synthesis

The different primers used are summarized in Supplemental Table S1 and were obtained from Integrated DNA Technologies. For each construct, 0.1 pmol of template was introduced in 100 µL of a PCR mixture containing 0.2 mM of each dNTP, 0.2 µM of each primer (forward and reverse), 0.1 U/µL of DreamTaq (ThermoFisher), and the corresponding buffer at the recommended working concentration. The mixture was then thermocycled starting with an initial denaturation step of 30 sec at 95°C followed by 25 cycles of 5 sec at 95°C and 30 sec at 60°C. Upon thermocycling, the 100 µL of PCR mixture were introduced in 1 mL of in vitro transcription mixture containing 2 mM of each NTP, 25 mM $MgCl_2$, 40 mM Tris–HCl pH 8.0, 5 mM DTT, 1 mM Spermidine, 70 µg/mL T7 RNA polymerase, and 5 µg/mL inorganic pyrophosphatase (Roche). After 6 h of incubation at 37°C, 10 units of Baseline-Zero DNase (Epicentre) and the corresponding buffer were added to the mixture and a second incubation of 1 h at 37°C was performed. RNA was recovered by phenol extraction followed by an ethanol precipitation in the presence of 300 mM sodium acetate pH 5.5 (Sigma-Aldrich) and 3 volumes of ice cold 100% ethanol overnight at −20°C. After centrifugation and a wash in 70% ethanol, the pellets were dissolved in denaturing buffer (0.05% bromophenol blue, 0.05% xylene cyanol, 20% glycerol, TBE 1×, 8 M urea) and the solution loaded on a 12% denaturing 8 M urea acrylamide/bisacrylamide gel. The piece of gel containing the RNA was identified by UV shadowing, sliced from the gel, and transferred into a BioTrap device (GE Healthcare) filled with TBE buffer. RNA was eluted for 3 h at 200 V and 4°C prior to being precipitated in the presence of 300 mM sodium acetate pH 5.5 (Sigma-Aldrich) and 3 volumes of ice cold 100% ethanol overnight at −20° C. RNA was then recovered by centrifugation, the pellet washed with 70% ethanol prior to being resuspended in deionized DEPC treated-water (Roth). RNA was finally quantified by OD measurement at 260 nm using a NanoDrop device.

As a quality control, the RNA was analyzed on an 8 % denaturing 8 M urea acrylamide/bisacrylamide gel. Moreover, 2 µM of RNA were renaturated in water by 2 min of warming at 85°C followed by a cooling of 5 min at 25°C using a thermocycler before adding a volume of 2× incubation buffer (80 mM Tris–HCl pH 7.5, 2 mM $MgCl_2$, and 200 mM KCl) and DFHBI (Lucerna Technologies) to a final concentration of 1.2 µM. Finally, the fluorescence was measured in a real-time Thermocycler (Mx3005P, Agilent) on the FAM channel (ex: 492 nm/em: 516 nm) of the device.

RNA/DFHBI cocrystallization

A solution of 210 µM of RNA was renaturated as described above and mixed 1:1.2 with DFHBI in a solution containing 40 mM Tris–HCl pH 7.5, 1 mM $MgCl_2$, and 100 mM KCl.

In crystallization conditions consisting of Li_2SO_4 (1.9–1.6 M) MES pH 5.4 $MgCl_2$ (15–5 mM) at 20°C, both iSpinach-D5 and iSpinach-D5-G30–A32 crystallized with the same shape: lentils/ball shape. A second set of crystals was obtained in MPD (28%–34%), sodium spermine (10–14 mM), 12 mM Tris–HCl pH 7.4, and NaCl (40–90 mM) at 20°C and 30°C, but with two different shapes: For iSpinach-D5, only needles were observed, while iSpinach-D5-G30–A32 had a laminar shape crystal. In this crystallization condition, small elliptic crystals were also found with iSpinach-D5-G30–G32.

A solution of 2.0 M Li_2SO_4, 50 mM MES pH 5.4, 5 mM $MgCl_2$ and 20% glycerol was used as cryo-protector for the first crystallization conditions; while a solution of 36% MPD, 12 mM sodium spermine, 40 mM sodium cacodylate HCl pH 7.0, 60 mM NaCl and 10% glycerol was used for the second one. Samples in both conditions were frozen in liquid ethane and stored in liquid nitrogen.

Data collection and processing

Data were collected at the X06DA beamline at the SLS using a PILATUS 2M detector. The wavelength was set to 1.8 Å. Several data sets were collected for "lentils/ball shape" crystals for iSpinach-D5 and iSpinach-D5-G30–A32 constructs as well as for iSpinach-D5-G30–A32 with "laminar" crystal shape. Whereas most of the crystals had very poor (or no) diffraction (Supplemental Table S2), the "laminar" crystal obtained with iSpinach-D5-G30–A32 allowed the structure of the complex to be solved. All data sets collected from the "laminar" crystal were processed and scaled independently with XDSGUI (Kabsch 2010). All of them showed the same space group and cell dimension, therefore all were merged together in one data set for the refinement process with XDSGUI software. For phase determination, one of the Spinach structures solved previously (PDB code: 4Q9R.pdb, Huang et al. 2014) was used as a search model for molecular replacement using PHASER (PHENIX) (Adams et al. 2010). The final structure was obtained after alternative cycles of refinement with PHENIX and manual refinement with COOT (Supplemental Table S3; Emsley and Cowtan 2004). The final refined model was deposited on the PDB under the reference 5OB3.

iSpinach fluorescence comparison between solution and crystal

The crystals obtained with iSpinach-D5-G30–A32 construct in both crystallographic conditions (Li_2SO_4 and MPD) were washed 3–4 times in respective crystallographic condition from the reservoir (Supplemental Table S2) in order to remove RNA molecules left in the solution. A single crystal was loaded into a Crystal Clear Special Glass 10 capillary (Hampton Research, 0.4 mm inner diameter).

For complexes kept in solution, 2 µM of iSpinach-D5-G30–A32 was mixed with 2.4 µM DFHBI in 40 mM Tris–HCl pH 7.5, 1 mM $MgCl_2$ and 100 mM KCl and, afterwards, mixed at 50% with their respective crystallogenesis buffer (Supplemental Table S2). The mixture was then emulsified into 100 pL water-in-oil droplets using Novec 7500 (3M) fluorinated oil supplemented with 2% of fluorosurfactant and using a microfluidic device as described previously (Autour et al. 2016). The collected emulsions were then loaded into the same capillary as those used with crystals. Fluorescence intensity decay was monitored by placing the loaded capillaries onto the platform of an epi-fluorescence microscope (TiE, Nikon) and exposing it to a constant illumination at 475 nm at the maximum intensity of the light source (Lumencor). An Orca Flash 4 camera using a proper filter set was used to collect the light emitted at 490–530 nm and the fluorescence was quantified using custom-made analysis software written in MatLab.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Redmond Smyth for his help in manuscript proofreading and Guillaume Bec for his assistance in using the Mosquito liquid handling robot. This work has been published under the framework of the LABEXs: ANR-10-LABX-0036_NETRNA and ANR-11-LABX-0057_MITOC ROSS, and benefits from funding by the state-managed French National Research Agency as part of the Investments for the Future program. We also received the financial support of the Agence Nationale de la Recherche (project BacNet: ANR-10-BINF-02-01), the Université de Strasbourg, and the Centre National de la Recherche Scientifique.

Received July 16, 2017; accepted September 5, 2017.

REFERENCES

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213–221.

Ageely EA, Kartje ZJ, Rohilla KJ, Barkau CL, Gagnon KT. 2016. Quadruplex-flanking stem structures modulate the stability and metal ion preferences of RNA mimics of GFP. *ACS Chem Biol* **11:** 2398–2406.

Anna SL, Bontoux N, Stone HA. 2003. Formation of dispersions using "flow focusing" in microchannels. *Appl Phys Lett* **82:** 364–366.

Autour A, Westhof E, Ryckelynck M. 2016. iSpinach: a fluorogenic RNA aptamer optimized for in vitro applications. *Nucleic Acids Res* **44:** 2491–2500.

Aw SS, Tang MX, Teo YN, Cohen SM. 2016. A conformation-induced fluorescence method for microRNA detection. *Nucleic Acids Res* **44:** e92.

Babendure JR, Adams SR, Tsien RY. 2003. Aptamers switch on fluorescence of triphenylmethane dyes. *J Am Chem Soc* **125:** 14716–14717.

Bose D, Su Y, Marcus A, Raulet DH, Hammond MC. 2016. An RNA-based fluorescent biosensor for high-throughput analysis of the cGAS-cGAMP-STING pathway. *Cell Chem Biol* **23:** 1539–1549.

Davis AR, Znosko BM. 2007. Thermodynamic characterization of single mismatches found in naturally occurring RNA. *Biochemistry* **46:** 13425–13436.

Dolgosheina EV, Unrau PJ. 2016. Fluorophore-binding RNA aptamers and their applications. *Wiley Interdiscip Rev RNA* **7:** 843–851.

Emsley P, Cowtan K. 2004. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132.

Ennifar E, Nikulin A, Tishchenko S, Serganov A, Nevskaya N, Garber M, Ehresmann B, Ehresmann C, Nikonov S, Dumas P. 2000. The crystal structure of UUCG tetraloop. *J Mol Biol* **304:** 35–42.

Enterina JR, Wu L, Campbell RE. 2015. Emerging fluorescent protein technologies. *Curr Opin Chem Biol* **27:** 10–17.

Filonov GS, Moon JD, Svensen N, Jaffrey SR. 2014. Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J Am Chem Soc* **136:** 16299–16308.

Grate D, Wilson C. 1999. Laser-mediated, site-specific inactivation of RNA transcripts. *Proc Natl Acad Sci* **96:** 6131–6136.

Guet D, Burns LT, Maji S, Boulanger J, Hersen P, Wente SR, Salamero J, Dargemont C. 2015. Combining Spinach-tagged RNA and gene localization to image gene expression in live yeast. *Nat Commun* **6:** 8882.

Hallberg ZF, Wang XC, Wright TA, Nan B, Ad O, Yeo J, Hammond MC. 2016. Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3′, 3′-cGAMP). *Proc Natl Acad Sci* **113:** 1790–1795.

Han KY, Leslie BJ, Fei J, Zhang J, Ha T. 2013. Understanding the photophysics of the spinach-DFHBI RNA aptamer-fluorogen complex to improve live-cell RNA imaging. *J Am Chem Soc* **135:** 19033–19038.

Huang H, Suslov NB, Li NS, Shelke SA, Evans ME, Koldobskaya Y, Rice PA, Piccirilli JA. 2014. A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat Chem Biol* **10:** 686–691.

Huang K, Doyle F, Wurz ZE, Tenenbaum SA, Hammond RK, Caplan JL, Meyers BC. 2017. FASTmiR: an RNA-based sensor for in vitro quantification and live-cell localization of small RNAs. *Nucleic Acids Res* **45:** e130.

Kabsch W. 2010. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66:** 133–144.

Kellenberger CA, Wilson SC, Sales-Lee J, Hammond MC. 2013. RNA-based fluorescent biosensors for live cell imaging of second messengers cyclic di-GMP and cyclic AMP-GMP. *J Am Chem Soc* **135:** 4906–4909.

Kellenberger CA, Chen C, Whiteley AT, Portnoy DA, Hammond MC. 2015a. RNA-based fluorescent biosensors for live cell imaging of second messenger cyclic di-AMP. *J Am Chem Soc* **137:** 6432–6435.

Kellenberger CA, Wilson SC, Hickey SF, Gonzalez TL, Su Y, Hallberg ZF, Brewer TF, Iavarone AT, Carlson HK, Hsieh YF, et al. 2015b. GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. *Proc Natl Acad Sci* **112:** 5383–5388.

Ketterer S, Fuchs D, Weber W, Meier M. 2015. Systematic reconstruction of binding and stability landscapes of the fluorogenic aptamer spinach. *Nucleic Acids Res* **43:** 9564–9572.

Mishin AS, Belousov VV, Solntsev KM, Lukyanov KA. 2015. Novel uses of fluorescent proteins. *Curr Opin Chem Biol* **27:** 1–9.

Molinaro M, Tinoco I Jr. 1995. Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: thermodynamic and spectroscopic applications. *Nucleic Acids Res* **23:** 3056–3063.

Nakayama S, Luo Y, Zhou J, Dayie TK, Sintim HO. 2012. Nanomolar fluorescent detection of c-di-GMP using a modular aptamer strategy. *Chem Commun (Camb)* **48:** 9059–9061.

Ong WQ, Citron YR, Sekine S, Huang B. 2017. Live cell imaging of endogenous mRNA using RNA-based fluorescence "Turn-On" probe. *ACS Chem Biol* **12:** 200–205.

Ouellet J. 2016. RNA fluorescence with light-up aptamers. *Front Chem* **4:** 29.

Paige JS, Wu KY, Jaffrey SR. 2011. RNA mimics of green fluorescent protein. *Science* **333:** 642–646.

Paige JS, Nguyen-Duc T, Song W, Jaffrey SR. 2012. Fluorescence imaging of cellular metabolites with RNA. *Science* **335:** 1194.

Pauff S, Withers JM, McKean IJ, Mackay SP, Burley GA. 2017. Synthetic biological approaches for RNA labelling and imaging: design principles and future opportunities. *Curr Opin Biotechnol* **48:** 153–158.

Ryckelynck M, Baudrey S, Rick C, Marin A, Coldren F, Westhof E, Griffiths AD. 2015. Using droplet-based microfluidics to improve the catalytic properties of RNA under multiple-turnover conditions. *RNA* **21:** 458–469.

Sato S, Watanabe M, Katsuda Y, Murata A, Wang DO, Uesugi M. 2015. Live-cell imaging of endogenous mRNAs with a small molecule. *Angew Chem Int Ed Engl* **54:** 1855–1858.

Sharma S, Zaveri A, Visweswariah SS, Krishnan Y. 2014. A fluorescent nucleic acid nanodevice quantitatively images elevated cyclic adenosine monophosphate in membrane-bound compartments. *Small* **10:** 4276–4280.

Song W, Strack RL, Jaffrey SR. 2013. Imaging bacterial protein expression using genetically encoded RNA sensors. *Nat Methods* **10:** 873–875.

Strack RL, Disney MD, Jaffrey SR. 2013. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat Methods* **10:** 1219–1224.

Su Y, Hickey SF, Keyser SG, Hammond MC. 2016. In vitro and in vivo enzyme activity screening via RNA-based fluorescent biosensors for *S*-adenosyl-L-homocysteine (SAH). *J Am Chem Soc* **138:** 7040–7047.

Trachman RJ III, Demeshkina NA, Lau MWL, Panchapakesan SSS, Jeng SCY, Unrau PJ, Ferré-D'Amaré AR. 2017a. Structural basis for high-affinity fluorophore binding and activation by RNA Mango. *Nat Chem Biol* **13:** 807–813.

Trachman RJ III, Truong L, Ferré-D'Amaré AR. 2017b. Structural principles of fluorescent RNA aptamers. *Trends Pharmacol Sci* **38:** 928–939.

Wang XC, Wilson SC, Hammond MC. 2016. Next-generation RNA-based fluorescent biosensors enable anaerobic detection of cyclic di-GMP. *Nucleic Acids Res* **44:** e139.

Warner KD, Chen MC, Song W, Strack RL, Thorn A, Jaffrey SR, Ferré-D'Amaré AR. 2014. Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat Struct Mol Biol* **21:** 658–663.

You M, Litke JL, Jaffrey SR. 2015. Imaging metabolite dynamics in living cells using a Spinach-based riboswitch. *Proc Natl Acad Sci* **112:** E2756–E2765.

Zhang J, Fei J, Leslie BJ, Han KY, Kuhlman TE, Ha T. 2015. Tandem spinach array for mRNA imaging in living bacterial cells. *Sci Rep* **5:** 17295.

Zheng G, Lu XJ, Olson WK. 2009. Web 3DNA—a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res* **37:** W240–W246.

Zheng H, Cooper DR, Porebski PJ, Shabalin IG, Handing KB, Minor W. 2017. CheckMyMetal: a macromolecular metal-binding validation tool. *Acta Crystallogr D Struct Biol* **73:** 223–233.

The work presented in this publication allowed us to explain the positive contribution of 4 out of the 5 mutations displayed by iSpinach. Indeed, and as anticipated earlier, changing Us for Cs in basal and apical stems converts weak G-U into stronger G=C base pairs with the net effect of reinforcing the stability of the helices.

More interestingly, this work showed that exchanging the weak AoA base pair located underneath the G-quadruplex domain by the stronger GoG pair may strongly stabilize the overall structure of the G-quadruplex domain. Indeed, both Gs form extensive contact partly involving water molecules that also coordinate the potassium ion stabilizing the two bottom G-quartet layers. This may result in the complete locking of the G-quartet structure and is likely to be one of the main factors explaining the improved folding efficiency displayed by iSpinach. Such an effect would have been difficult to predict *a priori* and further demonstrates how powerful ultrahigh-throughput screening can be in identifying mutations acting synergistically.

The last question left to be answered is the origin of the fantastic fluorescence increase observed in the presence of sodium. We obtained crystals in the presence sodium. However, they were only poorly diffractive giving a very low resolution at the level of the G-quartets core region whereas both the basal and the apical stems were well resolved. Interestingly, the A17U mutation was found to play a role in the higher fluorescence observed in the presence of sodium, since reverting this nucleotide to the wild type suppresses “sodium effect” (see Figure 3E on **Original Article 1**). Moreover, this mutation could also work in cooperation with the GoG pair discussed in the previous paragraph since its reversion to AoA also suppresses “sodium effect” (see Figure 3G on **Original Article 1**). Beside its involvement in the “sodium effect”, the A17U mutation was also found to be involved in crystal packing since this U nucleotide establishes key interactions within the crystal that holds the central (G-quartet core) region in place and makes it resolved in the final structure. Therefore, and based on these observations, it is tempting to propose that in the presence of sodium a structural change involving U17 occurs making the residue unavailable to form contacts in the crystal that are key to hold the G-quartet domain in a fixed position. As a consequence, the central region of the aptamer its partly disordered and is not resolved when exposed to X-ray.

To conclude, this crystal structure partly explains the superior folding properties of iSpinach in the presence of potassium and suggests that the basal stem of the aptamer is properly locked into a very stable structure. This robust bottom part of the aptamer can now serve as a platform for the development of allosteric biosensors in which a sensing aptamer together with a communication module would be inserting into the apical part of iSpinach (see the next section).

5.4. Engineering iSpinach in a semi-automated way

The great stabilization of the basal part of iSpinach revealed by the crystal structure should confer an overall folding efficiency to the rest of the molecule and potentially act as a scaffold. In the next step of my work, I focused my attention on the apical part of the molecule and started to engineer it into a fluorogenic biosensor of metabolites.

5.4.1. RNA-based sensors of metabolites

Metabolites are important biological molecules involved in plenty of key cellular functions such as energy production and storage, signaling, cell-to-cell communication or even gene expression regulation. Therefore, being able to image metabolites would make possible characterizing these different functions in a dynamic and non-invasive way. Chemical and protein-based fluorescent probes of metabolite have been described but they are extremely challenging to design (Jaffrey, 2017). For example, calcium ions can be sensed using the fluorogenic Fura-2 chemical (Grynkiewicz et al., 1985). Alternatively, the cellular calcium content can be measured using a fusion protein made of a calmodulin domain surrounded by two fluorescent proteins (e.g. CFP and YFP) forming a FRET pair (see §2.2.1.1.). Calcium binding to the calmodulin domain leads to an increased FRET signal. However, developing such sensors is labor-intensive and time-consuming as it requires long synthetic chemistry or molecular engineering to be performed. Moreover, the lack of efficient informatic folding predictive tools further complexifies the development of protein-based probes. On the contrary, several bioinformatic algorithms have been developed to predict RNA folding. This, together with the great capacity of RNA aptamers to specifically recognize small molecules and potentially change their structure upon ligand binding (see Riboswitch section §1.2.2.2.), as well as the existence of light-up RNA aptamers, makes RNA good candidate for the development of fluorogenic biosensors of metabolites.

5.4.1.1. Ribozyme-based sensors

RNA-based sensors share the overall architecture of riboswitches as they are made of three parts: i) a sensor aptamer, ii) a communication (also called actuator) module and iii) an effector (Figure 31.A). The first sensors used hammerhead ribozyme (a catalytic RNA) as effector module (Breaker, 2002). Here, the communication module plays a key role. Indeed, in the absence of metabolite, the communication module has to adopt a folding that destabilizes the ribozyme and prevents its self-cleaving activity. Conversely, the presence of metabolites should induce a conformational change of the communication module leading to ribozyme activation. Inserting such self-cleaving molecules into target mRNA, makes possible adjusting gene expression to the presence of a target metabolite(Win and Smolke, 2007). Moreover, ribozymes can be engineered to be active *in trans* on a fluorogenic RNA substrate as this was done in our group with the X-motif (Ryckelynck et al., 2015). Nevertheless, ribozyme-based sensors are unidirectional since the reaction they catalyze is not reversed upon ribozyme deactivation (i.e. metabolite removal). However, significant gain in dynamics can however been obtained by using light-up aptamers as effector module.

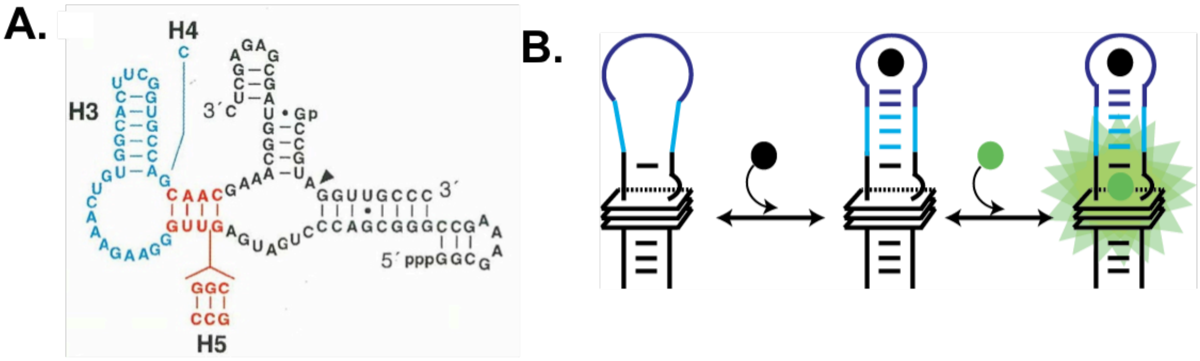


Figure 31: RNA biosensor containing 3 parts: Sensing part, communication module and an effector. (A) ATP-sensor based on hammerhead ribozyme. The ATP aptamer (in light blue) acting as sensing part is linked to a communication module (in red). Upon ATP sensing, the communication module transduces the signal to the ribozyme (in black) leading to its self-cleavage (the black arrow indicates the cleavage site). Figure adapted from Tang and Breaker, 1997. **(B)** Schematic representation of an RNA fluorogenic sensor. A light-up domain (iSpinach in black) is coupled to a sensing aptamer (dark blue) via a communication module (in light blue). When the sensed molecule (black dot) is present, the communication module gets stabilized an restore the fluorogen-binding capacity of the light-up aptamer, leading to fluorescence emission.

5.4.1.2.RNA-based fluorogenic sensors

Sensors using light-up RNA aptamers as effector domain are expected to circumvent limitations of ribozyme-based one while sharing similar architecture (Figure 31.B). In this scheme, the communication module stabilizes the structure of a light-up aptamer which becomes then able to interact and activate the fluorescence of its cognate fluorogen. Interestingly, one can imagine that a decrease in the cellular concentration of the sensed metabolite may lead to its release from the sensor, which in turn may lead to the release of the fluorogen and sensor extinction. The first fluorogenic sensor was based on MG aptamer coupled with FMN specific aptamer (Stojanovic and Kolpashchikov, 2004). Later on, Jaffrey's lab engineered Spinach and converted it into a sensor of S-adenosyl-methionine and showed for the first time that such sensors could indeed be used *in vivo* (Paige et al., 2012). Since then, many of Spinach-derived fluorogenic sensors have been developed especially the groups of Jaffrey (Jaffrey, 2017) and Hammond (Hallberg et al., 2017).

There are two main difficulties when developing such sensors. First, the sensing aptamer needs to undergo a structural remodeling upon ligand binding to be able to transmit the recognition event to the effector domain. Such structural plasticity is naturally found in riboswitches aptamers. Therefore, these molecules have been widely used in the design of fluorogenic sensors (Hallberg et al., 2017). Artificial aptamers are more complex to adapt. Indeed, they are usually developed by SELEX (see §2.2.2.2.) and are isolated only for their capacity to interact with a target molecule. Therefore, such aptamers are usually rigid structures that fits the shape of the target and do not undergo structural remodeling. However, using alternative SELEX procedure such as Capture-SELEX makes possible isolating structurally flexible aptamer (Stoltenburg et al., 2012). Capture-SELEX was initially developed to isolate DNA aptamers but can be adapted to the selection of RNAs. In this scheme, the RNA library is made of two randomized regions interspaced by a constant region (Figure 32).

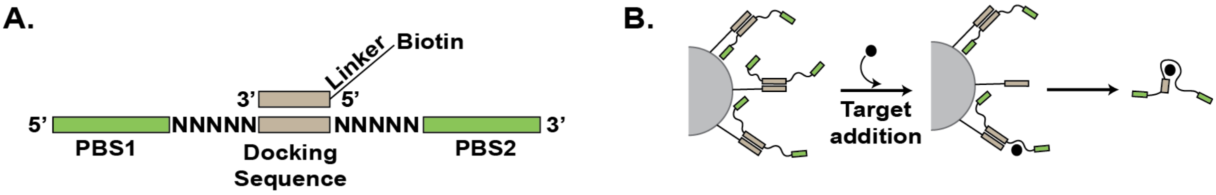


Figure 32: Principle of capture-SELEX. (A) Library containing primer binding sequence (PBS in green) and random region (pictured by the N repetition) anneals to capture primers conjugated via docking sequence (in light brown). **(B)** Capture primers conjugated with biotin are immobilized onto bead with the library. After adding the target of interest, only the sequences able to switch their structure upon target recognition are eluted and enriched.

The library is attached to a resin *via* an oligonucleotide annealed to the central constant region. When incubated in the presence of the target metabolite, the aptamers able to interact with the metabolite will form a complex and, if the binding event triggers a conformational change, will be released and eluted from the resin. The second difficulty is linked to the proper design of the communication module. Indeed, to be optimal, this module should keep the effector domain completely extinct in absence of the target, whereas it should turn it completely on in the presence of the target. This domain is usually developed through a "trial-and-error" strategy in which sequences are more or less arbitrarily chosen based on thermodynamic consideration. Even though valuable, such an approach is time-consuming and a may not return the optimal sequence since only a handful of communication modules are assayed. However, the throughput of μIVC makes it well suited to exhaustively assay the whole sequence space of communication modules made of 4 (65,536 (4^8) different sequences) of 5 (1,048,576 (4^{10}) different sequences) base pairs. Moreover, coupling such screening with NGS should make it possible to automate the post-selection analysis.

5.4.1.3.Nucleic acids selection assisted by Next Generation Sequencing

Next Generation Sequencing (NGS) technologies have already proven to be very useful in assisting nucleic acids selection. Indeed, monitoring SELEX process by NGS makes possible the early detection of aptamers of interest and this way significantly shorten SELEX duration to only a few rounds (Quang et al., 2016). This strategy was initially applied to RNA by Pitt and Ferré-D'Amaré who showed that NGS profiling of SELEX-enriched libraries could be used to construct the empirical fitness landscape of an RNA ligase ribozyme without requiring further biochemical measurement of the phenotype. More recently, the group of Yokobashi used a similar approach to profile the phenotype of several thousands of self-cleaving ribozyme mutants in a single experiment (Akter and Yokobayashi, 2015; Kobori and Yokobayashi, 2016) but also to isolate the most representative communication module to be used with a ribozyme (Kobori et al., 2015). The abovementioned examples involved self-modifying ribozymes, which represents only a small sub-set of existing functional RNAs. Interestingly, μIVC was also coupled with NGS in a proof-of-concept work aiming at profiling mutations tolerated by the protein enzyme (β-glucosidase) and allowed isolating beneficial mutations (Romero et al., 2015). Here, a mutant gene library was generated and expressed into bacteria later encapsulated and lyzed into water-in-oil droplets in the presence of a fluorogenic substrate of the enzyme. Then, droplets were sorted according to their fluorescence and both starting and sorted libraries were sequenced by NGS. Measuring sequence enrichment and depletion allowed identifying negative and positive mutations and doing so to identify a few mutations that allowed improving enzyme thermostability.

5.4.2.iSpinach allosteric library design

According to what precedes, a significant post-selection analysis time could be saved by analyzing μIVC output using NGS. I decided to verify the assertion by functionally and exhaustively exploring the whole sequence space of a communication module made of 4 base pairs (~ 65,000 variants) connecting the theophylline-sensing aptamer and iSpinach. Theophylline aptamer (Jenison et al., 1994) is frequently used for such proof-of-concept experiment because it is small (25 nucleotides), well characterized, it undergoes a structural change upon ligand binding and enables precise control *in vivo* since this system is bio-orthogonal (i.e. the molecule and its aptamer do not cross react with cell compounds).

5.4.2.1. Optimizing iSpinach sequence for sensor discovery

As highlighted above, iSpinach is expected to have a pretty stable basal part thanks to the GoG non-canonical pair that we proposed to lock the G-quadruplex domain. Therefore, I decided to insert the theophylline aptamer and the communication module at the apical part of iSpinach as previously done with other biosensors based on Spinach (Hallberg et al., 2017; Song et al., 2013). Moreover, as stated above, it is important not only that a sensing event leads to the full activation of the light-up domain but also that the latter is kept completely extinct in the absence of target. Therefore, to further weaken the apical part of the DFHBI-binding pocket, without compromising iSpinach properties, I tested several permutations of the strong C=G base pair found immediately on the top of the pocket (Figure 33.A).

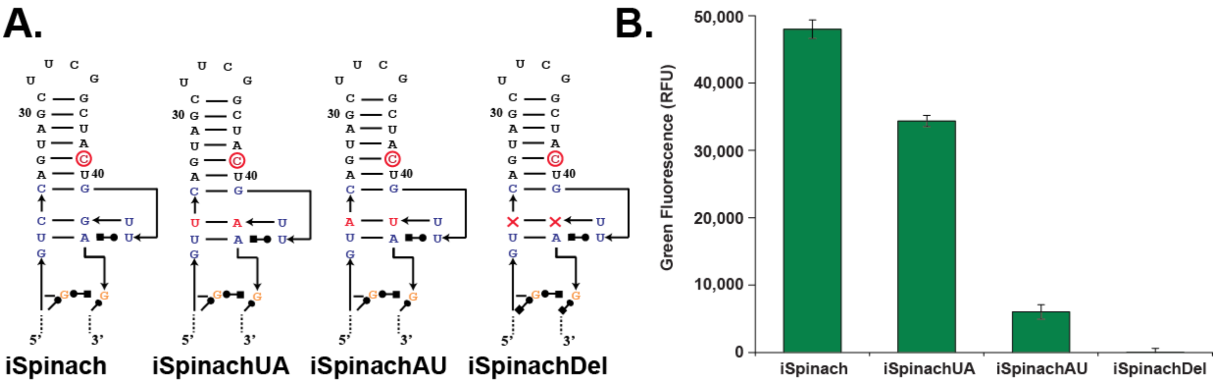


Figure 33: Introducing weakening mutations in iSpinach. (A) Sequence of the upper of three iSpinach variants, C=G base pair (C24&G44) was mutated or deleted (residues in red). **(B)** Comparisonof the fluorescence capacity of iSpinach and the three mutants. 500 nM RNA were renatured in water prior to being incubated in a buffer containing 10 mM $MgCl_2$, 10 μM DFHBI-1T, 100 mM KCl, 40 mM Tris-HCl pH 8.0. Then, the green fluorescence was measured at 25 °C. Errors bars represent standard deviation of 3 different experiments.

Changing the C=G for a weaker U-A still preserve ~ 70% of wild type fluorescence. Interestingly, introducing an A-U pair or deleting the base pair had a dramatic effect by almost entirely compromising DFHBI-binding capacity of the aptamer. We explain these effects by the disturbance of the U-A-U triplex that forms the ceiling of the DFHBI-binding pocket. Therefore, I decided to continue the design of the theophylline sensor using the U-A iSpinach variant and to connect it to the theophylline aptamer *via* a communication module made of 4 completely randomized base pairs (Figure 34). This made a library containing a diversity of 65,536 different sequences that needed to be functionally screened for its switching capacity in the presence of theophylline.

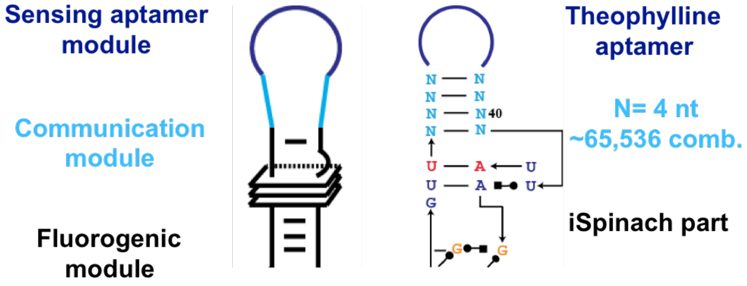


Figure 34: Schematic representation of iSpinach allosteric library. The fluorogenic module iSpinach-UA has been connected to the theophylline aptamer by communication module consisting of 4 randomized base pairs.

5.4.2.1. Experimental strategy

Since I also wanted to automate the post-selection analysis using NGS, I first had to barcode each DNA molecule of the library prior to screening them using a μIVC pipeline allowing for sorting properly switching sensors from those unresponsive. A material and methods section describing the process is available as Annex 1.

5.4.2.1.1. Gene library barcoding

To gain in measurement accuracy during NGS analysis, a Unique Droplet Identifier (UDI) sequence was appended to each gene prior to its encapsulation into PCR droplets using a primer displaying a randomized region of 20 nucleotides. The high theoretical diversity of UDIs (4^{20}, so 10^{12} different combinations) makes each molecule unique in the library (Figure 35). Therefore, every molecule originating from the same droplet are expected to share the same UDI. As a consequence, and similarly to what is done with UMI (see §3.2.2.2.), upon droplet selection and DNA sequencing, counting the number of different UDI for each sequence will give access to the number of sorted droplets containing this sequence, and so to precisely compute an enrichment factor for each sequence (see below). Moreover, this barcode was reinitialized prior to starting a new round of μIVC.

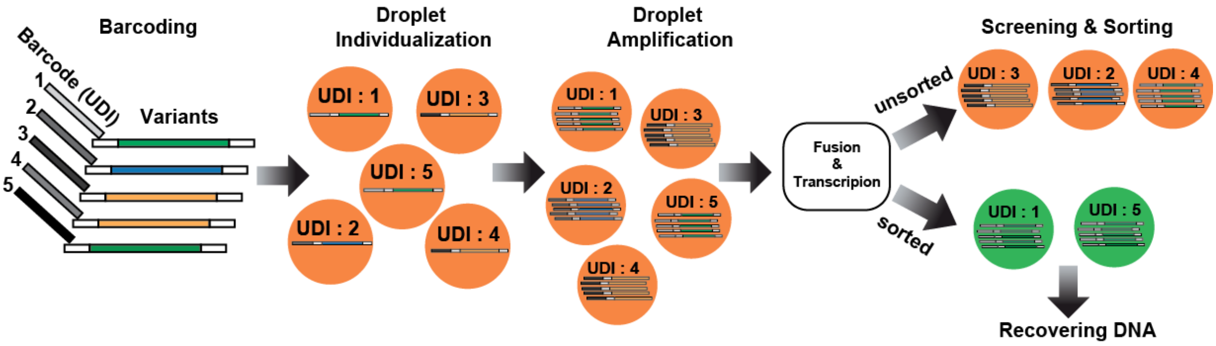


Figure 35: Schematic barcoding strategy. Prior to encapsulation, a Unique Droplet Identifiers (UDI) containing a 20 nucleotide-long randomized sequences was appended to each variants by PCR. Barcoded DNA were individualized into droplets, amplified and transcribed; and the droplets were sorted. Upon microfluidic screening, droplets of interest were broken and the DNA recovered. Before each new round a novel barcoding step was performed to re-initialize the UDI.

5.4.2.1.2. μIVC strategy for selecting switching molecules

The μIVC-screening process was similar to the one I used to isolate iSpinach (see §5.2.1.). Each barcoded variant of the library was diluted into a PCR mixture prior to being individualized into water-in-oil droplets and being thermocycled. Then, amplified DNA-containing emulsion was re-injected into the second chip that allowed merging each droplet with a second one containing an *in vitro* transcription mixture (IVT) supplemented with the fluorogenic dye and optionally theophylline. Upon an off-chip incubation allowing the transcription step to proceed, the droplets were analyzed and sorted according to their fluorescence intensity using the last chip (FADS).

To search for switching RNAs, such screening proceeds in two different regimes: i) a positive selection during which the metabolite, here the theophylline, is added to the IVT mixture and the most fluorescent droplets are sorted. This should allow selecting for sensors that turn on in the presence of theophylline. ii) a negative selection during which the metabolite is omitted from the IVT mixture and the least fluorescent droplets are sorted. This should allow selecting for sensors that stay off in the absence of theophylline. Alternating rounds of positive and negative selections is expected to gradually favor the isolation of theophylline-responding sensors. Finally, at the end of the process the DNA-content of the different libraries is analyzed by NGS and the sequences of interest are identified by bioinformatics.

5.4.2.2. iSpinach-based sensor library screening

The barcoded library of iSpinach-based sensors was subjected to five rounds of screening starting with a first positive selection, followed by three rounds of negative selections and concluded by a round of positive selection (Figure 36 and Table 2). Upon each round, I validated proper library enrichment by expressing the resulting library *in vitro* using an IVT mixture supplemented in DFHBI-1T and in presence or absence of theophylline. Real-time DFHBI fluorescence monitoring then allowed assessing library responsiveness to theophylline (Figure 36.A). This showed that, upon each round the library, responsiveness followed the tendency expected from the selection mode used. Indeed, whereas the starting library was poorly fluorescent, a single round of positive screening allowed to strongly enrich it in molecules fluorescing in presence and absence of theophylline. Yet, the overall fluorescence of library slightly decreased in the absence of theophylline. This suggests that in addition of the over-stabilized molecules that stay fluorescent even in the absence of theophylline and are therefore unresponsive, the pool should also contain a fraction of theophylline responsive sensors. To further enrich the library in responsive RNAs, three rounds (round 2 to 4 on Figure 36) of negative screening were performed. Upon each negative round, the population kept its capacity to fluoresce in the presence of theophylline but gradually lost its fluorogenicity in the absence of theophylline (see black curves on Figure 36.A). Finally, I performed a last round of positive screening to get rid of the residual over-extinct unresponsive molecules.

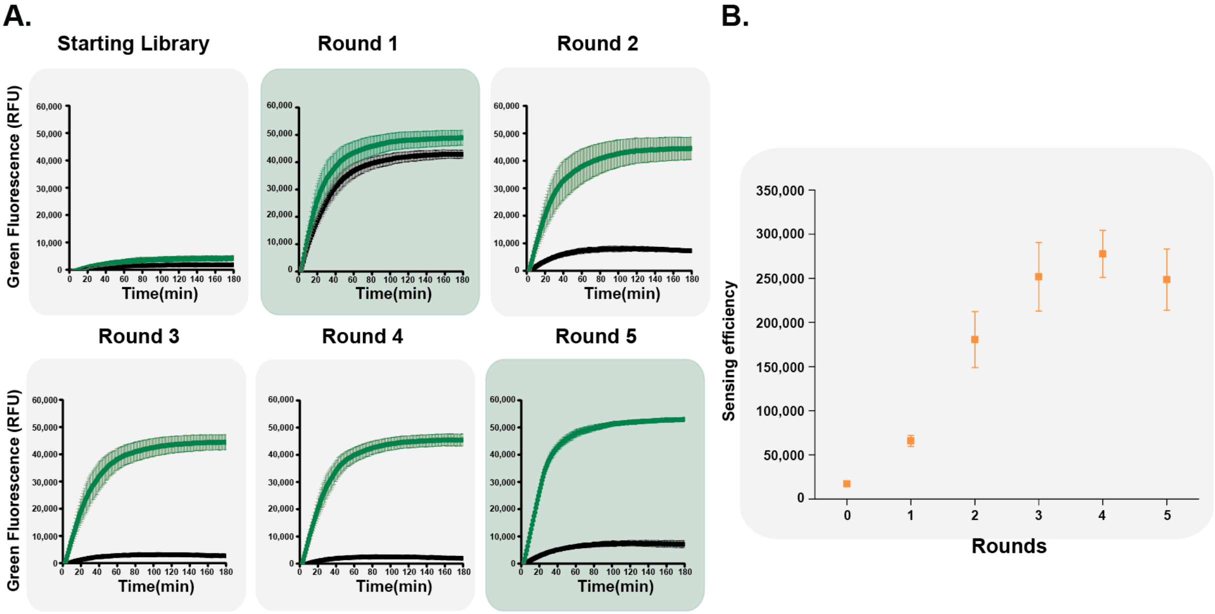


Figure 36: Analysis of library responsiveness to theophylline. (A) Monitoring by IVT transcription of gene libraries in presence of DFHBI-1T as well as in the presence (green dots) or in the absence (black dots) of theophylline. Rounds (1 and 5) performed in positive selection regime are shaded in green, whereas those performed in negative selection mode are shaded in gray. **(B)** Sensing efficiencies were computed. The values are the mean of three independent experiments and error bars correspond to ±1 standard error.

***Table 2** – Metrics of the µIVC screening used to search for theophylline responsive aptamers*

Round	**Temp. (°C)**	**λ value**	**Fusion efficiency (%)**	**Number of analyzed droplets**	**Number of analyzed variants**	**Number of sorted droplets**	**Droplet recovered**
1	37	0.10	90	3,500,000	315,000	3,500	1.11 %
2	37	0.10	87	555,000	48,285	17,645	36.54 %
3	37	0.12	89	700,000	74,760	15,000	20.06 %
4	37	0.06	95	1,300,000	74,100	17,000	22.94 %
5	37	0.09	90	800,000	64,800	8,700	13.42 %

To better appreciate and quantify the screening process, I computed a sensing efficiency factor of each library using equation (Eq. 3):

$$\text{Eq. 3: } Sensing\ Efficiency\ (S.E) = \left({Fluo_{wTheo}}/{Fluo_{w/oTheo}} \right) * Fluo_{wTheo}$$

where $Fluo_{wTheo}$ is the maximum fluorescence of the library in the presence of theophylline and $Fluo_{w/oTheo}$ is the maximum fluorescence of the library in the absence of theophylline. One can see that the S.E first increases and starts to plateau upon the third round (Figure 36.B). The two additional rounds did not significantly modify the phenotype of the library. Therefore, the screening was stopped after the fifth round, libraries were indexed using Nextera kit (Illumina) and analyzed by NGS on a MiSeq device.

5.4.2.3. Sequence analysis

Upon sequencing, 1 to 3 million reads were obtained for each library. Sequences were then extracted, filtered and analysed using a custom Python-written bioinformatic pipeline working in three main steps:

- 1. Extraction step & QC step: sequences were recovered from fastq files and only those fulfilling three criteria were conserved for the rest of the analysis: i) both T7 promoter and an UDI region must be present; ii) UDI should have a length of 20 nucleotides and iii) mutation should occur neither in the theophylline aptamer nor in the iSpinach moiety. Next, UDIs and communication modules were extracted and conserved for the next step.
- 2. Clustering & counting: Since UDI diversity largely exceeds the number of reads (UDI diversity of 1.10^{12} for a maximum of $3.6.10^{6}$ reads), we assumed that each UDI was unique in the starting library. Therefore, sequences sharing the same UDI were considered as coming from the same droplet and counted as a unique copy. Counting the number of different UDIs for a given sequence of Communication Module then gave access to the number of sorted droplets containing this sequence.
- 3. Noise removal, sequence clustering: Noise can arise in sequencing files as a result of contamination by air-borne DNA molecules lying in the lab, recombination events between PCR products or simply appearance of point mutations during DNA amplification. However, since these events are expected to occur at low frequency, they can easily be detected in data files as being low occurrence events. Therefore, a simple way of removing the noise is to conserve only the sequences present above a certain threshold (adjusted for each library). Applying this threshold allowed to conserve only relevant sequences. Next, we removed any sequence present after a round N but absent in round N-1. Indeed, since no mutagenesis was performed between two rounds, the appearance of such sequences would likely correspond to a point mutation event during the amplification step and was not considered.

Next, the occurrence frequency of each sequence was determined in each library and an enrichment rate was computed for each sequence of Communication Module as being the ratio between the occurrence frequency of the sequence at round N over the occurrence frequency of the sequence at round N-1. Finally, depending on how the enrichment factor evolved along the process, sequences were distributed into four different classes: i) sequences constantly enriching all along the process were clustered as Class I; ii) sequences constantly enriching only until the end of the negative selections were clustered as Class II; iii) sequences constantly enriching only until the first round of negative selection were clustered as Class III; and sequences that never really enriched were clustered as Class IV. Since sequence of interest were expected to enrich all along the screening process, optimal theophylline sensors were likely to be found in the Class I.

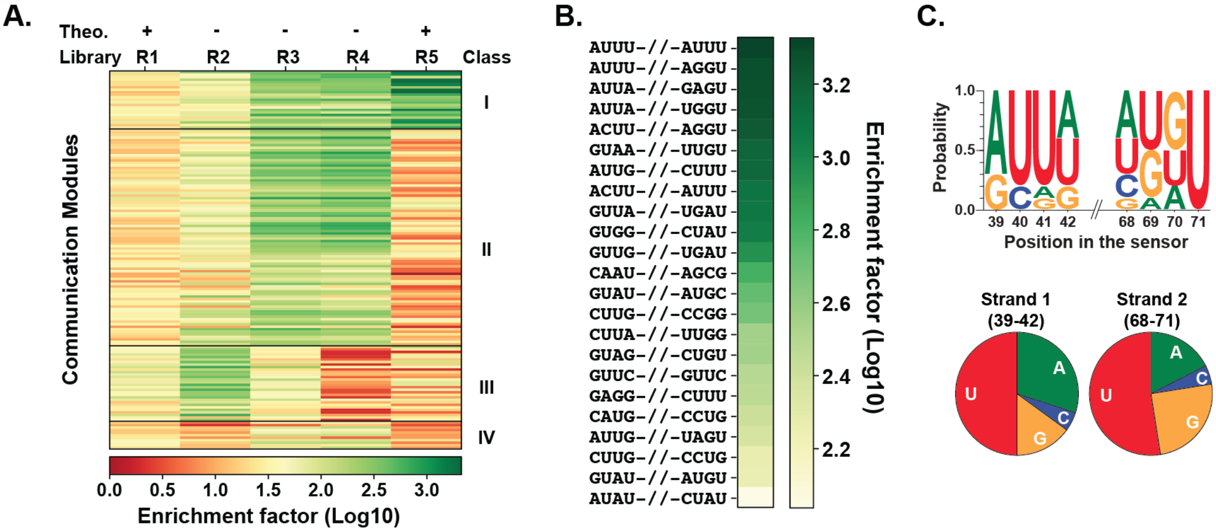


Figure 37: Bioinformatic analysis of NGS data from the selection of theophylline responsive aptamers. (A) Heat-map showing the enrichment evolution of the 150 variants isolated at the end of round 5. Enrichment factors were calculated using R0 as reference. Variants were clustered according to their enrichment profile. Class I contains variants that continuously enriched all along the process. (B) Sequences of Class I communication modules. (C) Logo representation and pie charts summarizing base composition of each strand of the communication module.

Closer analysis of Class I sequences revealed a significant bias in their composition (Figure 37.C). Indeed, on both strands of the communication module, at least 50 % of the residues were Us. Moreover, whereas As and Gs were found in similar proportions (30-18 % and 15-25 % respectively), Cs were strongly under-represented (5 %). Furthermore, there was a strong enrichment in non-canonical interactions (especially GU/UG/UU). Interestingly, such sequence bias is often found in thermosensors (Kortmann and Narberhaus, 2012), a class of structurally flexible RNAs involved in the temperature-controlled adjustment of gene expression (see §1.2.3.2.1.).

Moreover, it is interesting to note that, despite the fact that the Sensing Efficiency plateaus after the third round (Figure 36.B), among the 150 sequences found in the last round, only 23 properly responded all along the process (class I) and a strong sequence enrichment was still observed upon the last round.

5.4.2.4. Validation of the selected fluorogenic theophylline responsive sensors

To validate bioinformatics predictions, several sensors from Class I and II were prepared by transcription, gel-purified and their fluorescence was analyzed in presence or absence of theophylline. Finally, the Sensing Efficiency of each variant was calculated (see Eq. 3).

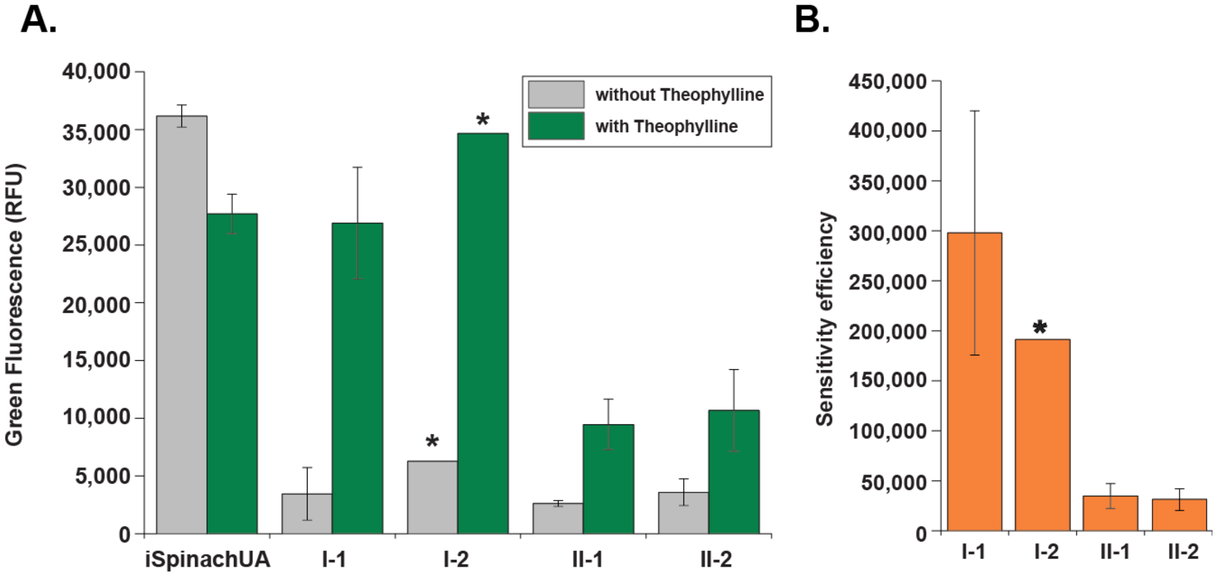


Figure 38: Responsiveness and sensing efficiency from R5 variants. **(A)** iSpinach-derived sensors fluorescence in the presence and in the absence of theophylline. RNAs were gel-purified and renatured in water, prior to being incubated in a buffer containing 40 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$ 5 μM DFHBI-1T, 100 mM KCl and ,optionally 50 μM theophylline. Green fluorescence was recorded every minute during 15 minutes at 25 °C. **(B)** Sensing efficiency was compute according to Eq.3. The values are the mean of two independent experiments and error bars correspond to ±1 standard error. Variants labeled with an asterisk were tested only once.

According to NGS analysis, Class I sensors should be the most efficient sequences since they enriched all along the process. Conversely, Class II should contain sensors that are efficiently extinct in the absence of theophylline but that poorly activate in the presence of theophylline since these sequences stopped enriching after Round 4. Indeed, fluorescence analysis indicates that Class I variants were be more efficient sensors that Class II (Figure 38). These data are still preliminary and need to be reproduced. However, we were pleased to see that the two most abundant Class I sensors (I-1 and I-2) were also those with the highest sensing efficiency. When looking closer, one can see that both sensors are almost completely extinct in the absence of theophylline whereas, in the presence of the co-factor they become almost as fluorescent as the unmodified iSpinach (Figure 38).

5.4.3. Conclusion on sensor selection

Even though this is still an on-going work, I showed that μIVC application scope can be extended to the selection of switching RNAs. Indeed, alternating rounds of positive and negative selections allows selecting molecule able to turn on in the presence of a target, whereas unresponsive ("always on" and "always off") molecules are progressively depleted from the pool. Beside the screening, using NGS profiling also makes possible sub-dividing molecules on the basis of their responsiveness by using their enrichment profile. This approach showed to be of good predictivity and makes the long post-selection cloning and hand-screening step dispensable. Indeed, starting from the ~ 65,000 possible sequence permutations of a four base-pairs communication module, I found that only a few dozens of them were likely to be optimal. Finally, only a handful of sequences had to be prepared and tested and, as suggested from their enrichment profiles, those ranked at the first position revealed to be the very efficient. Whereas in this work I analyzed ~ 65,000 variants with high coverage, our current screening limit of ~ 5 millions of droplets still leave room for increasing the sequence diversity to explore. For instance, we could pool libraries containing randomized communication modules of various lengths (2, 3 or 4 base pairs) and this way search for both optimal sequence and length.

Used in tandem with selection strategies such as Capture-SELEX, this new screening strategy would make possible the really fast discovery of optimized fluorogenic sensing sequences but it could also be extended to the design and development of new artificial riboswitches to could find, for instance, applications in synthetic biology.

5.5. Conclusive remarks on Spinach chapter

To conclude this first results chapter, my work demonstrated for the first time how powerful μIVC can be at isolating improved light-up RNA aptamers and derived biosensors. Indeed, using this approach I first developed iSpinach, which is, to the best of our knowledge, the most efficient DFHBI-binding aptamer in term of brightness, folding efficiency and thermal stability described to date. This robustness can partly be attributed to the exchange of the weak AoA base pair located underneath the G-quadruplex motif for the stronger GoG pair that establishes a tight interaction network involving water molecules and that locks the structure of the bottom part of the aptamer. Furthermore, using μIVC in tandem with NGS, my work demonstrated that an additional level of automation can be reached in the identification of optimal RNA sequences through bioinformatics making the time-consuming and labor-intensive work of post-selection cloning and hand-screening dispensable.

However, despite their superior performances, both Broccoli and iSpinach seems to be limited by the photostability of the DFHBI. Indeed, it was shown that DFHBI-binding aptamers only loosely accommodate DFHBI that is free to rapidly photo-isomerize into its non-fluorescent isomer upon illumination (Han et al., 2013). This reduced photostability leads to short half-life of the complex ($t_{1/2}$ ~ 1 second), making imaging of low abundant RNAs extremely challenging unless sophisticated pulsed-illumination (Han et al., 2013) and image analysis algorithms (Guet et al., 2015) are used. These features motivated me to turn my efforts toward the more photostable RNA-based fluorogenic module TO1/Mango.

6. Improvement, characterization and application of Mango RNA aptamers

6.1. The original Mango/TO1-Biotin module

In 2014, the team of Peter Unrau has isolated Mango, a new light-up RNA aptamer that specifically recognize and light-up the thiazole orange derivative TO1-Biotin (Dolgosheina et al., 2014). Thiazole orange (TO) is known to be a non-specific nucleic acid intercalating dye (Lee et al., 1986). However, adding an acetate group (Figure 39.B) converts TO into TO1 and suppresses the nonspecific intercalating ability of the compound. Furthermore, extending TO1 by a biotinylated PEG chain converts it into TO1-Biotin, a compound that can be immobilized on streptavidin-substituted resin. 12 rounds of SELEX were then performed using this immobilized form of TO1-Biotin to search for binding RNAs. Moreover, the use of stringent wash conditions selecting for rapid association (using short incubation time) and slow dissociation (applying long and stringent washing steps) allowed enriching the pool in high affinity RNA variants. Among them, Mango RNA (Figure 39.A) was particularly interesting because of its high affinity (K_D ~3 nM) for TO1-Biotin. Interestingly, structural characterization showed that Mango is organized around a three-tier G-quadruplex structure on the top of which TO1-Biotin is accommodated (Figure 39.C) and that was proposed to efficiently fold in a variety of sequences context (Dolgosheina et al., 2014; Trachman et al., 2017). These different features together with the small size of Mango (39 nucleotides) make it a good candidate for being inserted into small RNAs without disturbing their function.

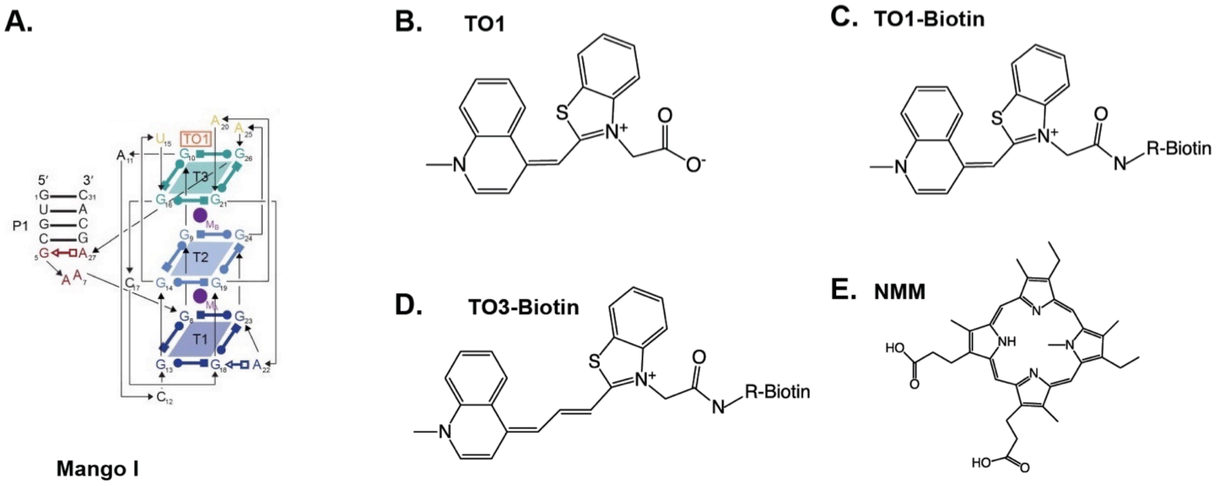


Figure 38: Mango RNA aptamers and its ligands. (A) Mango RNA folds as a G-quadruplex structure. Figure adapted from Trachman et al. 2017. (B) Thiazole Orange substituted with an acetate group (TO1). (C) TO1 conjugated with a PEG-biotin linker. (D) TO3-Biotin has a methyl bridge made of two supplemental carbon and has a red-shifted fluorescence emission compare to TO1-Biotin. (E) N-methyl mesoporphyrin IX (NMM), a general G-quadruplex binder.

As a first application, this high affinity module can be used as an affinity tag, as this is often done with the MS2-GFP system. Indeed, since TO1 is biotinylated, the fluorogen can be immobilized on a resin a later used to pull-down Mango-labelled target RNA and its associated partners (e.g. proteins) in native conditions while still maintaining its fluorogenic capacities (Panchapakesan et al., 2015).

On an imaging point of view, the high affinity of Mango for TO1-Biotin makes it possible to use low concentration of the fluorogen and, this way, to increase imaging contrast by reducing background fluorescence of the unbound dye. Indeed, in contrast to DFHBI-binding aptamers (K_D ~500 nM) that require using high fluorogen concentration (~ 200 μM for imaging in bacteria (Alam et al., 2017; Paige et al., 2012; Song et al., 2013)), as few as 200 nM of TO1-Biotin is enough to image Mango module (see **Original Article 3**). Moreover, Mango revealed to be much more specific than Spinach(Jeng et al., 2016). Indeed, whereas Mango is highly specific for TO1-biotin and does not complex DFHBI, Spinach is able to light-up both compounds. Moreover, Spinach/TO1-Biotin complex is even brighter than the cognate Mango/TO1-Biotin indicating that the original Mango aptamer sub-optimally activates TO1-Biotin fluorescence. Based on this observation, a collaboration was set-up with the team of Peter Unrau to search for Mango sequences with increased light-up capacity using the screening pipeline I previously validated with Spinach aptamer.

6.2. Isolation of improved Mango aptamers

Mango was identified from a library enriched in TO1-binding aptamers after 12 rounds of SELEX. Data collected by Unrau's lab during the hand-screening of this library (R12) (Dolgosheina et al., 2014)suggested that the pool still contained a large sequence diversity. Moreover, all the assayed R12 variants displayed an affinity ranging from 1 to 100 nM (Dolgosheina et al., 2014)suggesting a high overall affinity of this reservoir of variants. Given this high theoretical diversity whom only a very low fraction has been hand-screened in the previous work, R12 was very likely to contain a significant number of variants at least as good as Mango, if not better. Consequently, we decided to use R12 library as starting point of a μIVC functional screening using the approach validated above with Spinach.

6.2.1.Setting-up R12 library screening

Prior to starting μIVC screening, I first verified that R12 library contained indeed potentially interesting aptamers. To do so, ~ 200,000 genes from R12 were analyzed by μIVC using similar conditions than with iSpinach (50mM potassium, 20 mM $MgCl_2$) in the presence of 100 nM TO1-Biotin. Upon gene individualization, amplification and transcription, droplet fluorescence analysis revealed that, indeed, a significant fraction of R12 gene encoded aptamers more fluorogenic than the original Mango (see Figure 2.b below in **Original article 3**).

The main challenge here was to preserve a selection pressure for high affinity aptamers while being able to identify the brightest ones. Indeed, since droplets spend a short time (typically a few milliseconds) in front of the sensors during the sorting step, a minimal concentration of 100 nM TO1-Biotin is required to distinguish signal from noise. This concentration largely exceeding Mango K_D for TO1-Biotin, it could lead to the isolation of aptamers of much lower affinity. To limit this adverse effect, I decided to supplement the reaction mixture with a second dye able to compete with TO1-Biotin. Progressively increasing competitor concentration along the screening process was expected to challenge the stability of low-affinity complexes, whereas those RNAs interacting specifically and strongly with TO1-Biotin should still be able to form a fluorescent complex with the TO1-Biotin. Two competitors were chosen and used (Figure 39.D & 39.E). First, the N-methyl mesoporphyrin IX (NMM) was used as a non-specific competitor. Indeed, previous work indicated that this general G-quadruplex binding molecule displayed a high affinity for Mango and was an efficient competitor if used at high concentration (Jeng et al., 2016). As a second competitor, we used TO3-Biotin, a red-emitting analogue of TO1-Biotin also recognized by Mango but with a lower affinity. Both complexes (Mango/ TO1-Biotin and Mango/ TO3-Biotin) can easily be discriminated during the screening since they emit fluorescence at different wavelengths. Finally, and still in the view of limiting the selection of low affinity aptamers, the transcription rate in the droplets was also reduced to produce ~ 300 nM (instead of the ~ 8 μM used with Spinach) of RNA.

6.2.2. Isolating new Mango variants with improved performances

Knowing that R12 library was very likely to contain Mango RNAs with superior lighting-up properties and having defined a new competitor-based μIVC screening strategy, I next applied it to R12 to search for aptamer variants with improved performances. This allowed me to identify three new aptamers (Mango II, III and IV): one of them (Mango II) displaying an affinity even 10 times higher than that of the original Mango (hereafter called Mango I), whereas the two others (Mango III and IV) were able to form three times brighter complexes with TO1-Biotin as compared to Mango I, making them brighter than eGFP. Finally, we showed that these new Mango variants were well suited for imaging small non-coding RNA in mammalian cells. Altogether, these results allowed me to sign as a co-first author an article published in early 2018 in the journal Nature communication and reproduced hereafter (**Original article 3**).

Original Article 3:

Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells

Alexis Autour, Sunny C. Y. Jeng, Adam D. Cawte, Amir Abdolahzadeh, Angela Galli, Shanker S. S. Panchapakesan, David Rueda, Michael Ryckelynck & Peter J. Unrau

2018
Nature Communications

nature
COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-02993-8

OPEN

Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells

Alexis Autour[1], Sunny C. Y. Jeng[2], Adam D. Cawte [3,4], Amir Abdolahzadeh[2], Angela Galli[2], Shanker S.S. Panchapakesan[2], David Rueda [3,4], Michael Ryckelynck[1] & Peter J. Unrau [2]

Despite having many key roles in cellular biology, directly imaging biologically important RNAs has been hindered by a lack of fluorescent tools equivalent to the fluorescent proteins available to study cellular proteins. Ideal RNA labelling systems must preserve biological function, have photophysical properties similar to existing fluorescent proteins, and be compatible with established live and fixed cell protein labelling strategies. Here, we report a microfluidics-based selection of three new high-affinity RNA Mango fluorogenic aptamers. Two of these are as bright or brighter than enhanced GFP when bound to TO1-Biotin. Furthermore, we show that the new Mangos can accurately image the subcellular localization of three small non-coding RNAs (5S, U6, and a box C/D scaRNA) in fixed and live mammalian cells. These new aptamers have many potential applications to study RNA function and dynamics both in vitro and in mammalian cells.

---

[1] Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, 67000 Strasbourg, France. [2] Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada. [3] Single Molecule Imaging Group, MRC London Institute of Medical Sciences, Du Cane Road, London W12 0NN, UK. [4] Department of Medicine, Molecular Virology, Imperial College London, Du Cane Road, London W12 0NN, UK. Alexis Autour, Sunny C. Y. Jeng and Adam D. Cawte contributed equally to this work. Correspondence and requests for materials should be addressed to D.R. (email: david.rueda@imperial.ac.uk) or to M.R. (email: m.ryckelynck@unistra.fr) or to P.J.U. (email: punrau@sfu.ca)

Since their introduction, fluorogenic RNA aptamers that enhance the fluorescence of an unbound fluorophore have sparked significant interest and hold great potential to enable the visualization of RNA molecules within a cell[1–4]. However, developing high contrast aptamer-fluorophore systems with brightness comparable to existing fluorescent proteins has posed a significant experimental challenge. In an ideal system, unbound fluorophores with high extinction coefficients and low quantum yields become highly fluorescent when bound by an RNA aptamer whose tertiary structure correctly positions the fluorophore into an orientation that maximizes its brightness[1,5–7]. While reported aptamer-fluorophore complexes make use of fluorophores with high extinction coefficients, notably RNA Mango[8] and the cytotoxic Malachite Green binding aptamer[5], these systems suffer from low quantum yields. Conversely, systems with high quantum yields such as the GFP-mimic aptamers[1,2,9,10] have intrinsically low extinction coefficients. As a consequence, such complexes are all significantly less bright than enhanced GFP[11], diminishing their utility

High-affinity aptamers, with the notable exception of RNA Mango, have also been difficult to develop. While not important for a perfect fluorophore with zero unbound quantum yield, high-affinity fluorophore aptamer complexes allow lower free fluorophore concentrations to be used during imaging, effectively decreasing background fluorescent signal[12]. Despite the inability to simultaneously optimize aptamer-fluorophore brightness and binding affinity, existing fluorogenic systems have achieved some notable successes in bacteria, yeast and mammalian cells[1,2,13–15]. This suggests that using newly developed screening methodologies to select brighter fluorogenic RNA aptamers either by FACS[9] or droplet-based microfluidics platforms[10] can provide powerful and easy to use fluorescent RNA imaging tags to study cellular RNAs.

Here, we have used a competitive ligand binding microfluidic selection to isolate three new aptamers (Mango II, III and IV) with markedly improved fluorescent properties, binding affinities, and salt dependencies compared to the original Mango I aptamer[8]. These aptamers all contain a closing RNA stem, which isolates a small fluorophore-binding core from external sequence, making them easy to insert into arbitrary biological RNA. Unexpectedly several of these constructs are resistant to formaldehyde, allowing their use in live-cell imaging and also in conventional fixed cell methodologies. Stepwise photobleaching in fixed cell images indicate that as few as 4–17 molecules can be detected in each foci, and photobleaching rates in live cells or in vitro were at least an order of magnitude slower than found for Broccoli. These new aptamers work well with existing fluorescence microscopy techniques and we demonstrate their applicability by imaging the correct localization of 5S, U6 and the box C/D scaRNA (mgU2-47) in fixed and live mammalian cells. Together, these findings indicate that the new Mango aptamers offer an interesting alternative to existing fluorogenic aptamers[12].

Results

Microfluidic isolation of new and brighter Mango aptamers. Mango I is an RNA aptamer that was initially selected from a high diversity random sequence library for its TO1-Biotin (TO1-B) binding affinity rather than for its fluorescent properties, which may have precluded the enrichment of the brightest aptamers in the library[8]. Its structure consists of a three-tiered G-quadruplex with mixed parallel and anti-parallel connectivity (Fig. 1)[16]. The observation that the RNA Spinach aptamer can form a 4.5-fold brighter complex with TO1-B than Mango I, in spite of its significantly lower affinity[17], also suggests that more fluorogenic Mango-like folds may exist in the library. To address this, we rescreened the original round 12 Mango I library (R12) using microfluidic-assisted in vitro compartmentalisation (μIVC, Fig. 2a)[10]. Interestingly, the initial screening shows that a significant fraction of molecules in the R12 library are brighter than Mango I (Fig. 2b).

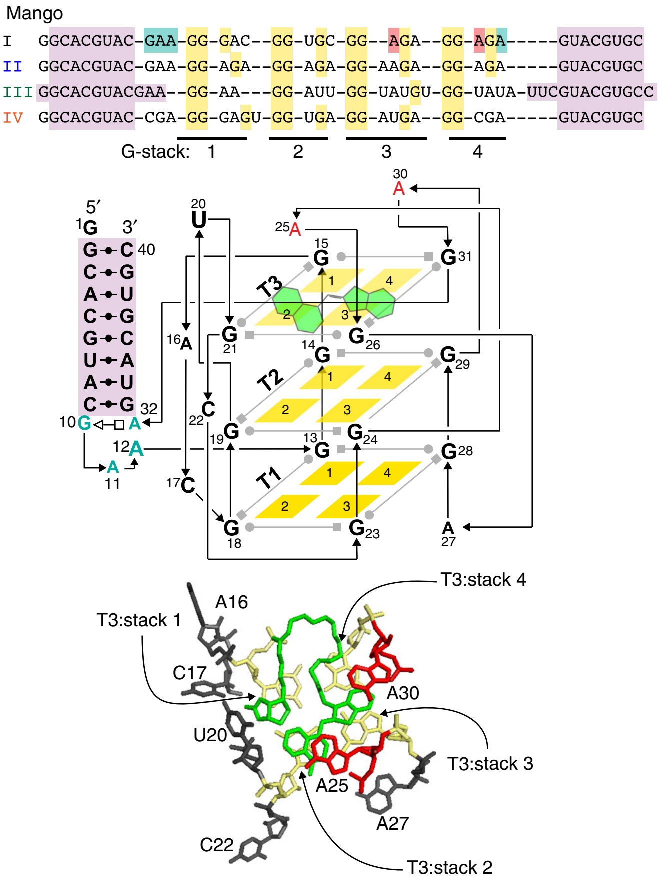

Mango
I GGCACGUAC-GAA-GG-GAC--GG-UGC-GG--AGA--GG-AGA-----GUACGUGC
II GGCACGUAC-GAA-GG-AGA--GG-AGA-GG-AAGA--GG-AGA-----GUACGUGC
III GGCACGUACGAA--GG-AA---GG-AUU-GG-UAUGU-GG-UAUA-UUCGUACGUGCC
IV GGCACGUAC-CGA-GG-GAGU-GG-UGA-GG-AUGA--GG-CGA-----GUACGUGC
G-stack: 1 2 3 4
5′ 3′
1G
C 40
G 32
10G
A 11
12A
16A
17C
18G
19G
20U
21G
22C
13G
14G
15G
23G
24G
25A
26G
27A
28G
29G
30A
31G
T1
T2
T3
A16
C17
U20
C22
A25
A27
A30
T3:stack 1
T3:stack 2
T3:stack 3
T3:stack 4

Fig. 1 RNA Mango aptamers core sequences. Colour-coded alignment of RNA Mango I, II, III and IV. G residues in yellow are protected from dimethyl sulfate (DMS) cleavage when folded in the presence of fluorophore. Quadruplex stacks and their associated propeller sequences are numbered 1 through 4. The GAAA isolation motif of Mango I, together with two adenines essential for binding, are shown in green and red respectively. Purple shading represents a flanking stem region for all four Mango aptamers. Schematic: a tertiary structure of Mango I, showing tier 1, 2 and 3 of its quadruplex structure (T1, T2 and T3) and colour-coded as in top panel. TO1-B is shown in green. Bottom: top view of the Mango I core (PDB ID: 5V3F[16]), showing the T3 tier of the quadruplex and relevant propeller residues, colour coding matches the schematic and top panel

A potential limitation of μIVC is the requirement of high TO1-B concentrations (~100 nM, due to the requirement for high speed fluorescent sorting) during the screening. Such high values would greatly exceed the Mango I K_D (~3 nM) and could potentially prevent the selection of high-affinity aptamers. To mitigate this, we supplemented the in vitro transcription (IVT) mixture with TO1-B competitors NMM (*N*-methyl mesoporphyrin IX)[18] and TO3-Biotin[8] (Supplementary Fig. 1), which are both known to interact with G-quadruplexes. As expected, the NMM supplemented IVT mixture significantly reduces TO1-B/Mango I fluorescence (Supplementary Fig. 2). The NMM concentration was progressively increased during the first four screening rounds, therefore, any brightness increase at each round presumably results from the selection of brighter aptamers in the library, while retaining high affinity and selectivity for

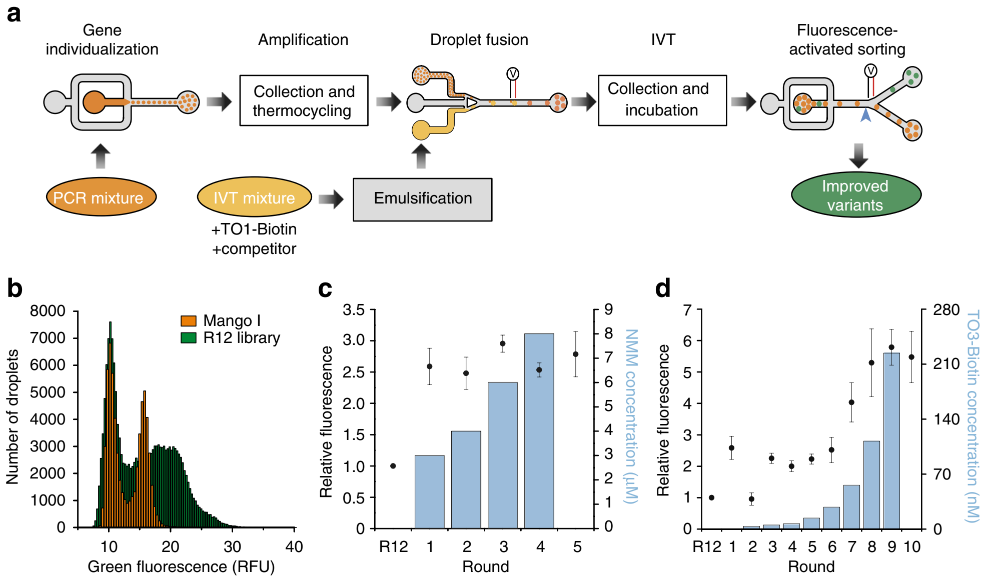


Fig. 2 Competitive selection of TO1-B-binding variants using droplet-based microfluidics fluorescence screening. **a** Experimental workflow for microfluidic-assisted fluorescence screening. Ovals and boxes represent on- and off-chip steps, respectively. Three microfluidic devices were used for gene individualization in 2.5 pL droplets containing PCR mixture; after thermocycling, fusing each PCR droplet with a droplet containing an in vitro transcription (IVT) mixture supplemented with TO1-B and competitor (NMM or TO3-Biotin); and, after incubation, analysing the fluorescence profile of each droplet and sorting them accordingly. **b** Fluorescence profile of droplets containing Mango I or the initial R12 library (~200,000 variants, Supplementary Table 1). Droplets containing no DNA have a fluorescence of 10 RFUs. **c** Improvement in fluorescence enhancement of aptamer libraries during the screening process in the presence of increasing amounts of NMM. The fluorescence (black dots) of the RNA libraries in complex with TO1-B was determined by mixing 2 µM RNA and 100 nM TO1-B in the absence of NMM. These values were normalized to that of the starting library (R12). The values are the mean of three independent experiments and error bars correspond to ±1 standard error. **d** Enhancement in fluorescence resulting from selection with TO3-Biotin competitor. The fluorescence (black circles) was determined after each round by mixing 300 nM RNA and 100 nM TO1-B in the absence of TO3-Biotin. The values were normalized to that of the starting library (R12). The blue bars represent the concentration of competitor used in each round of selection. The values are the mean of three independent experiments and error bars correspond to ±1 standard error, and for each sort, the gated populations can be found in Supplementary Fig. 3

TO1-B (Fig. 2c). The stability of the RNA/TO1-B complex was further challenged by sorting the droplets at 45 °C as previously described[10]. The relative fluorescence of the library increased 2.5-fold in the first round (~3 million variants analysed, Supplementary Table 2), but it did not increase further over the later rounds (Fig. 2c, Supplementary Fig. 3a). However, the ability of NMM to compete against TO1-B binding decreased progressively with each round (Supplementary Fig. 4a), indicating that the aptamers in the later rounds have higher affinity and/or selectivity for TO1-B. The last screening round shows that, in the absence of NMM, the fluorogenic properties of the enriched library remained unchanged (Fig. 2c). From the final enriched library, we cloned and sequenced 32 pool RNAs, and analysed their fluorogenic capacity (Supplementary Fig. 5a, b). While the brightest clone was R5-NMM-20, 6 of the 13 brightest aptamers exhibited an almost identical sequence to clone R5-NMM-5 (Supplementary Fig. 5c).

In a second set of screenings, we increased the selection stringency by using the Mango I specific competitor TO3-Biotin, which differs from TO1-B by having two additional carbons in the methine bridge of TO1-B (Supplementary Fig. 1). To further increase the selection pressure for TO1-B binding, we also decreased the RNA concentration in the droplets to 0.3 µM (from 8 µM with NMM). TO3-Biotin competitor was introduced in the second round of screening to ensure that positive droplets were not missed in the first round (Supplementary Fig. 3b). In subsequent rounds, TO3-Biotin concentration was gradually increased (Fig. 2d and Supplementary Table 3). While the relative fluorescence of the population increased in the first screening round, it decreased upon addition of competitor in round two, likely due to the elimination of brighter but weaker binding aptamers (Fig. 2d). In later rounds, the relative fluorescence increased progressively until the TO3-Biotin concentration exceeded TO1-B by 2.2-fold (220 nM and 100 nM, respectively). Beyond this ratio, the competition was too high and the selection process collapsed. The final round shows that the enriched library maintains its fluorescent properties in the absence of competitor. RNA molecules from each of the final rounds were cloned and sequenced. Interestingly, further characterization of the different libraries indicates that aptamers were first selected for their capacity to discriminate TO1-B from TO3-Biotin (Supplementary Fig. 4b, c). Surprisingly, we found that, starting from round three, the libraries were dominated by a single cluster of sequences (cluster E, Supplementary Fig. 6) that was attributed to the TO3-resistant aptamers discussed above. In the last four rounds, this sequence was progressively replaced by the point mutant C66U, best represented by the aptamer R10-17 and identical in sequence to R5-NMM-20 found in the NMM competitive screen. The progressive domination by R10-17 was likely the origin of the fluorescence improvement observed in the last rounds of selection. Finally, among the remaining clusters identified in the early rounds, cluster D, represented by R2-1, was found to have particularly high binding affinity (Supplementary Fig. 7a).

Each new Mango variant is unique in structure and function. Based on the parental sequence isolates R2-1, R5-NMM-20 (R10-17), and R5-NMM-5 (Supplementary Figs. 5, 6 and 7), we engineered the minimal reference constructs Mango II, Mango III and Mango IV (Fig. 1), respectively, by truncation and sequence manipulation while maintaining the binding and fluorescent

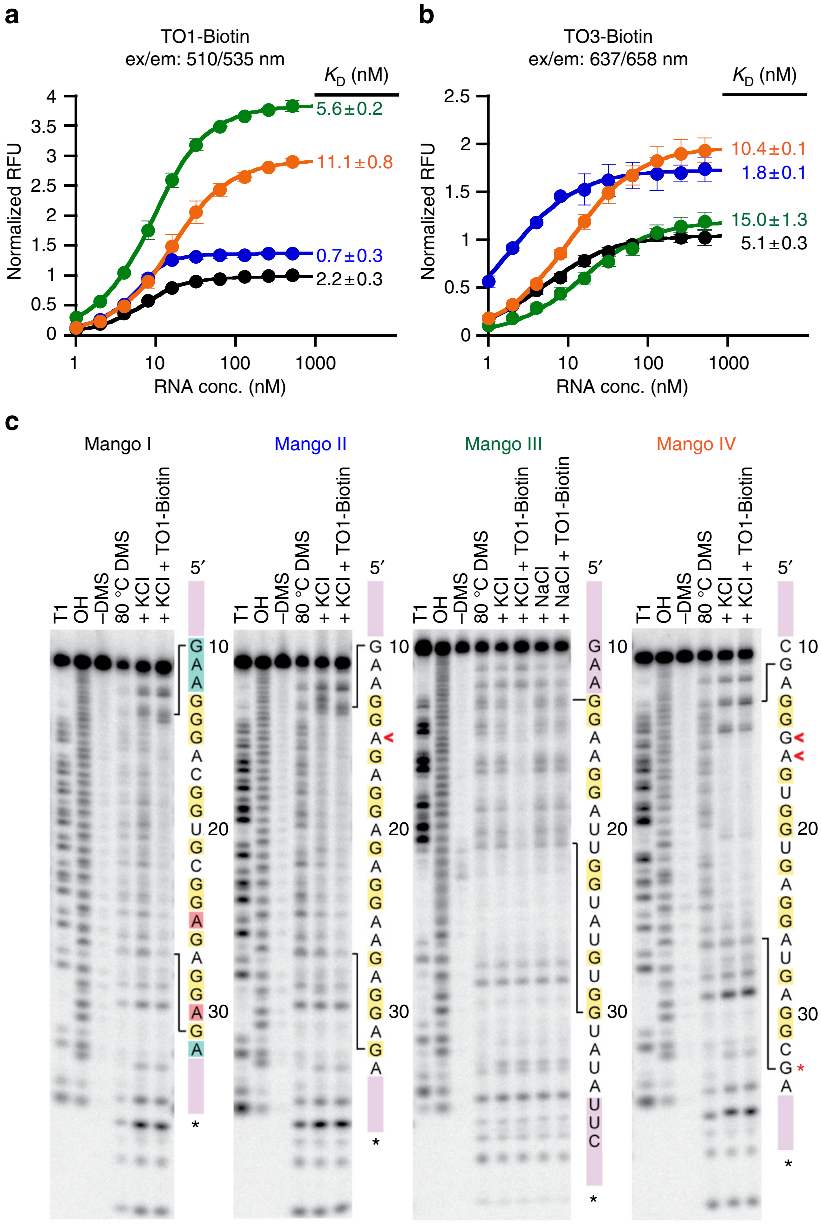


Fig. 3 RNA Mango aptamers fluorophore-binding and DMS protection. **a** Fluorescence binding curves for each Mango aptamer determined by titrating RNA aptamer concentration while holding TO1-B fluorophore constant at 10 nM. K_D values are shown next to each titration. **b** Same as for **a** but using 1.4 nM TO3-Biotin. Data for **a** and **b** have been normalized such that Mango I has a maximum RFU of 1 in each case. Error bars are standard deviation of three replicates. **c** DMS chemical protection patterns for the four Mango aptamers. 3′ end-labelled RNA (^{32}P pCp shown as a black asterisk) was subjected to DMS chemical modification followed by reduction by $NaBH_4$ and aniline cleavage as described in the methods. RNA sequences are displayed to the right of each set of lanes with stem portions represented as purple blocks. Legend: T1—denatured T1 ladder; OH—partial alkaline hydrolysis ladder; (−) DMS—denatured reaction with ddH_2O added in place of DMS; 80 °C DMS—denatured DMS ladder; remaining lanes are native DMS reactions with addition of potassium to 140 mM final (+KCl), addition of sodium to 140 mM final (+NaCl), with or without 500 nM TO1-B (+TO1-B). Red asterisk indicates a notably unprotected G in Mango IV. Red daggers in Mango II and Mango IV indicate nucleotides presumed to be looped out in the T3 layer relative to the Mango I G stack shown in Fig. 1

properties of the parental constructs (Supplementary Fig. 7). Mango II, III and IV were found to be 1.5-, 4- and 3-fold brighter than Mango I, respectively (Fig. 3a). Mango II binds TO1-B with subnanomolar affinity, while Mango III and IV had slightly weaker affinities than Mango I (Fig. 3a). Mango II and IV also demonstrated improved fluorescence response when bound to TO3-Biotin relative to Mango I while exhibiting nanomolar binding affinities to this strongly red-shifted fluorophore (Fig. 3b). Notably, the brightness of the Mango III and Mango IV TO1-B bound complexes are 43,000 M^{-1} cm^{-1} and 32,000 M^{-1} cm^{-1}, respectively, making Mango III 1.3 times brighter than enhanced GFP (EGFP) a common benchmark for the characterization of improved fluorescent proteins[11].

Mango I binds TO1-B by sandwiching it between the T3 layer of the G-quadruplex and A25 and A30 (Fig. 1, yellow and red residues, respectively)[16]. This fluorophore-binding core is isolated from an arbitrary RNA duplex (Fig. 1, purple residues) by a GAAA tetraloop-like adaptor[19] that inserts the TO1-B binding core between the third and fourth residues of the tetranucleotide motif (Fig. 1, cyan residues)[16]. Like Mango I, all of the new

Mango aptamers are contained within an arbitrary closing stem (Supplementary Fig. 7). Dimethyl sulfate (DMS) probing, which correctly confirmed the three-tiered quadruplex structure of Mango I, indicates that Mango II also contains a three-tiered quadruplex structure (Fig. 3c). In distinct contrast to Mango I, which has a contiguous run of protected three guanine residues forming G-stack 1 (Fig. 1, G13-G15 of Mango I), Mango II has an A15 insertion between G14 and G16. Since G13, G14 and G16 are DMS-protected in Mango II, this implies a structural rearrangement of the T3 level of the aptamer (Fig. 3c, '<' symbol). In addition to this change, Mango II has a dinucleotide adenine in its third propeller loop region (Fig. 3c), whereas Mango I has a single adenine in this location. In Mango I, this adenine is stacked on top of the methylquinoline heterocycle of the TO1-B (Mango I A25, Fig. 1 and Fig. 3c, first red residue) implicating an additional structural change in Mango II relative to that of Mango I. Indeed, either of these changes, individually or together, were shown to play an important role in the improved affinity and brightness of Mango II (Supplementary Fig. 7a).

Mango IV has a different fold than Mango I or II and is considerably brighter than either. Unexpectedly Mango IV, which like Mango I contains three contiguous G residues (G13-G15), lacked N-7 protection of residue G15, but like Mango II has strong DMS protection of G17 after A16 (Fig. 3c, '<' symbols). This implies that the dinucleotide G15 and A16 of Mango IV must be in a distinct conformation relative to either Mango I or II. Interestingly, Mango IV also lacks DMS protection of residue G33 (Fig. 3c, * symbol), which in Mango I plays an instrumental role in forming the T3 layer and that is DMS-protected in Mango II. In addition, the GAAA linker motif of Mango I (Fig. 1, blue residues), which is also apparently found in Mango II, does not appear to be present in Mango IV, as replacing the 5′ CGA sequence of the Mango IV core sequence with GAA resulted in a four-fold decrease in binding affinity (Supplementary Fig. 7b, variant 20). These data, together with additional point mutational analysis, indicate that, while Mango IV appears likely to contain T1 and T2 tiers of guanine tetraplexes in common with Mango I and Mango II, its T3 tier is likely to differ considerably in structure from either that of Mango I or II.

The folding of the Mango II and IV constructs was characterized further by examining their salt dependence, thermal melting properties, and CD spectra. Both Mango II and IV have Hill coefficients and affinities for potassium similar to Mango I (Supplementary Table 4 and Supplementary Fig. 8). Mango II showed a limited fluorescence response in the presence of sodium ions, while Mango I and Mango IV did not fluoresce appreciably with this ion. Most notably, in the presence of potassium, these aptamers were resistant to Mg^{2+} levels >10-fold higher than the selection concentration (22 mM), whereas Mango I fluorescence was strongly inhibited at such high concentrations (Supplementary Fig. 8). This indicates that Mango II and Mango IV are substantially more stably folded than Mango I, likely as a result of being selected for fluorescence at elevated temperature and in the presence of high levels of free magnesium. While Mango I and II both displayed a change in DMS protection upon addition of TO1-B to aptamers pre-incubated in potassium buffer, the DMS protection pattern of Mango IV was largely unchanged upon addition of TO1-B (Fig. 3c). Mango II, had thermal melting properties that were largely unchanged whether or not TO1-B was present, while Mango IV exhibited hysteresis in the unbound melting curve. Both thermal melts were considerably different from that of Mango I, which changes its A_{260} thermal melt profile depending on the presence or absence of TO1-B[17] (Supplementary Fig. 9). Consistent with the formation of a G-quadruplex structure in Mango I, II and IV, the ligand bound CD spectra for each aptamer were quite similar (Supplementary Fig. 10).

Mango III, the brightest of the three constructs, contains only nine guanines in its core, and is therefore unable to form a three-tiered G-quadruplex. All nine core guanines are DMS-protected (Fig. 3c). Mutant analysis suggests that the helical region of Mango III is likely to extend an additional 3 bp into its core (Figs. 1 and 3c, light purple), as changing this putative 3 bp duplex had only a modest impact on binding affinity. In contrast, removing either the 5′ GAA or 3′ UUC sequence completely ablated binding (Supplementary Fig. 7c, variants 9, 10 and 11). Consistent with the hypothesis that the C66U mutation observed during selection played a role in higher fluorescence, reverting this mutation in the truncated Mango III context reduced fluorescence by 40% (Supplementary Fig. 7c, variant 12), suggesting that this nucleotide plays an important role in conferring fluorescence. Mango III contains much longer A/U rich propeller regions than any of the other Mangos (Fig. 1) and has a ~100-fold higher affinity for potassium and sodium, while being only modestly inhibited by high levels of magnesium (Supplementary Fig. 8). Its sigmoidal fluorescent melting curve resembles the melting of RNA Spinach2[2] and not the more linear melting curves observed for Mango I, II and IV (Supplementary Fig. 9). Similarly, the CD spectrum of the bound Mango III complex is different in the 270–300 nm region from the other Mango constructs tested (Supplementary Fig. 10). Correlated with this distinct CD spectra, Mango III lacks an excitation shoulder found to be in common for all the other Mango constructs in the 270–300 nm region (Supplementary Fig. 11). This and other differences in the excitation and emission spectra all suggest that Mango III binds TO1-B differently than Mango I, II and IV aptamers. Detailed X-ray structure analysis will be required to uncover further details of this interesting aptamer.

Cellular imaging of Mango-tagged RNAs. To test the newly isolated aptamers in cells, we tagged the small and well-characterized human 5S ribosomal RNA with each Mango variant by incorporating an F30 folding scaffold (Supplementary Fig. 12a) previously shown to improve cellular fluorescence and RNA stability[1,20]. Each Mango-tagged RNA, with or without the folding scaffold or terminator hairpin, exhibited comparable fluorescence intensities in vitro (Supplementary Fig. 12b). No appreciable fluorescence was observed in the absence of the TO1-B fluorophore or with the control constructs that contained either the F30 folding scaffold alone, or a G-quadruplex mutant Mango sequence (Supplementary Fig. 12b and Supplementary Table 5).

To image the tagged RNA, we transfected in vitro-transcribed 5S-F30-Mango RNA into HEK293T cells, fixed the cells on ice and stained with TO1-B (Online Methods). This protocol is based on the observation that, in vitro, Mango I, II and IV-fluorophore complexes are substantially resistant to formaldehyde at room temperature (Supplementary Fig. 13a). Up to ~10 bright RNA Mango foci could be readily detected per cell with a fluorescence microscope, but not in control transfections (Fig. 4a). A time course of this process (Supplementary Fig. 14a) shows the initial delivery of lipofectamine particles to the cell membrane (5 min after transfection) followed by dispersal of the RNA in the cytoplasm (15–30 min) and foci formation (30–60 min), indicating that the observed foci are not intact lipofectamine-RNA particles that remain after transfection. Furthermore, these foci cannot correspond to 5S-F30-Mango RNA in late endosomes, given that Mango fluorescence decreases significantly at low pH (Supplementary Fig. 13b). Contrary to 5S-F30-Mango I and III, transfections with 5S-F30-Mango II and IV RNA consistently exhibit visible foci. A possible explanation is that Mango II and IV fold correctly both in the presence and absence of TO1-B, unlike the other Mangos (Supplementary Fig. 9). Consistent with

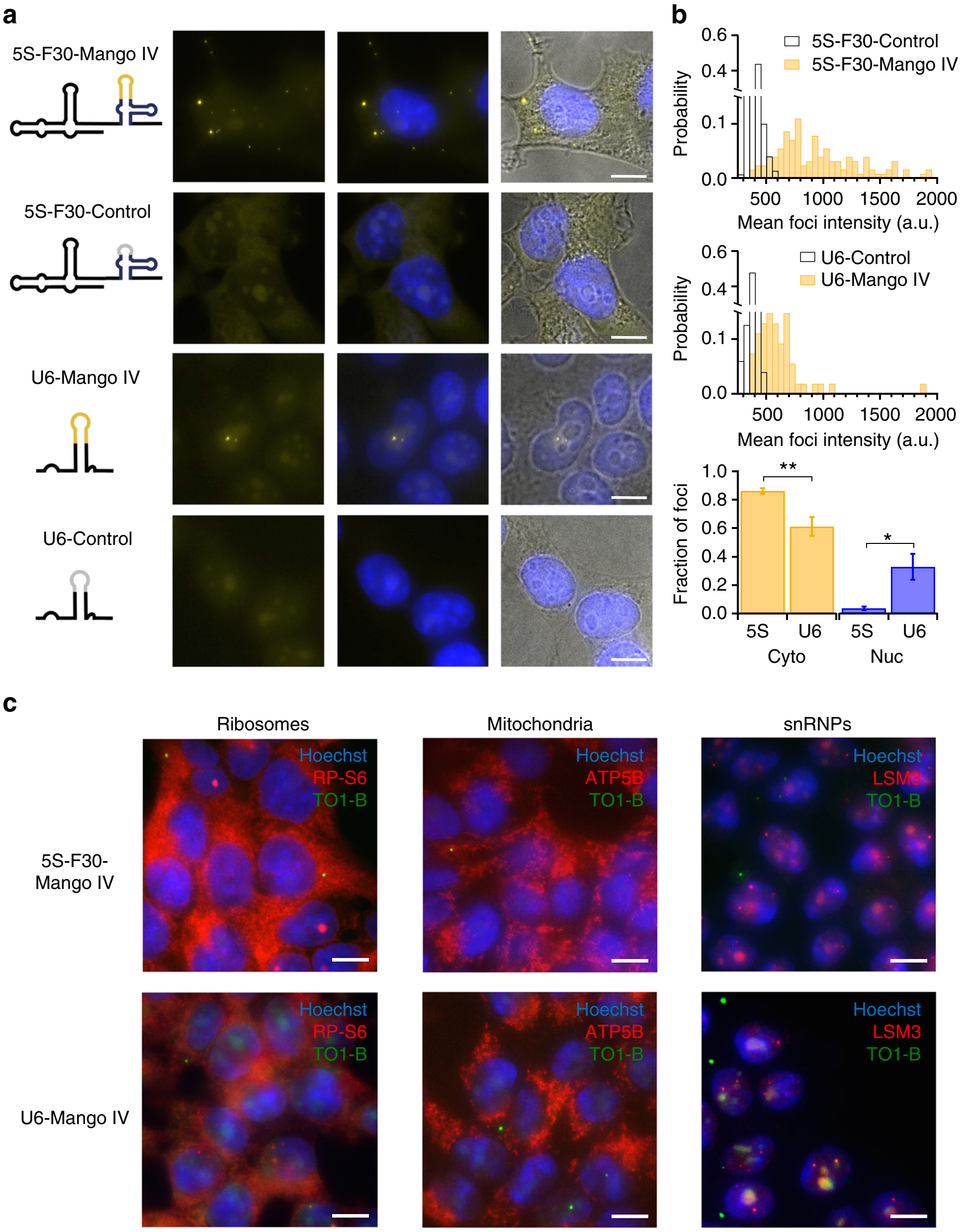

a
5S-F30-Mango IV
5S-F30-Control
U6-Mango IV
U6-Control
b
Probability
0.6
0.4
0.1
0.0
☐ 5S-F30-Control
🟨 5S-F30-Mango IV
500 1000 1500 2000
Mean foci intensity (a.u.)
☐ U6-Control
🟨 U6-Mango IV
Fraction of foci
1.0
0.8
0.6
0.4
0.2
0.0
**
*
5S U6 5S U6
Cyto Nuc
c
Ribosomes
Mitochondria
snRNPs
5S-F30-Mango IV
U6-Mango IV
Hoechst
RP-S6
TO1-B
ATP5B
LSM3

Fig. 4 Cellular imaging of Mango IV tagged RNAs. **a** Maximum projections of fixed cells containing Mango IV tagged 5S and U6 RNAs stained with 200 nM TO1-B (yellow) and 1 μg/mL Hoechst 33258 (blue)—construct diagrams shown as RNA-Mango (yellow) or non-fluorescent control RNA (grey), F30 folding scaffold (blue) and remaining RNA sequence (black). **b** Mean intensity distributions of 5S-Mango IV and U6-Mango IV foci (yellow) compared to low intensity foci detected in control experiments (black). Fraction of foci observed in the cytoplasm and nucleus for 5S and U6-Mango IV RNAs (bottom panel, $*p < 0.05$ and $**p < 0.01$ calculated using a *t* test). The number of cells for 5S-F30-Control, 5S-F30-Mango IV, U6-Control and U6-Mango IV were 57, 114, 131 and 183, respectively. Error bars depict standard error in the mean. **c** Localization of 5S-Mango IV and U6-Mango IV relative to immunostained ribosomes (RP-S6), mitochondria (ATP-5B) and snRNPs (LSm3). Scale bars are 10 μm. All images are maximum projections except in **c**, which show a single focal plane

this, similar levels of DMS protection in dye bound and unbound samples were most clearly seen in Mango IV and in part for Mango II (Fig. 3c). The mean intensity of the 5S-F30-Mango IV foci was two- to three-fold higher than 5S-F30-Control background (Fig. 4b). The majority of 5S-F30-Mango IV foci (~85%) are cytoplasmic, a small fraction (~5%) are clearly nuclear and the remaining foci appeared on the nuclear boundary.

To precisely determine the subcellular localization of the 5S-F30-Mango IV foci, we combined Mango-based imaging with immunostaining, which is made feasible by the ability of Mango IV to withstand formaldehyde fixation. As expected, cytoplasmic 5S-Mango IV foci overlap with antibody staining against the ribosomal protein S6 (RP-S6, Fig. 4c). In addition, we observe that the cytoplasmic 5S-Mango IV foci overlap with immunostained mitochondria (ATP5B, Fig. 4c and Supplementary Fig. 15a), and not with constructs lacking Helix IV (Supplementary Fig. 17c, e), consistent with the observation that Helix IV of 5S ribosomal RNA (rRNA) mediates its import into mitochondria[21]. Conversely, we do not observe overlap with other subcellular compartments, such as P-bodies, endosomes or stress granules, where the transfected RNA could be processed for degradation (Supplementary Fig. 15b).

To confirm that the observed foci are specific, we tagged and transfected an RNA that localizes to a different cellular compartment. The U6 snRNA (small nuclear RNA) is expected to associate with snRNP (ribonuclear protein) complexes in the nucleus. We tagged U6 snRNA by incorporating Mango IV directly into an internal stem loop (Supplementary Fig. 12), known to be amenable to modification[22–24]. The resulting construct exhibits comparable fluorescence intensity to Mango IV alone in vitro (Supplementary Fig. 12, Supplementary Table 5). Direct transfection of U6-Mango IV snRNA yields fluorescent foci comparable to 5S-F30-Mango IV (Fig. 4a), albeit with lower intensity (Fig. 4b). As expected, the fraction of nuclear foci increased approximately ninefold, while cytoplasmic foci decrease significantly ($p < 0.01$, Fig. 4b). As opposed to 5S-F30-Mango IV, cytoplasmic U6-Mango IV foci did not significantly overlap with mitochondria or ribosomes, whereas nuclear U6-Mango IV foci do overlap with snRNP protein LSm3 (Fig. 4c), as expected. Similar to 5S-F30-Mango IV, U6-Mango IV foci do not overlap with other subcellular compartments, such as P-bodies, endosomes or stress granules (Supplementary Fig. 15b). To further quantify the proportion of nuclear and cytoplasmic Mango-tagged RNA, we created two-dimensional (2D) intensity plots for each pixel from multiple images (Supplementary Fig. 16). Intensity thresholds were set above the observed profile of control RNAs and the apparent nuclear boundary. These plots highlight the preferential cellular location of Mango-tagged 5S and U6 RNAs, with 5S predominantly cytoplasmic (~64%) and U6 predominantly nuclear (~73%). We performed a similar analysis to correlate the normalized immunostaining and Mango signals by setting appropriate thresholds above the signal observed in each control (Supplementary Fig. 17a, b). Co-localization with the mitochondrial marker ATP5B is only observed with 5S-F30-Mango IV and not with U6-Mango IV or the helix IV mutant 5S-Δ78-98-F30-Mango IV shown to be deficient in mitochondrial import[21] (Supplementary Fig. 17c–e). Also, consistent with the images shown in Fig. 4, a high level (90%) of co-localization between U6-tagged RNAs and the snRNP marker LSm3 is observed compared to the U6-Control and 5S-F30-Mango IV RNAs (Supplementary Fig. 17f–i). Taken together, these fixed cell data show that Mango IV can be used to label and image small cellular RNAs via direct transfection of in vitro-transcribed RNAs, without affecting their expected subcellular localization.

To test whether Mango-tagged RNA molecules can be imaged in live cells, we took advantage of the aptamer's high affinity for TO1-B, and transfected in vitro-transcribed 5S-F30-Mango RNAs pre-incubated with TO1-B. After transfection, cells exhibit bright foci only in the presence of each 5S-F30-Mango RNA due to the pre-incubation with TO1-B stabilizing efficient fluorescence (Supplementary Fig. 14b and Supplementary Movies 1 and 2). The foci observed were similar to those in Fig. 4a, albeit with a lower background fluorescence in the nucleolus of fixed cells. The foci can be readily tracked revealing three distinct diffusive behaviours (fast, slow and static) and their respective root mean square displacement coefficients could be quantified (Supplementary Fig. 14c and Supplementary Movie 3). Interestingly, the Mango-based aptamers stably fluoresce under constant or pulsed illumination. Photobleaching curves of constantly illuminated live cells containing 5S-F30-Mango I, IV and dBroccoli show >10-fold improvement in photostability compared with dBroccoli (Supplementary Fig. 14d). These data are in good agreement with in vitro photobleaching analysis in aqueous droplets (Supplementary Fig. 18a). Under pulsed illumination (200 ms, 0.2 Hz), the lifetime of Mango RNA aptamers in fixed cells increases by

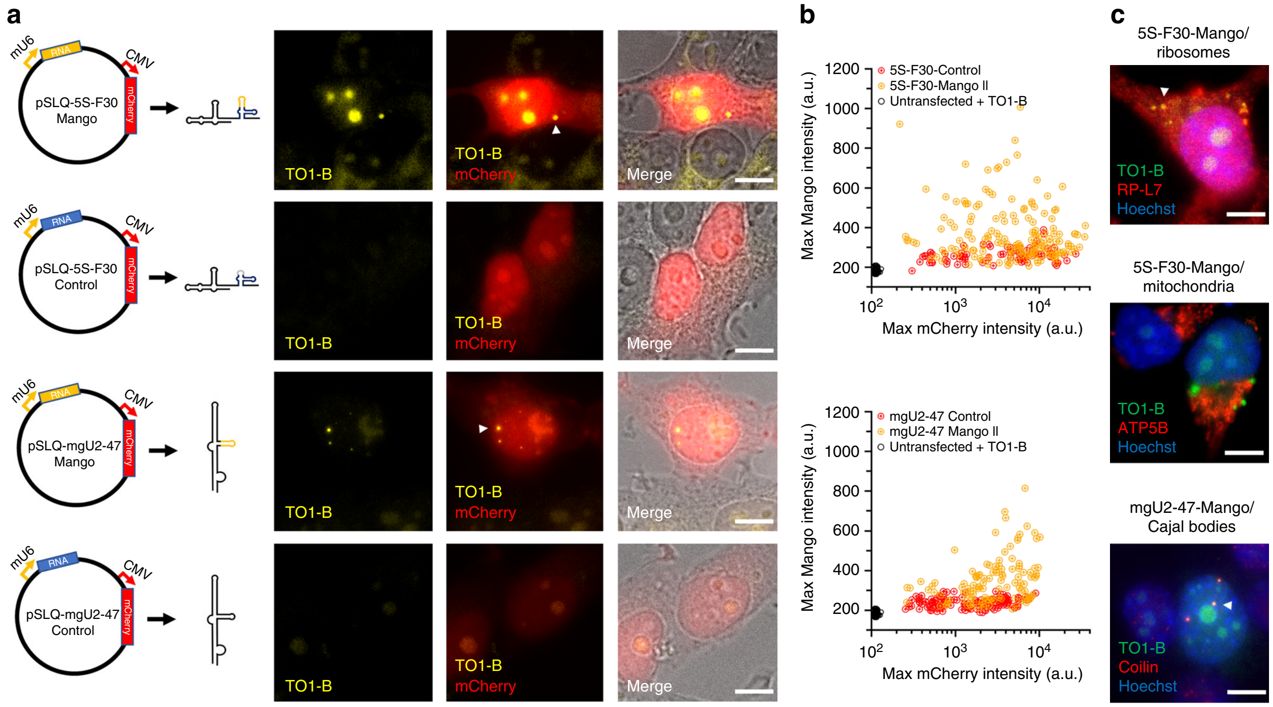


Fig. 5 Cellular imaging of genetically encoded Mango II-tagged RNAs. **a** Diagram of plasmid constructs with the 5S rRNAs and mgU2-47 scaRNAs under the control of a murine U6 promoter (Pol III) and co-expression of a mCherry reporter gene (CMV promoter). Shown adjacent are images of individual slices of fixed cells either expressing Mango II-tagged RNAs (top) or control RNAs (bottom) with the TO1-B (200 nM) signal in yellow, mCherry in red and brightfield image in greyscale. Arrows depict significant cellular and nuclear foci. Scale bar = 10 µm. **b** 2D maximum intensity plots of individual nucleoli and Mango II specific foci for both the TO1-B signal (y axis) and mCherry signal (x axis—$\log_{10}$ scale). The number of cells for 5S-F30-Control, 5S-F30-Mango II, untransfected cells + TO1-B, mgU2-47 Control and mgU2-47 Mango II were 89, 167, 98, 130 and 117 respectively. **c** Maximum projections of cytoplasmic 5S-F30-Mango IV foci and nuclear mgU2-47 foci from plasmid expression in conjunction with immunostained ribosomes (RP-L7), mitochondria (ATP5B) and Cajal bodies (Coilin). Arrows depict significantly co-localized foci, scale bar = 10 µm

>60-fold (from 11.7 s to >10 min, Supplementary Fig. 18b), as previously observed for the Spinach aptamer[25].

To estimate the number of fluorescent 5S-F30-Mango IV molecules in each foci, we performed photobleaching-assisted microscopy on fixed cells (Supplementary Fig. 19a). A maximum likelihood estimate analysis of the photobleaching trajectories[26,27] revealed between 4 and 17 photobleaching steps per foci. In addition, the photobleaching step distribution reveals two peaks corresponding to either one or two molecules (Supplementary Fig. 19b). The number of observed steps correlates linearly with the initial foci intensity below 2000 intensity units (Supplementary Fig. 19c). Altogether, these results indicate that each foci contains at least 4–17 fluorescent molecules, consistent with the observed range of experimental intensities, and raises the interesting possibility of imaging single molecules in live cells via the incorporation of a small number of Mango repeats.

Finally, to test whether the new Mangos have the ability to function as genetically encoded tags expressed in cells, we constructed plasmids that express the 5S rRNA under the control of a RNA pol III promoter in conjunction with an mCherry reporter gene to identify successfully transfected cells (Fig. 5a). Upon fixation, we observed that cells expressing the pSLQ-5S-F30-Mango II and IV constructs exhibit an increased fluorescent signal in nucleolar compartments as well as forming distinct cytoplasmic foci when compared with the pSLQ-5S-F30-Control construct (Fig. 5a, Supplementary Fig. 20a). The analysis of the peak Mango and mCherry intensities for multiple cells expressing the pSLQ-5S-F30-Mango II plasmid shows a population of cells with a high Mango specific signal, not seen in cells expressing the pSLQ-5S-F30-Control plasmid (Fig. 5b). Interestingly, we observe that cells exhibiting lower mCherry intensities can also show higher Mango signal, consistent with RNA transcription preceding mCherry translation. In agreement with this, reducing plasmid expression time, from 24 to 12 h, increased the number of observed cytoplasmic foci (Supplementary Fig. 20a). Under the same conditions of fixation and staining, signal was not observed in untransfected cells or in cells expressing the 5S-F30-Broccoli construct (Supplementary Fig. 20b). The robust cytoplasmic signal observed after 12 h of pSLQ-5S-Mango IV expression enabled us to combine Mango imaging with immunofluorescence (Fig. 5c). As expected, the observed Mango foci co-localize significantly with Ribosomal Protein L7. However, no significant co-localization was observed with the mitochondrial stain ATP5B. The absence of co-mitochondrial localization, in this case, is likely due to the fact that most nucleolar expressed 5S rRNA will assemble into ribosomes in the nucleus, whereas 5S rRNA molecules transfected directly in the cytoplasm will not, and are more readily available for mitochondrial import. The observed cytoplasmic foci did not co-localize with immunostaining for stress granules, P-bodies or endosomes (Supplementary Fig. 20c).

To confirm that the observed 5S rRNA foci are specific, we expressed a Mango II-tagged small Cajal-body-specific RNA (mgU2-47) that mediates the 2′-O-methylation of the U2 snRNA[28]. Upon expression, the Mango-tagged mgU2-47 RNA formed well-defined nuclear foci that were absent in the mgU2-47 Control RNA (Fig. 5a, b). The nuclear foci also co-localized with immunostained Cajal bodies (Fig. 5c). Taken together, these results demonstrate the ability of Mango tags to function as efficient genetically encoded reporters of RNA subcellular location.

Discussion

To be of broad utility, fluorescent RNA aptamers should be bright, bind their ligands with high affinity, and be compatible with existing methodologies to image proteins in live and fixed cells. To achieve this, we have developed a novel, competition-based, ultrahigh-throughput fluorescent screening approach that takes advantage of microfluidic-assisted in vitro compartmentalization to select three new and highly effective RNA Mango fluorogenic aptamers. The broad range of novel photophysical and biochemical properties in the new Mango aptamers promises to make them highly competitive with existing aptamer-fluorophore systems[1,5,8,10,29,30].

The new Mango aptamers when bound to TO1-B are very bright. The original Mango I construct when bound to TO1-B exhibits a quantum yield of ~0.14. This quantum yield is similar to that typically observed when thiazole orange is rigidified by intercalation into a duplex nucleic acid[17,31]. Mango III and IV, when bound to TO1-B, have quantum yields of ~0.56 and 0.42, respectively. Such high quantum yields are comparable or exceed that of TO1-activating proteins[32]. When combined with the high absorbance of thiazole orange, these quantum yields result in fluorescence as bright or brighter than EGFP, a significant milestone for aptamer-fluorophore systems.

While a brighter aptamer system has been recently discovered[29], the high brightness of the Mango aptamers combined with their extremely high binding affinity toward TO1-B and the very low unbound fluorescence of TO1-B strongly enables high contrast imaging. The nanomolar binding affinity of these aptamers allows very low concentrations of fluorophore ligand to be used (50-fold less than typically used by the GFP-mimic aptamers). Indeed, transfecting RNA bound to stoichiometric amounts of TO1-B resulted in fluorescent RNA foci that could be tracked in cells, implicating that fluorophore-bound RNA Mango complexes are both stable and robustly fluorescent in living cells, even when at very low effective concentrations. Likewise, staining fixed cells with low concentrations of TO1-B (200 nM) readily generated Mango-tagged fluorescence with only modest levels of nonspecific nucleolar staining being observed, strongly suggesting that nonspecific fluorescence induced by intercalation of TO1-B into nonspecific nucleic acids can be effectively controlled using the Mango system. As modifying thiazole orange to TO1-B considerably destabilizes the weak intercalation of thiazole orange into dsDNA[31] and RNA[8,17], further optimization of the imaging approaches presented here appears likely to enable higher contrast RNA Mango imaging in the future.

Equally important for RNA imaging is the photostability and biological compatibility of the Mango system. The RNA Mango aptamers are at least an order of magnitude more photostable than the Broccoli systems and comparable to the recently published Corn aptamer, albeit Corn's requirement for dimerization seems less compatible with its simple biological utilization[30,33]. Furthermore, pulsed illumination dramatically enhances the imaging time possible with RNA Mango, opening up the interesting possibility to track biological RNAs for 10 min or longer. Just as critical, our results with three small non-coding RNAs demonstrate that Mango tags can be incorporated either into non-essential stems or as 3′-tags, without significantly interfering with RNA subcellular localization and we do not believe a folding scaffold is required for the Mango system as evidenced by the success of the U6 and scaRNA constructs. Indeed, using either genetically encoded Mango-tagged RNA or direct RNA transfections correctly recapitulates the expected localization patterns.

There are many other applications of the Mango systems that appear likely based on the established biochemistry and photophysics of these aptamer complexes. First, as we have demonstrated, via the observation of quantized photobleaching of RNA foci, as few as four RNA Mango can be robustly imaged within one puncta. In the future, tagging biological RNAs with a small number of Mango repeats could, therefore, enable robust single-molecule RNA imaging. Second, our ability to sort the latest Mango constructs with microfluidics, together with our observation of Mango I dependent fluorescence in bacteria via FACs[8] strongly suggests, at least for highly expressed biological RNAs, that the RNA Mango system can be used to enable RNA-based FACs experiments. Third, the high affinity of Mango I-based tags to TO1-desthiobiotin has been used to recover native RNP complexes from streptavidin beads[34]. Such pulldowns should be readily applicable with our new Mango constructs potentially enabling a unique combination of RNA imaging and RNP pulldown experiments via the insertion of a single Mango tag into a biological RNA of interest. Finally and while not utilized here, the broad salt tolerance of the new Mango aptamers in contrast to that of Mango I and other fluorophore aptamer systems, makes the Mango system compatible with a range of enzymatic reactions. This implies that in addition to the in-cell demonstrations given here, in the future, in vitro fluorescent applications can be developed that make use of the Mango systems high brightness and fluorophore-binding properties.

While the full biological compatibility of the Mango aptamers is still not completely explored, their small size relative to other fluorogenic RNA aptamers, their ability to fold correctly into monomers at physiological temperatures[16], combined with their unusual ability to withstand formaldehyde fixation all promise to be very useful for investigating biological systems in the future.

Methods

High-throughput screening. High-throughput screening proceeds in three major stages:

(i) Digital droplet PCR: DNA libraries were diluted in 200 µg/mL yeast total RNA solution (Ambion) as described before[35] to have a final average number of DNA molecule per droplet (λ) of ~0.2–1 (Supplementary Tables 1–3). One microlitre of this solution was introduced in 100 µL of a PCR mixture containing 0.2 µM of forward primer (5′–CTTTAATACGACTCACTATAGGAACCCGCAAGCCATC), 0.2 mM of reverse primer (5′–CAGAATCTCACACAGCC), 0.2 mM of each dNTP, 0.67 mg/mL Dextran-Texas Red 70 kDa (Molecular Probes), 0.1% Pluronic F68, 2 µL Phire II DNA polymerase (Thermo Scientific, concentration unavailable) and the supplied buffer (proprietary to Thermo Fisher) to recommended concentrations. The mixture was loaded in a length of polytetrafluoroethylene (PTFE) tubing (I.D. 0.75 mm tubing; Thermo Scientific) and infused into a droplet generator microfluidic device[10] where it was dispersed into 2.5 pL droplets (production rate of ~12,000 droplets per s) carried by HFE 7500 fluorinated oil (3 M) supplemented with 3% of a fluorosurfactant[35]. Droplet production frequency was monitored and used to determine droplet volume by adjusting pumps flow rates (MFCS, Fluigent). Emulsions were collected in 0.2 µL tubes as described[35] and subjected to an initial denaturation step of 2 min at 95 °C followed by 30 PCR cycles of: 30 s at 95 °C, 30 s at 55 °C, 1 min 30 s at 72 °C.

(ii) Droplet fusion: PCR droplets were then injected into a fusion device[35] at a rate of ~1500 droplets per s, spaced by a stream of HFE 7500 fluorinated oil supplemented with 2% fluorosurfactant. Each PCR droplet was synchronized with a 16 pL IVT droplet containing 2.2 mM of each NTP (Larova), 24 mM MgCl$_2$, 44 mM Tris-HCl pH 8.0, 50 mM KCl, 5 mM DTT, 1 mM Spermidine, 35 µg/mL of Dextran-Texas Red 70 kDa (Molecular Probes), 0.1% Pluronic F68, 3500 U T7 RNA polymerase (purified in the laboratory and estimated to have an activity around 2500 U/µL by comparing it with commercial enzyme), 100 nM TO1-B[8], 5 ng/µL inorganic pyrophosphatase (Roche) supplemented with the desired concentration of NMM. For the screenings performed in the presence of TO3-Biotin[8], the T7 RNA polymerase (New England Biolabs) concentration we reduced to 70 U per reaction. The IVT mixture was loaded in a length of PTFE tubing (I.D. 0.75 mm tubing; Thermo Scientific) that was kept on ice during all the experiment. IVT droplets of 16 pL were produced at a rate of ~1500 droplets per s and paired to one PCR droplet. Pairwise droplets were then fused by electrocoalescence while passing between a pair of electrodes subjected to an AC electric field of 400 V (30 kHz) via high-voltage amplifier (Model 623b, Trek)[10]. The resulting emulsion was collected off-chip and incubated for 120 min (high concentration of T7 RNA polymerase, NMM screenings) or 30 min (low concentration of T7 RNA polymerase, TO3-Biotin screenings) at 37 °C.

(iii) Droplet analysis and sorting: the emulsion was finally re-injected into an analysis and sorting microfluidic device mounted on a Thermo plate (Tokai Hit) holding the temperature at 45 °C as previously described[10]. Droplets were re-injected at a frequency of ~200 droplets per s, spaced with a stream of surfactant-free HFE 7500 fluorinated oil. The green fluorescence (TO1-B in complex with the aptamer) of each droplet was analysed. Between 1 and 2% green fluorescence droplets were gated for each round of selection. The gated droplets were deflected into a collecting channel by applying a 1 ms AC field (1200 V, 30 kHz) and were collected into a 1.5 mL tube. Collected droplets were recovered by flushing 200 µL of surfactant-free HFE 7500 fluorinated oil (3 M) through the tubing. 1H, 1H, 2H, 2H-perfluoro-1-octanol of 100 µL (Sigma-Aldrich) and 200 µL of 200 µg/mL yeast total RNA solution (Ambion) were then added, the droplets broken by vortexing the mixture and DNA-containing aqueous phase was recovered.

Quantification of RNA produced in droplets. A PCR mixture supplemented with DNA coding for RNA Mango (starting with 10 copies of template DNA molecules per droplet to ensure that all the droplets were occupied) was emulsified in 2.5 pL droplets and the DNA amplified as above. The droplets were paired and fused with droplets of IVT mixture containing either a low (70 U of enzyme from New England Biolabs) or a high (20 µg/mL of enzyme purified in the lab) concentration of T7 RNA polymerase and the resulting emulsions were incubated for respectively 30 min or 120 min at 37 °C. After incubation, the RNA-containing phase was recovered using 1H, 1H, 2H, 2H-perfluoro-1-octanol (Sigma-Aldrich) and the transcription was stopped by a phenol extraction followed by an ethanol precipitation in the presence of 300 mM sodium acetate pH 5.5 (Sigma-Aldrich). After centrifugation and a wash in 70% ethanol, the pellets were re-suspended in water. Baseline-Zero™ DNase of 10 U (Epicentre) and the corresponding buffer were added and a second incubation of 60 min at 37 °C was performed. The DNase was removed by phenol extraction and RNA recovered by ethanol precipitation.

Recovered RNAs were reverse-transcribed for 60 min at 55 °C, followed by 5 min at 95 °C, in a mixture containing 1 µM of reverse primer, 0.5 mM of each dNTP, 8 U/µL RT Maxima (Thermo Scientific) and the supplied buffer according to recommended concentrations. The complementary DNA (cDNA) was amplified using SsoFast™ Evagreen supermix (Bio-Rad) supplemented with 0.2 µM of each primer (forward and reverse) using a CFX96 Touch™ real-time PCR detection system (Bio-Rad). Finally, the cDNA was quantified using the calibration curve obtained with reactions performed with purified Mango II RNA.

Enrichment measurement. The pool molecules contained in 2 µL recovered from the sorted fractions were introduced into 100 µL of PCR mixture containing 0.1 µM of each primer (fwd and rev), 0.2 mM of each dNTP, 0.05 U/µL of DreamTaq™ and its corresponding buffer (Fermentas). The mixture was then subjected to an initial denaturation step of 30 s at 95 °C, followed by 20 cycles of: 5 s at 95 °C and 30 s at 60 °C. 20 µL of PCR products were then in vitro-transcribed in 250 µL of mixture containing 2 mM of each NTP, 25 mM MgCl$_2$, 40 mM Tris-HCl pH 8.0, 5 mM DTT, 1 mM Spermidine and 70 µg/mL T7 RNA polymerase. After 4 h of incubation at 37 °C, 10 U of Baseline-Zero™ DNase (Epicentre) and the corresponding buffer were added and the mixture was incubated for 60 min at 37 °C. RNAs were recovered by phenol extraction followed by an ethanol precipitation in the presence of 300 mM sodium acetate pH 5.5 (Sigma-Aldrich). After centrifugation and a wash in 70% ethanol, the pellets were dissolved in denaturing loading buffer (0.05% bromophenol blue, 0.05% xylene cyanol, 20% glycerol, 1× TBE, 8 M urea) and the solution loaded onto a 12% denaturing 8 M urea polyacrylamide gel. The piece of gel containing RNA was identified by ultraviolet (UV) shadowing, sliced from the gel and transferred into dialysis tubing (molecular weight cut-off (MWCO) = 3500, Spectrum Lab) filled with TBE. RNA was electro-eluted by placing the montage in TBE for 60 min at 100 V. Eluted RNA were filtered in centrifuge tube (porosity 0.45 µm, VWR) and ethanol precipitated in the presence of 300 mM sodium acetate pH 5.5. After centrifugation and a wash in 70% ethanol, the pellets were dissolved in DEPC-Treated water and quantified with Nanodrop (Thermo Scientific).

In the case of NMM screenings, 2 µM of RNA were incubated with 100 nM of TO1-B in 40 mM Tris-HCl pH 8.0, 50 mM KCl, and 22 mM MgCl$_2$ and TO1-B fluorescence (ex. 492 nm/em. 516 nm) was measured. In the case of NMM resistance measurement, the mixture was further supplemented with 3 µM NMM. In the case of TO3-Biotin screenings, 300 nM of RNA and 100 nM of TO1-B were used with or without 110 nM of TO3-Biotin. Both green (ex. 492 nm/em. 516 nm) and red (ex. 635 nm/em. 665 nm) fluorescence were measured.

TA cloning and colony screening. Genes contained in the libraries were diluted in a PCR mixture as immediately above and thermocycled 25 times using a final extension step of 10 min at 72 °C. PCR products were inserted in pTZ57R/T vector following manufacturer's instruction (InsTAclone PCR cloning Kit, Thermo Scientific). Ligation products were recovered by phenol/chloroform extraction and ~100 ng of DNA used to transform Electro-10 blue bacteria (Agilent) placed in a 2 mm electroporation (MicroPulser, Bio-Rad). After an hour of recovery at 37 °C under agitation, bacteria were plated on Luria broth (LB)-ampicillin agar plate and incubated overnight at 37 °C. The colonies were picked, used to inoculate liquid LB and grown at 37 °C until saturation. Plasmids DNA were extracted using 'GeneJet Plasmid Miniprep kit' (Thermo Scientific), and sequences determined by Sanger approach (GATC Biotech).

Single colonies were introduced in 10 µL of a PCR mixture identical to that used for TA cloning and the DNA amplified as above. Two microlitres of PCR product added to 18 µL of IVT mixture containing 2 mM of each NTP, 25 mM MgCl$_2$, 40 mM Tris-HCl pH 8.0, 50 mM KCl, 5 mM DTT, 1 mM Spermidine, 70 µg/mL T7 RNA polymerase and 100 nM TO1-B. The mix was then split in two and one aliquot was supplemented with 3 µM of NMM. The reaction was incubated in a real-time thermocycler (Mx 3005 P, Agilent) for 2 h at 37 °C and the green fluorescence (ex. 492 nm/em. 516 nm) measured every minute.

DMS probing of Mangos. DMS probing consists of four main steps: (i) DMS (denaturing): protocol is adapted from Lorsch and Szostak[36]. RNA of 50 nM was 3′-end-labelled with ^{32}P pCp and gel-purified. The resulting RNA was incubated in 50 mM HEPES pH 7.5 (volume 50 µL) at room temperature for 30 min. After incubation, 10 µg carrier RNA was added. The sample was then heated to 90 °C for 3 min before the addition of 0.5 µL of 25% DMS (diluted in ethanol) and heated to 80 °C for 1 min, 150 µL ice cold ethanol + 5 µL 3 M NaCl was then immediately added and the sample moved to −20 °C for 30 min. DMS-modified RNA was pelleted by centrifuge at 16,300 RCF at 4 °C for 20 min.

(ii) DMS (native): 50 nM 3′-end-labelled RNA was incubated in 50 mM HEPES pH 7.5, 1 mM $MgCl_2$, 140 mM either KCl or NaCl, with or without 500 nM TO1-B (final volume 50 µL) at room temperature for 30 min. After incubation, 10 µg carrier RNA was added. The sample was then incubated at room temperature for 15 min after the addition of 0.5 µL of 100% DMS. Ice cold ethanol of 150 µL + 5 µL 3 M NaCl was then immediately added and pelleted as for the denaturing DMS protocol.

(iii) Reduction: pellets were re-suspended in 10 µL 1 M Tris buffer pH 8 and 10 µL of freshly prepared 0.2 M sodium borohydride was added. Reaction was carried out on ice and in the dark for 30 min. Reactions were stopped by ethanol precipitation as above.

(iv) Aniline cleavage: to the resulting pellet, 20 µL (1 part Aniline, 7 parts ddH_2O, 3 parts glacial acetic acid) were added and incubated at 60 °C for 15 min in the dark. Samples were flash-frozen by placing tubes in liquid nitrogen and lyophilized by speed vacuum centrifuge. Once dry, 20 µL ddH_2O was added, the sample was refrozen and lyophilized once again. The pellet was re-suspended in a 50% formamide denaturing solution before being loaded on a 15% polyacrylamide gel (19:1 acrylamide:bis).

T1 RNase ladder and alkaline hydrolysis ladder. 3′-end-labelled RNA of 200 pmol was incubated in 20 mM sodium citrate, 6.3 M urea, and 1 U/µL T1 RNase (Thermo Scientific) at 50 °C for 10 min. Samples was flash-frozen in liquid nitrogen for 5 min, heat denatured in denaturing solution at 95 °C for 5 min prior to gel loading. Hydrolysis ladders were generated by incubating in 50 mM $NaHCO_3$ at 90 °C for 20 min and neutralizing using 1 M Tris-HCl.

Screening for minimal functional Mango motifs. To identify the minimal functional motif of each Mango, truncated constructs were designed as shown in Supplementary Fig. 7. DNA constructs (IDT) were transcribed by run-off transcription using T7 RNA polymerase. RNA was gel-purified on 10% 19:1 acrylamide:bis polyacrylamide gels. RNA concentrations were determined by NanoDrop readings at A_{260}, where extinction coefficients were estimated based on an average 11,000 M^{-1} cm^{-1} per base.

Affinity measurements of Mango variants. Fluorescence data were gathered using a Varian Cary Eclipse Spectrophotometer unless otherwise stated. Fluorescent titrations were measured in a buffer mimicking cellular conditions (WB: 140 mM KCl, 1 mM $MgCl_2$, 10 mM NaH_2PO_4 pH 7.2, 0.05% Tween-20) to determine binding affinities. Fluorescence was measured at the maximum excitation and emission wavelengths of each complex (Supplementary Fig. 11). Curves were fitted using least squares (Kaleidagraph 4.5) using the following equation for TO1-B:

$$F = F_0 + \frac{F_{\max}}{2}\left(K_D + [\mathrm{RNA}] + [\mathrm{TO1\text{-}B}] - \sqrt{([\mathrm{RNA}] - [\mathrm{TO1\text{-}B}])^2 + K_D(K_D + 2[\mathrm{RNA}] + 2[\mathrm{TO1\text{-}B}])}\right), \qquad (1)$$

where F is the fluorescence at a given (RNA), F_0 is the unbound fluorescence and $F_{\max}$ the maximal complex fluorescence, respectively. When F_0 was undetectable, it was set to zero.

Or to the following equation for TO3-Biotin and NMM experiments:

$$F = F_0 + \frac{F_{\max}[\mathrm{RNA}]}{K_D + [\mathrm{RNA}]}, \qquad (2)$$

$F_{\max}$ was determined using Eq. 1 or Eq. 2, as appropriate.

Temperature-dependent fluorescence and UV-melting curves. Temperature-dependence measurements were started at 90 °C decreasing at a rate of 1 °C per min until 20 °C, then returned at 1 °C per min until 90 °C was reached. Fluorescence measurements were obtained at the maximum excitation/emission of the fluorescent complex used and were measured in WB buffer using 1 µM RNA either with or without 5 µM TO1-B. Temperature dependence of fluorescence and absorbance were measured using a Varian Cary Eclipse Fluorescence Spectrophotometer at excitation and emission peaks and a Varian Cary 100 Bio UV-visible spectrophotometer monitoring at 260 nm.

Circular dichroism. Circular dichroism spectra were obtained on an Applied Photophysics Chirascan Circular Dichroism Spectrometer using 5 µM RNA, 140 mM monovalent salts and 7 µM TO1-B. Spectra were scanned in 1 nm steps with a bandwidth of 1 nm. Data shown is the average of three repeats. Samples were measured using a 1 mm pathlength quartz cuvette (Starna Cells Inc.).

Formaldehyde resistance assay. RNA Mango aptamers were incubated with TO1-B in WB buffer for at least 1 h until equilibrium fluorescence was reached. Formaldehyde was then added such that final concentrations after dilution were 50 nM RNA, 100 nM TO1-B, and 0, 2, 4 or 8% formaldehyde. Fluorescence was measured as a kinetic run at a rate of 2 readings per min using a Varian Cary Eclipse Fluorescence Spectrophotometer, ex/em = 510 ± 2.5/535 ± 5 nm.

Cell culture and maintenance. HEK293T cells (293T-ATCC® CRL-3216™) were grown in Dulbecco Modified Eagle's Medium containing 10% foetal bovine serum, 2 mM D-Glucose, 2 mM L-Glutamine, 1 mM sodium pyruvate and 100 U/mL penicillin/streptomycin (Thermo Fisher) and maintained at 37 °C with 5% CO_2 in a humidified incubator. Cells used for imaging were cultured in Ibidi glass-bottomed eight-well chamber slides (Ibidi GmbH).

Plasmid and RNA synthesis. DNA encoding the F30 folding scaffold[20] was modified to incorporate the Mango RNA sequences (Supplementary Table 5) and ordered from (Integrated DNA Technologies). The DNA was amplified by PCR to incorporate 5′ SalI and 3′ XbaI restriction sites (Supplementary Table 6). PCR products were digested using Fast Digest enzymes (Thermo Fisher) and ligated into SalI/XbaI linearized and shrimp alkaline phosphatase (NEB)-treated pAV5S-F30-2xdBroccoli (Addgene plasmid 66845, a gift from Dr. S. Jaffrey). DNA for both the 5S rRNAs and the scaRNAs were amplified by PCR to incorporate a 5′ BstXI site and a 3′ XhoI site (Supplementary Table 6). PCR products were digested using Fast Digest enzymes (Thermo Fisher) and ligated into BstXI/XhoI linearized and shrimp alkaline phosphatase (NEB) treated pSLQ1651-sgTelomere(F + E) (a gift from Bo Huang and Stanley Qi, Addgene plasmid # 51024). For RNA synthesis, DNA encoding the full 5S-F30-Mango/control sequences were PCR-amplified to include a 5′ T7 RNA polymerase promoter. DNA was transcribed in vitro with T7 RNA polymerase (NEB) at 37 °C for 16 h in 40 mM Tris-HCl, 30 mM $MgCl_2$, 2 mM spermidine, 1 mM dithiothreitol, 5 mM rNTPs, 1 U/µl *Escherichia coli* inorganic pyrophosphatase, 4 U/µl T7 RNA polymerase (pH 7.9). RNA was purified from an 8 M urea, 12% denaturing polyacrylamide gel using 29:1 acylamide:bis solution (Fisher Scientific). The RNA was excised and eluted in RNA elution buffer (40 mM Tris-HCl pH 8.0, 0.5 M sodium acetate, 0.1 mM EDTA) followed by ethanol precipitation. Fluorescence measurements were taken for each of the RNA constructs using a Varian Cary Eclipse Fluorescence Spectrophotometer (Agilent) containing 40 nM TO1-B, 200 nM RNA, 10 mM sodium phosphate, 100 mM KCl and 1 mM $MgCl_2$ at pH 7.2. Similar measurements were also taken with a limiting amount of RNA (40 nM) in an excess of TO1-B (200 nM) and the results showed a similar trend. U6-Mango/control RNA was synthesized by the PCR amplification of a 5′ T7 sequence to each construct followed by IVT and purification as described above.

Plasmid and RNA transfection. RNA was transfected directly into eight-well chamber slides using the lipofectamine-based CRISPRMAX reagent following the manufacturers guidelines (Invitrogen). A final concentration of 62.5 nM RNA in 10 mM sodium phosphate buffer (pH 7.2), 100 mM KCl, and 1 mM $MgCl_2$ was incubated at room temperature with a 1:1 dilution in OptiMEM prior to transfection. The RNA transfected was incubated at 37 °C for 1 h in complete growth medium. FuGene 6 was used to transfect 400 ng of the pSLQ--Mango and control plasmids directly in the eight-well chamber slides following the manufacturer's instructions. Plasmids were left to express between 12–48 h before fixation as described below.

Cell fixation and immunostaining. Cells were fixed in PBS containing 4% paraformaldehyde for 10 min on ice followed by permeabilization in 0.2% Triton X-100 for 10 min at room temperature. Cells not requiring immunostaining were washed three times for 5 min each with PKM buffer (10 mM Sodium Phosphate, 100 mM KCl and 1 mM $MgCl_2$) followed by a 10 min incubation in 200 nM TO1-B diluted in PKM buffer before replacing with imaging media (10 mM sodium phosphate, 100 mM KCl and 1 mM $MgCl_2$ 1 µg/mL Hoechst 33258). For immunostaining, cells were first blocked (2% BSA in PBS) for 30 min followed by primary antibody (1:50 – 1:500 dilutions) incubation for 120 min in blocking solution.

Primary antibodies used here were: anti-ribosomal protein S6 (MAB5436, R&D Systems–8 µg/mL), anti-ribosomal protein L7 (ab72550, Abcam—1 µg/mL), anti-ATP5B (ab14730, Abcam—1 µg/mL), anti-GW182 (ab7052, Abcam—5 µg/mL), anti-EEA-1 (ab70521, Abcam—1 µg/mL), anti-LSm3 (NBP2-14206, Novus Biologicals—1 µg/mL), anti-TIAR (sc-398372, Santa Cruz—4 µg/mL). Secondary antibodies used were donkey anti-mouse and donkey anti-rabbit Alexa Fluor 680 (Molecular Probes). Primary antibodies were washed three times for 20 min each in blocking solution followed by incubation with secondary antibody at 1:500 dilution for 60 min, which was subsequently washed as above. After immunostaining the cells were washed and stained in TO1-B and Hoechst 33258 as described previously.

Fluorescence microscopy and live-cell imaging. Live and fixed cell images were taken directly in the eight-well chamber slides using a Zeiss Elyra wide-field microscope by exciting at 405 nm (blue), 488 nm (green), 561 nm (red) and 642 nm (far-red) and detecting emission at 420–480 nm, 495-550 nm, 570–640 nm and >650 nm, respectively. Image acquisition for the Mango signal (488 nm laser) used 5 mW of power and 200 ms exposure time, except in the photobleaching-assisted microscopy experiments where 50 ms was used. Due to the observation that the Mango signal is stabilized under pulsed illumination in both fixed and live cells, Z-stacks and time series experiments containing more than one colour were acquired by alternating between each channel for an individual frame, leading to recovery and minimal loss of the Mango signal throughout the acquisition. To visualize the nuclear boundary in live cells, a plasmid expressing a fluorescently tagged histone protein (EBFP2-H2B-6, Addgene plasmid 55243) was transfected using FuGene 6 (Promega) 24 h prior to RNA transfection. RNA was transfected directly into eight-well chamber slides (Ibidi GmbH) as described above, with an additional pre-incubation step with 125 nM of TO1-B prior to the addition of the CRISPRMAX transfection reagent. Upon addition to the imaging chambers, the final effective concentrations of RNA and fluorophore were 10 and 20 nM, respectively. Following incubation of the RNA transfection, the cells were washed once with PBS and replaced with live-cell imaging media (fluorobrite DMEM supplemented with 10% FBS, 2 mM D-Glucose, 2 mM L-Glutamine, 1 mM sodium pyruvate and 10 mM HEPES, Invitrogen). Live cells were maintained at 37 °C with 5% CO_2 in a stage top incubator (Tokai Hit).

Photobleaching-assisted microscopy. To image the photobleaching of 5S-F30-Mango IV foci as compared with dBroccoli in live cells, 5S-F30-Mango I and IV were directly transfected as described above, whereas 5S-F30-dBroccoli was expressed from pAV5S as previously described[13,20]. Live-cell photobleaching was conducted with a 200 ms frame rate and 5% wide-field laser illumination at 488 nm for 150 frames. To determine the half-life of fluorescence, each photobleaching curve was fit to an exponential decay function. In order to obtain the appropriate signal to noise ratio and time resolution for the analysis of single-step photobleaching, 5S-F30-Mango IV foci were imaged in fixed cells with a 50 ms frame rate, 5% wide-field laser illumination at 488 nm for 800 frames. Maximum likelihood estimation was used to determine each of the photobleaching steps within a trace as previously described[26,27]. The step sizes were subsequently binned and the histogram was fit to a double Gaussian equation.

Image processing and quantification. Images were processed using FIJI and spot detection analysis was performed on each maximum projection by the spot detector plugin in the ICY image analysis software, which detects significant foci with a pixel area ≥3 × 3 pixels and intensity ≥300 a.u. A lower threshold of ≥150 a.u. was used to create a population of the apparent background intensities in the control RNA transfections (Fig. 4b). To create the 2D co-localization plots (Supplementary Fig. 16, 17), five to six images for each condition (~100 cells) were processed and the normalized intensity (max = 1, min = 0) of each pixel in both the TO1 channel (y axis) and the AlexaFluor 680 channel (x axis) was plotted. To append a Z-axis density of pixels, Igor Pro was used to carry out a bivariate histogram upon corresponding pixels for both channels and displayed as a heatmap.

pH titrations. Mango of 50 nM was incubated with 100 nM TO1-B in the presence of 140 mM KCl, 1 mM $MgCl_2$, and varying pH (50 mM sodium citrate buffer for pH 3–6, 50 mM sodium phosphate for pH 6–8, 50 mM Tris for pH 8–9) for 1 h at room temperature. Mango fluorescence was measured with excitation and emission at 485 nm and 535 nm, respectively.

In vitro photobleaching measurement. 5S-F30-Mango I, II, III and IV as well as 5S-F30-Broccoli template regions were placed under the control of T7 RNA polymerase promoter. Genes were PCR-amplified, in vitro-transcribed, purified and quantified as before (see Enrichment measurement section). Then 1 volume of 3 µM RNA solution was added to 1 volume of 2-times concentrated buffer (280 mM KCl, 2 mM $MgCl_2$, 20 mM NaH_2PO_4 pH 7.5) supplemented with 3.6 µM of fluorogenic dye (DFHBI-1T for Broccoli aptamer, and TO1-B for Mango aptamers). The mixture was incubated for an hour at room temperature prior to being loaded into a length of PTFE tubing (Thermo Fisher) and infused into a droplet generator microfluidic device where it was dispersed into 100 pL droplets carried by HFE 7500 fluorinated oil (3 M) supplemented with 3% of a fluorosufactant as described previously[37]. The resulting emulsion was then loaded into 5 µL capillary (Corning) and the montage was imaged on an epifluorescence microscope (TiE, Nikon). Depending on the dye, the emulsion was exposed to a constant illumination wavelength of 470 nm (DFHBI-1T) or 508 nm (TO1-B) at the maximum intensity of the light source (Spectra X, Lumencor), and the emitted fluorescence (respectively, 514 nm ± 24 and 540 nm ± 12) was collected by an Orca-Flash IV camera for 200 ms every 100 ms with ×20 objective (numerical aperture (NA) 0.45). Fluorescence intensity of each picture was then extracted using NiS software (Nikon) and the data were fit to an exponential decay equation to compute the fluorescence half-life.

Data availability. The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Received: 20 November 2017 Accepted: 11 January 2018
Published online: 13 February 2018

References

1. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA mimics of green fluorescent. *Protein Sci.* **333**, 642–646 (2011).
2. Strack, R. L., Disney, M. D. & Jaffrey, S. R. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat. Methods* **10**, 1219–1224 (2013).
3. You, M. & Jaffrey, S. R. Structure and mechanism of RNA mimics of green fluorescent protein. *Annu. Rev. Biophys.* **44**, 187–206 (2015).
4. Ouellet, J. RNA fluorescence with light-up aptamers. *Front. Chem.* **4**, 29 (2016).
5. Babendure, J. R., Adams, S. R. & Tsien, R. Y. Aptamers switch on fluorescence of triphenylmethane dyes. *J. Am. Chem. Soc.* **125**, 14716–14717 (2003).
6. Warner, K. D. et al. Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat. Struct. Mol. Biol.* **21**, 658–663 (2014).
7. Huang, H. et al. A G-quadruplex–containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.* **10**, 686–691 (2014).
8. Dolgosheina, E. V. et al. RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking. *ACS Chem. Biol.* **9**, 2412–2420 (2014).
9. Filonov, G. S., Moon, J. D., Svensen, N. & Jaffrey, S. R. Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc.* **136**, 16299–16308 (2014).
10. Autour, A., Westhof, E. & Ryckelynck, M. iSpinach: a fluorogenic RNA aptamer optimized for in vitro applications. *Nucleic Acids Res.* **44**, 2491–2500 (2016).
11. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat. Methods* **2**, 905–909 (2005).
12. Dolgosheina, E. V. & Unrau, P. J. Fluorophore-binding RNA aptamers and their applications. *Wiley Interdiscip. Rev. RNA* **7**, 843–851 (2016).
13. Zhang, J. et al. Tandem spinach array for mRNA imaging in living bacterial cells. *Sci. Rep.* **5**, 17295 (2015).
14. Pothoulakis, G., Ceroni, F., Reeve, B. & Ellis, T. The spinach RNA aptamer as a characterization tool for synthetic biology. *ACS Synth. Biol.* **3**, 182–187 (2014).
15. Guet, D. et al. Combining spinach-tagged RNA and gene localization to image gene expression in live yeast. *Nat. Commun.* **6**, 8882 (2015).
16. Trachman, R. J. III et al. Structural basis for high-affinity fluorophore binding and activation by RNA Mango. *Nat. Chem. Biol.* **13**, 807–813 (2017).
17. Jeng, S. C. Y., Chan, H. H. Y., Booy, E. P., McKenna, S. A. & Unrau, P. J. Fluorophore ligand binding and complex stabilization of the RNA Mango and RNA Spinach aptamers. *RNA* **22**, 1884–1892 (2016).
18. Nicoludis, J. M. et al. Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. *J. Am. Chem. Soc.* **134**, 20446–20456 (2012).
19. Krasilnikov, A. S., Yang, X., Pan, T. & Mondragón, A. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421**, 760–764 (2003).
20. Filonov, G. S., Kam, C. W., Song, W. & Jaffrey, S. R. In-gel imaging of RNA processing using broccoli reveals optimal aptamer expression strategies. *Chem. Biol.* **22**, 649–660 (2015).
21. Smirnov, A., Entelis, N., Martin, R. P. & Tarassov, I. Biological significance of 5S rRNA import into human mitochondria: role of ribosomal protein MRP-L18. *Genes Dev.* **25**, 1289–1305 (2011).
22. Sun, J. S. & Manley, J. L. The human U6 snRNA intramolecular helix: structural constraints and lack of sequence specificity. *RNA* **3**, 514–526 (1997).
23. Guo, Z., Karunatilaka, K. S. & Rueda, D. Single-molecule analysis of protein-free U2–U6 snRNAs. *Nat. Struct. Mol. Biol.* **16**, 1154–1159 (2009).
24. Karunatilaka, K. S. & Rueda, D. Post-transcriptional modifications modulate conformational dynamics in human U2–U6 snRNA complex. *RNA* **20**, 16–23 (2014).
25. Han, K. Y., Leslie, B. J., Fei, J., Zhang, J. & Ha, T. Understanding the photophysics of the spinach–DFHBI RNA aptamer–fluorogen complex to improve live-cell RNA imaging. *J. Am. Chem. Soc.* **135**, 19033–19038 (2013).
26. Watkins, L. P. & Yang, H. Detection of intensity change points in time-resolved single-molecule measurements. *J. Phys. Chem. B* **109**, 617–628 (2005).
27. Liao, Y., Li, Y., Schroeder, J. W., Simmons, L. A. & Biteen, J. S. Single-molecule DNA polymerase dynamics at a bacterial replisome in live cells. *Biophys. J.* **111**, 2562–2569 (2016).

28. Marnef, A., Richard, P., Pinzón, N. & Kiss, T. Targeting vertebrate intron-encoded box C/D 2′-O-methylation guide RNAs into the Cajal body. *Nucleic Acids Res.* **42**, 6616–6629 (2014).
29. Tan, X. et al. Fluoromodules consisting of a promiscuous RNA aptamer and red or blue fluorogenic cyanine dyes: selection, characterization, and bioimaging. *J. Am. Chem. Soc.* **139**, 9001–9009 (2017).
30. Song, W. et al. Imaging RNA polymerase III transcription using a photostable RNA-fluorophore complex. *Nat. Chem. Biol.* **13**, 1187–1194 (2017).
31. Lee, L. G., Chen, C.-H. & Chiu, L. A. Thiazole orange: a new dye for reticulocyte analysis. *Cytometry* **7**, 508–517 (1986).
32. Szent-Gyorgyi, C. et al. Fluorogen-activating single-chain antibodies for imaging cell surface proteins. *Nat. Biotechnol.* **26**, 235 (2008).
33. Warner, K. D. et al. A homodimer interface without base pairs in an RNA mimic of red fluorescent protein. *Nat. Chem. Biol.* **13**, 1195–1201 (2017).
34. Panchapakesan, S. S. S. et al. Ribonucleoprotein purification and characterization using RNA Mango. *RNA* **23**, 1592–1599 (2017).
35. Ryckelynck, M. et al. Using droplet-based microfluidics to improve the catalytic properties of RNA under multiple-turnover conditions. *RNA* **21**, 458–469 (2015).
36. Lorsch, J. R. & Szostak, J. W. In vitro selection of RNA aptamers specific for cyanocobalamin. *Biochemistry* **33**, 973–982 (1994).
37. Fernandez-Millan, P., Autour, A., Ennifar, E., Westhof, E. & Ryckelynck, M. Crystal structure and fluorescence properties of the iSpinach aptamer in complex with DFHBI. *RNA* **23**, 1788–1795 (2017).

Acknowledgements

We thank Eric Westhof (Univesité de Strasbourg) and Dipankar Sen (Simon Fraser University) for critical reading of the manuscript, Zewei Ding (Simon Fraser University) for help with plasmid construction, and Julie Biteen (University of Michigan) for help with the step photobleaching analysis. We thank the microscopy core facility at the MRC London Institute of Medical Sciences. We acknowledge support by the MRC core grant (MC-A658-5TY10 to DR) and a NSERC operating grant (P.J.U.). This work has been published under the framework of the LABEX: ANR-10-LABX-0036_NETRNA and benefits from funding from 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 (M.R.).

Author contributions

A.A., S.J., A.C., D.R., M.R. and P.J.U. designed the experiments, analysed the data and wrote the manuscript. A.A. performed the fluorescence activated microfluidic sorting of the R12 library. S.J., A.Ab., A.G. and S.S.S.P. performed subsequent aptamer characterisation and DMS probing. A.C. performed the cellular imaging of the small non-coding RNAs.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-02993-8.

Competing interests: P.J.U., M.R., S.J., A.A., A.Ab. and S.S.S.P. have filed a provisional patent on aspects of this work. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018

The work presented in the **Original article 3**, allowed me to isolate three new Mango variants, all being able to light-up TO1-Biotin to different degree while forming high affinity complexes since the lowest affinity variant (Mango IV) still displayed a K_D of ~ 10 nM, a value far below that of most the other RNA-based fluorogenic modules described to date (Bouhedda et al., 2017). This demonstrates that, even though being able to measure fluorescence in rapidly circulating droplets requires using high fluorogen concentration, introducing competitors in the assay allows balancing the potential adverse selection of low affinity aptamers. Moreover, since μIVC selects aptamers for their light-up capacity, the new variants were either as fluorogenic as (Mango II) or even much more fluorogenic (Mango III and IV) than Mango I. Therefore, used in combination with competitors, μIVC has the potential to isolate light-up aptamers both for their affinity and their fluorogenic capacity. This is nicely demonstrated by Mango III, the variant that dominated the pool at the end of the screening process. Indeed, whereas Mango III displays a nearly intact affinity for TO1-Biotin (K_D of ~ 5 nM) it also generates a complex ~ 4-fold more fluorescent than Mango I, reaching a fluorescence 1.3 times higher than that of the eGFP. Moreover, even though Mango III is also able to interact with and activate the fluorescence of TO3-Biotin, it displays a 3-fold lower affinity for this fluorogen than Mango I does, showing that Mango III also gained in specificity.

The efficiency of μIVC at identifying aptamers of interest is further supported by the fact that none of the improved sequences isolated through this work belonged to one of the sequence family identified during the original hand screening of R12 library (Dolgosheina et al., 2014) (Figure 40). This indicates that the most represented variants found in R12 were probably very good binders of TO1-Biotin but poor activator of its fluorescence. This, once again, illustrates the "you get what you select for" concept established by You and Arnold (You and Arnold, 1996) and support the idea already discussed above (see §2.2.2.2) that, even though SELEX is an efficient way of pre-enriching libraries in molecule of interest, it is less efficient at identifying optimal light-up aptamers. An ideal scenario for the efficient development of light-up aptamers should therefore combine rounds of stringent pre-enriching SELEX together with rounds of μIVC functional screening.

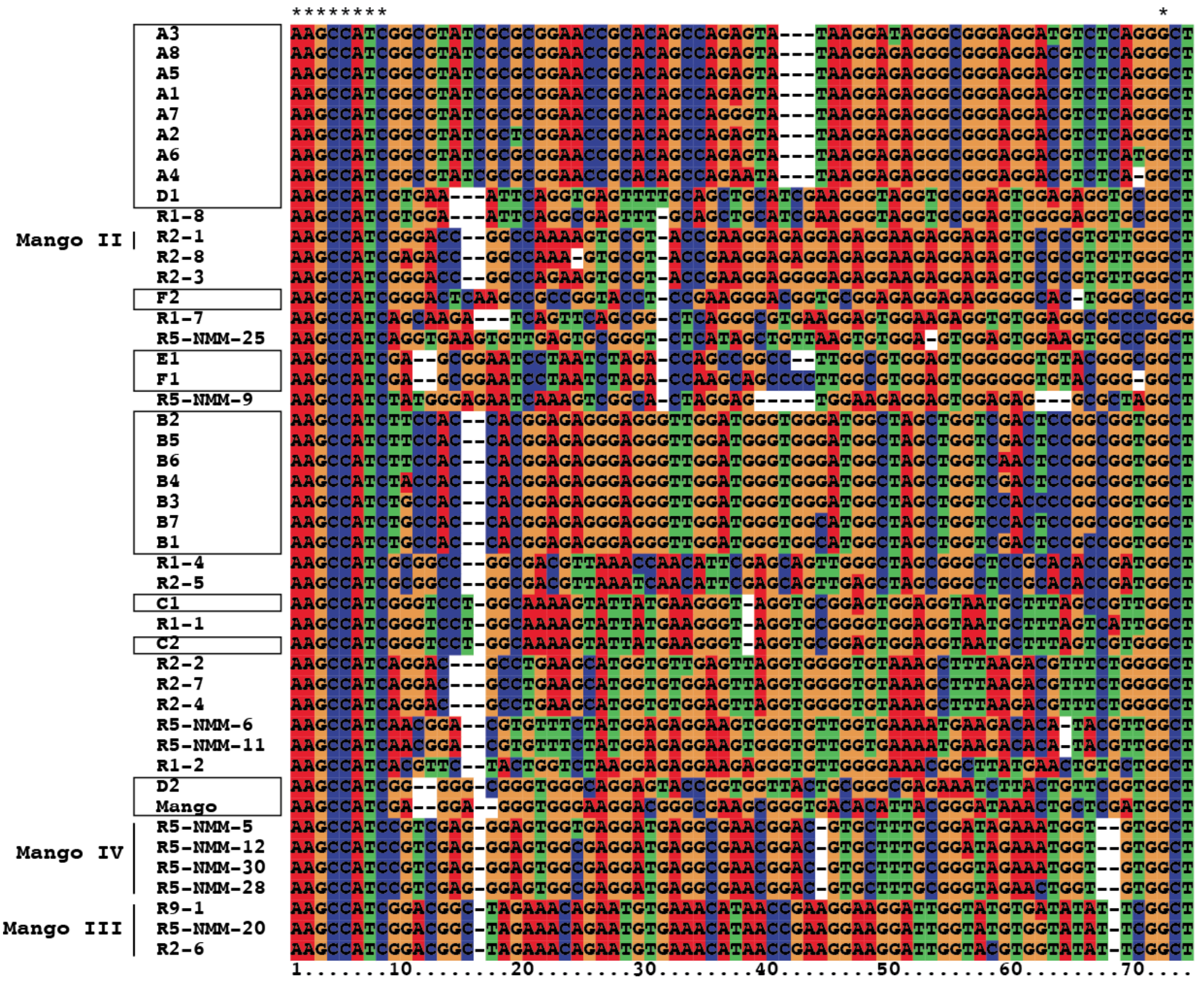

Mango II
Mango IV
Mango III
A3
A8
A5
A1
A7
A2
A6
A4
D1
R1-8
R2-1
R2-8
R2-3
F2
R1-7
R5-NMM-25
E1
F1
R5-NMM-9
B2
B5
B6
B4
B3
B7
B1
R1-4
R2-5
C1
R1-1
C2
R2-2
R2-7
R2-4
R5-NMM-6
R5-NMM-11
R1-2
D2
Mango
R5-NMM-5
R5-NMM-12
R5-NMM-30
R5-NMM-28
R9-1
R5-NMM-20
R2-6
1.......10........20........30........40........50........60........70....

Figure 40: Sequences alignment comparing molecules identified during the original hand screening of R12 with those identified upon *μ*IVC screening. Sequences from the original hand screening (Dolgosheina et al., 2014) are boxed and labeled with a letter (A to D). Other sequences starting with “R” are those identified in the **Original Article 3**.

Using this new generation of Mango aptamers enabled us to image several small non-coding RNAs. This was possible thanks to the improved fluorogenic capacity and folding efficiency of the new Mango variants, but also thanks to the 10-fold higher photostability of Mango/TO1-Biotin complex with respect to that of Spinach/DFHBI (see Figure S18 of **Original Article 3**). Each labelled RNA was found at its expected sub-cellular location, confirming our starting hypothesis that the small size of Mango together with its capacity to fold in various context does not significantly affect the functionality of the labelled RNA. As a consequence, Mango tags now make possible imaging experiments aiming at monitoring the synthesis and the location of small non-coding RNAs and future work in which small regulatory RNAs would be labelled by Mango should now be feasible. Such experiment would then make possible monitoring RNA-based gene expression regulatory events.

Since TO1-Biotin and TO3-Biotin emit fluorescence at different wavelengths, future work aiming at selecting orthogonal modules could also be undertaken using the mixed SELEX/*μ*IVC procedure discussed above. In this scheme, aptamer forming highly fluorescent complex with TO1-Biotin and interacting only poorly with TO3 could be rapidly identified. Moreover, an orthogonal module in which the aptamer would be able to efficiently activate TO3-Biotin fluorescence while not recognizing TO1-Biotin could also be developed in parallel. Such a pair of orthogonal modules would then make possible to simultaneously monitor two target RNAs in live cells. Moreover, since TO1 emission and TO3 excitation wavelengths partly overlap, these modules could serve as FRET pair to monitor RNA/RNA interaction or even structural modification within an RNA labeled by both modules like this was recently done using Spinach and Mango aptamers (Jepsen et al., 2018).

6.3. Structural characterization and structure-guided improvement of Mango aptamers

As this was the case with iSpinach, explaining the origin of aptamer properties improvement is greatly facilitated by establishing the crystal structure of the aptamer in complex with its fluorogen. Mango I/TO1-Biotin crystal structure was reported in 2017 (Trachman et al., 2017) and showed that Mango I is organized as a three-tier G-quadruplex linked to a closing helix by a split GAAA loop (Figure 41). This kite-like structure accommodates TO1-Biotin at the surface of the top quadruplex (T3) with additional interactions established with flap residues. Establishing the crystal structure of the other Mango variants in complex with TO1-Biotin and comparing it with that of Mango I/TO1-Biotin complex was then expected to make possible explaining, at least partly, the origin of aptamers improvement. Indeed, whereas Mango IV/TO1-Biotin structure has not been determined yet, solving that of Mango II/TO1-Biotin and Mango III/TO1-Biotin gave first insight on their improved properties and allowed me to isolate a new Mango III-derived variant.

6.3.1.Structural characterization of Mango II

The crystal structure of Mango II/TO1-Biotin complex has been recently solved by the group of Adrian Ferré d'Amaré (Trachman et al., 2018). As expected from the high degree of sequence similarity, Mango I and Mango II share similar folding. Yet interesting differences were observed. First, additional hydrogen bonding was observed at the level of the GAAA split-loop, suggesting an increased stability of this loop probably impacting that of the above quadruplex, an observation reminiscent to what happened with iSpinach. Moreover, the bottom tier (T1) of the quadruplex module is augmented to an hexad whereas it was a pentad in Mango I. These two elements were proposed to act together to increase Mango II overall folding stability. Second, the TO1-Biotin-binding site was found to be better suited to accommodate the fluorogen. Indeed, whereas in Mango I a G of the top G-quartet (T3) buckles at 30° out of plane, in Mango II the four Gs of the quartet are in a perfect planar conformation, forming a flat binding surface better accommodating the dye. Moreover, Mango II displays 5 flap adenines that accommodate the fluorogen much better than Mango I for which binding involves only 3 flap residues. As a net effect, these different subtle local changes allow Mango II to much better accommodate TO1-Biotin leading to a 10 times higher affinity and a 1.5 times more fluorescent complex. The structure also allowed to identify a position (A22) at which a steric clash occurred with TO-1 and the PEG linker. Interestingly, mutating this residue into a smaller pyrimidine alleviated this clash, further improving the affinity between Mango II and the TO1-Biotin.

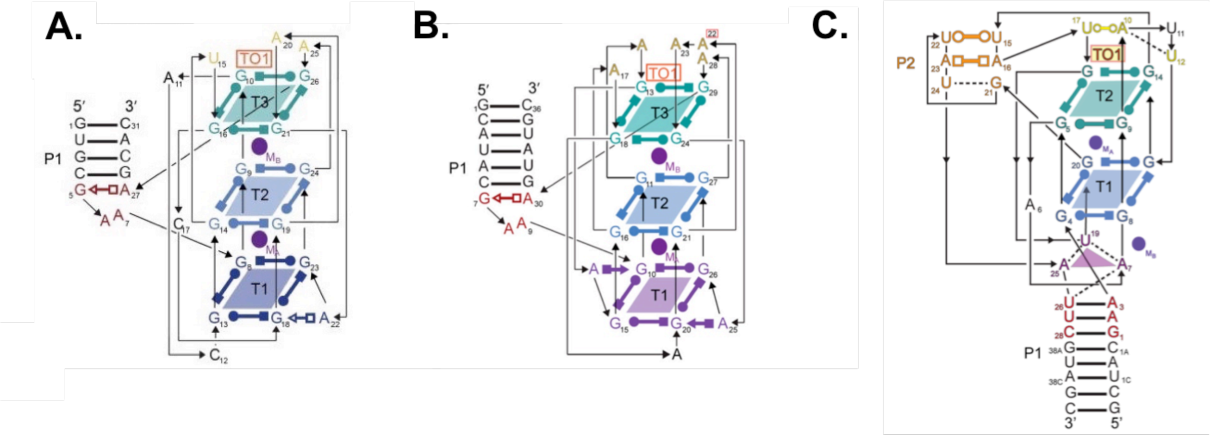


Figure 41: Secondary structures models of Mango I, II and III. **(A)** Mango I structure (adapted from (Trachman et al., 2017)) is showed side-by-side with that of Mango II **(B)**(from and adapted (Trachman et al., 2018) and Mango III (data provided by Robert Trachman)**(C)**.

6.3.2. Structural characterization and optimization of Mango III

Among all the new Mango sequences isolated above, Mango III, the dominant specie found at the end of the screening process, is the most intriguing. Indeed, Mango III is the variant displaying the least homology with Mango I and the others. As it contains only 9 Gs in its sequence, it was unlikely that Mango III can adopt the three-tier G-quadruplex folding shared by the other Mangos, suggesting a significant modification in its structure. Furthermore, Mango III is the brightest Mango variant, it displays a very high tolerance to salt and strongly binds potassium ions (Autour et al., 2018). Last but not least, Mango III displays an interesting sigmoidal melting temperature curve, similar to that observed with Spinach but shared with none of the other Mangos (see Supplementary Figure 9 in **Original Article 3**). Altogether, these observations suggest that Mango III should adopt a folding and a TO1-Biotin recognition mode quite different to those of other Mango variants. This assumption has been recently confirmed by the group of Adrian Ferré d'Amaré who recently solved the structure of Mango III in complex with TO1-Biotin (manuscript in preparation).

6.3.2.1. Crystal structure of Mango III/TO1-Biotin complex

Early 2018, the group of Adrian Ferré d'Amaré has solved the structure of Mango III in complex with TO1-Biotin which revealed some interesting and unexpected features of the aptamer (manuscript in preparation). First, and as anticipated before, Mango III is made of a two-tier G-quadruplex (Figure 41.C). However, the G-quadruplex is extended by an atypical triplex (A_7-U_{19}-A_{25}, also interacting with U_{26}) that stacks under the basal tier (T1). A second atypical feature of Mango III is the presence of the P2 stem proposed to rigidify the connection between P1 and the G-quadruplex and, doing so, increases the overall stability of the aptamer. Interestingly, P2 is made of two parallel non-canonical base pairs (U_{15}oU_{22} and A_{16}oA_{23}) and a third antiparallel G_{21}oU_{24} base-pair. Finally, last but not least, the TO1-

Biotin-binding site displays two main differences with respect to Mango I: i) as for Mango II, the top G-quadruplex layer (T2) is rather flat and offers a nice interaction area with the fluorogen and ii) a $A_{10}oU_{17}$ base pair forms a lid, further stabilized by an interaction with U_{12}, that covers the fluorogen and encloses it. This forms a closed binding site proposed to be responsible of the observed increased fluorogenic capacity of the aptamer.

The μIVC process used during Mango reselection did not involve any mutagenesis step. Therefore, one cannot exclude that some regions of Mango III, especially the atypical ones, are still sub-optimal; as this was the case with the residue A22 in Mango II. To address this question, we prepared a semi-randomized library in which 11 positions were randomized based on the crystal structure (Figure 42):

- The three residues (A_{10}, U_{17} and U_{12}) forming the cap covering the fluorogen
- The three residues of the triplex (A_7-U_{19}-A_{25})
- The bottom base pair ($G_{21}oU_{24}$) of P2 because of its direct involvement in T1/triplex connection
- The first base-pair (U_{26}-A_3) underneath the triplex and that interact with the triplex *via* U_{26}
- An additional randomized nucleotide inserted into the triplex area

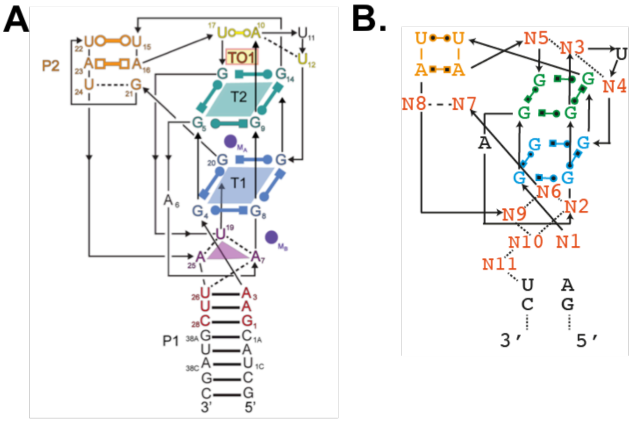


Figure 42: Mango III-derived library design. (A) Mango III structure model (data provided by Robert Trachman). (**B**) Localization of the eleven Mango III positions randomized.

The resulting library was then subjected to the semi-automatized NGS-coupled μIVC screening process I previously developed and validated with Spinach aptamer (see §5.4.1.3).

6.3.2.2. Screening Mango III-derived mutant library

Randomizing 11 positions leads to a theoretical sequence diversity of ~ 4.10^6 variants. Moreover, because every variant is not equally represented in the library, we usually prefer to screen the library using an at least 10-fold coverage, meaning that ~ 10^8 DNA molecules should be analyzed. However, μIVC is currently restricted to the analysis of 10^7 molecules. To solve this paradox, a first round of SELEX was performed by the group of Peter Unrau to decrease the overall diversity to explore, especially by removing sequence unable to bind TO1-Biotin. Upon a single round of SELEX performed using stringent wash conditions, the enriched gene library was posted to us to perform a μIVC screening.

In view of later NGS analysis, a Unique Droplet Identifier (UDI) made of 20 random nucleotides was appended to each gene prior to its encapsulation into PCR droplets, as previously described with theophylline-responsive iSpinach variant (see §5.4.2.2.1.). This pre-enriched barcoded library was named R0.

R0 was subjected to 4 rounds of screening using the same conditions described in **Original article 3**. In particular, the concentration of produced RNA was limited to ~ 300 nM and a concentration of 100 nM of TO1-Biotin was used. The first round was performed without adding further selection pressure and only the most fluorescent droplets (top 0.18%) were recovered (Table 3). When looking at the droplets fluorescence profile, one can see that, even though ~ 64 % of the droplets was occupied by a DNA molecule (calculated assuming a Poisson distribution and a λ value of 1), only a small fraction of the population (~ 0.18 %) was indeed fluorescent (Figure 43). This indicates either that the SELEX was not as efficient as expected or, as already stated before, that a large fraction of the binding RNAs were not fluorogenic. Starting from round 2, TO3-Biotin was introduced in the mixture and gradually increased until the end of the process to select for the affinity. Finally, at round 4, a selection pressure for RNA folding stability was introduced by warming the sorting chip a 45°C.

Table 3 *– Metrics of the µIVC screening used to search for improved Mango III-derived aptamers*

Round	**Temp. (°C)**	**TO3-Biotin (nM)**	**TO1-Biotin (nM)**	**λ value**	**Fusion efficiency (%)**	**Number of analyzed droplets**	**Number of analyzed variants**	**Number of sorted droplets**	**Droplet recovered**
1	25	0	100	1.00	87	4,200,750	3,650,000	7,774	0.18 %
2	25	35	100	0.20	95	1,096,125	208,264	443	0.04 %
3	25	70	100	0.25	90	779,751	175,443	966	0.12 %
4	45	140	100	0.07	98	1,171,778	80,383	789	0.07%

At the end of the process, a single discrete population of positive droplets was observed. The nice gaussian shape of this population was indicative of a highly uniform phenotype, confirming the success of the process (compare profiles Round1 and Round 4-analysis on Figure 43), suggesting that the selection reached its end and that only a few different sequences should be found in R4 library (enriched library obtained after round 4). To characterize the content of each library, the 5 libraries (R0, R1, R2, R3 and R4) were prepared and indexed using Nextera kit (Illumina) prior to being analyzed on a MiSeq NGS device.

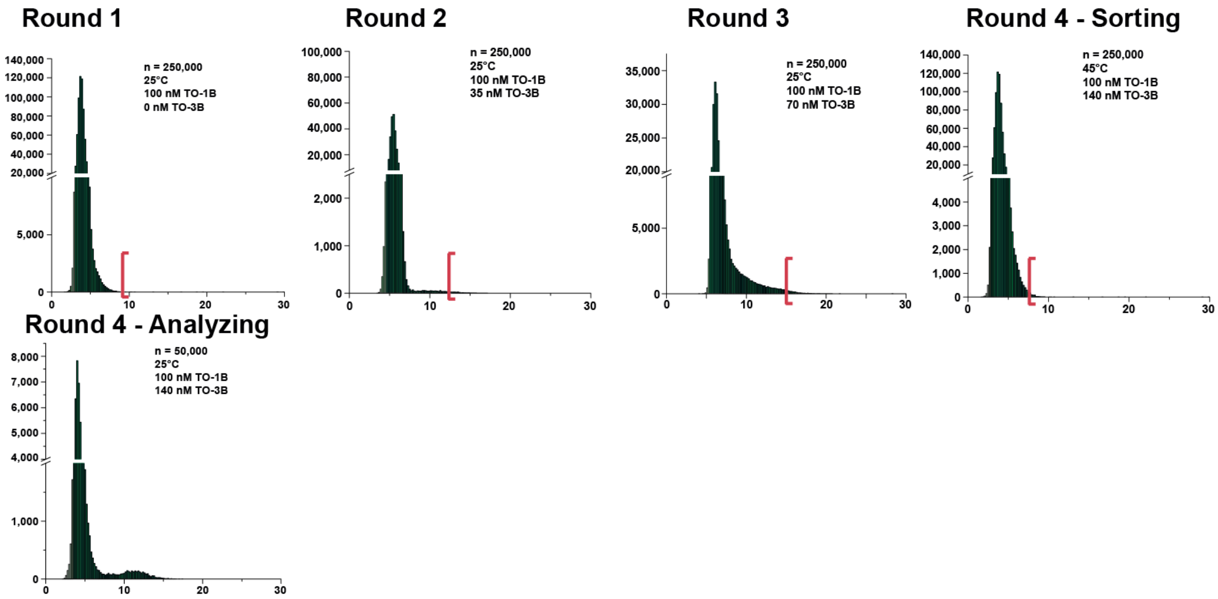


Figure 43: Green fluorescence profiles of μIVC screenings. The green fluorescence of 250,000 droplets was used to build R0-R4 fluorescence profile, and 50,000 for R4-Analyzing. Red bars indicate the threshold over which droplets were gated and sorted. R4-sorting was performed at 45°C.

6.3.2.3. Sequence analysis

A total of ~ 20 million reads were collected from the sequencing run and analyzed using the Python-written analysis pipeline presented hereafter.

6.3.2.3.1. Bioinformatics algorithm

Using Python language, I wrote an analysis script allowing for extracting, QC filtering and analyzing the sequences through a pipeline in 4 main steps (Figure 44):

- 1. Extraction step: sequences were recovered from fastq files. Identical reads (UDI + Mango coding sequence) were clustered and their occurrence determined (later used for noise filtering, see below). Since each UDI is unique in the population the clustering step should reduce the number of different unique UDI/Mango close to the number of sorted droplets (Table 1). However, the obtained values far exceeded the expected ones (e.g. 1,729,282 unique UDI/Mango found in R1, while only 7,774 droplets were sorted). Even though double encapsulation may account for a small fraction of the extra sequences found, the large values obtained definitely point that a significant level of noise is still present in the data and needs to be removed.
- 2. Sequence QC filtering step: UDI/Mango sequences displaying mutations outside the regions expected to be variable (i.e. UDI and Mango randomized positions) were removed (sub-step A and B). Moreover, the UDIs and the 11 studied nucleotides were extracted from the rest of the sequence and used to pursue the analysis.
- 3. UDI QC filtering step: Sequences with UDI of the wrong length were removed. Such UDI could arise from inefficient synthesis step.

• 4. Noise removal step: This was the most crucial step. Each UDI is expected to be unique to each template molecule individualized into droplets at the beginning of the process. Therefore, each droplet should contain only one or a very few different UDIs that will be amplified a large number (hundreds of thousands) of times. Consequently, and provided the sequencing goes deep enough, true UDI sequences coming from droplets should be found in much larger amount than noisy sequences resulting from: contamination by air-borne DNA molecules lying in the lab, recombination events between PCR products, appearance of point mutations… Therefore, a cut-off was applied to eliminate poorly represented UDIs and only the most represented were conserved. The number of different UDIs was strongly reduced and found to barely correspond to the number of sorted droplets (compare values in Table 3 and Droplet number shown in the box of Step 4 on Figure 44). At the end of the analysis only 70 different Mango sequences were conserved in R4 library.

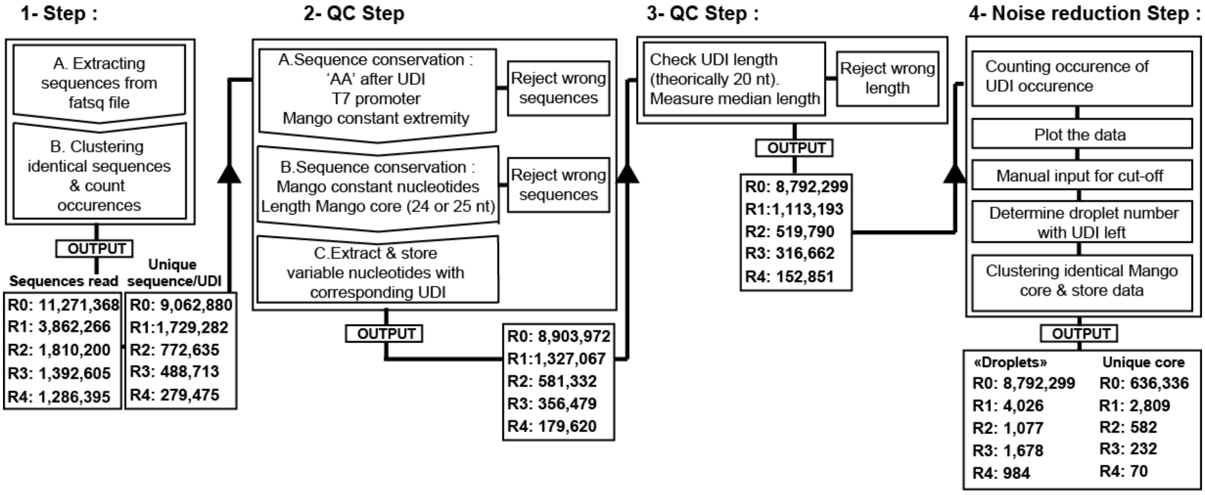

1- Step :
A. Extracting sequences from fatsq file
B. Clustering identical sequences & count occurences
OUTPUT
Sequences read
Unique sequence/UDI
R0: 11,271,368
R1: 3,862,266
R2: 1,810,200
R3: 1,392,605
R4: 1,286,395
R0: 9,062,880
R1:1,729,282
R2: 772,635
R3: 488,713
R4: 279,475
2- QC Step
A.Sequence conservation : 'AA' after UDI T7 promoter Mango constant extremity
Reject wrong sequences
B.Sequence conservation : Mango constant nucleotides Length Mango core (24 or 25 nt)
Reject wrong sequences
C.Extract & store variable nucleotides with corresponding UDI
OUTPUT
R0: 8,903,972
R1:1,327,067
R2: 581,332
R3: 356,479
R4: 179,620
3- QC Step :
Check UDI length (theorically 20 nt). Measure median length
Reject wrong length
OUTPUT
R0: 8,792,299
R1:1,113,193
R2: 519,790
R3: 316,662
R4: 152,851
4- Noise reduction Step :
Counting occurence of UDI occurence
Plot the data
Manual input for cut-off
Determine droplet number with UDI left
Clustering identical Mango core & store data
OUTPUT
«Droplets»
Unique core
R0: 8,792,299
R1: 4,026
R2: 1,077
R3: 1,678
R4: 984
R0: 636,336
R1: 2,809
R2: 582
R3: 232
R4: 70

Figure 44: Python-based analysis pipeline and metrics.

6.3.2.3.2. *Sequence analysis*

The pool of 70 sequences identified in R4 library was largely dominated by a few sequences (Figure 45), consistent with the Gaussian-shaped form of the population observed during this round (Figure 43, Round4-Analysis). In particular, one variant largely dominated the pool by representing more than half of the occurrences (~570/984), while the four next sequences represented between 2 and 13 % of the population. The dominating variant will from now be called Super-Mango III throughout the rest of the document. Moreover, the different variants were really close in sequences. Indeed, all of the 5 most abundant variants possess a U at each of the three positions of the lid (N3, N4 and N5 on Figure 44). Interestingly, one of these Us (U occupying the position N3 on Figure 44.A) was originally an A in Mango III. Also, a GoG pair (N7-N8 on Figure 44) is also found at the bottom of P2 stem in the five variants, whereas it was a GoU pair in Mango III. Beside these two sets of fixed mutations, the rest of the tested residues accepted more variations.

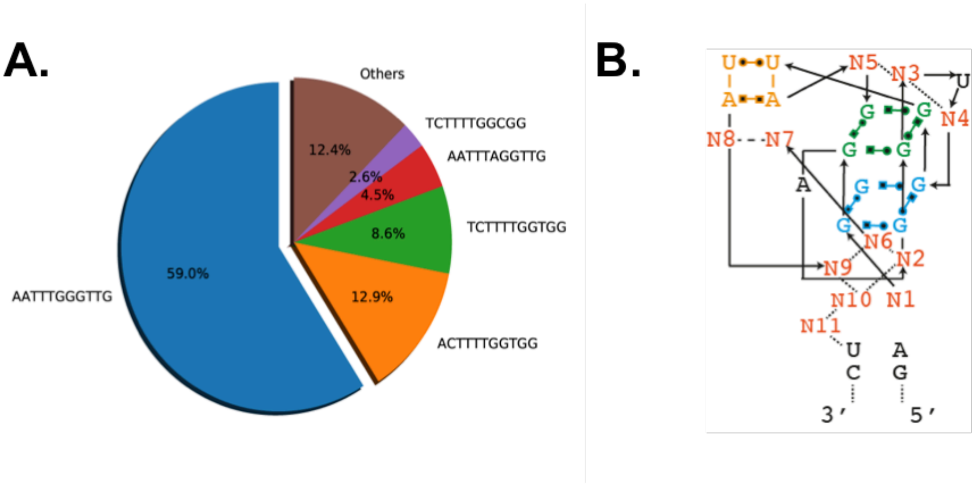


Figure 45: Sequence content of the Round 4 library. (A) The nucleotides shown correspond to the residues found at each position originally randomized in Mango core region. (B) The position of each residue in the structure is shown.

This trend was further confirmed by establishing a sequence correlation matrix (Figure 46) that unambiguously revealed that some residues were more represented than others and that they were always found associated together (red squares on Figure 46). Based on this correlation matrix, a consensus sequence was established (Figure 46) to which the 5 most abundant variants conform.

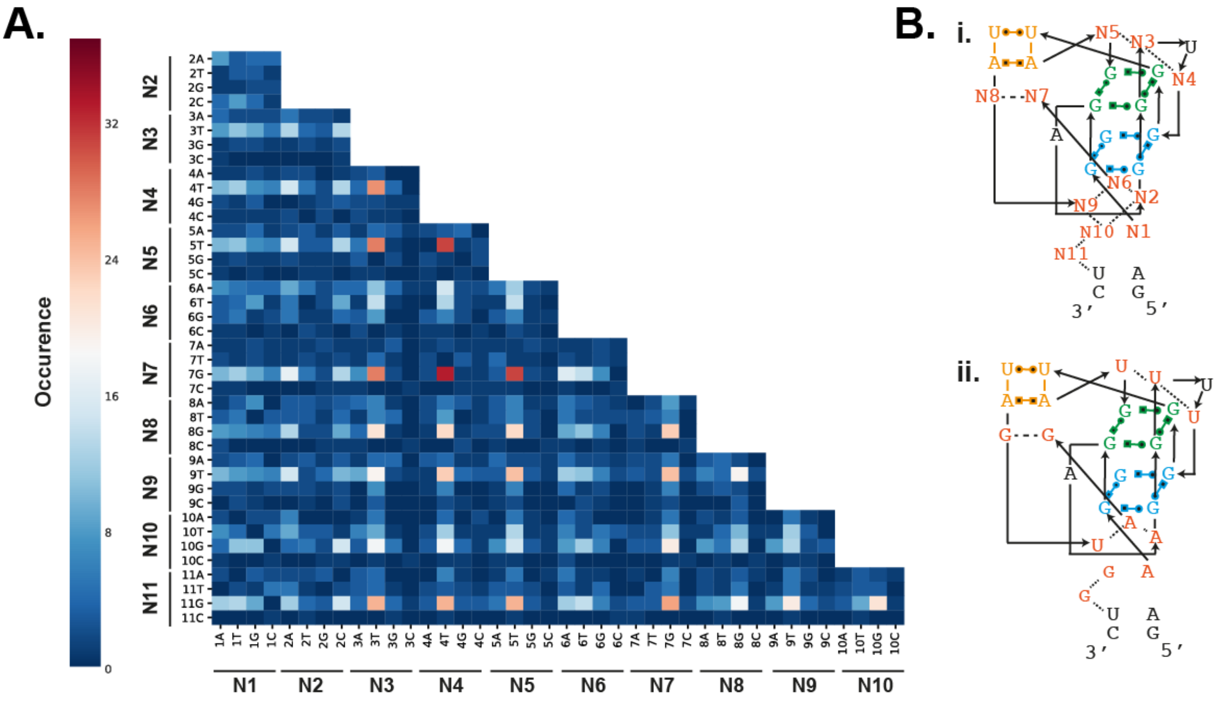


Figure 46: Sequences correlation matrix. (A) The occurrence frequency of each nucleotide was computed at each position and the results are summarized as a heat-map. **(B) i.** Mango III library showing the randomized nucleotides. **ii.** Consensus sequence of most represented R4 variant.

Comparing the sequence content of the different rounds and computing the enrichment rate of each sequence allowed to gain further insight on the selection process. First, I filtered out any sequence appearing in a round while being absent in the previous one (therefore likely to be contaminant introduced during library preparation, keeping in mind that no mutagenesis step was performed during the whole process). This allowed further reducing the diversity of the final pool down to 37 sequences (still containing the 5 most abundant sequences). Next, I computed the enrichment rate of each of these 37 sequences along the screening process with respect to their occurrence in R0 library (Figure 47). This allowed me to confirm that the three most abundant variants of R4 gradually enriched along the screening process (Figure 47). Interestingly, in Round 4, the application of the highest selection pressure (45°C, 140 nM TO3B) led to the depletion of most of the sequences with the exception of the 3 most abundant ones (Figure 47), suggesting that these 3 variants should, in principle, best fit to the criteria fixed for the screening: brightness (gating the most fluorescent droplets), folding efficiency (resistance at 45°C) and selectivity of TO1 over TO3.

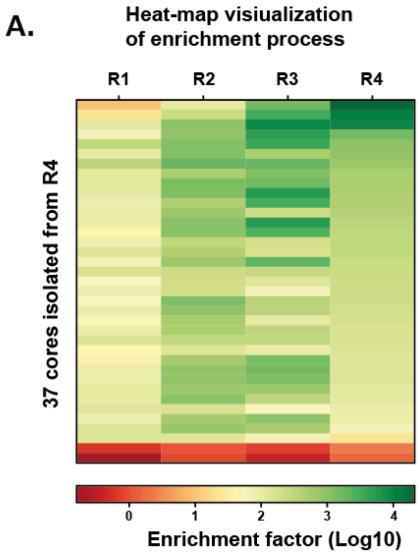


Figure 47: Enrichment monitoring of the 37 sequences validated at the end of the screening process. Heat-map representation of variant enrichment over the screening process.

6.3.2.3.3. Fluorogenic properties of the new Mango III-derived aptamers

I next evaluated the performances of the 3 most abundant variants noted Core A (Super-Mango III), B and C as well as two minor variants called Core-D and E (Figure 48.A). RNAs were synthesized, purified and their fluorogenic properties were assayed. Whereas Core-E (the least represented variant) did not yield any fluorescence (data not shown), the 4 other molecules activated TO1-Biotin fluorescence (Figure 48.B) with the most abundant (Core-A, B and C) being significantly more efficient than the less represented Core-D. This further highlights the predictive capacity of our bioinformatic approach.

Interestingly, an excess of RNA gives a similar level of fluorescence for Mango III Wt, Core-A, B and C, whereas using an excess of dye clearly pointed Core A (Super-Mango III) as having superior properties (Figure 48.B and C). According to data from the literature, such behavior is indicative of better folding molecules and was already described for Spinach2 (Strack et al., 2013) and iSpinach (**Original Article 1**). This therefore suggests that Super-Mango III could fold more efficiently than its parent.

This work is still on-going and deeper characterization of the three most enriched variants (Super-Mango III, Core B and C) is in progress in collaborators' lab. In particular, crystallization experiments of Super-Mango III/TO1-Biotin are in progress in Adrian Ferré D'Amaré's group. Preliminary data shared by the group of Peter Unrau confirmed the positive contribution of $A_{10}U$ mutation I found in the region covering TO1-Biotin. A manuscript involving the three groups is currently in preparation and should be submitted upon the collection of the last results cited before.

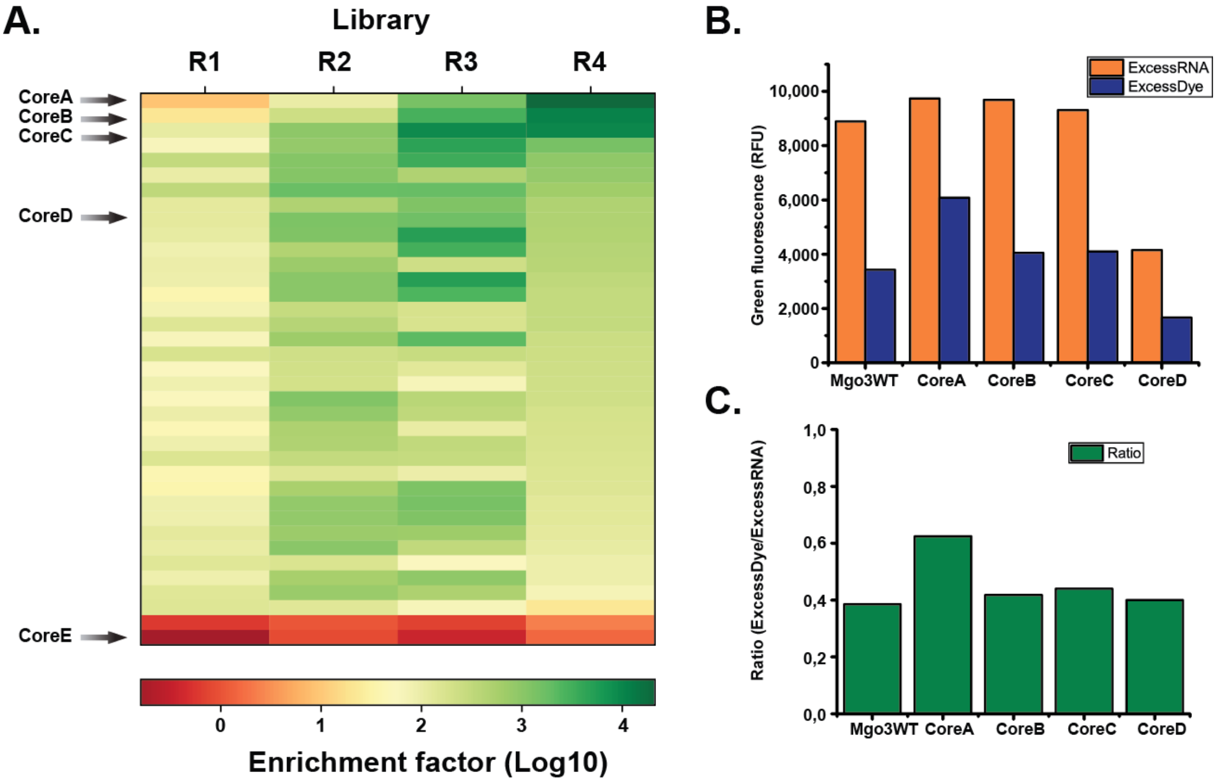


Figure 48: Performances of MangoIII-derived variants. **(A)** Heat-map showing the 37-isolated Mango cores identified in R4 library. The enrichment factor computed relative to R0. The arrows indicate the Mango variant assayed: the 3 most enriched and abundant. **(B)** Gel purified RNAs were pre-incubated for 20 min at room temperature in a buffer (140 mM KCl, 1 mM MgCl2 10 mM NaH2PO4 pH 7.5) prior to measuring green fluorescence at 25 °C. "Excess RNA" condition corresponds to 500nM purified RNA mixed with 50 nM TO-1Biotin, whereas "Excess dye" condition corresponds to 50 nM RNA mixed with 500 nM TO1-Biotin. **(C)** Bar chart representing the ratio of the fluorescence measured in "Excess RNA" and "Excess dye" conditions. This informs on RNA folding efficiency. Since Core E did not emit fluorescence signal, data were not included. Note that Core-A corresponds to Super-Mango III.

6.4. Conclusive remarks on the Mango chapter

The work I performed on re-exploring R12 library allowed the discovery of 3 new Mango variants, all surpassing the performances of the original molecule. The improved fluorogenic capacity and folding efficiency of these variants together with the good photostability of Mango/TO1-Biotin complex makes possible imaging small non-coding RNAs both in live and fixed mammalian cells. The latter scenario makes possible to detect Mango-labelled RNA concomitantly to the immuno-detection of other targets. Moreover, since these new variants have conserved the high affinity of Mango I, they can still be used for purification purposes. Finally, re-exploiting the screening μIVC in tandem with NGS allowed me to identify an even better Mango III-derived variant named Super-Mango III. The great properties of these new light-up aptamers convinced the CNRS to file a patent (PCT/IB2018/052808) application (of which I'm co-inventor) to protect this new generation of molecules.

This new set of applications further validate the technologies (μIVC and NGS-coupled μIVC) I originally developed with the DFHBI/Spinach module and demonstrates the robustness of these procedures. However, R12 was re-explored using an *in vitro* selection strategy in which no mutagenesis step was performed. Therefore, it is not impossible that even better aptamers could be identified by reproducing the experiment while introducing mutagenesis steps. This would give the opportunity to evolve Mango *in vitro* by testing a wider range of mutation combinations. Using μIVC in tandem with SELEX would then allow exploring pretty large libraries in an efficient way while combining the use of NGS analysis would considerably simplify the post-selection work as demonstrated with Super-Mango III isolation.

General discussion, conclusion and outlook

7. Screening technology

8. RNA-based fluorogenic modules: current and future applications

III. General discussion, conclusion and outlook

The goal of this thesis was to adapt and use μIVC for improving and discovering new light-up RNA aptamers. During this work, I first developed a robust screening procedure using the well-characterized DFHBI/Spinach module prior to applying it to the optimization of other fluorogenic systems. Moreover, I started to integrate μIVC with other technological bricks to increase the overall throughput and further automated the process. In this last chapter, I will sum-up and discuss the main achievements of this thesis work before concluding it by giving a rapid overview of the next possible achievements made possible by this work.

7. Screening technology

My thesis work was focused on the use of droplet-based microfluidics, a technology consisting in producing and manipulating droplets by the serial use of dedicated devices. Using this technology allows to quantitatively analyze up to 10 million mutants in a single day using a three-step procedure during which DNA genes are: i) individualized and amplified into small PCR droplets prior to being ii) transcribed into RNA following a droplet fusion step bringing IVT mixture and fluorogen to each PCR droplets and iii) sorted based on droplet fluorescence. Since the beginning of this thesis, alternative technologies have been reported in the literature and need to be discussed prior to showing how my own technology could be further improved, developed and applied.

7.1. Benefits of μIVC over competitive screening technologies

As highlighted in the **Review Article 2**, biological reactions need to be miniaturized far beyond the Microtiter Plate gold standard format if one wants to be able to screen large mutant libraries *in vitro* in a cost and time-effective way.

In this view, late 2015 the group of Meier reported on a Large Scale Integration (LSI, see §3.1) microfluidic device in which clusters of known DNA sequences are spotted onto the chip prior to being transcribed and assayed into small nanoliter wells. This system offers the great advantage of making possible the fluorescence monitoring of each chamber while varying its composition (e.g. fluorogen concentration). This gives access to thermodynamic parameters (e.g. K_D) on a large number of mutants (several hundreds) in a single experiment. However, whereas this technology is extremely powerful for the characterization of an aptamer as well as for screening libraries of low genetic diversity (less than a thousand different sequences), it is not competitive with μIVC for the ultrahigh-throughput screening of larger libraries made of million variants (Ketterer et al., 2015).

Reaching higher throughput requires exchanging solid reaction vessels for soft matter-made ones, so to switch to emulsion-based technologies. Recently, the group of Tom Soh as reported on a technology called Particle Display (Wang et al., 2017) combining the use of IVC, beads and FACS and that is somehow reminiscent to the Compartmentalized Bead Tagging initially reported by the group of Phil

Holliger (Wochner et al., 2011). In these technologies, genes are individualized, amplified and expressed at the surface of the beads. The clonability of each step is ensured by individualizing the beads into water-in-oil droplets using conventional IVC stirring emulsification process. During a first emulsification step, DNA are individualized and amplified onto the beads. Then, the emulsion is broken, beads are recovered and re-emulsified into a second set of droplets where they are *in vitro* transcribed into RNA. Finally, beads displaying RNA molecules of interest are fluorescently labelled prior to being sorted in ultrahigh-throughput regime using a FACS machine. In the case of light-up aptamers selection, beads carrying efficient aptamer become strongly fluorescent making the sorting straightforward (Gotrik et al., 2018). Moreover, the possibility of incubating the beads with a low concentration of fluorogen makes it possible to select aptamers for both their fluorogenicity and their affinity. Using this approach, the group of Tom Soh managed to isolate new variants of Malachite Green-binding aptamer endowed with improved brightness and affinity. Particle Display is therefore competitive with μIVC; both technologies having close throughputs. Yet, relying on stirring-based emulsification, Particle Display also suffers from IVC limitations: strong emulsion polydispersity leading to potential bead-to-bead variations as well as difficulty to control multiple encapsulation events. Moreover, Particle Display requires the RNA to be physically attached to bead surface which may affect RNA folding properties. Finally, last but not least, Particle Display does not allow for selecting fast interacting aptamers. Indeed, significant time (tens of minutes to hours) can separate the beginning of the incubation of beads with the fluorogen and their FACS analysis. On the contrary, μIVC makes possible to apply very short incubation times (milliseconds to minute) using an integrated device in which the reagents are added a few seconds before droplet sorting (see below). Therefore, in comparison with Particle Display, μIVC makes possible to generate highly monodisperse emulsions in which RNA are synthesized in their native form. Moreover, it makes possible to search for rapid binders by using a dedicated integrated chip making μIVC performances superior to that of Particle Display. The only current disadvantage of μIVC is its requirement for high fluorogen concentration (at least 100 nM) because of the short time droplets spends in front of the detector. Nevertheless, using brighter dyes should make possible using lower amount of fluorogen and *de facto* to apply a selection pressure for the affinity. Furthermore, I demonstrated that aptamers could also be selected for their affinity if competitors are included into the mixture (see **Original Article 3**).

Therefore, we can conclude from what precedes that, nowadays, μIVC is likely the most efficient screening technology for isolating light-up aptamers and derivatives. Yet, the current μIVC still suffers limitations that could be overcome in the near future as explained below.

7.2. Improvement of the μIVC procedure

Up to now, μIVC was exclusively applied to the improvement of existing aptamer, making a throughput of 10^7 sequences analyzed per day absolutely acceptable. However, should someone be interested in identifying a brand-new RNA module, then a much larger sequence space (at least 10^{14}

different sequences) must be analyzed, which require adapting the current μIVC procedure. Moreover, using new microfluidic devices should allow further increasing selection accuracy, while integrating NGS analysis in the pipeline will considerably reduce the post-selection work.

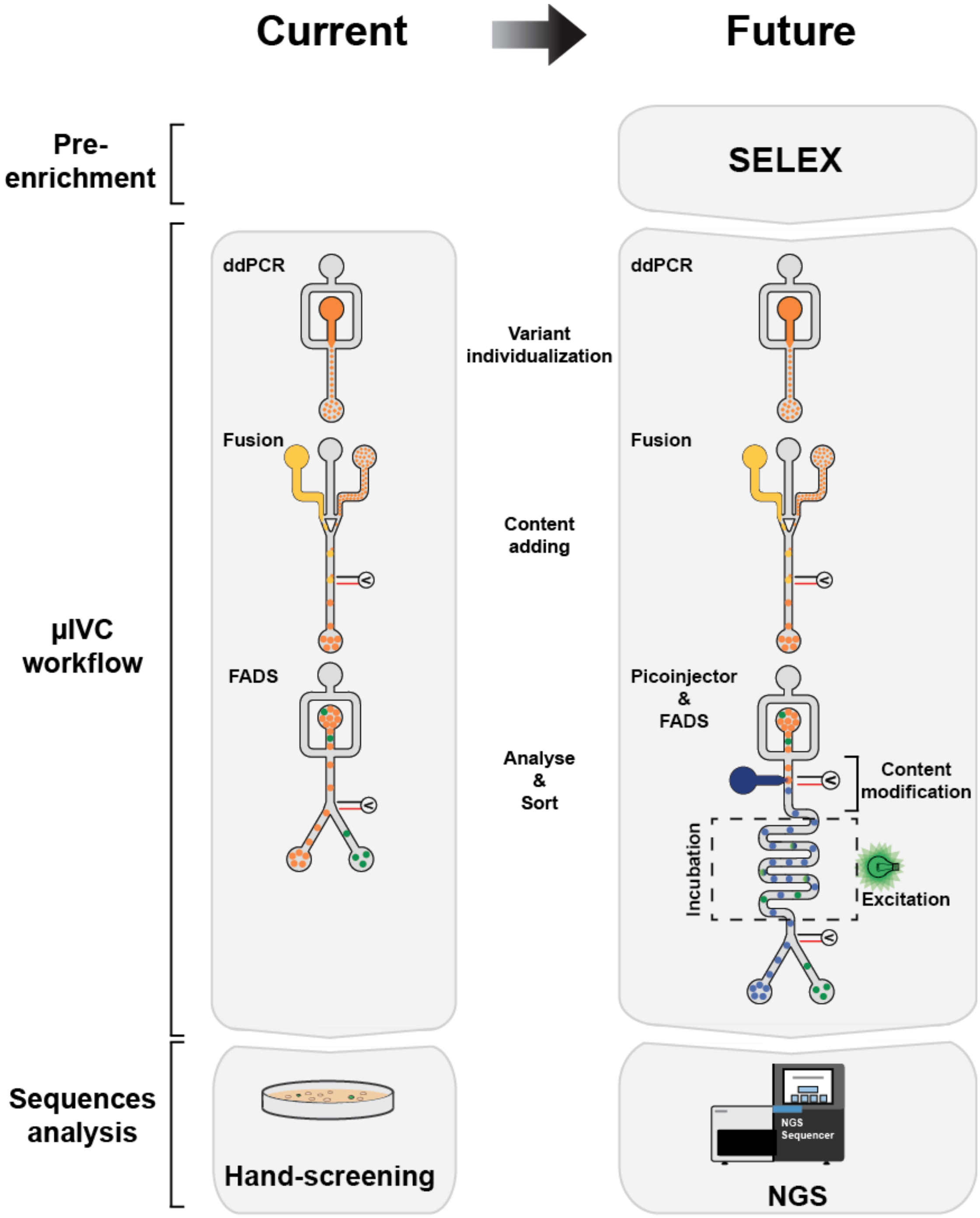

Current
Future
Pre-enrichment
SELEX
µIVC workflow
ddPCR
ddPCR
Variant individualization
Fusion
Fusion
Content adding
FADS
Picoinjector & FADS
Analyse & Sort
Content modification
Incubation
Excitation
Sequences analysis
Hand-screening
NGS Sequencer
NGS

Figure 49: Current and future μIVC workflow. **Left**: first screening process based on µIVC including 3 steps (ddPCR, fusion and FADS) and followed by manual screening which allowed the isolation of iSpinach and Mango. **Right**: future screening process encompassing new and future improvement. This workflow now encompasses a pre-enrichment step (SELEX) to reduce the diversity to one explorable by μIVC. Analysis and sorting integrated chip will allow for modifying droplet content *via* a pico-injection. Then, upon a short incubation time during which the photostability of the complex will be challenged, droplets will be sorted. Finally, a NGS-based analysis will allow identifying the most promising variants in a semi-automated way.

7.2.1.Increasing the screening throughput

It is complicated to handle more than 10^7 sequences with μIVC, due to the overall throughput of key microfluidic devices (droplet-fusion and sorting device) that is limited to 1,000-2,000 events per seconds, as well as by the volume that can be handled (e.g. limited PCR tubes capacity). Nevertheless, it is possible to strongly increase the size of the gene library analyzed by pre-enriching it using one (or several) step(s) of SELEX. Doing so, libraries can be pre-enriched in ligand-binding RNAs and, once the diversity approaches 10^7 different sequences, selections can be pursued using μIVC functional screening. Hence, in our lab we started to set-up SELEX process upstream of μIVC workflow and we already validated this strategy with Super-Mango III by subjecting a starting library of ~5.10^6 variants to a round of SELEX prior to performing μIVC screening (see §6.3.2). However, depending on the type molecule one wants to develop in the end (RNA endowed with structure switching capacity or not), care should be taken to properly choose the SELEX approach (see below).

The efficiency of the current pipeline could also be improved by modifying the sorting device used at the end of the process. Indeed, the FADS device we currently use is limited to the analysis of 300 droplets per second, so 1 million droplets per hour. As a consequence, analyzing/sorting 10^7 variants takes some 10 hours and this step is currently the bottleneck of the screening process. However, an optimized version of the FADS based on gapped divider in the sorting fork that prevents droplet breaking was reported as being able to handle up 30,000 droplets analyzed per second (Sciambi and Abate, 2015), enabling to save a substantial amount of time. Increasing the sort frequency by only a factor of 10, would make possible analysis up to 10^8 molecules in a day.

Therefore, integrating a SELEX step upstream the μIVC screening and modifying the sorting module could be two ways of increasing the overall efficiency of the process (Figure 49).

7.2.2.Using an integrated device to better tailor optimized aptamers

Significant improvement of the technology could also come from the use of an additional integrated device allowing for uncoupling the *in vitro* expression from the functional assay, while making possible to apply new selection pressures.

Indeed, in the current screening workflow, the co-factors (the fluorogen and the molecule to be sensed) are brought together with the *In Vitro Transcription* (IVT) mixture. Therefore, the functional assay is performed at high magnesium concentration (~ 20 mM) and the incubation time is primarily dictated by the transcription reaction (at least 30 minutes). This may have two major drawbacks. First, RNA found to be optimal at 20 mM magnesium may later behave poorly *in vivo* because of the much lower ~ 1 mM free magnesium concentration available in the cell. Moreover, using incubation times as long as 30 minutes precludes the preferential selection of very efficient binders (i.e. those having very low k_{on}). However, both limitations can be overcome by using an integrated device in which: i) droplets would be reinjected upon transcription completion prior to being ii) picoinjected with a solution containing the fluorogen, optionally the molecule to sense and some EDTA. Here, EDTA

would deplete the mixture in free magnesium and lead to the dissociation of complexes requiring elevated magnesium concentrations. iii) Right upon pico-injection, the droplets would circulate into a channel whom the length will impose a short incubation time (from seconds to minutes) before the droplets reach the analysis and sorting area. Applying such a short incubation would allow selecting for very fast binders and could not be performed with screening technologies like Particle Display.
Even though these data were not included into this thesis manuscript, I actually already validated such an integrated chip on a microfluidic point of view. This chip allows for performing a short 1-minute incubation between the pico-injection and the analysis/sort events. This chip was found to be robust over the time without significant fluctuation at the pico-injection level.

Interestingly, getting the droplets circulating into an incubation channels would also allow to apply a selection pressure for the complex photostability. Indeed, one could imagine shining an intense excitation light on the circulating droplets and, this way, photobleach the dyes that would not be efficiently protected by the RNA. Indeed, recent report on DFHO/Corn module revealed that the RNA interacts with the fluorogen in such a way that the aptamer protects the dye from rapid photoinactivation and forms with it a complex of an extraordinarily high photostability (Song et al., 2017; Warner et al., 2017).
Therefore, adding such an integrated chip into the screening pipeline (Figure 49) would drastically improve the process by making possible to apply a whole bunch of relevant selection pressures that would allow to select RNAs with elevated brightness and photostability while needing only low amount of magnesium and being able to quickly interact with their ligand.

7.2.3.Optimizing the post-screening analysis by systemizing NGS analysis

My work on the theophylline/iSpinach sensor (see §5.4.1) and Super Mango III (see §6.3.2) demonstrated that analyzing μIVC-enriched libraries using NGS/bioinformatics allows to precisely predict the sequences that effectively respond to selection pressures and, doing so, to identify the most relevant variants without further experiment to be performed. This allowed to drastically shorten the post-screening work by restricting the hands-on experiments to only those variants the most likely to display the properties we are searching for. The efficiency of the process (wet experiment and bioinformatic analysis) I pioneered encouraged the group to systematically incorporate this analytical strategy into the screening pipeline.

Therefore, in the very near future, combining μIVC with a SELEX pre-enrichment step, the use of the integrated analysis device and NGS post-selection analysis should significantly increase the performances of the μIVC pipeline I set-up and make it highly efficient and competitive.

8. RNA-based fluorogenic modules: current and future applications

Using the μIVC workflow, I developed iSpinach, an improved DFHBI-binding aptamer (**Original Article 1**) as well as four new Mango-related aptamers (Mango II, III, IV and Super Mango III, **Original Article 3**) all endowed with superior fluorescence and folding capacities. Indeed, the crystal structures solved to date unambiguously show that improved aptamers adopt much more stable folding than the original molecule. This increased stability involves new direct (Mango II (Trachman et al., 2018), or water-mediated (iSpinach, **Original Article 2**) interactions stabilizing the basal part of the G-quadruplex motif. Moreover, some fancy new motifs connecting the TO1-biotin-binding site and the basal part of the aptamer appeared in Mango III, presumably to rigidify the overall structure of the aptamer and increase its stability. This overall stabilization is likely the result of some of the applied selection pressures (e.g. exchange of K^+ for Na^+ and warming of the sorting device) and can be expected to have a benefic effect for experiments aiming at embedding the aptamer within another RNA sequence. Finally, the improved brightness of the different variants also found some explanation from the structural studies since, besides the gain in folding efficiency that necessarily contributes to the fluorogenic capacity of the molecule, better accommodation of the dye as observed with Mango II and III variants does also contribute to the overall improvement of the molecule. These improved folding and fluorescence capacities makes iSpinach and Mangos highly amenable to plenty of applications.

8.1. Live-cell imaging applications

As deeply exposed in the introduction (see chapter 1), RNA is at the center of gene expression due to its role of transient carrier of the genetic information (messenger RNA, mRNA) but it is also a key player of the regulatory mechanisms (regulatory RNAs) controlling this expression. Therefore, being able to monitor and track RNAs in living cells with single-cell resolution is of key importance to properly understand the system as a whole.

8.1.1.Imaging messenger RNAs

mRNAs were already successfully imaged in living cells by inserting several dozens of tandem repeats of light-up aptamers in the 3' untranslated region of the target RNA (Zhang et al., 2015). Due to their long size, inserting a long array into mRNA does not interfere with mRNA function. However, it was also noticed that, with the original Spinach, the different aptamers of an array tend to interfere with the folding of each other, leading to a sub-optimal labelling of the target RNA. The low fluorescence efficiency of the original aptamers made also challenge the detection of RNA labelled with too few copies of the aptamer, unless a long and complex post-acquisition image treatment was performed (Guet et al., 2015).

Considering the improved brightness and folding efficient of the new aptamers developed during this thesis, one can expect a significant gain of signal due to both increased brightness and decreased folding interference within the array. However, firmly concluding on the improved performances of the new variants will require a side-by-side comparative study with the gold standard MS2-GFP system (see §2.2.1.2) to be done.

8.1.2.Imaging small non-coding RNAs

The work presented in this thesis allowed imaging abundant small non-coding RNAs (i.e. 5S rRNA and U6 snRNA) labelled with a single copy of a Mango aptamer in living cells (**Original Article 3**). The uniqueness of the tag is here extremely important since, with such small and highly structured RNAs, inserting an array of aptamer would likely disturb proper biological function of the target RNA. Unfortunately, and despite their improved properties, the aptamers developed during this thesis will probably not allow to visualization of less abundant but biologically more relevant small RNAs. This limitation is mainly due to the still too low brightness of these modules together with their limited photostability (fluorescence half-life of 1 and 30 seconds for iSpinach and Mangos respectively). Therefore, proper imaging of RNA-based gene expression regulation will stay out of reach until much more efficient fluorogenic modules will be developed.

8.1.3.Toward the development of a new generation of efficient fluorogenic modules

Ideally, optimal modules should possess elevated brightness, affinity and photostability, while the aptamer should possess high folding efficiency and stability and the fluorogen should readily enter the cell without being toxic. Such module is currently under development in the lab considering the following points.

Ideally, the fluorogen should be built around an organic dye known to be very bright (i.e. displaying a high absorption coefficient and a quantum yield as close as possible to 1) and highly photostable. Moreover, the dye should possess features like flat rings, charged groups as well as H-bond donors/acceptors in order to being able to interact with the RNA *via* staking, ionic or hydrogen interactions. Compounds like sulforhodamine (Sunbul and Jäschke, 2013), Pyronin (Sunbul and Jäschke, 2018)and cyanine like DIR (Tan et al., 2017) have such properties and were shown to efficiently interact with RNA. Unfortunately, they form either complexes of low affinity (sulforhodamine B), they display a strong fluorescence background (Pyronin) or they fail entering the cell (DIR). Moreover, the way the cognate aptamer will be selected will also be of prime importance. Considering the explanations given above, it is likely that the integrated selection scheme depicted above (Figure 49) should make possible isolating RNA allowing to form bright, photostable, high affinity and stable complex with the new generation fluorogen currently under development.

8.2. Other application of the optimized light-up aptamers

Even though they will be of limited utility to image small non-coding RNAs, iSpinach and Mangos will find plenty of other applications by serving, for instance, as building block of mRNA and small fluorogenic sensors or by assisting the development of artificial riboswitches.

8.2.1.Developing optimized trans-acting RNA sensors

Even though the recent advent of Crispr-Cas9 technology significantly eased genome modification (Jiang and Doudna, 2017), the targeted insertion of a sequence coding for an aptamer (or an array of aptamers) is still time-consuming and labor-intensive. Consequently, trans-acting RNA sensors synthesized from a plasmid transiently expressed into the cell represents an interesting alternative. A trans-acting RNA sensor is typically organized the same way a sensor of metabolite is, except that the sensing aptamer is exchanged for a pair of sequences complementary to the target RNA molecule (see Figure 4b in **Review Article 1**). Therefore, in its free state, the loose structure of the sensor prevents the aptamer to interact with its fluorogen. However, in the presence of the target RNA, the annealing of the aptamer-appended sequences to their complementary sequences in the target stabilizes the light-up domain leading to its activation and the emission of fluorescence. As this was the case with the theophylline/iSpinach sensor (see §5.4.1), the communication module is the key element of this sensor. Therefore, applying the same selection principle I developed above (see 5.4.3.1), it should be possible to isolate optimal communication modules in a short period of time and with a minimum of hands-on experiments.

Even more extreme type of sensors made of split aptamers could also be selected using this μIVC workflow (Figure 49). These split molecules are made of two independent strands that reassociate to form a functional light-up aptamer only upon very specific annealing with the target RNA (see Figure 4d in **Review Article 1**). Working with two separated strands is very easy when using μIVC since the proper confinement is maintained from the beginning to the end of the process, whereas doing the same experiment with other screening methods, especially Particle Display where emulsions need to be transiently broken, would be extremely difficult if not impossible.

8.2.2.Developing new biosensors of small molecules

Following the same strategy used to develop theophylline/iSpinach biosensor, any optimal biosensor should now be developed in a fast and efficient way provided and structurally flexible sensing aptamer is available. Indeed, whereas riboswitch-aptamers are naturally endowed with the required flexibility (see §1.2.2.2), this is less obvious for artificial aptamers developed *in vitro* since, in general, they were not selected for that particular function. Therefore, should one want to develop a completely artificial sensor, the procedure depicted on Figure 49, should then adapted and a SELEX-derivative such as Capture-SELEX (see §5.4.1.2) should be used instead of conventional SELEX to maximize the chances of isolating an aptamer with switching properties. This critical importance of properly choosing the right SELEX methodology is well exemplified by the Neomycin aptamer. Indeed, the Neomycin-binding aptamer isolated by conventional SELEX was found to bind its target really tightly whereas it had extremely poor switching, so regulatory, capacity. On the contrary, the N1 Neomycin-binding aptamer isolated by SELEX in tandem with *in vivo* sensing screening (Weigand et al., 2008), or more recently the ciprofloxacin-binding aptamer, (Groher et al., 2018), displayed both very good switching capacity. Finally, upon selection of the proper switching sensing aptamer, its incorporation with the light-up module can be performed in a semi-automated way using the NGS-assisted *μ*IVC pipeline I developed.

8.2.3.Developing synthetic riboswitches

A switching sensing aptamer could not only be inserted into a light-up aptamer, but it may also serve as a building block in the development of synthetic riboswitches. In that case, the sensing aptamer would be integrated with the regulatory platform of another riboswitch and the communication module connecting them would be adapted using the NGS-assisted *μ*IVC pipeline (Figure 49).

The regulatory function of the synthetic riboswitch could then be monitored by placing the sequence of iSpinach or of a Mango downstream the riboswitch. Expressing the construct *in vitro* using purified *E. coli* RNA polymerase holoenzyme would then allow screening libraries for optimal transcriptional riboswitch (see §1.2.2.2.). Moreover, further inserting the region coding for a fluorescent protein between the riboswitch and the light-up aptamer and expressing the construct using coupled *in vitro* transcription/translation system would even make possible searching for optimal synthetic translational riboswitches.

Finally, besides using *de novo* isolated switching RNA modules, artificial riboswitches could also be developed starting with a preexisting riboswitch and reprogramming it by targeted insertion of randomized region (or by random mutagenesis) followed by NGS-assisted *μ*IVC. Being able to develop such artificial riboswitches would have a strong impact in the fields of biosensing and synthetic biology.

To conclude this thesis, the work I performed allowed not only to demonstrate how powerful the *μ*IVC can be for the development of new optimized molecules, but it also allowed me to develop new optimized variants of Spinach and Mango aptamers. In addition, I further pushed the limits of the technology by integrating it with NGS and by starting its integration with SELEX. Altogether, these new molecular, microfluidic and bioinformatic tools should open the way to the development of plenty of new artificial molecules and workflows with potential applications in gene expression imaging, biosensing and synthetic biology.

Annex 1 - Materials and methods

IV. Annex 1 – Materials and methods

Supplementary Material and methods corresponding to unpublished data presented throughout the manuscript are given hereafter.

9. Fluorescence measurements of RNA purified

9.1. iSpinach, fluorescence measurement into different buffers.

Prior to fluorescence measurement, purified RNAs were renatured by warming them 2 min at 85°C prior to cooling them at 25°C for 5 min. Then, 1 volume of renatured RNA (at desired concentration ranging from 1 to 2 μM) was incubated with 1 volume of buffer (described below). The green fluorescence (ex: 492 nm/em: 516 nm) was then monitored for 5 min at desired temperature on a real-time thermocycler (Mx 3005P, Agilent).

i) Potassium or sodium buffers:

One volume of buffer containing 80 mM Tris-HCl pH 7.5, 2 mM $MgCl_2$, 20 µM DFHBI and 200 mM of salt (KCl or NaCl) was incubated with 2 μM RNA.

ii) HL3 Buffer:

One volume of HL3 buffer containing 220 mM sodium, 10 mM potassium, 20 mM NaHCO3, 10 mM HEPES pH 7.5, 60 mM sucrose, 10 mM Threalose, and 20 mM $MgCl_2$ was also incubated with 2μM of RNA.

iii)Theophylline buffer:

For theophylline-response measurements, the buffer contained 80 mM Tris-HCl pH 8.0, 20 mM $MgCl_2$, 10 μM DFHBI-1T and 200 mM KCl, and it was optionally supplemented with 100 μM of theophylline. The mixture was then supplemented with 1 μM of RNA.

The sensing efficiency factor was computed using the equation:

$$Sensing\ Efficiency\ (S.E) = \left({Fluo_{wTheo}} \Big/ {Fluo_{w/oTheo}} \right) * Fluo_{wTheo}$$

Where $Fluo_{wTheo}$ is the maximum fluorescence of the library in the presence of theophylline and $Fluo_{w/oTheo}$ is the maximum fluorescence of the library in the absence of theophylline.

9.2. Mango fluorescence measurements

In the case of Mango variants fluorescence measurement, purified RNAs (50 or 500 nM) were pre-incubated for 20 min at room temperature in a buffer containing 140 mM KCl, 1 mM $MgCl_2$ 10 mM NaH2PO4 pH 7.5 prior to recording green fluorescence (ex: 492 nm/em: 516 nm) at 25 °C on a real-time thermocycler (Mx 3005P, Agilent). Excess RNA represents the condition with 500nM purified RNA mixed with 50 nM TO-1Biotin, and 50 nM RNA for 500 nM TO1-Biotin in case of excess of dye.

10. *In vivo E.coli* imaging

10 ng of construction under T7 promoter and T7 terminator control and inserted into pET28a plasmid were used to transform BL21 (DE3) Star *E. coli* cells later plated on 2YT Kanamycin plates supplemented with 2% of glucose. A colony was then picked and inoculated into a fresh liquid media and placed at 37°C overnight at 160 rpm. A new culture was inoculated the next day and the O.D$^{600\text{ nm}}$ monitored. Once the O.D$^{600\text{ nm}}$ has reached ~ 0.6, cultures were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 2h at 37°C at 160 rpm. Induced cell were pelleted, washed and resuspended in PBS supplemented with 200 μM DFHBI-1T, 2 mM $CaCl_2$, 1 mM $MgCl_2$, 5 MM $MgSO_4$ for 30 min at 37°C. Prior imaging, the cell suspension were washed with PBS and 10 μL were placed between a glass slide and cover a slip. The green fluorescence (ex. 475 nm/em.514 nm ± 24) was imaged on an epifluorescence microscope (TiE, Nikon).

11. Gel-imaging and staining

Prior in-gel imaging following the protocol described by (Filonov et al., 2015), total RNA was extracted. Upon imaging, 1 mL of culture was pelleted and 1 mL of Trizol reagent (Euromedex) was added and vigorously and incubated for 10 min at RT. Then 200 μL of chloroform (Sigma-Aldrich) were added and incubated 10 min at RT prior to being centrifuged at 4°C and retrieving total RNA. Recovered RNA was ethanol precipitated in the presence of 300 mM sodium acetate pH 5.5 (Sigma-Aldrich). After centrifugation and a wash in 70% ethanol, the pellets were dissolved into 50 μL of ddH_2O.

12. Barcoding and μIVC selections for allosteric iSpinach and Mango III

12.1. Barcoding

Barcoding was performed before each μIVC round. The barcodes were added and reinitialized using PCR mixture containing 0.2 μM of forward barcode and reverse primers (Rev-MangoIII or Rev-iSpi), 0.2 mM of each dNTP, and Q5 DNA polymerase (NEB) and the supplied buffer (proprietary to NEB) at the recommended concentrations. The mixture was thermocycled starting with an initial step of denaturation of 1 min at 98°C followed by 30 cycles of: 10 sec at 98°C, 30 sec at 55°C and 30 sec at 72°C, and PCR products were purified by wizard kit.

12.2. Digital PCR

iSpinach and re-selection of Mango III followed the same μIVC workflow described before (see **Original Articles 1 and 3**) and subdivided into 3 chips (digital PCR, Fusion, and sorting). Digital PCR were identical for both selections except for primers, and the thermo cycling program.

i) Digital PCR:

Barcoded DNA libraries were diluted into a 200 μg/mL yeast total RNA solution (Ambion) as described elsewhere (Autour et al., 2016). One microliter of this solution was introduced in 100 μL of a PCR mixture containing 0.2 μM of forward primer (Fwd-Ampli-Barcode) and reverse primer (Rev-MangoIII or Rev-iSpi), 0.2 mM of each dNTP, 0.67 mg/mL Dextran-Texas Red 70 kDa (Molecular Probes), 0.1% Pluronic F68, 2 μL Q5 DNA polymerase (NEB) and the supplied buffer (proprietary to NEB) at the recommended concentrations. The mixture was loaded into a length of PolyTetraFluoroEthylene (PTFE) tubing.

The PCR mixture was infused into the PCR droplet generator microfluidic chip and dispersed into droplets as previously described (**Original Articles 1 and 3**).

Concerning theophylline-responsive iSpinach sensor selection, PCR emulsion was collected into 200 μL tubes and subjected to an initial denaturation step of 30 sec at 95 °C followed by 30 PCR cycles of: 5 s at 95 °C, 30 s at 60 °C.

In the case of Mango III, emulsion were thermocycled with a initial step of denaturation of 1 min at 98°C followed by 30 cycles of: 10 sec at 98°C, 30 sec at 55°C and 30 sec at 72°C

12.3. Fusion and sorting for theophylline-responsive iSpinach sensor selection

ii) Droplet fusion:

PCR droplets were reinjected into a fusion as described before (Autour et al., 2016). Each PCR droplet was synchronized with a 16 pL IVT droplet containing 2.2 mM of each NTP, 24 mM $MgCl_2$, 44 mM Tris-HCl pH 8.0, 50 mM KCl, 5 mM DTT, 1 mM Spermidine, 35 μg/mL of Dextran-Texas Red 70 kDa (Molecular Probes), 0.1% Pluronic F68, 3500 U T7 RNA polymerase (purified in the laboratory) 20 μM DFHBI-1T, 5 ng/μL inorganic pyrophosphatase (Roche) supplemented optionally with 1 mM of theophylline.

iii) Droplet analysis and sorting:

The emulsion was finally re-injected into an analysis and sorting microfluidic device mounted on a Thermo plate (Tokai Hit) holding the temperature at 37 °C. Droplets were re-injected at a frequency of ~200 droplets per second and spaced with a stream of surfactant-free HFE 7500 fluorinated oil.

For positive rounds of screening, the green fluorescence (DFHBI-1T in complex with iSpinach upon theophylline recognition) of each droplet was analyzed and the 1 to 10% brightest droplets were gated for positive round of selection. Concerning the negative rounds, the dimmest green droplets were gated (~25% total of droplets) in the absence of theophylline. The gated droplets were deflected into a collecting channel by applying a 1 ms AC field (1200 V, 30 kHz) and were collected into a 1.5 mL tube. Collected droplets were recovered by flushing 200 μL of surfactant-free HFE 7500 fluorinated oil (3 M) through the tubing. 1H, 1H, 2H, 2H-perfluoro-1-octanol of 100 μL (Sigma-Aldrich) and 200 μL of 200 μg/mL yeast total RNA solution (Ambion) were then added, the droplets broken by vortexing the mixture and DNA-containing aqueous phase was recovered.

12.4. Fusion and sorting for Super Mango III selection

ii) Droplet fusion:

The protocol was similar as described for TO-3 selection in **Original Article 3**.

iii) Droplet analysis and sorting:

The fused and incubated emulsion was re-injected into an analysis and sorting microfluidic device mounted on a Thermo plate (Tokai Hit) holding the desired temperature as previously described (see **Original Article 3**). Then, the green fluorescence of the droplets was analyzed and the 1% most fluorescent droplets were gated at each round of selections.

12.5. NGS analysis

Prior to NGS sequencing, DNA pool of each round was subjected to two PCRs. First, libraries were introduced into a PCR mixture containing 0.2 µM of forward (Ampli-Barcode) and reverse primers (Rev-Tag-MangoIII or R-Tag-iSpiTA8nTheo), 0.2 mM of each dNTP, and Q5 DNA polymerase (NEB) and the supplied buffer (proprietary to NEB) and thermocycled using an initial step of 1 min at 98°C followed by 30 cycles of: 10 sec at 98°C, 30 sec at 55°C and 30 sec at 72°C. Tagged DNA were then indexed through a second PCR following Illumina recommendations. Then, indexed libraries were gel purified on 1% agarose gel-TBE 1x and DNA recovered using NEB monarch® kit.

Upon indexing, and MiSeq Illumina sequencing, sequences were analysed using a Python-written bioinformatic pipeline and described in the manuscript (see §5.4.2.3 and §6.3.2.3.1 for iSpinach and Mango respectively).

13. Enrichment measurement by IVT monitoring

Prior to IVT monitoring-based enrichment measurement, 1 µL of the enriched pool was amplified into 100µL of PCR mixture containing 0.1 µM of each primer (Fwd-iSpi and Rev-iSpi), 0.2 mM of each dNTP, 0.05 U/µL of Q5 and its corresponding buffer (NEB). The PCR products were purified using Wizard kit (Promega) and quantified with a Nanodrop (Thermo Scientific). 1.5 pmol of purified PCR were introduced into an *in vitro* transcription mixture (IVT) containing 40 mM Tris-HCl pH 8.0, 22 mM MgCl2, 5 mM DTT, 1 mM Spermidine, 1.6 mM of each NTP, 20 µM DFHBI-1T, 50 mM KCl, 5 ng/µL inorganic pyrophosphatase (Roche), 1400 U T7 RNA polymerase (purified in the laboratory) and optionally supplemented with 1 mM of theophylline.

The green fluorescence (ex: 492 nm/em: 516 nm) was then monitored every minute for 2h30 min at 37°C on a real-time thermocycler (Mx 3005P, Agilent)

14. Sequences of the primers used

Table 4 : Primers used for PCR and barcoding.

Primer	Sequence (5’-3’)
Fwd-iSpi	GCTAATACGACTCACTATAGGAAGACGTAGCAAGGCGACTACGGTGAGGGTCGGG
Rev-iSpi	TGTGTTCTCTGTCATCTCTCTGGCGACTACGGAGCCCACACTC
Fwd-Barcode-iSpiTA8nTheo	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAANNNNNNNNNNNNNNNNNNNNAATATCTAATACGACTCACTATAGGAAGACGTAGCAAG
Fwd-Ampli-Barcode	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
R-Tag-iSpiTA8nTheo	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGTGTTCTCTGTCATCTCTCTG
Fwd-MangoIII	TAATACGACTCACTATAGGAAGAATGA
Fwd-MgoIII-barcode	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAANNNNNNNNNNNNNNNNNNNNAACAGCTAATACGACTCACTATAGGAAGAATG
Rev-MangoIII	CAAACATGGGAAGAATGA
Rev-Tag-MangoIII	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAAACATGGGAAGAATGA

Annex 2 – French summary

V. Annex 2 – French Summary

Introduction

Les ARN (Acide RiboNucléique) remplissent de nombreuses fonctions clés dans le vivant. Ils peuvent être support de l'information génétique, régulateurs de celle-ci, ou même posséder des activités catalytiques. Visualiser ces molécules au sein d'une cellule permet d'apprécier leur dynamique d'expression et leur distribution au cours du temps, apportant ainsi des informations fondamentales notamment sur la régulation de l'expression des gènes.

En 2011, l'équipe de Samy Jaffrey a synthétisé et décrit le composé fluorogène 3,5-difluoro-4-hydroxybenzylideneimidazolinone (DFHBI) mimant le 4-hydroxybenzlidene imidazolinone (HBI), le fluorophore naturel de la GFP (Chudakov *et al.* 2010). Le DFHBI est peu fluorescent à l'état libre et il diffuse facilement à travers les membranes biologiques sans toutefois être toxique. Un aptamère (Spinach) reconnaissant spécifiquement le DFHBI a ensuite été isolé par sélection *in vitro* (Paige *et al.*, 2011). La reconnaissance entre les deux molécules conduit à la formation d'un complexe possédant les mêmes propriétés spectrales que l'eGFP (ex: 490 nm, em: 520 nm). D'autre part, à l'instar de la GFP, cet outil est universel et fonctionne aussi bien chez les procaryotes que chez les eucaryotes (Strack and Jaffrey, 2013). Cette méthode permet de visualiser les ARN aussi bien *in vitro* qu'*in vivo* tout comme les techniques conventionnelles telles que le MS2-GFP (Bertrand et al. 1998) ou encore le FISH (Fluorescent *In Situ* Hybridation). Outre l'observation directe de l'ARN, cet aptamère a également a servi de point de départ à la conception de bio-senseurs fluorogènes de protéines et de métabolites; comme par exemple la S-AdénosylMethionine (Paige *et al.,* 2012).

Cependant Spinach a été initialement isolé par SELEX (Systematic Evolution of Ligands by Exponential Enrichment), une méthode qui, comme son nom l'indique, est dédiée à l'identification d'ARN pour leur capacité à reconnaître un ligand (Stoltenburg *et al.* 2007) et non pour leur capacité à générer des complexes fluorescents. La conséquence directe a été que, bien que l'ensemble des variants isolés était capable de lier le DFHBI, moins de 1% des ARN sélectionnés conduisaient à la formation de complexes fluorescents (Filonov *et al.* 2014). De plus, une autre étude (Strack et al. 2013) a révélé que la fluorescence du complexe DFHBI/Spinach décroissait rapidement avec la température et était réduite de près de 50% à 37°C. Cette limite est d'autant plus importante, qu'il s'agit de la température de développement de la plupart des organismes modèles de laboratoire. Ainsi, les variants alors connus de Spinach étaient vraisemblablement sous-optimaux et des améliorations du système étaient encore possibles.

Une seconde limite importante du système Spinach/DFHBI est liée à la faible photostabilité du complexe. En effet, le DFHBI est sujet à une photosiomérisation rapide (Han et al. 2013) conduisant à l'extinction rapide du complexe compromettant un suivi prolongé du devenir d'un ARN sous illumination constante. Ainsi, l'exploration d'autres systèmes a été envisagée au cours de cette thèse afin de développer de nouveaux systèmes d'aptamères fluorogènes plus performants.

Stratégie expérimentale

Afin d'isoler un variant amélioré s'affranchissant de la plupart des limites des formes actuelles de Spinach, nous avons choisi une stratégie de développement par criblage fonctionnel basée sur la compartimentation *in vitro* (CIV). Appliquée à notre problématique, cette méthode permet de sélectionner les molécules non pas pour leur capacité à interagir avec le DFHBI comme en SELEX, mais pour leur capacité à former un complexe fluorescent avec le DFHBI à une température définie. La CIV consiste en l'encapsulation de réactions biologiques dans des microgouttes d'eau dans l'huile (Tawfik and Griffiths, 1998). Pour augmenter le débit d'analyse et contrôler la taille des microgouttes tout en pouvant en modifier à volonté le contenu après leur formation, nous utilisons la μCIV (CIV assistée par microfluidique) qui repose sur la production et la manipulation de microgouttes en systèmes microfluidiques. Pour cela, une phase aqueuse est poussée au travers d'une phase d'huile et se retrouve dispersée en microgouttes de quelques picolitres et de tailles identiques ; l'émulsion étant alors dite monodisperse (Anna *et al.,* 2003). Cette production de gouttes à haut-débit (plusieurs milliers de gouttes par seconde) se fait dans une huile fluorée (limite les échanges entre gouttes) contenant un tensio-actif qui stabilise les gouttes (Holtze *et al*. 2008). Une fois les gouttes formées, il est possible d'en modifier le contenu en les fusionnant avec d'autres gouttes (Mazutis *et al*. 2009) ou en les pico-injectant avec une autre phase aqueuse (Abate *et al*. 2010). Après incubation, il est également possible d'analyser la fluorescence des gouttes et de les trier en fonction de leur profil de fluorescence (Baret *et al*. 2009). Cette grande flexibilité de manipulation des gouttes associée au parfait contrôle de leur taille et de leur contenu fait de la µCIV un outil de choix pour entreprendre l'identification de variants améliorés d'aptamères fluorogènes tels que Spinach par criblage à très haut débit.

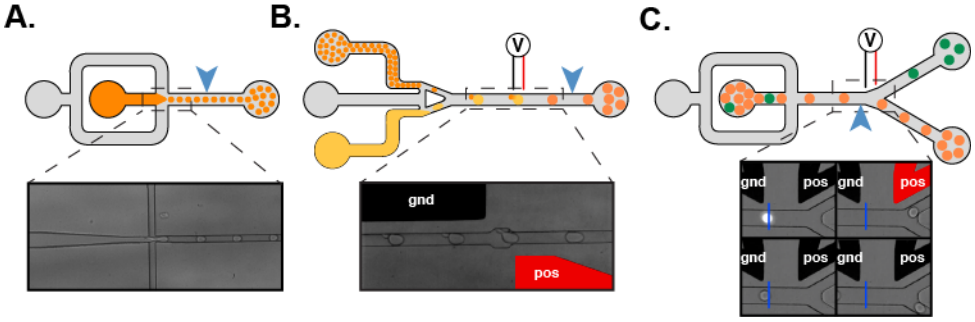


Figure 1 : Représentation schématique des modules microfluidiques. (A) Le générateur de gouttelettes pousse la phase aqueuse de PCR (en orange) vers l'huile fluoronée (en gris) pour former des gouttelettes. La flèche bleue indique un point de mesure pour surveiller la monodispersité de l'émulsion par fluorescence. **(B)** Représentation du module de fusion. L'émulsion de PCR est réinjectée pendant qu'une deuxième émulsion est générée dans la puce (en jaune). Les deux émulsions sont synchronisées et fusionnées entre deux électrodes. **(C)** Schéma de la puce de tri des gouttelettes par leurs fluorescences (FADS). La fluorescence de chaque gouttelette est mesurée au niveau de la flèche bleue. Lorsque la gouttelette d'intérêt est détectée, un champ électrique est appliqué et extrait la gouttelette dans le canal de tri, sinon toutes les gouttelettes sont dirigées vers un autre canal.

Amélioration de l'aptamère Spinach par évolution *in vitro*

L'évolution *in vitro* consiste en la réalisation de tours de mutagénèse en tandem avec des tours de sélection (ou criblage dans notre cas). L'utilisation d'un système de sélection performant combinée à l'application de pressions de sélection appropriées permet de sélectionner les molécules et de les faire évoluer vers des variants ayant les propriétés souhaitées. Ainsi, des banques de mutants du gène Spinach ont été générées par PCR mutagène et criblées par μCIV. Lors des criblages, diverses pressions de sélection ont été appliquées : tout d'abord, seules les molécules les plus fluorescentes ont été sélectionnées. D'autre part, des pressions de sélections non compatibles avec des cellules vivantes (environnement sans potassium, chauffage des puces à 45°C) ont pu être appliquées. Une de ces pressions de sélection (utilisation de sodium en lieu et place de potassium) s'est relevée particulièrement judicieuse car Spinach présente un repliement particulier autour d'un G-quadruplex (Warner *et al.*, 2015). Cette structure G-quadruplex comporte trois plateaux dont deux composés de 4 guanines. Ces structures sont stabilisées par l'intercalation d'un ion potassium s'intercale entre chaque paire de plateaux. Le remplacement des ions potassium par des ions sodium plus petits, réduit la stabilité globale de la structure secondaire G-quadruplex, ce qui impose à la molécule en cours d'évolution d'adapter sa structure notamment en en augmentant l'efficacité de repliement. Grâce à ce principe, j'ai réussi à d'isoler des variants avec une structure beaucoup plus stable et près de 10 fois plus fluorescents que la molécule parentale après seulement 4 tours de criblage dans un environnement contenant du sodium. J'ai ensuite combiné les mutations bénéfiques et éliminé les régions non essentielles par ingénierie moléculaire pour générés un variants minimal (iSpinach) présentant des propriétés nettement améliorées. En effet, iSpinach a gagné près de 12°C en thermostabilité et est respectivement 4 à 7 fois plus fluorescent que la molécule parentale en présence de sodium et potassium. L'originalité de l'approche expérimentale a permis à ces travaux d'être publiés dans le journal Nucleic Acids Research en 2016.

De plus, dans le but de mieux comprendre l'origine de l'amélioration d'iSpinach, j'ai participé en association avec Pablo Fernandez-Millan à l'établissement de la structure cristallographique du complexe Spinach/DFHBI en présence de potassium. Cette étude nous a permis de voir qu'une paire AoA située à la base du G-quadruplex dans la molécule parentale était remplacée par une paire GoG capable d'interagir fortement par de nombreuses liaisons et impliquant notamment des molécules d'eau. Ces travaux ont été publiés dans le journal RNA en 2017.

Criblage de banques de l'aptamère Mango

Suite au développement d'iSpinach, je me suis intéressé à la recherche de nouveaux variants de l'ARN fluorogène Mango au travers d'une en collaboration avec l'équipe du Prof. Peter Unrau (Simon Fraser University, Vancouver – Canada). De par la forte affinité entre l'ARN et son co-facteur (100 fois supérieure à celle de Spinach/DFHBI) et la bonne photostabilité du complexe, le système Mango/TO-1 représente une alternative attractive au couple Spinach/DFHBI. Cependant, le couple Mango/TO-I original souffrait d'une brillance modérée. J'ai donc appliqué la même approche de criblage utilisée avec Spinach à une banque préalablement enrichie par SELEX. Toutefois, afin de préserver la haute affinité de Mango pour son co-facteur (TO-1), j'ai également introduit dans le milieu de sélection des concentrations croissantes en compétiteur du TO-1 (soit le NMM, soit le TO-3 tous deux fluoresçant dans le rouge) exerçant une pression de sélection pour le maintien de l'affinité. J'ai ainsi réalisé un grand nombre de cycles de criblage au terme desquels j'ai été en mesure d'isoler 3 nouveaux ARN Mango (dénommés II, III et IV) présentant une affinité encore meilleure pour le TO-1 (Mango-II) et plus fluorescents que la GFP (Mango III et IV). Ces nouveaux variants ont ensuite été caractérisés sur un plan structural et fonctionnel, et se sont notamment révélés très performants pour l'imagerie de petits ARN non codants dans des cellules mammifères fixées ou vivantes. Comme cela avait été le cas pour iSpinach l'originalité de l'approche expérimentale et la qualité des résultats a permis à ces travaux d'être publiés dans le journal Nature Communications en 2018 et ont conduit le CNRS à procéder à un dépôt de brevet mondial qui vient d'atteindre son extension PCT (*PCT/IB2018/052808*).

La caractérisation structurale de certains de ces nouveaux variants par l'équipe d'Adrian Ferré D'Amare, en particulier celle de Mango III, a conduit à l'identification de positions sous-optimale en terme structural basé sur la structure cristallographique de Mango-III en complexe avec le TO-1. A partir de ces données, j'ai réalisé des nouveaux criblages de ces résidus en utilisant l'approche de criblage μCIV en tandem avec du séquençage à haut débit. Ainsi j'ai pu isoler un nouveau variant (iMango-III) aux propriétés surpassant encore celle de Mango-III. Les données issues de ce travail font l'objet d'un manuscrit en cours de préparation.

Bio-senseurs

Les ARN fluorogènes peuvent non seulement servir pour l'imagerie directe d'ARN mais ils peuvent également être ingénieurés en biosenseurs. Ainsi, iSpinach a servi de base pour mettre au point une sélection allostérique senseur de la théophylline en μCIV. Ces ARN se composent de trois parties. La première est l'aptamère reconnaissant le fluorophore. La seconde est un transducteur, cette partie assure que le signal fluorescent est uniquement présent en présence du métabolite, et enfin la troisième partie est l'aptamère reconnaissant le cofacteur voulu. Habituellement, la partie transducteur est en général obtenue par conception rationnelle et est composé de 3 à 5 paires de nucléotides choisis de façon plus ou moins empirique. Ma technologie me permettant d'analyser de façon quantitative un très grand nombre de séquences, j'ai donc décidé de l'appliquer au criblage de banques de biosenseurs dans le module de communication a été totalement randomisé. En utilisant à nouveau le criblage μIVC en tandem avec le séquençage haut débit, j'ai analysé de manière exhaustive les quelques 65 000 combinaisons de séquence possible de ce module et isoler les quelques modules les plus performants ; c'est à dire permettant un allumage total de la sonde en présence du métabolite cible (la théophylline) et son extinction complète en absence du composé. Ces travaux font l'objet d'un manuscrit en cours de rédaction.

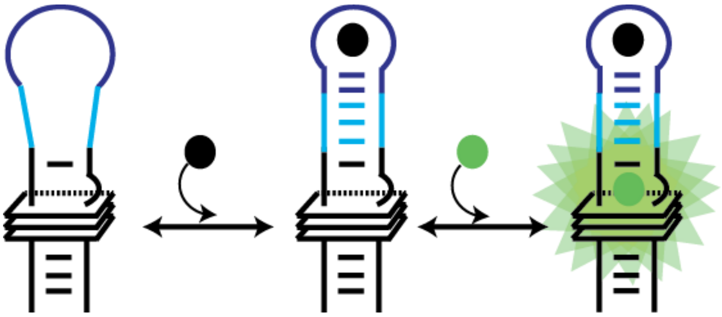


Figure 2 : Représentation schématique d'un senseur d'ARN fluorogène. La partie fluorogénique (iSpinach en noir) est couplée à un module de communication (en bleu clair) ainsi que à un aptamère de détection (en bleu foncé). Lorsque le co-facteur (point noir) est ajouté, les modules de communication sont stabilisés de manière à reconstituer l'aptamère fluorogène pour déclencher la fluorescence.

Conclusions

La μCIV a démontré son efficacité et se présente aujourd'hui comme une approche complémentaire au SELEX. Mes travaux ont permis l'adaptation de cette technologie à l'amélioration d'aptamères fluorogènes, ce qui a conduit à la mise en place de nombreuse collaboration. Utilisée en tandem avec le séquençage haut débit, cette technologie permet à présent l'optimisation très rapide et semi-automatisée à la fois d'aptamères mais aussi de biosenseurs fluorogènes. Ces nouveaux outils permettent à présent d'entrevoir des applications étendues à beaucoup d'autres domaines tels que par exemple la biologie de synthèse.

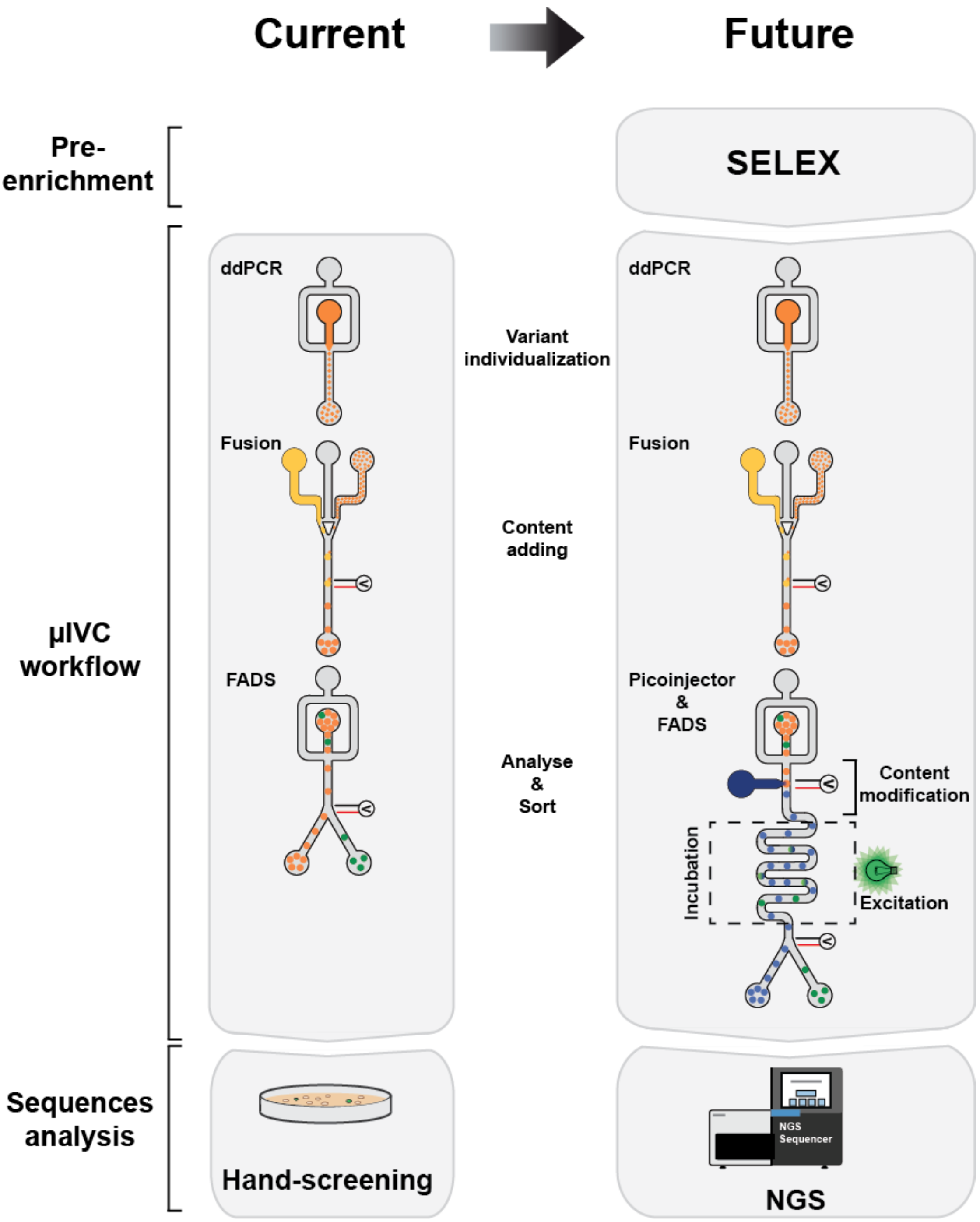

Current
Future
Pre-enrichment
SELEX
µIVC workflow
ddPCR
ddPCR
Variant individualization
Fusion
Fusion
Content adding
FADS
Picoinjector & FADS
Analyse & Sort
Content modification
Incubation
Excitation
Sequences analysis
Hand-screening
NGS Sequencer
NGS

Figure 3 : Figure conclusive des travaux de thèses. A gauche le premier processus de criblage se basant sur la µCIV comprenant 3 puces (PCR digitale, Fusion et de Tri) et suivi d'un criblage manuel qui a permis l'isolation d'iSpinach et des Mango. A droite, l'amélioration général du processus englobant les nouvelles et les futures. Celui ci a été supplémenté en amont par une étape de pré-enrichissement (SELEX) permettant de réduire la diversité à explorer pour la partie μCIV. Cette partie pourrait être amélioré par l'utilisation d'une puce intégrée. Les gouttelettes seraient modifiées par l'intermédiaire d'un pico-injecteur. Suite à l'ajout de contenu, les gouttelettes seraient incubées dans une ligne de délai pouvant être optionnellement éclairé pour sélectionner des aptamères plus photo-stable. Enfin pour augmenter le nombre de mutants analysés et semi-automatisé le processus, les séquences sont envoyé sur un NGS.

VI. References

A

Abate, A.R., Hung, T., Mary, P., Agresti, J.J., and Weitz, D.A. (2010). High-throughput injection with microfluidics using picoinjectors. Proc. Natl. Acad. Sci. U. S. A. *107*, 19163–19166.

Abatemarco, J., Sarhan, M.F., Wagner, J.M., Lin, J.L., Liu, L., Hassouneh, W., Yuan, S.F., Alper, H.S., and Abate, A.R. (2017). RNA-aptamers-in-droplets (RAPID) high-throughput screening for secretory phenotypes. Nat. Commun. *8*.

Ageely, E.A., Kartje, Z.J., Rohilla, K.J., Barkau, C.L., and Gagnon, K.T. (2016). Quadruplex-Flanking Stem Structures Modulate the Stability and Metal Ion Preferences of RNA Mimics of GFP. ACS Chem. Biol. *11*, 2398–2406.

Aharoni, A., Amitai, G., Bernath, K., Magdassi, S., and Tawfik, D.S. (2005). High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. Chem. Biol. *12*, 1281–1289.

Akter, F., and Yokobayashi, Y. (2015). RNA Signal Amplifier Circuit with Integrated Fluorescence Output. ACS Synth. Biol. *4*, 655–658.

Alam, K.K., Tawiah, K.D., Lichte, M.F., Porciani, D., and Burke, D.H. (2017). A Fluorescent Split Aptamer for Visualizing RNA-RNA Assembly in Vivo. ACS Synth. Biol. *6*, 1710–1721.

André, G., Even, S., Putzer, H., Burguière, P., Croux, C., Danchin, A., Martin-Verstraete, I., and Soutourina, O. (2008). S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of Clostridium acetobutylicum. Nucleic Acids Res. *36*, 5955–5969.

Ausländer, S., Fuchs, D., Hürlemann, S., Ausländer, D., and Fussenegger, M. (2016). Engineering a ribozyme cleavage-induced split fluorescent aptamer complementation assay. Nucleic Acids Res. *44*.

Autour, A., Westhof, E., and Ryckelynck, M. (2016). ISpinach: A fluorogenic RNA aptamer optimized for in vitro applications. Nucleic Acids Res. *44*, 2491–2500.

Autour, A., Jeng, S.C.Y., Cawte, A.D., Abdolahzadeh, A., Galli, A., Panchapakesan, S.S.S., Rueda, D., Ryckelynck, M., and Unrau, P.J. (2018). Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells. Nat. Commun. *9*.

B

Babendure, J.R., Adams, S.R., and Tsien, R.Y. (2003). Aptamers Switch on Fluorescence of Triphenylmethane Dyes. J. Am. Chem. Soc. *125*, 14716–14717.

Balaban, N.Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004). Bacterial Persistenceas a Phenotypic Switch. Science (80-.). *305*, 1622–1625.

Baret, J.C. (2012). Surfactants in droplet-based microfluidics. Lab Chip *12*, 422–433.

Baret, J.-C., Miller, O.J., Taly, V., Ryckelynck, M., El-Harrak, A., Frenz, L., Rick, C., Samuels, M.L., Hutchison, J.B., Agresti, J.J., et al. (2009). Fluorescence-activated droplet sorting

(FADS): efficient microfluidic cell sorting based on enzymatic activity. Lab Chip *9*, 1850–1858.

Bartel, D.P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. Cell *136*, 215–233.

Battesti, A., Majdalani, N., and Gottesman, S. (2011). The RpoS-Mediated General Stress Response in *Escherichia coli*. Annu. Rev. Microbiol. *65*, 189–213.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M. (1998). Localization of ASH1 mRNA particles in living yeast. Mol. Cell *2*, 437–445.

Bigger, J.W. (1944). Treatment of Staphylococcal Infections With Penicillin By Intermittent Sterilisation. Lancet *244*, 497–500.

Boersma, A.J., Zuhorn, I.S., and Poolman, B. (2015). A sensor for quantification of macromolecular crowding in living cells. Nat. Methods *12*, 227–229.

Bouhedda, F., Autour, A., and Ryckelynck, M. (2017). Light-Up RNA Aptamers and Their Cognate Fluorogens: From Their Development to Their Applications. Int. J. Mol. Sci. *19*.

Braselmann, E., Wierzba, A.J., Polaski, J.T., Chromiński, M., Holmes, Z.E., Hung, S.-T., Batan, D., Wheeler, J.R., Parker, R., Jimenez, R., et al. (2018). A multicolor riboswitch-based platform for imaging of RNA in live mammalian cells. Nat. Chem. Biol.

Bratu, D.P., Cha, B.-J., Mhlanga, M.M., Kramer, F.R., and Tyagi, S. (2003). Visualizing the distribution and transport of mRNAs in living cells. Proc. Natl. Acad. Sci. *100*, 13308–13313.

Breaker, R.R. (2002). Engineered allosteric ribozymes as biosensor components. Curr. Opin. Biotechnol. *13*, 31–39.

Butala, M., Žgur-Bertok, D., and Busby, S.J.W. (2009). The bacterial LexA transcriptional repressor. Cell. Mol. Life Sci. *66*, 82–93.

C

Cabantous, S., Terwilliger, T.C., and Waldo, G.S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat. Biotechnol. *23*, 102–107.

Cai, L., Friedman, N., and Xie, X.S. (2006). Stochastic protein expression in individual cells at the single molecule level. Nature *440*, 358–362.

Cain, J.A., Solis, N., and Cordwell, S.J. (2014). Beyond gene expression: The impact of protein post-translational modifications in bacteria. J. Proteomics *97*, 265–286.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W., and Prasher, D. (1994). Green fluorescent protein as a marker for gene expression. Science (80-.). *263*, 802–805.

Choi, K., Ng, A.H.C., Fobel, R., and Wheeler, A.R. (2012). Digital microfluidics. Annu. Rev. Anal. Chem. (Palo Alto. Calif). *5*, 413–440.

Choi, P.J., Cai, L., Frieda, K., and Xie, X.S. (2008). A Stochastic Single-Molecule Event Triggers Phenotype Switching of a Bacterial Cell. Science (80-.). *322*, 442–446.

Chudakov, D., Matz, M., Lukyanov, S., and Lukyanov, K. (2010). Fluorescent Proteins and

Their Applications in Imaging Living Cells and Tissues. Physiol. Rev. *90*, 1103–1163.

Cobb, M. (2017). 60 years ago, Francis Crick changed the logic of biology. PLoS Biol. *15*, 1–8.

Cody, C.W., Prasher, D.C., Westler, W.M., Prendergast, F.G., and Ward, W.W. (1993). Chemical Structure of the Hexapeptide Chromophore of the Aequorea Green-Fluorescent Protein. Biochemistry *32*, 1212–1218.

Crick, F. (1958). On Protein Synthesis. Symp. Soc. Exp. Biol. 138–166.

Crosetto, N., Bienko, M., and van Oudenaarden, A. (2015). Spatially resolved transcriptomics and beyond. Nat. Rev. Genet. *16*, 57–66.

Cui, C., Shu, W., and Li, P. (2016). Fluorescence In situ Hybridization: Cell-Based Genetic Diagnostic and Research Applications. Front. Cell Dev. Biol. *4*, 1–11.

D

Darmostuk, M., Rimpelova, S., Gbelcova, H., and Ruml, T. (2014). Current approaches in SELEX: An update to aptamer selection technology. Biotechnol. Adv. *33*, 1141–1161.

Delvigne, F., Zune, Q., Lara, A.R., Al-Soud, W., and Sørensen, S.J. (2014). Metabolic variability in bioprocessing: Implications of microbial phenotypic heterogeneity. Trends Biotechnol. *32*, 608–616.

Dolgosheina, E. V., Jeng, S.C.Y., Panchapakesan, S.S.S., Cojocaru, R., Chen, P.S.K., Wilson, P.D., Hawkins, N., Wiggins, P.A., and Unrau, P.J. (2014). RNA Mango aptamer-fluorophore: A bright, high-affinity complex for RNA labeling and tracking. ACS Chem. Biol. *9*, 2412–2420.

E

Eldar, A., and Elowitz, M.B. (2010). Functional roles for noise in genetic circuits. Nature *467*, 167–173.

Ellington, A.D., and Szostak, J.W. (1990). In vitro selection of RNA molecules that bind specific ligands. Nature *346*, 818–822.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183–1186.

F

Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of Single RNA Transcripts in Situ - Femino wt al. 1998. *280*, 585–590.

Filonov, G.S., Moon, J.D., Svensen, N., and Jaffrey, S.R. (2014). Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. J. Am. Chem. Soc. *136*, 16299–16308.

Filonov, G.S., Kam, C.W., Song, W., and Jaffrey, S.R. (2015). In-gel imaging of RNA processing using broccoli reveals optimal aptamer expression strategies. Chem. Biol. *22*, 649–660.

Fisher, R.A., Gollan, B., and Helaine, S. (2017). Persistent bacterial infections and persister cells. Nat. Rev. Microbiol. *15*, 453–464.

G

Geisler, S., and Coller, J. (2013). RNA in unexpected places: Long non-coding RNA functions in diverse cellular contexts. Nat. Rev. Mol. Cell Biol. *14*, 699–712.

Ghadessy, F.J., Ong, J.L., and Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. Proc. Natl. Acad. Sci. U. S. A. *98*, 4552–4557.

Gotrik, M., Sekhon, G., Saurabh, S., Nakamoto, M., Eisenstein, M., and Soh, H.T. (2018). Direct Selection of Fluorescence-Enhancing RNA Aptamers. J. Am. Chem. Soc. *140*, 3583–3591.

Grate, D., and Wilson, C. (1999). Laser-mediated, site-specific inactivation of RNA transcripts. Proc. Natl. Acad. Sci. *96*, 6131–6136.

Grimbergen, A.J., Siebring, J., Solopova, A., and Kuipers, O.P. (2015). Microbial bet-hedging: The power of being different. Curr. Opin. Microbiol. *25*, 67–72.

Groher, F., Bofill-Bosch, C., Schneider, C., Braun, J., Jager, S., Geißler, K., Hamacher, K., and Suess, B. (2018). Riboswitching with ciprofloxacin-development and characterization of a novel RNA regulator. Nucleic Acids Res. *46*, 2121–2132.

Grynkiewicz, G., Poenie, M., and Tsien, R.Y. (1985). A new generation of Ca2+ indicators with greatly improved fluorescence properties. J. Biol. Chem. *260*, 3440–3450.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell *35*, 849–857.

Guet, D., Burns, L.T., Maji, S., Boulanger, J., Hersen, P., Wente, S.R., Salamero, J., and Dargemont, C. (2015). Combining Spinach-tagged RNA and gene localization to image gene expression in live yeast. Nat. Commun. *6*, 8882.

H

Hallberg, Z.F., Su, Y., Kitto, R.Z., and Hammond, M.C. (2017). Engineering and In Vivo Applications of Riboswitches. Annu. Rev. Biochem. *86*, 515–539.

Han, K.Y., Leslie, B.J., Fei, J., Zhang, J., and Ha, T. (2013). Understanding the Photophysics of the Spinach–DFHBI RNA Aptamer–Fluorogen Complex To Improve Live-Cell RNA Imaging. J. Am. Chem. Soc. *135*, 19033–19038.

Harms, A., Maisonneuve, E., and Gerdes, K. (2016). Mechanisms of bacterial persistence during stress and antibiotic exposure. Science (80-.). *354*.

Heim, R., Prasher, D.C., and Tsien, R.Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc. Natl. Acad. Sci. *91*, 12501–12504.

Huang, H., Suslov, N.B., Li, N.-S., Shelke, S.A., Evans, M.E., Koldobskaya, Y., Rice, P.A., and Piccirilli, J.A. (2014). A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. Nat. Chem. Biol. *10*, 686–691.

Iliuk, A.B., Hu, L., and Tao, W.A. (2011). Aptamer in bioanalytical applications. Anal. Chem.

83, 4440–4452.

I

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.B., Lönnerberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. *21*, 1160–1167.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods *11*, 163–166.

J

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

Jaffrey, S.R. (2017). RNA-Based Fluorescent Biosensors for Detecting Metabolites in vitro and in Living Cells (Elsevier Inc.).

Jeng, S.C.Y., Chan, H.H.Y., Booy, E.P., McKenna, S.A., and Unrau, P.J. (2016). Fluorophore ligand binding and complex stabilization of the RNA Mango and RNA Spinach aptamers. Rna *22*, 1884–1892.

Jenison, R.D., Gill, S.C., Pardi, A., and Polisky, B. (1994). High-resolution molecular discrimination by RNA. Science *263*, 1425–1429.

Jepsen, M.D.E., Sparvath, S.M., Nielsen, T.B., Langvad, A.H., Grossi, G., Gothelf, K. V, and Andersen, E.S. (2018). Development of a genetically encodable FRET system using fluorescent RNA aptamers. Nat. Commun. *9*, 18.

Jiang, F., and Doudna, J.A. (2017). CRISPR-Cas9 Structures and Mechanisms. Annu. Rev. Biophys. *46*, 505–529.

Johansson, J., Mandin, P., Renzoni, A., Chiaruttini, C., Springer, M., and Cossart, P. (2002). An RNA thermosensor controls expression of virulence genes in Listeria monocytogenes. Cell *110*, 551–561.

Johnsen, J.M. (2013). Massively parallel sequencing : the new frontier of hematologic genomics Massively parallel sequencing. Blood *122*, 3268–3275.

K

Karunanayake Mudiyanselage, A.P.K.K., Yu, Q., Leon-Duque, M.A., Zhao, B., Wu, R., and You, M. (2018). Genetically Encoded Catalytic Hairpin Assembly for Sensitive RNA Imaging in Live Cells. J. Am. Chem. Soc. *140*, 8739–8745.

Kent, K.P., Childs, W., and Boxer, S.G. (2008). Deconstructing Green Fluorescent Protein. J. Am. Chem. Soc. *130*, 9664–9665.

Ketterer, S., Fuchs, D., Weber, W., and Meier, M. (2015). Systematic reconstruction of binding and stability landscapes of the fluorogenic aptamer spinach. Nucleic Acids Res. *43*, 9564–

9572.

Kilgard, R., Heim, A.B., and Tsien, R.Y. (1995). Improved green fluorescence. Nature *373*, 663–664.

Klein, D.J., and Ferré-D'Amaré, A.R. (2006). Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. Science *313*, 1752–1756.

Kobori, S., and Yokobayashi, Y. (2016). High-Throughput Mutational Analysis of a Twister Ribozyme. Angew. Chem. Int. Ed. Engl. *55*, 10354–10357.

Kobori, S., Nomura, Y., Miu, A., and Yokobayashi, Y. (2015). High-throughput assay and engineering of self-cleaving ribozymes by sequencing. Nucleic Acids Res. *43*, e85.

Köhler, O., Jarikote, D.V., and Seitz, O. (2005). Forced intercalation probes (FIT Probes): Thiazole orange as a fluorescent base in peptide nucleic acids for homogeneous single-nucleotide-polymorphism detection. ChemBioChem *6*, 69–77.

Kolb, A., Busby, S., Buc, I.I., Garges, S., and Adhya, S. (1993). Transcriptional Regulation by cAMP and its Receptor Protein. Annu. Rev. Biochem. *62*, 749–797.

Kortmann, J., and Narberhaus, F. (2012). Bacterial RNA thermometers: Molecular zippers and switches. Nat. Rev. Microbiol. *10*, 255–265.

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene *361*, 13–37.

Kummer, S., Knoll, A., Socher, E., Bethge, L., Herrmann, A., and Seitz, O. (2012). PNA FIT-probes for the dual color imaging of two viral mRNA targets in influenza H1N1 infected live cells. Bioconjug. Chem. *23*, 2051–2060.

L

Lee, L.G., Chen, C.H., and Chiu, L.A. (1986). Thiazole orange: a new dye for reticulocyte analysis. Cytometry *7*, 508–517.

Levine, J.H., Lin, Y., and Elowitz, M.B. (2013). Functional Roles of Pulsing in Genetic Circuits. Science (80-.). *342*, 1193–1200.

Levy, M., Griswold, K.E., and Ellington, A.D. (2005). Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. RNA *11*, 1555–1562.

Locke, J.C.W., and Elowitz, M.B. (2009). Using movies to analyse gene circuit dynamics in single cells. Nat. Rev. Microbiol. *7*, 383–392.

Lubeck, E., and Cai, L. (2012). Single-cell systems biology by super-resolution imaging and combinatorial labeling. Nat. Methods *9*, 743–748.

M

Maas, W.K. (1994). The Arginine Repressor of Escherichia-Coli. Microbiol. Rev. *58*, 631–640.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell *161*, 1202–1214.

Mandal, M., and Breaker, R.R. (2004a). Gene regulation by riboswitches. Nat. Rev. Mol. Cell Biol. *5*, 451–463.

Mandal, M., and Breaker, R.R. (2004b). Adenine riboswitches and gene activation by disruption of a transcription terminator. Nat. Struct. Mol. Biol. *11*, 29–35.

Marcus, J.S., Anderson, W.F., and Quake, S.R. (2006). Microfluidic single-cell mRNA isolation and analysis. Anal. Chem. *78*, 3084–3089.

Mastrobattista, E., Taly, V., Chanudet, E., Treacy, P., Kelly, B.T., and Griffiths, A.D. (2005). High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. Chem. Biol. *12*, 1291–1300.

Mazutis, L., Araghi, A.F., Miller, O.J., Baret, J.C., Frenz, L., Janoshazi, A., Taly, V., Miller, B.J., Hutchison, J.B., Link, D., et al. (2009). Droplet-based microfluidic systems for high-throughput single DNA molecule isothermal amplification and analysis. Anal. Chem. *81*, 4813–4821.

Mellin, J.R., and Cossart, P. (2015). Unexpected versatility in bacterial riboswitches. Trends Genet. *31*, 150–156.

Metzker, M.L. (2010). Sequencing technologies the next generation. Nat. Rev. Genet. *11*, 31–46.

Miller, O.J., Bernath, K., Agresti, J.J., Amitai, G., Kelly, B.T., Mastrobattista, E., Taly, V., Magdassi, S., Tawfik, D.S., and Griffiths, A.D. (2006). Directed evolution by in vitro compartmentalization. Nat. Methods *3*, 561–570.

Miyawaki, a, Llopis, J., Heim, R., McCaffery, J.M., Adams, J. a, Ikura, M., and Tsien, R.Y. (1997). Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. Nature *388*, 882–887.

N

Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E., and Altuvia, S. (2009). A pH-responsive riboregulator. Genes Dev. *23*, 2650–2662.

Nienhaus, K., and Nienhaus, G.U. (2016). Chromophore photophysics and dynamics in fluorescent proteins of the GFP family. J. Phys. Condens. Matter *28*.

Nissen, P., Nissen, P., Hansen, J., Ban, N., and Moore, P.B. (2000). The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. Sciences (New. York). *289*, 920–930.

Novick, A., and Weiner, M. (1957). Enzyme Induction As an All-or-None Phenomenon. Proc. Natl. Acad. Sci. *43*, 553–566.

O

Okuda, M., Fourmy, D., and Yoshizawa, S. (2017). Use of Baby Spinach and Broccoli for imaging of structured cellular RNAs. Nucleic Acids Res. *45*, 1404–1415.

Österberg, S., Peso-Santos, T. del, and Shingler, V. (2011). Regulation of Alternative Sigma Factor Use. Annu. Rev. Microbiol. *65*, 37–55.

P

Paige, J.S., Wu, K.Y., and Jaffrey, S.R. (2011). RNA mimics of green fluorescent protein. Science *333*, 642–646.

Paige, J.S., Nguyen-Duc, T., Song, W., and Jaffrey, S.R. (2012). BREVIA Fluorescence Imaging of Cellular Metabolites with RNA. Science (80-.). *335*, 1194.

Panchapakesan, S.S.S., Jeng, S.C.Y., and Unrau, P.J. (2015). RNA complex purification using high-affinity fluorescent RNA aptamer tags. Ann. N. Y. Acad. Sci. *1341*, 149–155.

Pastan, I., and Perlman, R. (1970). Cyclic adenosine monophosphate in bacteria. Science *169*, 339–344.

Pédelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nat. Biotechnol. *24*, 79–88.

Pekin, D., Skhiri, Y., Baret, J.C., Le Corre, D., Mazutis, L., Ben Salem, C., Millot, F., El Harrak, A., Hutchison, J.B., Larson, J.W., et al. (2011). Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. Lab Chip *11*, 2156–2166.

Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. (2002). Xist RNA and the Mechanism of X Chromosome Inactivation. Annu. Rev. Genet. *36*, 233–278.

Potvin-Trottier, L., Luro, S., and Paulsson, J. (2018). Microfluidics and single-cell microscopy to study stochastic processes in bacteria. Curr. Opin. Microbiol. *43*, 186–192.

Prakadan, S.M., Shalek, A.K., and Weitz, D.A. (2017). Scaling by shrinking: Empowering single-cell "omics" with microfluidic devices. Nat. Rev. Genet. *18*, 345–361.

Q

Quan, P.L., Sauzade, M., and Brouzes, E. (2018). DPCR: A technology review. Sensors (Switzerland) *18*.

Quang, N.N., Perret, G., and Ducongé (2016). Applications of high-throughput sequencing for in vitro selection and characterization of aptamers. Pharmaceuticals *9*, 1–15.

R

Rauhut, R., and Klug, G. (1999). mRNA degradation in bacteria. FEMS Microbiol. Rev. *23*, 353–370.

Ren, G.X., Guo, X.P., and Sun, Y.C. (2017). Regulatory 3' untranslated regions of bacterial mRNAs. Front. Microbiol. *8*, 1–6.

Romero, P.A., Tran, T.M., and Abate, A.R. (2015). Dissecting enzyme function with microfluidic-based deep mutational scanning. Proc. Natl. Acad. Sci. *112*, 7159–7164.

Ryckelynck, M., Baudrey, S., Rick, C., Marin, A., Coldren, F., Westhof, E., and Griffiths, A.D. (2015). Using droplet-based microfluidics to improve the catalytic properties of RNA under multiple-turnover conditions. Rna *21*, 458–469.

S

Sciambi, A., and Abate, A.R. (2015). Accurate microfluidic sorting of droplets at 30 kHz. Lab Chip *15*, 47–51.

Sedlyarova, N., Shamovsky, I., Bharati, B.K., Epshtein, V., Chen, J., Gottesman, S., Schroeder, R., and Nudler, E. (2016). sRNA-Mediated Control of Transcription Termination in E. coli. Cell *167*, 111–121.e13.

Shembekar, N., Chaipan, C., Utharala, R., and Merten, C.A. (2016). Droplet-based microfluidics in drug discovery, transcriptomics and high-throughput molecular genetics. Lab Chip *16*, 1314–1331.

Sherlock, M.E., Sudarsan, N., and Breaker, R.R. (2018). Riboswitches for the alarmone ppGpp expand the collection of RNA-based signaling systems. Proc. Natl. Acad. Sci. 201720406.

Shimomura, O. (2005). The discovery of green fluorescent protein. J. Microsc. *217*, 3–15.

Shine, J., and Dalgarno, L. (1974). The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U. S. A. *71*, 1342–1346.

Solopova, A., van Gestel, J., Weissing, F.J., Bachmann, H., Teusink, B., Kok, J., and Kuipers, O.P. (2014). Bet-hedging during bacterial diauxic shift. Proc. Natl. Acad. Sci. *111*, 7427–7432.

Song, W., Strack, R.L., and Jaffrey, S.R. (2013). Imaging bacterial protein expression using genetically encoded RNA sensors. Nat. Methods *10*, 873–875.

Song, W., Strack, R.L., Svensen, N., and Jaffrey, S.R. (2014). Plug-and-play fluorophores extend the spectral properties of Spinach. J. Am. Chem. Soc. *136*, 1198–1201.

Song, W., Filonov, G.S., Kim, H., Hirsch, M., Li, X., Moon, J.D., and Jaffrey, S.R. (2017). Imaging RNA polymerase III transcription using a photostable RNA-fluorophore complex. Nat. Chem. Biol. *13*, 1187–1194.

Stojanovic, M.N., and Kolpashchikov, D.M. (2004). Modular aptameric sensors. J. Am. Chem. Soc. *126*, 9266–9270.

Stoltenburg, R., Reinemann, C., and Strehlitz, B. (2007). SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. Biomol. Eng. *24*, 381–403.

Stoltenburg, R., Nikolaus, N., and Strehlitz, B. (2012). Capture-SELEX: Selection of DNA aptamers for aminoglycoside antibiotics. J. Anal. Methods Chem. *1*.

Strack, R.L., Disney, M.D., and Jaffrey, S.R. (2013). A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. Nat. Methods *10*, 1219–1224.

Sunbul, M., and Jäschke, A. (2013). Contact-mediated quenching for RNA imaging in bacteria with a fluorophore-binding aptamer. Angew. Chemie - Int. Ed. *52*, 13401–13404.

Sunbul, M., and Jäschke, A. (2018). SRB-2: a promiscuous rainbow aptamer for live-cell RNA imaging. Nucleic Acids Res. 1–10.

Svensen, N., and Jaffrey, S.R. (2016). Fluorescent RNA Aptamers as a Tool to Study RNA-Modifying Enzymes. Cell Chem. Biol. *23*, 415–425.

T

Tan, X., Constantin, T.P., Sloane, K.L., Waggoner, A.S., Bruchez, M.P., and Armitage, B.A. (2017). Fluoromodules Consisting of a Promiscuous RNA Aptamer and Red or Blue Fluorogenic Cyanine Dyes: Selection, Characterization, and Bioimaging. J. Am. Chem. Soc. *139*, 9001–9009.

Tang, J., and Breaker, R.R. (1997). Rational design of allosteric ribozymes. Chem. Biol. *4*, 453–459.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382.

Tawfik, D.S., and Griffiths, A.D. (1998). Man-made cell-like compartments for molecular evolution. Nat. Biotechnol. *16*, 652–656.

Thorsen, T. (2002). Microfluidic Large-Scale Integration. Science (80-.). *298*, 580–584.

Trachman, R., Abdolahzadeh, A., Andreoni, A., Cojocaru, R., Knutson, J.R., Ryckelynck, M., Unrau, P.J., and Ferré-D'Amaré, A. (2018). Crystal structures of the Mango-II RNA aptamer reveal heterogeneous fluorophore binding and guide engineering of variants with improved selectivity and brightness. Biochemistry.

TrachmanIii, R.J., Demeshkina, N.A., Lau, M.W.L., Panchapakesan, S.S.S., Jeng, S.C.Y., Unrau, P.J., and Ferré-D'Amaré, A.R. (2017). Structural basis for high-affinity fluorophore binding and activation by RNA Mango. Nat. Chem. Biol. *13*, 807–813.

Tuerk, C., Eddy, S., Parma, D., and Gold, L. (1990). Autogenous translational operator recognized by bacteriophage T4 DNA polymerase. J. Mol. Biol. *213*, 749–761.

Tyagi, S. (2009). Imaging intracellular RNA distribution and dynamics in living cells. Nat. Methods *6*, 331–338.

U

Urbanek, M.O., Galka-Marciniak, P., Olejniczak, M., and Krzyzosiak, W.J. (2014). RNA imaging in living cells – methods and applications. RNA Biol. *11*, 1083–1095.

V

Veening, J.-W., Smits, W.K., and Kuipers, O.P. (2008a). Bistability, Epigenetics, and Bet-Hedging in Bacteria. Annu. Rev. Microbiol. *62*, 193–210.

Veening, J.-W., Stewart, E.J., Berngruber, T.W., Taddei, F., Kuipers, O.P., and Hamoen, L.W. (2008b). Bet-hedging and epigenetic inheritance in bacterial cell development. Proc. Natl. Acad. Sci. *105*, 4393–4398.

Vera, M., Biswas, J., Senecal, A., Singer, R.H., and Park, H.Y. (2016). Single-Cell and Single-Molecule Analysis of Gene Expression Regulation. Annu. Rev. Genet. *50*, 267–291.

Vogelstein, B., and Kinzler, K.W. (1999). Digital PCR. Proc. Natl. Acad. Sci. *96*, 9236–9241.

Vorobyeva, M., Davydova, A., Vorobjev, P., Pyshnyi, D., and Venyaminova, A. (2018). Key Aspects of Nucleic Acid Library Design for in Vitro Selection. Int. J. Mol. Sci. *19*, 470.

W

Wagner, E.G.H., and Romby, P. (2015). Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It (Elsevier Ltd).

Wang, J., Yu, J., Yang, Q., McDermott, J., Scott, A., Vukovich, M., Lagrois, R., Gong, Q., Greenleaf, W., Eisenstein, M., et al. (2017). Multiparameter Particle Display (MPPD): A Quantitative Screening Method for the Discovery of Highly Specific Aptamers. Angew. Chemie Int. Ed. *56*, 744–747.

Wang, Z., Luo, Y., Xie, X., Hu, X., Song, H., Zhao, Y., Shi, J., Wang, L., Glinsky, G., Chen, N., et al. (2018). In Situ Spatial Complementation of Aptamer-Mediated Recognition Enables Live-Cell Imaging of Native RNA Transcripts in Real Time. Angew. Chemie - Int. Ed. *57*, 972–976.

Warner, K.D., Chen, M.C., Song, W., Strack, R.L., Thorn, A., Jaffrey, S.R., and Ferré-D'Amaré, A.R. (2014). Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. Nat. Struct. Mol. Biol. *21*, 658–663.

Warner, K.D., Sjekloća, L., Song, W., Filonov, G.S., Jaffrey, S.R., and Ferré-D'Amaré, A.R. (2017). A homodimer interface without base pairs in an RNA mimic of red fluorescent protein. Nat. Chem. Biol. *13*, 1195–1201.

Weigand, J.E., Sanchez, M., Gunnesch, E.-B., Zeiher, S., Schroeder, R., and Suess, B. (2008). Screening for engineered neomycin riboswitches that control translation initiation. RNA *14*, 89–97.

Win, M.N., and Smolke, C.D. (2007). A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc. Natl. Acad. Sci. *104*, 14283–14288.

Winkler, W.C., Cohen-Chalamish, S., and Breaker, R.R. (2002). An mRNA structure that controls gene expression by binding FMN. Proc. Natl. Acad. Sci. U. S. A. *99*, 15908–15913.

Wochner, A., Attwater, J., Coulson, A., and Holliger, P. (2011). Ribozyme-Catalyzed Transcription of an Active Ribozyme. Science (80-.). *332*, 209–212.

Wu, B., Chen, J., and Singer, R.H. (2014). Background free imaging of single mRNAs in live cells using split fluorescent proteins. Sci. Rep. *4*, 3615.

Y

You, L., and Arnold, F.H. (1996). Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide. Protein Eng. *9*, 77–83.

Z

Zaccolo, M., Williams, D.M., Brown, D.M., and Gherardi, E. (1996). An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. J. Mol. Biol. *255*, 589–603.

Zaher, H.S., and Unrau, P.J. (2007). Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. RNA *13*, 1017–1026.

Zhang, J., Fei, J., Leslie, B.J., Han, K.Y., Kuhlman, T.E., and Ha, T. (2015). Tandem Spinach Array for mRNA Imaging in Living Bacterial Cells. Sci. Rep. *5*, 17295.

Zhou, J., Ellis, A.V., and Voelcker, N.H. (2010). Recent developments in PDMS surface modification for microfluidic devices. Electrophoresis *31*, 2–16.

Alexis AUTOUR

Amélioration et criblages de propriétés d'ARN aptamères fluorogènes en systèmes microfluidiques

Résumé

Les ARN (Acide RiboNucléique) remplissent de nombreuses fonctions clés dans le vivant. Ils peuvent être support de l'information génétique, régulateurs de celle-ci. Visualiser ces molécules au sein d'une cellule représenterait une étape importante vers une meilleure compréhension de la régulation de l'expression des gènes. Les ARN fluorogènes tels que Spinach et Mango sont des outils extrêmement prometteurs pour atteindre cet objectif. Cependant ces deux ARN fluorogènes présentent une brillance limitée. La Compartimentation *in vitro* assistée par microfluidique (µCIV) est un outil très prometteur dont notre groupe a démontré l'efficacité pour l'évolution d'ARN.

Dans le cadre de cette thèse, la µCIV a été adaptée à la sélection d'aptamères d'ARN fluorogènes (Spinach et Mango) pour en améliorer les propriétés (en particulier la brillance). De plus, l'utilisation conjointe du séquençage haut débit a permis l'optimisation très rapide et semi-automatisée à la fois d'aptamères mais aussi de biosenseurs fluorogènes. Ainsi, cette thèse a permis de mettre en place et d'exploiter des technologies de criblage robustes pour la découverte de nouveaux aptamères d'ARN ainsi que de biosenseurs de métabolites.

Mots-clés : ARN fluorogènes, Biosenseur, Microfluidique en gouttelettes, Sélection, Criblage haut débit, séquençage haut-débit

Résumé en anglais

RNA is a key molecule in gene expression and its regulation. Therefore, being able to monitor RNA through live-cell imaging would represent an important step toward a better understanding of gene expression regulation. RNA-based fluorogenic modules are extremely promising tools to reach this goal. To this end, two light-up RNA aptamers (Spinach and Mango) display attractive properties but they suffer from a limited brightness. Since previous work in the group demonstrated the possibility to evolve RNA using microfluidic-assisted *in vitro* compartmentalization (µIVC), this technology appeared to be well suited to improve light-up aptamers properties by an evolution strategy.

Therefore, the µIVC procedure was adapted to fluorogenic RNA aptamers (Spinach and Mango) to improve their properties (especially the brightness). Finally, using µIVC in tandem with high-throughput sequencing (NGS) allowed further developing the technology into a more integrated and semi-automatized approach in which RNAs and biosensors are selected by µIVC screening and the best variants identified by a bioinformatics process upon NGS analysis. To summarize, this thesis allowed to establish robust µIVC screening workflows for the discovery of novel efficient light-up RNA aptamers as well as metabolites biosensors.

Keywords: Light-up RNA aptamers, Biosensors, Droplet-based microfluidics, Selection, Ultrahigh-throughput, High-throughput sequencing