Advances in fluorogenic RNA aptamer systems for live cell imaging: Towards orthogonality and multicolour applications

by

Sunny Jeng

B.Sc., Simon Fraser University, 2013

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

in the
Department of Molecular Biology & Biochemistry
Faculty of Science

© Sunny Jeng 2018
SIMON FRASER UNIVERSITY
Fall 2018

Copyright in this work rests with the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.
Approval

Name: Sunny Jeng
Degree: Doctor of Philosophy
Title: Advances in fluorogenic RNA aptamer systems for live cell imaging; Towards orthogonality and multicolour applications
Examining Committee: Chair: Edgar Young
Associate Professor

Peter Unrau
Senior Supervisor
Professor
Dipankar Sen
Supervisor
Professor
David Vocadlo
Supervisor
Professor
Andrew Bennet
Internal Examiner
Professor
Department of Chemistry
Joseph Piccirilli
External Examiner
Professor
Department of Chemistry
The University of Chicago

Date Defended/Approved: October 25, 2018
Abstract

Recent developments in the field of RNA biology continue to demonstrate the importance of RNA in regulating cellular processes. However, directly imaging biologically important RNAs has been hindered by a lack of live cell fluorescent tools. As such, aptamers that bind and enhance the brightness of fluorogenic dyes are promising tools to improve fluorescent RNA imaging. The Unrau laboratory developed RNA Mango I, a small, 39-nt aptamer that binds to a modified thiazole orange fluorophore (TO1-Biotin) with nanomolar affinity. This binding is accompanied by an 1,100-fold increase in its green channel fluorescence. To further improve the Mango aptamers, in collaboration with the Ryckelynck laboratory, we used microfluidics-based selection methods to isolate three brighter, high affinity RNA Mango fluorogenic aptamers (Mango II, III and IV). Together with the Rueda laboratory, we show that these new Mangos can accurately image the sub-cellular localization of three small non-coding RNAs in fixed and live mammalian cells. These new Mangos are unique in structure. Unlike Mango I and II, Mango III rigidly connects its ligand binding core to an external helix. As the Spinach/Broccoli aptamer family, which binds GFP-like chromophores (DFHBI, DFHBI-1T), also share this property, the Broccoli/DFHBI-1T aptamer complex was used as a FRET donor paired with the far red-shifted Mango III/YO3-Biotin complex as a FRET acceptor. Interestingly, the high affinity Mango I, II, III aptamers can discriminate between TO1/YO3-Biotin and DFHBI/DFHBI-1T by at least a 102-fold difference in affinity. In contrast, Spinach binds many fluorophores indiscriminately and weaker. With this, concentrations could be determined to obtain appropriate binding for Mango III/YO3-Biotin and Broccoli/DFHBI-1T when in the same system. FRET efficiency was measured using an RNA duplex of variable length between the two aptamers. FRET signal depended on the length of the duplex, and oscillated in intensity precisely with the predicted twist of the helix, demonstrating strong orientation dependence. While this pair of aptamers enable in vitro FRET studies, there are no truly orthogonal fluorescent aptamer systems. To that end, I discuss an in vitro selection to develop orthogonal aptamers for a red fluorophore, TO3-Biotin, that can potentially be paired with existing Mango aptamers.

Keywords: RNA, Aptamer, Fluorescence, FRET, In vitro selection, SELEX, microfluidics
Acknowledgements

My graduate studies have been a wonderful and enlightening experience and would not have been possible without the people I have had the privilege of interacting with.

First, I thank Dr. Peter Unrau. I would be nowhere near the place I am at today without his support, guidance and the opportunities he has given me throughout my years as a graduate student.

Thank you to my committee members, Dr. Dipankar Sen and Dr. David Vocadlo for their support by their intellectual input into my projects, encouragement, and all the time they have spent to help me in my career through things like reading my materials, and acting as my references. I’d also like to thank our current department chair Dr. Nancy Hawkins for similar support despite not being a member of my committee.

I extend my appreciation to my internal examiner, Dr. Andrew Bennet and my external examiner Dr. Joseph Piccirilli for taking the time to read my thesis and attend my defence.

I also would like to thank all the collaborators I have interacted with these past years. Particularly, I want to mention Alexis Autour and Dr. Michael Ryckelynck in Strasbourg, and Adam Cawte and Dr. David Rueda in London. I am grateful for the opportunity to work with them these last years, and thankful for the hospitality they provided me and the time they spent for me when I visited them one summer.

I want to thank the co-workers and friends I have met in my time here including: Members of the Unrau lab for all the fun we had together in and out of the lab: Dr. Elena Dolgosheina who first guided me as an undergraduate, Dr. Shyam Panchapakesan, Razvan Cojocaru, Amir Abdolahzadeh, Lyssa Martin, Dr. Mariana Oviedo and all of the undergraduates who have come and gone. Members of the department office: Nancy Suda, Mimi Fourie, Nagisa Inoue, Christine Beauchamp and for support in ways of mundane things like paperwork, but also fun chats that could bring smiles to even the toughest days. Lastly, thank you also to all the professors, graduate students, post-docs and lab technicians that I encountered daily for various forms of support and that made life in the south science building most enjoyable.
I’d like to thank my family for their most valued support. To my parents, no words can describe the amount they have done to support, raise and encourage me my entire life. I will be forever grateful.

And Eugean, I thank her for all she has given me. All the joy, patience, love and support she provided in times of happiness, despair, anxiousness or excitement has been my pillar of strength. With you by my side I look forward to taking on any challenge in the next chapter of our lives.
# Table of Contents

Approval .......................................................................................................................... ii  
Abstract .......................................................................................................................... iii  
Acknowledgements ........................................................................................................ iv  
Table of Contents .......................................................................................................... vi  
List of Tables ................................................................................................................ ix  
List of Figures ................................................................................................................ x  
List of Acronyms .......................................................................................................... xiii  

## Chapter 1. Introduction .............................................................................................. 1  
1.1. RNA history and their functions ........................................................................ 1  
1.2. RNA structure ...................................................................................................... 2  
   1.2.1. Basic interactions ......................................................................................... 4  
   1.2.2. Higher order structures .............................................................................. 5  
Duplexes ....................................................................................................................... 6  
   Stem-loops, tetraloops ............................................................................................. 7  
   G quadruplexes .................................................................................................... 10  
1.3. In vitro selection and SELEX methods ................................................................ 12  
1.4. Fluorescent systems for cellular RNA imaging .................................................. 15  
   1.4.1. Fluorescence in situ hybridization (FISH) ................................................... 15  
   1.4.2. Fluorescence-based aptamers ..................................................................... 16  
      MS2/PP7 and other aptamer-protein systems ..................................................... 16  
      Fluorogenic dye aptamers .............................................................................. 17  
      Quencher-based fluorogenic dye aptamers ...................................................... 21  
1.5. Structural studies of fluorogenic dye-binding RNA aptamers ......................... 25  
   1.5.1. Malachite green aptamer .......................................................................... 25  
   1.5.2. Spinach aptamer ....................................................................................... 26  
   1.5.3. Corn aptamer ........................................................................................... 29  
   1.5.4. Mango aptamers ...................................................................................... 32  

## Chapter 2. Fluorophore ligand binding and complex stabilization of RNA Mango  
and RNA Spinach aptamers ...................................................................................... 34  
2.1. Introduction .......................................................................................................... 35  
2.2. Results .................................................................................................................. 36  
   2.2.1. RNA Mango and RNA Spinach fluorophore binding and fluorescence  
      enhancement .................................................................................................. 36  
   2.2.2. Fluorophore ligand discrimination by RNA Mango ................................... 41  
   2.2.3. Fluorescent efficiencies induced by RNA quadruplex binding are correlated  
      with ligand discrimination ............................................................................. 42  
   2.2.4. Temperature-dependent binding and fluorescent properties of RNA Mango  
      and RNA Spinach ......................................................................................... 43  
   2.2.5. Fluorescent lifetimes of TO1-Biotin:RNA Mango and TO1-Biotin:RNA  
      Spinach are distinct ....................................................................................... 47  
2.3. Discussion .............................................................................................................. 50
Chapter 3. Fluorogenic RNA Mango Aptamers for Imaging Small non-coding RNAs in Mammalian Cells

3.1. Introduction
3.2. Results
  3.2.1. Microfluidic isolation of new and brighter Mango aptamers
  3.2.2. Each new Mango variant is unique in structure and function
  3.2.3. Cellular imaging of Mango-tagged RNA
3.3. Discussion
3.4. Materials & Methods
  3.4.1. High throughput screening
  3.4.2. Quantification of RNA produced in droplets
  3.4.3. Enrichment measurement
  3.4.4. TA-cloning and colony screening
  3.4.5. DMS probing of Mango
  3.4.6. T1 RNase ladder and alkaline hydrolysis ladder
  3.4.7. Screening for minimal functional Mango motifs
  3.4.8. Affinity measurements of Mango variants
  3.4.9. Temperature-dependent fluorescence and UV melting curves
  3.4.10. Circular dichroism
  3.4.11. Formaldehyde resistance assay
  3.4.12. Cell culture and maintenance
  3.4.13. Plasmid and RNA synthesis
  3.4.14. Plasmid and RNA transfection
  3.4.15. Cell fixation and immunostaining
  3.4.16. Fluorescence microscopy and live cell imaging
  3.4.17. Photobleaching-assisted microscopy
  3.4.18. Image processing and quantification
  3.4.19. pH titrations
  3.4.20. In vitro photobleaching measurement

Chapter 4. Structure studies of Mango I and III, and studying orientation-dependent FRET with Mango III

4.1. Introduction
4.2. Results
  4.2.1. Structure of Mango I (excerpts from Trachman et al., Structural basis for high-affinity fluorophore binding and activation by RNA Mango)
4.2.2. Structure of Mango III (excerpts from Trachman et al., Mango-III is a compact fluorogenic RNA aptamer for unusual structural complexity) ........................................ 113
4.2.3. A rigid transition from ligand-binding core to external helix allows orientation FRET measurements between Mango III and Broccoli. ........................................... 114
4.2.4. Optimal conditions for FRET with DFHBI-1T and YO3-Biotin .......................... 116
4.2.5. Construction of M3-nbp-Broc models ............................................................. 119
4.2.6. FRET oscillates as a function of connecting duplex length ............................. 120
4.2.7. Calculation of FRET values using structure models ....................................... 124
4.3. Discussion ........................................................................................................... 127
4.4. Materials & Methods ....................................................................................... 128
  4.4.1. M3-nbp-Broc FRET measurements ............................................................... 128
  4.4.2. FRET spectrum fitting ................................................................................... 128
  4.4.3. Generating theoretical structure models ....................................................... 129
  4.4.4. Modelling of FRET parameters ................................................................... 130

Chapter 5. Selection for an orthogonal red fluorescent aptamer .................. 131
  5.1. Introduction ....................................................................................................... 132
  5.2. Results .............................................................................................................. 133
    5.2.1. Selection of an aptamer that discriminates TO3-Biotin from TO1-Biotin .... 133
    5.2.2. Diagnostics reveal increase in TO3 preference in late selection rounds ....... 134
    5.2.3. Studying individual clones from the TO3-Biotin selection ......................... 136
  5.3. Discussion ........................................................................................................ 137
  5.4. Materials & Methods ....................................................................................... 138
    5.4.1. In vitro selection .......................................................................................... 138
    5.4.2. Reverse-transcription PCR ......................................................................... 139
    5.4.3. Mutagenic PCR ......................................................................................... 139
    5.4.4. Fluorescence competition assay of libraries .............................................. 140
    5.4.5. Fluorescence competition assay of individual clones ............................... 140

References .............................................................................................................. 141
List of Tables

Table 1.1  Comprehensive list of fluorescent RNA aptamer systems ..................... 23
Table 2.1  Binding affinities in nM of each fluorophore:RNA combination at room
temperature ................................................................. 38
Table 2.2  Fluorescence enhancement (F_E) of each fluorophore:RNA combination at
room temperature ............................................................. 38
Table 2.3  RNA sequences used in binding affinity experiments .............................. 38
Table 2.4  Fluorescence enhancement (F_E) of each fluorophore:RNA combination at
room temperature ............................................................. 38
Table 2.5  Fluorescence efficiency of each fluorophore:RNA combination in units of
(nM)^{-1} at room temperature .............................................. 40
Table 2.6  RNA sequences used in binding affinity experiments .............................. 40
Table 2.7  Fluorescence efficiency of each fluorophore:RNA combination in units of
(nM)^{-1} at room temperature .............................................. 40
Table 2.8  Lifetimes of fluorescent decay experiments determined from single
exponential fit of data shown in Figure 2.7  ................................ 42
Table 2.9  Visible excitation and emission peaks for each fluorophore/RNA complex
used in this study (wavelengths in nm) ..................................... 42
Table 2.10  Extinction coefficients of unbound fluorophores used in this study in
conditions as cited ............................................................ 44
Table 3.1  Metrics of screenings in presence of NMM  .......................................... 61
Table 3.2  Metrics of screenings in the presence of TO3-Biotin ............................... 65
Table 3.3  Hill coefficients of RNA Mango/TO1-Biotin complexes ......................... 67
Table 3.4  Constructs used in in vivo experiments ................................................. 70
Table 3.5  Metrics of the fluorescence profiling of Mango and R12 .......................... 76
Table 3.6  Primers used for the construction of plasmids ...................................... 90
Table 4.1  Binding affinities and relative fluorescence of Mango I tetraloop motifs 112
Table 4.2  Binding affinity and excitation/emission peaks of Broccoli and Mango III
with DFHBI, DFHBI-1T, TO3-Biotin and YO3-Biotin ................................ 117
Table 4.3  Fitted emission peak shapes in Gaussian analysis of M3-nbp-Broc
constructs ............................................................................. 121
Table 4.4  Relative angle between normal vectors of the donor and acceptor
fluorophores in M3-nbp-Broc structures ....................................... 125
Table 5.1  Concentration conditions for each round of TO3-Biotin selection .......... 139
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Chemical structure of RNA</td>
<td>3</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Sugar puck er conformations.</td>
<td>4</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Nucleotide base interactions</td>
<td>5</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>B-form vs A-form double helix</td>
<td>7</td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>A GNRA tetraloop</td>
<td>8</td>
</tr>
<tr>
<td>Figure 1.6</td>
<td>A UNCG tetraloop</td>
<td>9</td>
</tr>
<tr>
<td>Figure 1.7</td>
<td>A CUUG tetraloop</td>
<td>9</td>
</tr>
<tr>
<td>Figure 1.8</td>
<td>G4 structure and topology</td>
<td>11</td>
</tr>
<tr>
<td>Figure 1.9</td>
<td>Generalized scheme of RNA in vitro selection procedure</td>
<td>12</td>
</tr>
<tr>
<td>Figure 1.10</td>
<td>Fluorogenic dyes for the malachite green aptamer</td>
<td>18</td>
</tr>
<tr>
<td>Figure 1.11</td>
<td>Dimethylindole red and Hoechst-1b fluorogenic dyes</td>
<td>19</td>
</tr>
<tr>
<td>Figure 1.12</td>
<td>Spinach and corn fluorophores based on fluorescent protein</td>
<td>20</td>
</tr>
<tr>
<td>Figure 1.13</td>
<td>Thiazole orange-based fluorophores for Mango aptamers</td>
<td>21</td>
</tr>
<tr>
<td>Figure 1.14</td>
<td>Fluorescent compounds for the SRB2 and DN aptamers</td>
<td>22</td>
</tr>
<tr>
<td>Figure 1.15</td>
<td>Cobalamin coupling to various fluorophores to form quencher-dye</td>
<td>23</td>
</tr>
<tr>
<td>Figure 1.16</td>
<td>Structure of the malachite green aptamer</td>
<td>26</td>
</tr>
<tr>
<td>Figure 1.17</td>
<td>Global structure of Spinach</td>
<td>27</td>
</tr>
<tr>
<td>Figure 1.18</td>
<td>The two-layer G4 in L(_{12}) of Spinach</td>
<td>27</td>
</tr>
<tr>
<td>Figure 1.19</td>
<td>The fluorophore binding site of Spinach.</td>
<td>28</td>
</tr>
<tr>
<td>Figure 1.20</td>
<td>The structure of Spinach RNA in the absence of DFHBI</td>
<td>29</td>
</tr>
<tr>
<td>Figure 1.21</td>
<td>Structure of Corn–DFHO complex</td>
<td>30</td>
</tr>
<tr>
<td>Figure 1.22</td>
<td>Structure of Corn protomer</td>
<td>31</td>
</tr>
<tr>
<td>Figure 1.23</td>
<td>Structural features of the Corn-DFHO complex</td>
<td>31</td>
</tr>
<tr>
<td>Figure 1.24</td>
<td>Architecture of the quasisymmetric DFHO binding site.</td>
<td>32</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Fluorophores used in binding and fluorescence enhancement experiments</td>
<td>36</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Binding affinity curves for fluorophore/RNA pairs in WB buffer</td>
<td>37</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>TO1-Biotin has high fluorescence enhancement relative to TO1 when bound to RNA Mango</td>
<td>39</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Temperature-dependent spectroscopy of TO1-Biotin complex with RNA Mango or RNA Spinach</td>
<td>39</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Binding affinity curves of RNA Mango + TO1-Biotin as a function of temperature</td>
<td>44</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>RNA Mango TO1-Biotin complex fluorescence and dissociation constant as a function of temperature</td>
<td>45</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>Fluorescence lifetimes RNA aptamer fluorophore complexes</td>
<td>48</td>
</tr>
</tbody>
</table>
Figure 2.8  Excitation and emission spectra for fluorophore/RNA pairs in WB buffer ........................................... 49
Figure 3.1  RNA Mango aptamers core sequences ................................................................. 58
Figure 3.2  Competitive selection of TO1-B binding variants using droplet-based microfluidics fluorescence screening ........................................................................... 59
Figure 3.3  Chemical structures of TO1-Biotin, TO3-Biotin and NMM ....................... 60
Figure 3.4  Effect of N-methyl mesoporphyrin IX (NMM) concentration on TO1-Biotin/Mango I fluorescence. ................................................................. 60
Figure 3.5  Green-fluorescence profile of the screenings performed in the presence of NMM (left-panel) or TO3-Biotin (right panel) ......................................................... 62
Figure 3.6  Resistance or selectivity of TO1-Biotin binding variants in presence of competitors after each round of screening ......................................................... 63
Figure 3.7  Analysis of the clones obtained at the end of the screening process performed in the presence of NMM ........................................................................... 64
Figure 3.8  Sequence analysis of the variants isolated across the rounds of screening performed in the presence of TO3-Biotin ................................................................. 66
Figure 3.9  Mutations and truncations of Mangos II, III, and IV ................................. 67
Figure 3.10  RNA Mango aptamers fluorophore binding and DMS protection ............ 69
Figure 3.11  Salt dependence of new Mango variants ............................................................. 71
Figure 3.12  Temperature-dependent fluorescence and UV absorbance spectroscopy of new Mango variants .................................................................................. 72
Figure 3.13  Circular dichroism spectra ................................................................................. 73
Figure 3.14  Excitation/emission of the new Mangos ............................................................. 74
Figure 3.15  RNA scaffold diagrams. ..................................................................................... 75
Figure 3.16  Effect of formaldehyde and pH on Mango fluorescence ............................. 77
Figure 3.17  Cellular imaging of Mango IV tagged RNAs .................................................... 79
Figure 3.18  Formation of 5S-Mango foci and their dynamics in live cells ....................... 80
Figure 3.19  Co-Immunofluorescence with direct RNA-Mango transfections .......... 81
Figure 3.20  2D immunostaining co-localization plots ......................................................... 82
Figure 3.21  2D nuclear co-localization plots ........................................................................ 84
Figure 3.22  Photostability of new Mango variants in in vitro and in fixed cells .......... 85
Figure 3.23  RNA Mango stoichiometry of individual foci ............................................... 86
Figure 3.24  Cellular imaging of genetically encoded Mango II tagged RNAs .......... 87
Figure 3.25  Imaging of genetically encoded aptamer tagged 5S rRNA ......................... 89
Figure 3.26  Summary of contrast attributes of RNA aptamer-based fluorescent tracking systems ........................................................................................................ 90
Figure 4.1  Vectors related to orientation factor $\kappa^2$ ...................................................... 107
Figure 4.2  Overall structure of RNA Mango in complex with TO1-Biotin .................. 110
Figure 4.3  The duplex–quadruplex junction of RNA Mango resembles a GAAA tetraloop .................................................................................................................. 112
Figure 4.4  Overall structure of the Mango III-TO1-Biotin complex ......................... 113
Figure 4.5  Comparison of connection between aptamer core and external helix between Mango I and Mango III. ................................................................. 115
Figure 4.6  Unimolecular Broccoli & Mango fusions .................................................. 115
Figure 4.7  FRET$_S$ as a function of duplex length for M1-nbp-Broc/TO3-Biotin+DFHBI (blue) and M3-nbp-Broc/TO3-Biotin+DFHBI constructs (red). .................. 116
Figure 4.8  FRET compatibility of Broccoli/DFHBI-1T and Mango III/YO3-Biotin. ...118
Figure 4.9  Workflow of determining biophysical FRET values for Broccoli and Mango III partners ................................................................. 119
Figure 4.10 Structural models of M3-nbp-Broc constructs ........................................... 120
Figure 4.11 Gaussian fits for M3-nbp-Broc constructs ................................................. 123
Figure 4.12 FRET efficiency as a function of duplex length ........................................ 124
Figure 4.13 Relationship between FRET, $k^2$ and $R_0$ .............................................. 126
Figure 5.1  Summary of selection conditions for 16 rounds of selection for a TO3-Biotin aptamer ................................................................. 134
Figure 5.2  TO3-Biotin selection, bead-binding diagnostic ........................................... 135
Figure 5.3  TO3-Biotin selection, fluorescence competition diagnostic ......................... 136
Figure 5.4  TO3:TO1 binding ratio for ligand preference. ............................................. 137
**List of Acronyms**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>Cbl</td>
<td>Cobalamin</td>
</tr>
<tr>
<td>DFHBI</td>
<td>3,5-difluoro-4-hydroxybenzylidene imidazolinone</td>
</tr>
<tr>
<td>DFHO</td>
<td>3,5-difluoro-4-hydroxybenzylidene imidazolinone-2-oxime</td>
</tr>
<tr>
<td>DIR</td>
<td>Dimethylindole Red</td>
</tr>
<tr>
<td>DMS</td>
<td>Dimethylsulfate</td>
</tr>
<tr>
<td>DN</td>
<td>Dinitroaniline</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>E</td>
<td>Fluorescent efficiency</td>
</tr>
<tr>
<td>E\textsubscript{FRET}</td>
<td>FRET efficiency</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-activated cell sorting</td>
</tr>
<tr>
<td>F\textsubscript{E}</td>
<td>Fluorescence enhancement</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence in situ hybridization</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy transfer</td>
</tr>
<tr>
<td>G4</td>
<td>G-quadruplex</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>NMM</td>
<td>N-methyl mesoporphyrin IX</td>
</tr>
<tr>
<td>PDB ID</td>
<td>Protein Data Bank identification number</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>r.m.s.</td>
<td>Root mean square</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>scaRNA</td>
<td>Small Cajal body-specific RNA</td>
</tr>
<tr>
<td>SELEX</td>
<td>Systematic evolution of ligands by exponential enrichment</td>
</tr>
<tr>
<td>SR</td>
<td>Sulforhodamine B</td>
</tr>
<tr>
<td>SRB-2</td>
<td>Sulforhodamine B binding aptamer</td>
</tr>
<tr>
<td>TAR</td>
<td>Trans-activating response element</td>
</tr>
<tr>
<td>ThT</td>
<td>Thioflavin T</td>
</tr>
<tr>
<td>TO</td>
<td>Thiazole orange</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
</tbody>
</table>
Chapter 1.

Introduction

*But Mr. McClure, what does DNA stand for?*

Billy

1.1. RNA history and their functions

Ribonucleic acid (RNA) is one of two types of nucleic acid molecules in the cell (the other being deoxyribonucleic acid or DNA). While DNA serves primarily as a genetic information carrier, RNA serve a multitude of functions inside the cell including acting as an intermediate genetic carrier between DNA and complex proteins, and catalyzing important activity such as translation and splicing.

The existence of nucleic acids have been known since its first isolation (originally termed ‘nuclein’) by Friedrich Miescher in 1871\(^1\). Since then, our understanding of RNA has grown with many milestones throughout history. Phoebus Levene proposed the now defunct tetranucleotide theory on the structure of RNA (called ‘yeast nucleic acids’ at the time), postulating that nucleic acids are tetraters consisting of A, C, G and U/T (his original hypothesis posed cyclic tetranucleotides, but was later revised to linear chains, though with an incorrect chemical linkage)\(^2\). Of course, in 1953, Watson and Crick, with the aid of Rosalind Franklin’s X-ray diffraction images of DNA, solved the correct structure of the DNA double helix\(^3\) and soon after, Alexander Rich determined that RNA could also form double helical structures with RNA or DNA as a hybrid\(^4\). In 1961, Brenner, Jacob, Meselson and others discovered messenger RNA, the genetic precursor for proteins that catalyze cellular processes\(^5\). From these historical discoveries amongst others, the Central Dogma of Molecular Biology was formed which posits that genetic information encoded in DNA is transferred to RNA, which is then transferred to proteins which ultimately perform the catalytic processes in a cell\(^6\). From here, our understanding of the role of RNA grew significantly including milestone discoveries such
as tRNA as the ‘adaptor’ molecule to transfer information from RNA to proteins\textsuperscript{7,8}, the catalytic role of RNA in the ribosome, and similarly, the spliceosome and mRNA splicing\textsuperscript{9,10}. Beyond modern biology, it has been postulated that RNA is the key component of prebiotic life and the precursor to modern cellular life, implicated by its ability to store genetic information like DNA while also being able to perform catalytic functions such as protein synthesis in the ribosome.\textsuperscript{11,12} Hence, the diverse roles and functions of RNA makes it a highly fascinating biomolecule to study that has implications on understanding prebiotic evolution, modern biological function and medicine.

1.2. RNA structure

Ribonucleic acids (RNA) are polymeric chains of nucleotides. A nucleotide consists of a nitrogenous base, a ribose sugar (this differs from the deoxyribose of DNA) and a phosphate backbone. Polymers are formed by phosphodiester linkages at the 5’ and 3’ carbons of the ribose. The base is attached to the sugar via a glycosidic bond between the ribose 1’ carbon and the N9 of purines or N1 of pyrimidines (Figure 1.1). Additionally, the base can be in one of two conformations relative to the sugar: anti or syn, which denotes whether the bulky groups in the base are facing towards, or away from the sugar-phosphate. The chemical nature of RNA (or any type of nucleic acid) strands allow formation of many complex secondary and tertiary structures that are functional. The formation of these structures is driven by intramolecular interactions between nucleotides.

The furanose ring of ribose is typically not in a planar conformation due to torsional strain. Rather, the sugar can adopt various more energetically favourable pucker conformations (Figure 1.2). The pucker is defined by either C3’ or C2’ in ‘endo’ or ‘exo’, referring to its position relative to the C5’ bond (endo, same side or exo, opposing side).
Figure 1.1  Chemical structure of RNA
A four-base chain of RNA is shown. Chain direction is 5′ to 3′ from top to bottom. Bases are shown in blue with numbering on the first purine (two membered rings, Guanine, abbreviated as ‘G’ and Adenine, abbreviated as ‘A’) and the first pyrimidine (Cytosine, abbreviated as ‘C’ and Uracil, abbreviated as ‘U’). Ribose sugar are shown in red with carbon numbering on the first ribose. The bases on the left are in anti conformation. Examples on the right show syn conformations of adenine and uracil.
Figure 1.2 Sugar pucker conformations.
Potential sugar pucker conformations are shown, viewed planar from the 3'2' side of the sugar. Other pucker conformations are also possible, only a subset is shown.

1.2.1. Basic interactions

The overall tertiary structure of a nucleic acid motif can be broken down into basic interactions. Nucleobase interactions are a key component of nucleic acid secondary structure. The basic level of these interactions is between two bases and is characterized by three faces of each base: The Watson-Crick face, the Hoogsteen face, and the sugar edge. Additionally, each pair of bases can interact in either the cis or trans orientation, referring to the relative orientations of the glycosidic bond on either base. Thus, if any face can interact with any other and can form in either cis or trans orientations, this totals 12 basic interactions, together with phosphodiester and sugar backbone hydrogen bonding with electronegative groups consist the overall structure of a nucleic acid motif. The base interactions were described in great detail by Leontis and Westhof who established the eponymously named Leontis-Westhof nomenclature for describing base interactions (Figure 1.3)\(^{13}\).
1.2.2. Higher order structures

The culmination of interactions between nucleobases within a nucleic acid strand lead to formation of higher order structures which are functional. The best examples of this are the ribosome and spliceosome which are RNA-protein complexes that accomplish cellular processes (the ribosome performs translation while the spliceosome performs mRNA splicing), catalyzed by their structured RNA component\textsuperscript{14,15}. These complex structures are composed of simpler secondary and tertiary motifs. Many
classes of these motifs exist. The following section describes three common motifs that are found in RNA structures.

**Duplexes**

The simplest form of nucleic acid structures is the double-stranded helix (dsDNA or dsRNA), also called duplexes. Duplexes can take on various geometries that is based on their compositions. These structures are well understood with established terminology that define aspects of the structure. The rise is the distance between each base pair (bp) stack along the helical axis while twist is the rotation of the base-pairs about this axis. Helices can be either right-handed, or left-handed in orientation, defined by the direction of the helical twist. One turn is the length of nucleic acid at which it has completed one complete 360° twist. The length of the turn is constrained by the sugar puckers conformation in the helix. The structure of the double helix possesses two types of gaps between each strand called the major and minor grooves, which relate to the size of groove between base pair stacks. The major groove is the wider, deeper groove while the minor groove is narrower and thinner.

Helices can adopt different conformations, the most common being A-form and B-form. DNA duplexes inside the cells are typically B-form which is a right-handed helix and has ~10.5 bp/turn with a helical rise of 3.4 Å per base pair and helical twist of 36° per base pair (Figure 1.4). In B-form helices the sugar puckers are in the 2’ endo conformation. In RNA, it is not favourable to form a B-form helix, due to the presence of the 2’ OH which sterically prevents the 2’ endo conformation. Rather, RNA is commonly in the A-form. This type of duplex is also right-handed, and has ~11 bp/turn with a rise of 2.56 Å and twist of 33° (Figure 1.4). A-form duplexes have 3’ endo sugar puckers which contributes to the shorter rise and can be adopted by both DNA and RNA.


Figure 1.4  B-form vs A-form double helix
Examples of B-form and A-form helices are shown. The B-form DNA is from a crystal structure of a synthetic dodecamer (PDB ID: 1BNA)\textsuperscript{16}. The A-form RNA is an NMR structure of a synthetic dsRNA (PDB ID: 2KYD)\textsuperscript{17}. Carbons on one chain are in green while the opposite chain is in cyan. Phosphates are coloured in orange, oxygens red and nitrogens blue.

Stem-loops, tetraloops

Stem-loops are RNA strands closed by some form of a loop\textsuperscript{18} which fold the strands back into a hairpin-like shape, facilitating intramolecular duplex formation. In many biological settings, tetraloops serve this function. Tetraloops are loop structures comprised of 4 bases, and are often found in large RNA structures such as the ribosome\textsuperscript{18}. These tetraloops can also serve as recognition motifs for various interactions. The three most common tetraloop conservations based on sequence and topology are GNRA, UNCG, and CUUG, where N is any base and R is any purine. Though general structure is conserved within tetraloop classes, sequence variations can contribute to deviations in structure.

GNRA tetraloops are the most frequently occurring tetraloop motifs in ribosomal RNA due to their thermodynamic stability\textsuperscript{19,20}. Overall, the first G of the tetraloop is
situated on one side of the connected helix, while the NRA residues are stacked atop each other on the other side (Figure 1.5). The G, R and A participate in an extensive hydrogen bond network\textsuperscript{21}, including a non-canonical G:A base pair, which together with the stacking interactions help stabilize the structure. The N-peptide-boxB is an example of a protein that recognizes the GNRA tetraloop\textsuperscript{22}.

![Figure 1.5  A GNRA tetraloop](image)

An NMR solution structure of a GNRA tetraloop from a motif in group I and II self-splicing introns. The tetraloop here is GAAA (blue) with a stem (gray) connected. PDB ID: 4FNJ\textsuperscript{23}.

UNCG are a second class of tetraloops commonly found in biological RNAs. Like GNRA, they are very thermostable\textsuperscript{24}. The U and G residues, closest to the helix form a U:G wobble base-pair while the C residue is stacked atop the G. In this tetraloop, the N residue is looped out and exposed to solvent (Figure 1.6)\textsuperscript{25–27}. One example of the UNCG tetraloop in biology is a recognition motif for the retroviral Rous sarcoma virus nucleocapsid protein\textsuperscript{28}.
**Figure 1.6  A UNCG tetraloop**
NMR solution structure of a UNCG tetraloop from a 14-mer hairpin RNA. UNCG tetraloop residues are shown in blue with the helix in gray. PDB ID: 2KOC.²⁷

The CUUG tetraloop consists of a canonical Watson-Crick C:G interaction of the two bases closes to the stem. The U² residue stacks atop the G, while the U¹, unlike the GNRA or UNCG tetraloops, folds back into the minor groove of the helix (Figure 1.7).²⁹ The CUUG tetraloop is found in loop 83 of the *E.coli* 16S ribosomal RNA.³⁰

**Figure 1.7  A CUUG tetraloop**
NMR solution structure of a CUUG tetraloop from loop 83 of bacterial 16S rRNA. CUUG tetraloop residues shown in blue with helix in gray. PDB ID: 1RNG.²⁹
**G quadruplexes**

G-quartets are formed from four guanines cyclically bonded by Watson-Crick/Hoogsteen interactions (Figure 1.8A). G-quadruplexes (G4) are motifs that consist of at least two layers of G-quartets. G4s fold most favourably in the presence of potassium as well as other monovalent cations between G-quartet stacks and generally form from strands with consecutive stretches of G residues. The strand orientation of G4s can be either in parallel, anti-parallel or a mixture of both (3+1) (Figure 1.8B). RNA G4s are not typically found in anti-parallel conformation, though anti parallel RNA quadruplexes have recently been discovered. Further, the flat, planar structure of the G-quartet provides an accessible hydrophobic surface for ligand binding. As such, G4s are known to bind a host of ligands such as fluorescent molecules for RNA imaging, and heme as catalysts for oxygen transfer reactions.

In biological settings, G4s are prevalent in telomeres. They have also been shown to be present in biological RNAs and are important regulatory components. The first example of a regulatory RNA G4 was found on the 5′ UTR of N-ras proto-oncogene, which its presence was demonstrated to inhibit translation by ~80%. G4s have also been reported to be involved in modulation of splicing and alternative splicing. Due to their prominent roles in regulation, and their ability to bind ligands efficiently, G4s have become a drug target for medical applications.

There has been recent debate on the presence of RNA G4s in cells. Though many studies have proven that RNAs from cells can form G4s in vitro, a recent study from the Bartel lab demonstrated using an in vivo dimethylsulfate (DMS) probing assay that G4 content (by the metric of their DMS experiment) in eukaryotic cells were significantly lower relative to in vitro at least in the steady-state. It is hypothesized that G4s are perhaps transient in cells and would not be detected by the in vivo DMS experiment or that the G4s are actively binding regulatory proteins which would also mask them from chemical probing. Also, one counter argument is that RNA G4-binding probes have been visualized in cells, thus G4s must exist in the cell in some form. It remains to be seen how our understanding of the dynamics of cellular G4s will grow over time.
Figure 1.8  G4 structure and topology  
(A) A G-quartet layer is shown. $M^+$ represents a cation that stabilize two layers of G-quartets. Squiggly lines denote glycosidic bonds to the sugar.  
(B) Potential strand orientations of G4 structure.
1.3. *In vitro* selection and SELEX methods

*In vitro* selection and systematic evolution of ligands by exponential enrichment (SELEX) are experimental methods by which functional nucleic acid sequences (DNA or RNA) is isolated from a large pool to identify a sequence with a desired function. Simply, a pool of DNA or RNA of high diversity (typically >10^{14} unique sequences) is subject to selective conditions to promote a desired function or interaction, low activity sequences are removed, and remaining sequences are regenerated before proceeding to another round of selection. This process is cyclical, thus high activity sequences are exponentially enriched over each round until the best sequence or family of sequences dominates the pool (Figure 1.9).

![Generalized scheme of RNA in vitro selection procedure](image)

**Figure 1.9** Generalized scheme of RNA *in vitro* selection procedure

An example of an RNA *in vitro* selection to obtain an aptamer for a hypothetical ligand (star). 1) A high diversity DNA library is transcribed into RNA. 2) The RNA library is folded and bound to target ligand. 3) Wash steps by various procedures removes non-binders from the pool, and retains strong binders. 4) Strong binders are regenerated to DNA by RT-PCR. 5a) Newly regenerated DNA pool is transcribed again for another round of selection, or 5b) Sequences are analyzed for $K_D$, minimal motif, etc.
The first demonstration of *in vitro* selection/SELEX were by three independent groups in 1990. Tuerk and Gold, and Ellington and Szostak selected the first *in vitro* selected RNA aptamers for bacteriophage T4 DNA polymerase and small molecule dyes respectively, while Robertson and Joyce selected improved mutants of the Tetrahymena self-splicing group I intron ribozyme. These ground-breaking papers led to the growth of SELEX as a powerful technique to isolate functional nucleic acid sequences.

Classical selection methods generally utilize column-based methods where ligands are attached to a stationary support. Since then, selection methodologies have evolved and have been used to identify the vast number of functions of nucleic acid molecules. These techniques have been adapted into applications such as medicine, nanotechnology and discovery of novel biological processes. Below, I discuss several modifications to the classic selection approach that have been developed over the years.

The column-based methods are limited by the need to provide a linkage to a stationary support, thus the selective target is not the desired ligand, but the ligand with a covalently attached linker. Capillary electrophoresis (CE) provides a robust alternative. CE is an electrophoretic method of sample separation by size and charge. Thus, this separation technique is adaptable as a SELEX methodology. First described in a selection for IgG antibodies, capillary electrophoresis SELEX (CE-SELEX) utilizes electrophoretic mobility shifts to separate ligand-bound sequences from unbound which confers the advantage of not requiring linker conjugation of the ligand and can be done in fewer selection cycles due to the efficiency of separation. Further, it can also be interfaced with fraction collectors for further automation of the selection process. CE-SELEX has been used to select various aptamers for ligands such as proteins and porphyrins.

Many biological aptamers and ribozymes likely exist in the genome. Thus, a modification to the classic SELEX method which uses genomic material, rather than an artificial nucleic acid pool was introduced. This method enables the discovery of functional RNAs that are biologically relevant. The first use of genomic SELEX identified HDV-like self-cleaving ribozyme in the human CPEB3 gene, as well as protein-binding sequences.
Computational methods could also be applied in SELEX methodologies. Hence, in silico SELEX was first described for a number of small biomolecules\textsuperscript{64}. Here, an RNA pool of large diversity (but relatively low compared to normal SELEX, $10^8$ vs $10^{15}$) is analyzed by secondary structure analysis and selecting sequences that had minimal ‘simple’ structural motifs such as stem loops, while also selecting structures with minimal free energy once folded by a thermodynamic algorithm. The remaining sequences are then subject to 3D structure and molecular docking simulations to further reduce diversity. By this method, the group was able to reduce the $10^8$ pool diversity by $\sim 5$ orders of magnitude to $10^3$. From here, microarray methods or other low throughput selection methods can be applied to identify final RNA sequences. This method can greatly reduce SELEX time, but has obvious limitations. First, the throughput of the method is limited by the powers of computation which would not be expected to be able to reach pool diversities of $10^{15}$ easily. Secondly, the secondary structure thermodynamic models are often poor at identifying 3D motifs, thus may exclude valuable sequences in the initial screening.

Nucleic acid aptamers have a great number of \textit{in vivo} uses. As such, it would be advantageous to perform SELEX for these aptamers in environments that mimic the cell – or better yet, SELEX could be performed in cells themselves. \textit{In vivo} SELEX is a methodology that uses in cell expression of clones for the SELEX process. The first example of this, though with a limited pool diversity of just 64 clones, selected for mutants of the HIV-1 trans-activating response element (TAR) that promoted viral replication\textsuperscript{65}. In a more recent example, one study developed a method to select RNA aptamers that were able to penetrate the blood-brain barrier in live animals\textsuperscript{66}. Additionally, \textit{in vivo} SELEX can be interfaced with fluorescence-activated cell sorters (FACS) for fluorescence-based sorting and have been used to select DNA aptamers of CD19\textsuperscript{7} Burkitt's lymphoma cells as well as isolate variants of fluorogenic dye-binding aptamers\textsuperscript{67,68}. However, \textit{in vivo} SELEX, while it provides a technique which ensures that sequences will function in the cellular environment, is limited by the pool diversity that can be screened due to limitations in the size of the cell population as well as the vector library.

The advent of next generation sequencing (NGS) furthered the analysis of SELEX procedures\textsuperscript{69}. By combining NGS with SELEX, phylogenetic information can be gathered, revealing important information on library diversity throughout each selection.
round. This was first demonstrated by observing the genotype frequencies of a selection on a mutagenized library of the class II ligase\textsuperscript{70}. Since then, this technique has been applied to analysis of splicing protein-binding RNAs\textsuperscript{71}, the thrombin aptamer\textsuperscript{72}, and the Diels-Alderase ribozyme\textsuperscript{73} to name a few, which revealed how the most active sequences from a library are able to dominate the pool across multiple rounds of selection. To date, the throughput of NGS is not up to par with an entire SELEX procedure (whereas NGS can sequence millions of sequences, typical SELEX starting pools are \(\sim 10^{15}\) diversity), thus it is not possible to explore sequence information until the library has been significantly narrowed down. It remains to be seen whether sequencing technology or other workarounds will be able to capture this missing information.

1.4. **Fluorescent systems for cellular RNA imaging**

The biological importance of RNA localization in cells and the need for robust, high contrast tools to image these RNAs led to the development of various fluorescence systems to image RNA molecules.

1.4.1. **Fluorescence in situ hybridization (FISH)**

Methods to image RNA in non-living cells have been well developed. *In situ* hybridization is a technique which utilizes labelled nucleic acid probes that are complementary to a target biological sequence (RNA or DNA) such that in a tissue sample, or fixed cell samples, they will bind and generate a localized signal for the target molecule\textsuperscript{74}. A variety of labels can be used such as radioactivity, chemiluminescence and modern systems use fluorescence (FISH)\textsuperscript{75}. Fluorescence-based methods provide numerous advantages such as being multiplexed using sequence-specific probes and fluorophores in different emission channels, can be high resolution, are much less harmful compared to methods such as radioactivity, has much faster detection methods and is quantitative in its analysis.

The FISH technique has been adapted for numerous applications. Multicolour labels allowed multiplexing imaging by FISH\textsuperscript{76}, which led to the milestone development of simultaneous imaging of every human chromosome by computational analysis of five-colour labels\textsuperscript{77}. FISH also has clinical applications in detecting chromosome abnormalities by pre-natal screening\textsuperscript{78,79}, blood disorders\textsuperscript{80} and cancers\textsuperscript{81}.
FISH techniques have advanced over time. Super-resolution imaging allows detection of molecules beyond the resolution limit of standard light microscopy and the first demonstration of this was done using FISH probes. FISH has been demonstrated to allow single-molecule imaging of RNA transcripts. One method of doing so is to coat a target RNA with multiple probes, which increases the local concentration of the fluorescent probes on the target as well as suppress probe signals that happen to bind off-targets. Alternatively, FISH signals can be amplified using secondary antibodies, or enzymatic activation of pro-fluorescent compounds. Though FISH is an effective tool for fixed cells, live cell RNA imaging is more challenging.

1.4.2. Fluorescence-based aptamers

The development of fluorescent molecule-binding aptamers for both live and fixed cell RNA imaging has been a rapidly growing field. These are small aptamers that are genetically encodable, and bind a range of ligands including proteins and small molecules. For each system, several criteria should be considered to evaluate its efficacy. The fluorescent entity must be non-toxic and be able to enter the cell. In addition, two biophysical properties govern the overall contrast. These are the fluorescent enhancement of the fluorophore ($F_{E}$, Equation 2.3) and the binding affinity ($K_D$) of the complex. A strong fluorescent enhancement allows the fluorescent signal to overcome non-specific noise from unbound fluorophore. Likewise, low $K_D$ is necessary to ensure that the RNA-fluorophore binding is saturated at very low concentrations of both RNA and dye. Together, fluorescence enhancement and binding affinity define a fluorescent efficiency ($E$, Equation 2.4) for the complex that should be maximized to optimize the overall signal-to-noise of a system. Below, I discuss the current fluorescent aptamer systems.

**MS2/PP7 and other aptamer-protein systems**

The current most widely used system for fluorescence RNA imaging in live cells, is the MS2 or PP7 systems. Both MS2 and PP7 are stem-loop RNA aptamers that bind, with high affinity ($K_D$ in the range of $\sim 10^{-9} \text{ M}$), viral coat proteins. These aptamers were adapted for fluorescence imaging of RNAs by tagging the coat protein with fluorescent proteins and tagging the RNA with the MS2 stem-loop. This approach was first used in live cells in 1998 by Bertrand et al. Likewise, the PP7 acts similarly to MS2, but is
orthogonal in its binding (the PP7 stem-loop will not bind the MS2 coat protein, and vice versa). With this system, multi-colour fluorescence imaging of RNA polymerase II kinetics on a single gene has been shown using the MS2/PP7 in tandem.\textsuperscript{88}

Aside from MS2 and PP7, a series of aptamer-protein interactions have been implemented for fluorescence RNA imaging using similar principles of fluorescently-tagged proteins interacting with an RNA aptamer. Examples include the λN-GFP\textsuperscript{89}, BglG\textsuperscript{90} and U1A\textsuperscript{91,92} systems. Many of these systems are also orthogonal and provide alternatives for multi-colour imaging.

They key disadvantage of this type of aptamer-protein systems is the lack of a robust ‘off’-state for the fluorescent proteins ($F_E = 1$) – thus unbound molecules contribute to a relatively large background. This can be off-set by incorporating multiple RNA aptamer tags into a given RNA sequence. Alternatively, protein complementation using two orthogonal aptamer-protein interactions have been used to reduce this background\textsuperscript{93–96}. However, both solutions create a secondary problem of incorporating large, unnatural sequences into a biological RNA sequence (eg. a 20-nt MS2 aptamer of 24X repeat is a 480-nt insert).

**Fluorogenic dye aptamers**

In vitro selection methods have been applied in the development of aptamers that bind fluorogenic dyes. These types of system confer a significant advantage over protein-aptamer systems due to fluorescence enhancement of the dye upon binding to the aptamer, thus greatly improving signal-to-noise without inserting large repeats. Here, I discuss currently available fluorogenic aptamer systems. Structural details of these aptamers will be discussed in section 1.5.

The first RNA aptamer selected for a fluorogenic small molecule was the malachite green aptamer and was originally developed as a ‘bulls-eye’ for laser-mediated cleavage of mRNA\textsuperscript{97}. Interestingly, the fluorescence enhancement properties of this aptamer was not reported until 4 years after its first publication\textsuperscript{98}. The aptamer enhanced fluorescence of the normally dim malachite green as well as other related triphenylmethane compounds (Figure 1.10) by up to ~2360-fold with a binding affinities as low as 40 nM (Table 1.1), while rejecting sulfonated variants of these compounds\textsuperscript{98}. This property relies on the principle that molecular constraints of the fluorophore prevent
intramolecular motion, thus upon excitation, energy is released via a fluorescence pathway. Since this original development, numerous aptamers have been developed for fluorescence RNA imaging.

![Figure 1.10](image_url)

**Figure 1.10 Fluorogenic dyes for the malachite green aptamer**
The malachite green aptamer binds triphenylmethane dyes. Shown are two examples. Malachite green was the original ligand used in the selection of this aptamer.

In 2008, two fluorogenic dye-binding aptamers were reported. One aptamer was selected for the fluorogenic cyanine dye, dimethyl indole red (DIR, Figure 1.11)\(^99\). Although, these cyanine dyes often intercalate non-specifically into nucleic acids, two features of DIR prevent this non-specific interaction: The bulky dimethylindole group prevents stacking interactions with nucleic acid base pairs, and a propylsulfonate substituent possesses a negative charge that results in electrostatic repulsion from nucleic acid phosphate backbones\(^99\). Yet, an RNA aptamer was selected to bind to DIR with \(\sim 100\) nM \(K_D\) and a fluorescence enhancement of up to 20-fold (Table 1.1). A second aptamer was selected for a Hoechst dye (Figure 1.11)\(^100\). Typically, these dyes are used as a nuclear stain due their tendency to intercalate into AT-rich DNA\(^101\). In order to suppress intercalation, two ortho tert-butyl groups were added to the alkylated phenyl ring. Previous to this, a DNA aptamer was selected to bind the modified Hoechst dye\(^102\). The RNA aptamer selected for this Hoechst derivative enhanced the dye’s fluorescence by up to \(\sim 30\) fold and a \(K_D\) of 35 nM (Table 1.1). In both cases, the fluorescence enhancements are relatively low (on the order of 10-fold). Subsequent aptamers greatly improve upon this property.
Jaffrey and coworkers developed an aptamer that significantly improves upon the fluorescence enhancement of a fluorogenic dye when bound under in vivo mimicking conditions (i.e. high potassium, low magnesium, neutral pH). Spinach is an aptamer that binds a green fluorescent protein (GFP) chromophore-based compound, 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI, Figure 1.12) as well as other related compounds. In fact, Spinach possesses a level of ligand promiscuity, being able to bind many fluorescent dyes indiscriminately, which will be discussed in Chapter 2. Spinach enhances the fluorescence of DFHBI by ~2000 fold and binds with relatively weak affinity of ~500 nM and exhibits fluorescence in the GFP channel (emission maximum = ~500 nm) (Table 1.1). Further, iterations of the original Spinach aptamer improved upon a number of its deficiencies. Deliberate mutations of Spinach stabilized its stem loops yielding Spinach2, which exhibited improved thermal stability. A FACS-based selection methodology generated Broccoli, which is yet more thermal stable and ~1.5 fold improved binding affinity relative to Spinach2. Further, an independent group developed iSpinach using μIVC which adapted the aptamer to in vitro applications by minimizing potassium dependence. Most recently, a series of aptamers was selected.
to bind the chromophore of the DsRed family of fluorescent proteins, 3,5-difluoro-4-
hydroxybenzylidene imidazolinone-2-oxime (DFHO, Figure 1.12)\textsuperscript{107}; thus these aptamers
are up to 80 nm red-shifted from the Spinach family. Together, these aptamers comprise
the current fluorescent protein chromophore-binding family of fluorescent aptamers.

![Chemical structures of DFHBI, DFHBI-1T, DFHBI-2T, and DFHO](image)

**Figure 1.12 Spinach and corn fluorophores based on fluorescent protein chromophores**
The target fluorophores of Spinach and variants, and Corn aptamers are shown. DFHBI, DFHBI-
1T and DFHBI-2T are modified from the GFP chromophore while DFHO is modified from the
chromophore of the DsRed fluorescent protein family. Broccoli is known to bind all four of these
fluorophores while Corn is known to bind DFHO and is untested with the other three\textsuperscript{68,103,107}.

Independently, the Mango family of aptamers was developed in the Unrau lab. These aptamers bind to thiazole orange-based (TO) fluoromodules (Figure 1.13) with nanomolar affinity and fluorescence enhancement of up to ~4,000-fold (Table 1.1)\textsuperscript{54,108}. Since the publication of the original Mango aptamer in 2014, improved variants have been developed or are ongoing. The details of these aptamers are discussed in Chapters 3 and 5.
Thiazole orange-based fluorophores for Mango aptamers

The TO dyes are modified on the nitrogen of the benzothiazole ring to inhibit non-specific nucleic acid intercalation.

Quencher-based fluorogenic dye aptamers

An alternative strategy to generate fluorogenic compounds is by the fluorophore conjugated to a quencher molecule, which by contact quenching ‘dims’ the fluorophore. By selecting aptamers that sequester the quencher from the conjugated fluorophore, fluorescence emission is enhanced.

Jäschke and coworkers selected an aptamer for sulforhodanmine B (the dye abbreviated as SR and the aptamer called SRB-2)\textsuperscript{109}. When conjugated to a series of quencher molecules, contact quenching occurs to generate the fluorogenic probe (Figure 1.14). Binding of the aptamer to the fluorescent moiety displaced the quencher and enhanced fluorescence by up to ~100 fold but has relatively weak affinity of 1.4 μM (Table 1.1). The reciprocal case was also demonstrated where an aptamer for the quencher molecule dinitroaniline (DN) was selected rather than for the fluorophore. DN is conjugated to a series of rhodamine dyes which upon binding to the aptamer, fluorescence is enhanced in a similar mechanism as the previous case. This yielded fluorescence enhancement of up to ~37 fold and had much improved affinities as low as 350 nM\textsuperscript{110}. One advantage to quencher-based systems is it allows greater flexibility of fluorophore choice, including a large spectral range of dyes (Figure 1.14). Together, these two systems can be used in conjunction as orthogonal labelling systems and has been demonstrated in bacterial imaging\textsuperscript{110}.
Figure 1.14 **Fluorescent compounds for the SRB2 and DN aptamers.** Fluorescent moieties are in black. The SRB2 aptamer binds the fluorescent moiety, while the DN aptamer binds the quencher. Compounds are conjugated to quencher DN groups (blue) via polyethylene glycol (PEG) repeats (red) that can be $n = 0,1,2,3$. An example of Sulforhodamine-PEG$_1$-Dinitroaniline is shown. R groups are DN-PEG moieties bonded by via an amine linkage.

Along this trend of fluorophore-quencher systems, Braselmann et al. developed the ‘Riboglow’ system which uses cobalamin as the quencher, and adapts the cobalamin riboswitch as the aptamer tag. Several variants of the cobalamin riboswitch were used along with a series of cobalamin-dye molecules (Figure 1.15). Cobalamin effectively quenches fluorophores with a range of spectral channels, but is most efficient for green fluorophores such as FAM and ATTO 488 with up to 20-fold reduction in fluorescence while it suffers with red dyes such as ATTO 633 and Cy5 with quenching efficiencies as low as 3-fold. Given that dequenching via the aptamer is not 100% efficient, in the best case, fluorescence enhancement will be below 20-fold. Practically, the fluorescence enhancement of these aptamers range from 1.3-7.3 fold$^{111}$. Thus, the aptamer is greatly limited by background fluorescence *in vitro*. However, the system has shown effectiveness in practice, with comparable contrast in mammalian cells relative to other systems such as MS2.
**Figure 1.15** Cobalamin coupling to various fluorophores to form quencher-dye molecules


<table>
<thead>
<tr>
<th>Table 1.1</th>
<th>Comprehensive list of fluorescent RNA aptamer systems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aptamer</strong></td>
<td><strong>Dye</strong></td>
</tr>
<tr>
<td>Malachite Green</td>
<td>MG</td>
</tr>
<tr>
<td></td>
<td>IMG</td>
</tr>
<tr>
<td>Dimethylindole Red Hoechst</td>
<td>DIR</td>
</tr>
<tr>
<td></td>
<td>Hoechst-1b</td>
</tr>
<tr>
<td>Spinach</td>
<td>DFHBI</td>
</tr>
<tr>
<td></td>
<td>420 ± 40</td>
</tr>
<tr>
<td></td>
<td>TO1-Biotín</td>
</tr>
<tr>
<td>Spinach2</td>
<td>DFHBI</td>
</tr>
<tr>
<td></td>
<td>530</td>
</tr>
<tr>
<td>Sample Type</td>
<td>DFHBI-1T</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
</tr>
<tr>
<td>Broccoli iSpinach Corn</td>
<td>560</td>
</tr>
<tr>
<td>Orange Broccoli Red-Broccoli</td>
<td>830</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>29140</td>
</tr>
<tr>
<td></td>
<td>482/505</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample Type</th>
<th>TO1-Biotin</th>
<th>TO3-Biotin</th>
<th>TO3-Biotin</th>
<th>TO1-Biotin</th>
<th>TO3-Biotin</th>
<th>TO3-Biotin</th>
<th>TO3-Biotin</th>
<th>TO1-Biotin</th>
<th>TO3-Biotin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mango I</td>
<td>3.2</td>
<td>2.2 ± 0.3</td>
<td>5.1 ± 0.3</td>
<td>5.1 ± 0.3</td>
<td>3.2</td>
<td>2.2 ± 0.3</td>
<td>5.1 ± 0.3</td>
<td>5.1 ± 0.3</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>1100</td>
<td>1100</td>
<td>N.D.</td>
<td>N.D.</td>
<td>1100</td>
<td>1100</td>
<td>N.D.</td>
<td>N.D.</td>
<td>1100</td>
</tr>
<tr>
<td></td>
<td>343</td>
<td>500</td>
<td>637/658</td>
<td>637/658</td>
<td>343</td>
<td>500</td>
<td>637/658</td>
<td>637/658</td>
<td>343</td>
</tr>
<tr>
<td></td>
<td>10850</td>
<td>10850</td>
<td>10850</td>
<td>10850</td>
<td>10850</td>
<td>10850</td>
<td>10850</td>
<td>10850</td>
<td>10850</td>
</tr>
<tr>
<td>Mango II</td>
<td>0.7 ± 0.3</td>
<td>1.8 ± 0.1</td>
<td>5.6 ± 0.2</td>
<td>15.0 ± 1.3</td>
<td>0.7 ± 0.3</td>
<td>1.8 ± 0.1</td>
<td>5.6 ± 0.2</td>
<td>15.0 ± 1.3</td>
<td>0.7 ± 0.3</td>
</tr>
<tr>
<td></td>
<td>1650</td>
<td>1150</td>
<td>4400</td>
<td>N.D.</td>
<td>1650</td>
<td>1150</td>
<td>4400</td>
<td>N.D.</td>
<td>1650</td>
</tr>
<tr>
<td></td>
<td>2360</td>
<td>2360</td>
<td>790</td>
<td>N.D.</td>
<td>2360</td>
<td>2360</td>
<td>790</td>
<td>N.D.</td>
<td>2360</td>
</tr>
<tr>
<td></td>
<td>14100</td>
<td>14100</td>
<td>43000</td>
<td>43000</td>
<td>14100</td>
<td>14100</td>
<td>43000</td>
<td>43000</td>
<td>14100</td>
</tr>
<tr>
<td>Mango III</td>
<td>11.1 ± 0.8</td>
<td>10.4 ± 0.1</td>
<td>5.6 ± 0.2</td>
<td>15.0 ± 1.3</td>
<td>11.1 ± 0.8</td>
<td>10.4 ± 0.1</td>
<td>5.6 ± 0.2</td>
<td>15.0 ± 1.3</td>
<td>11.1 ± 0.8</td>
</tr>
<tr>
<td></td>
<td>3300</td>
<td>300</td>
<td>4400</td>
<td>N.D.</td>
<td>3300</td>
<td>300</td>
<td>4400</td>
<td>N.D.</td>
<td>3300</td>
</tr>
<tr>
<td></td>
<td>32000</td>
<td>32000</td>
<td>7900</td>
<td>7900</td>
<td>32000</td>
<td>32000</td>
<td>7900</td>
<td>7900</td>
<td>32000</td>
</tr>
<tr>
<td>Mango IV</td>
<td>11.1 ± 0.8</td>
<td>10.4 ± 0.1</td>
<td>5.6 ± 0.2</td>
<td>15.0 ± 1.3</td>
<td>11.1 ± 0.8</td>
<td>10.4 ± 0.1</td>
<td>5.6 ± 0.2</td>
<td>15.0 ± 1.3</td>
<td>11.1 ± 0.8</td>
</tr>
<tr>
<td></td>
<td>3300</td>
<td>300</td>
<td>4400</td>
<td>N.D.</td>
<td>3300</td>
<td>300</td>
<td>4400</td>
<td>N.D.</td>
<td>3300</td>
</tr>
<tr>
<td></td>
<td>32000</td>
<td>32000</td>
<td>7900</td>
<td>7900</td>
<td>32000</td>
<td>32000</td>
<td>7900</td>
<td>7900</td>
<td>32000</td>
</tr>
</tbody>
</table>

| Reference | 68,103 | 103 | 68 | 106 | 107 | 107 |

- There are 3 cobalamin aptamers that bind a large series of cobalamin-dye molecules. For comparison purposes, the aptamer/cobalamin-dye interaction of best $K_D$ and FE is listed.
- Reference refers to the citation that the values were obtained from. Some values were determined in more than one reference.
- 'N.D.' – Not determined.
- '>>' – $K_D$ was greater than detection limit of assay.
1.5. **Structural studies of fluorogenic dye-binding RNA aptamers**

The growing list of fluorogenic dye-binding RNA aptamers generated a flux of structural studies of these aptamers. Empirical understanding suggests that bright fluorescent compounds are generally in planar conformations, and rigidified to prevent non-radiative decay. Thus, it is interesting that aptamers that bind small molecule fluorophores have planar surfaces for ligand binding, and have mechanisms for rigidification. Many of the brightest fluorescent dye binders feature either G4 or other structures with a planar surface for this purpose. Below, I discuss crystallographic studies of several fluorogenic dye-binding aptamers.

### 1.5.1. Malachite green aptamer

The malachite green aptamer consists of two helices that flank an asymmetric loop which forms the binding pocket (Figure 1.16A). One of the two helices are closed by a GNRA tetraloop. The asymmetric loop consists of quadruplex stack, a duplex stack and two triplex stacks. The quadruplex is unique as it is comprised of two overlapping triplexes (A31:C7:G29 and G24:C7:G29), creating a large ~230 Å² surface (Figure 1.16B). The ligand (the crystal structure is bound to an analog, tetramethyl rosamine) situates planar between this quadruplex and a standard Watson-Crick duplex. The phenyl ring of the ligand is twisted to nearly perpendicular relative to the rest of the structure. This phenyl is closed by two adenine residues that are stacked on top of each other to pack the ligand. Together, this asymmetric loop comprises the binding pocket that rigidifies and enhances fluorescence of the ligand.
Figure 1.16  Structure of the malachite green aptamer.
(A) The global structure of the malachite green aptamer is shown. GNRA tetraloop is shown in blue. The core residues are shown in red (non-canonical quadruplex), orange (stacking duplex), and yellow (double triplex layers). The tetramethyl rosamine ligand is shown in green. (B) Zoomed in view of the core binding pocket. Colour schemes are as in (A). Gray residues are the double adenine that stack on the phenyl ring of the ligand. PDB ID: 1F1T.114

1.5.2. Spinach aptamer

The structure of the Spinach aptamer consists of two helices that flank a binding pocket. This pocket consists of a non-canonical quadruplex as well as a Hoogsteen triplex cap which together sandwich the ligand, DFHBI.

The quadruplex core is unique in that it consists of two G-quartets that are in an antiparallel strand orientation, which was unprecedented for an RNA quadruplex. This is facilitated by unusual loops between the G-quartet residues, including a long, 34-nt loop extending into a new A-form duplex, before returning to complete the quadruplex (Figure 1.17, P1).
Figure 1.17  Global structure of Spinach
Structure consists of three regions: P1 and P2 helices (gray), and the L_{12} core (red, blue and orange). The DFHBI ligand is shown in green. PDB ID: 4KZD.

Figure 1.18  The two-layer G4 in L_{12} of Spinach.

The core quadruplex also acts as the base of the fluorophore binding pocket. DFHBI sits planar atop one face of the quadruplex and is capped overttop by a Hoogsteen triplex, which consists of U50:A53:U29. DFHBI further hydrogen bonds with
surrounding residues: G28 and A58. Disruption of these residues significantly reduced fluorescence relative to wild-type. Together, the pocket rigidifies the fluorophore ligand and protects it from solvent exposure, thus directing energy release upon excitation in the fluorescence pathway. The triplex also serves the purpose of transitioning from the quadruplex into the loop-side stem.

**Figure 1.19 The fluorophore binding site of Spinach.**
(Left) Side view of the DFHBI binding site. U29 is omitted for clarity. Black dashes (with the distances numbered in Å) represent inferred hydrogen bonds from DFHBI to the G28 nucleobase and several 2'-hydroxyl groups. Potential hydrogen bonds from fluorine are in gray dashes with distances numbered in Å. The imidazolinone methyl groups of DFHBI may engage in hydrophobic interactions (orange dashes) with A58 (orange). Ligand atoms: lime green, C; blue, N; cyan, F; red, O. (Right) Top-down view showing DFHBI (lime green) and adjacent nucleotide G28 stacking between a U50:A53:U29 Hoogsteen base triple (blue) above and a layer of G4 (red) below. Modified and reprinted by permission from Springer Nature. Nature Chemical Biology. A G-quadruplex–containing RNA activates fluorescence in a GFP-like fluorophore. Huang et al. Copyright (2014).

Apo-crystal structures of Spinach revealed that the global fold of the aptamer is unchanged without the ligand (i.e. the G4 and flanking helices remain folded). However, within the binding pocket, the U50:A53:U29 triplex cap collapses into the quadruplex face to form a new A58:A53:U29 base triple.

The flanking helices of the Spinach aptamer are largely superfluous. Miniaturization of the Spinach aptamer by reducing the length of these helices to 5-6 base pairs yielded Baby Spinach, which retained 95% of the fluorescence of DFHBI bound to the wild-type Spinach aptamer. In a similar study, truncations of the closing helix showed that a 5 bp stem retained similar fluorescence to wild-type. Further, optimization of Spinach by FACS selection resulted in the Broccoli aptamer which is nearly identical in sequence with again, truncated helical regions.
Figure 1.20  The structure of Spinach RNA in the absence of DFHBI
(A) Overlay of Spinach structures obtained in the presence (green) and absence (yellow) of bound DFHBI fluorophore. DFHBI binding has minimal influence on global architecture (r.m.s. deviation = 0.92 Å on RNA). (B) Overlay of the G4 motifs in the presence (green) and absence (yellow) of bound DFHBI fluorophore (r.m.s. deviation = 0.72 Å). (C) The DFHBI binding site collapses in the absence of DFHBI ligand. The transparent, lime green structure indicates the position of DFHBI in the bound structure. The colour codes of other adjacent nucleotides and molecular orientation of the RNA match those in Figure 1.19. Modified and reprinted by permission from Springer Nature. Nature Chemical Biology. A G-quadruplex–containing RNA activates fluorescence in a GFP-like fluorophore. Huang et al. Copyright (2014).

1.5.3. Corn aptamer

The Corn aptamer, like Spinach, binds to a fluorescent protein chromophore (for Corn, the chromophore is from the DsRed proteins, DFHO)\textsuperscript{107}. The core fluorophore-binding pocket consists of a non-canonical 4-layer quadruplex containing two G-tetrads followed by two non-canonical quartets\textsuperscript{37}. Beneath the G4, the aptamer is closed by a single helix which together with the bottom quartet layer, loops and folds to form a ‘candy-cane’-like bend (Figure 1.21). Interestingly, the Corn aptamer binds to DFHO as a quasisymmetrical dimer in the crystal lattice and is corroborated by analytical ultracentrifugation studies as well as small-angle X-ray scattering (SAXS) measurements that reveal Corn binds to DFHO with 2:1 stoichiometry in solution\textsuperscript{37}. 
Figure 1.21 Structure of Corn–DFHO complex.


In the crystal structure, DFHO is sandwiched planar between two Corn aptamers at the face of the G4 (Figure 1.22). The two G-quartet layers (T1 & T2) and the first non-canonical quartet (T3: A10:U17:A21:U27) are coordinated by two potassium ions while the non-canonical quartets (T3 & T4:G9:C18:G20:C28) are-mediated by water molecules (Figure 1.22A). The quadruplex is connected to the P1 helix by a three ‘plane’ junction (J1) which includes two unbonded bases (G6 and U31) and two non-Watson-Crick base pairs (A7:C30 and G8:A29). The J1 riboses, along with ribose from T2, T3 and T4 (G9, A10, U27, G26) form zipper-shaped interactions to join the quadruplex core with the P1 stem (Figure 1.23).
Figure 1.22  Structure of the Corn protomer.

Figure 1.23  Structural features of the Corn-DFHO complex.
Though the global arrangement of the two Corn aptamers are apparently symmetrical, it is asymmetrical in the binding pocket. Three adenine residues (A11, A14, A24) from each protomer are positioned non-symmetrically, 4 of which are involved with DFHO interactions. A14 from one protomer and a11 and a24 from the second protomer form direct hydrogen bonds with the DFHO ligand. The remaining A11 stacks behind a24, which together are positioned perpendicular to the ligand while a14 and A24 do not appear to play a direct role in ligand binding. Mutational studies suggest that the most important residues are A14 and A24, both of which significantly reduce fluorescence when mutated. A11U mutants, however, still retain some fluorescence.

Figure 1.24 Architecture of the quasisymmetric DFHO binding site. (A) Side view of the DFHO binding site, colour-coded as in Figure 1.21. Water molecules are depicted as red spheres. Mesh represents a portion of the |Fo| − |Fc| electron density map, calculated before addition of DFHO and associated water molecules to the crystallographic model; green and pink mesh are contoured at 3 σ and 7 σ, respectively. (B) Axial views of the DFHO binding site, depicting DFHO above the T1 G-quartet, looking down on the T1 and t1 G-quartets (top and bottom, respectively). Modified and reprinted by permission from Springer Nature. Nature Chemical Biology. A homodimer interface without base pairs in an RNA mimic of red fluorescent protein. Warner et al. Copyright (2018).

1.5.4. Mango aptamers

Like other dye-binding aptamers, the Mango aptamers also consist of G4 cores. The crystal structures of Mangos I, II and III are known and are distinct from each other. Mango I and II contain a three-tiered G4 closed by a unique GNRA tetraloop-like motif that connects the core to an external helix. Though the G4 core is similar between Mango I and II, they differ in the structure of the ligand cap. Mango I
sandwiches the TO1-Biotin ligand between the G4 and a two-adenine cap, while Mango II utilizes a 5-adenine cage to serve this purpose\textsuperscript{38}. Mango III differs from both Mango I and II as it consists of a two-tiered G4, with its ligands sandwiched between itself and an A:U trans Watson-Crick base pair. The core of the Mango III aptamer transitions from the quadruplex to the helix by a triplex stack in between. The Mango I and Mango III crystal structure studies will be discussed in detail in Chapter 4.
Chapter 2.

Fluorophore ligand binding and complex stabilization of RNA Mango and RNA Spinach aptamers

Lisa, in this house…we obey the laws of thermodynamics!

H.J.S.

This chapter is based entirely on the following manuscript:


I (S.C.Y.J.) planned and performed most experiments and contributed to writing the manuscript. H.H.Y.C planned and performed experiments. E.P.B provided G-quadruplex RNA and intellectual input. S.A.M. provided G-quadruplex RNA, intellectual input and contributed to writing the manuscript. P.J.U. planned experiments and contributed to writing the manuscript.
2.1. Introduction

The \textit{in vitro} selection\textsuperscript{66–58} of high affinity RNA aptamers able to dramatically enhance the fluorescence of target fluorophores offers significant potential for the tracking and purification of cellular RNA complexes \textsuperscript{108,115,116}. Until recently, aptamers with sufficiently high affinity (low \(K_D\)) and fluorescence enhancement (the ratio of bound to unbound fluorophore fluorescence, \(F_E\)) have been lacking, but the \textit{in vitro} selection of RNA Spinach\textsuperscript{104}, and RNA Mango\textsuperscript{108} have provided much needed tools for studying biological RNAs. Curiously, RNA G4s formed by the Hoogsteen to Watson-Crick face pairing of guanines into planar tetrad structures \textsuperscript{31,117} are at the core of both RNA Mango and RNA Spinach. Based on the RNA Mango aptamer’s potassium ion dependence, circular dichroism, T1 RNase protection, and core sequence motif, we inferred that RNA Mango consists of a parallel-stranded G4 when bound to TO1-Biotin\textsuperscript{108}. Crystal structures demonstrate that RNA Spinach assembles one face of its fluorophore binding pocket out of an unusual, rigid antiparallel G4 platform and the opposite face out of a uniquely positioned RNA base triple\textsuperscript{32,33,104}. We, therefore, wondered how the artificially selected RNA Mango and RNA Spinach aptamers might differ in their ability to discriminate between their own fluorophores and a set of fluorophores known to interact with G4 structures (Figure 2.1).

We find that RNA Mango is able to distinguish its intended target ligand, TO1-Biotin, from other fluorophores while RNA Spinach and a set of biologically derived G4s are unable to do so. Simultaneously, the strong thermal stabilization observed when RNA Mango binds TO1-Biotin suggests a concerted binding mechanism for this aptamer. The ability of RNA Mango to distinguish ligands combined with its unusually high binding affinity for TO1-Biotin may allow the development of orthogonally binding fluorophore-aptamer systems as well as fluorophore-aptamer systems with improved fluorescent efficiency.
Figure 2.1  Fluorophores used in binding and fluorescence enhancement experiments.
RNA Spinach has DFHBI as a ligand, while RNA Mango was selected to bind TO1-Biotin.

2.2. Results

2.2.1. RNA Mango and RNA Spinach fluorophore binding and fluorescence enhancement

We initially wondered whether the RNA and fluorophore components of the RNA Spinach and RNA Mango systems could be orthogonal to each other. While RNA Mango did not detectably bind, or enhance the fluorescence of DFHBI, RNA Spinach was able to bind to TO1-Biotin ($K_D = 830 \pm 130$ nM) exhibiting only a two-fold weaker affinity for this ligand than its DFHBI binding partner ($K_D = 420 \pm 40$ nM) (Table 2.1, Figure 2.2). The finding that RNA Mango could strongly discriminate between TO1-Biotin and
DFHBI, while RNA Spinach was weak in this respect, immediately suggested that the mechanisms of binding between the two aptamers were distinct despite the fact that both contain G4 motifs\textsuperscript{32,33,108}.

Figure 2.2  Binding affinity curves for fluorophore/RNA pairs in WB buffer
(A) Binding affinity curves for 10 nM TO1-Biotin series. (B) 75 nM DFHBI (C) 20 nM NMM (D) 75 nM Thioflavin T (E) 75 nM Ethidium bromide. Error bars are standard deviations of 3 replicates unless no detectable fluorescence enhancement was measured, in which standard deviation is of 2 replicates.
Table 2.1  Binding affinities in nM of each fluorophore:RNA combination at room temperature

<table>
<thead>
<tr>
<th>Fluorophore:RNA</th>
<th>TO1-Biotin</th>
<th>DFHBI</th>
<th>NMM</th>
<th>Thioflavin T</th>
<th>EtBr</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA Mango</td>
<td>3.9 ± 1.0</td>
<td>&gt;&gt;</td>
<td>14 ± 11</td>
<td>620 ± 40</td>
<td>550 ± 60</td>
</tr>
<tr>
<td>RNA Spinach</td>
<td>830 ± 130</td>
<td>420 ± 40</td>
<td>50 ± 5</td>
<td>260 ± 10</td>
<td>110 ± 30</td>
</tr>
<tr>
<td>PITX1</td>
<td>1490 ± 350</td>
<td>&gt;&gt;</td>
<td>180 ± 20</td>
<td>&gt;</td>
<td>1650 ± 1200</td>
</tr>
<tr>
<td>hTR</td>
<td>1290 ± 230</td>
<td>&gt;&gt;</td>
<td>160 ± 50</td>
<td>&gt;</td>
<td>830 ± 90</td>
</tr>
<tr>
<td>SynTet</td>
<td>940 ± 350</td>
<td>&gt;&gt;</td>
<td>180 ± 160</td>
<td>1410 ± 370</td>
<td>320 ± 40</td>
</tr>
<tr>
<td>dsRNA</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;&gt;</td>
<td>&gt;</td>
<td>1260 ± 490</td>
</tr>
</tbody>
</table>

Errors are the standard deviation of three measurements. Affinities greater than 1000 nM are estimates from RNA titrations with end points at 1024 nM.

Surprisingly, once TO1-Biotin bound to RNA Spinach, the complex was remarkably bright. The TO1-Biotin:RNA Spinach complex exhibited a $F_E$ of 4,590 ± 770 (Table 2.2, Figure 2.2). This $F_E$ is significantly higher than that observed with TO1-Biotin:RNA Mango ($F_E = 1,020 ± 20$, quantum yield 0.13) or DFHBI:RNA Spinach ($F_E = 1,590 ± 60$, quantum yield 0.72$^{104}$). The TO1-Biotin:RNA Spinach $F_E$ is without precedence for a TO1-based fluorophore and exceeds that observed for unmodified thiazole orange (TO1) intercalation into duplex DNA$^{118}$. The $F_E$ for TO1-Biotin:RNA Spinach corresponds to an absolute brightness of ~50,000 M$^{-1}$cm$^{-1}$ making this complex ~3-fold brighter than DFHBI:RNA Spinach (Brightness of 17,500 M$^{-1}$cm$^{-1}$) and 1.4 fold brighter than eGFP (Brightness of 34,000 M$^{-1}$cm$^{-1}$$^{119}$). This finding may prove a significant advantage for RNA Spinach systems where ultimate brightness, and not binding affinity, are critical, and suggests that RNA Mango variants may be found in the future with considerably higher brightness.

Table 2.2  Fluorescence enhancement ($F_E$) of each fluorophore:RNA combination at room temperature

<table>
<thead>
<tr>
<th>Fluorophore:RNA</th>
<th>TO1-Biotin</th>
<th>DFHBI</th>
<th>NMM</th>
<th>Thioflavin T</th>
<th>EtBr</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA Mango</td>
<td>1020 ± 20</td>
<td>&gt;&gt;</td>
<td>11 ± 1</td>
<td>9.6 ± 0.5</td>
<td>4.7 ± 0.2</td>
</tr>
<tr>
<td>RNA Spinach</td>
<td>4590 ± 770</td>
<td>1590 ± 60</td>
<td>21 ± 1</td>
<td>57 ± 2</td>
<td>6.5 ± 0.3</td>
</tr>
<tr>
<td>PITX1</td>
<td>770 ± 90</td>
<td>&gt;&gt;</td>
<td>12 ± 1</td>
<td>&gt;</td>
<td>4.4 ± 0.5</td>
</tr>
<tr>
<td>hTR</td>
<td>1000 ± 50</td>
<td>&gt;&gt;</td>
<td>11 ± 1</td>
<td>&gt;</td>
<td>5.1 ± 0.4</td>
</tr>
<tr>
<td>SynTet</td>
<td>870 ± 130</td>
<td>&gt;&gt;</td>
<td>19 ± 2</td>
<td>36 ± 6.7</td>
<td>6.3 ± 0.2</td>
</tr>
<tr>
<td>dsRNA</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;&gt;</td>
<td>&gt;</td>
<td>8.9 ± 1.4</td>
</tr>
</tbody>
</table>

Fluorescence enhancements could not be determined for pairs with unknown dissociation constants and are indicated as in Table 1 (’>>’ or ‘>’).

TO1-Biotin was designed with modifications at the nitrogen of the benzothiazole ring as side chains at this location have previously been demonstrated to destabilize TO1 intercalation into duplex dsDNA by about 7 fold$^{108,120}$. To explore the importance of
the TO1-Biotin side chain for binding to RNA Mango we measured the $K_D$ and $F_E$ of unmodified TO1 binding to RNA Mango as well as to a 17-bp double-stranded RNA (dsRNA) duplex. TO1-Biotin was 6 to 7 times brighter at saturation than was TO1 when bound to RNA Mango (Figure 2.3A) while binding affinity was slightly weaker for TO1 than for TO1-Biotin. Conversely, TO1-Biotin significantly destabilized intercalation into dsRNA being ~20 fold less fluorescent than TO1 when measured at nearly $\mu$M concentrations of RNA (Figure 2.3B). The distinct binding and fluorescence enhancement properties for TO1 and TO1-Biotin to RNA Mango and to the dsRNA control, again, point to the ability of RNA Mango to distinguish ligands by a potentially unique mechanism.

![Graph A](image)

**Figure 2.3** **TO1-Biotin has high fluorescence enhancement relative to TO1 when bound to RNA Mango**

TO1-Biotin (blue) and TO1 (red) fluorophores are at the indicated concentrations where (A) RNA Mango and (B) dsRNA were titrated. In (B), linear fits were to the last four data points as instrumental scatter from the lower RFU points generated erroneous slopes. Fluorophore concentrations are labeled as follows: diamond – 10 nM, circle – 20 nM, triangle – 40 nM, and square – 80 nM. RFUs on panels A and B are to scale with each other.
RNA Mango and RNA Spinach are artificially selected aptamers that coincidentally adopt a G4 structure. We, therefore, compared these aptamers to a pair of biologically relevant parallel-stranded G4s, and a synthetic 4-stranded tetramolecular G4 in addition to the 17-bp duplex RNA control just mentioned (see Table 2.3 for sequences). These RNAs included PITX1 (PITX11901-1930), a G4 from the human PITX1 mRNA and hTR, a quadruplex from human telomerase RNA121,122. Based on a combination of UV/VIS spectroscopy, CD spectroscopy, small angle x-ray scattering, and nuclear magnetic resonance spectroscopy experiments, PITX1 and hTR RNA have both previously been found to adopt a parallel-stranded G4 structure122,123. Efficient staining with N-methyl mesoporphyrin IX (NMM, a parallel G4 ligand, Figure 2.1), and their interaction with a protein binding partner, RNA helicase associated with AU-rich element (RHAU), specific for parallel-stranded G4s, confirmed the parallel orientation typically observed for RNA quadruplexes121,122. SynTet (a synthetic tetrameric RNA quadruplex) is also a parallel-stranded quadruplex, and interacts with NMM124. Unsurprisingly, binding of NMM and Thioflavin T (ThT, Figure 2.1) to PITX1, hTR, and SynTet has been demonstrated125.

<table>
<thead>
<tr>
<th>Table 2.3 RNA sequences used in binding affinity experiments.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RNA Mango</strong></td>
</tr>
<tr>
<td><strong>RNA Spinach</strong></td>
</tr>
<tr>
<td><strong>Bimolecular dsRNA (17-nt)</strong></td>
</tr>
<tr>
<td><strong>PITX1-Q2</strong></td>
</tr>
<tr>
<td><strong>hTR10-43</strong></td>
</tr>
<tr>
<td><strong>SynTet</strong></td>
</tr>
</tbody>
</table>

All of these control quadruplex constructs were 300 to 400-fold worse than RNA Mango at binding to TO1-Biotin (Table 2.1, Figure 2.2), while the dsRNA control bound with a barely detectable linear increase in fluorescence enhancement (Figure 2.3B, blue). The micromolar affinities exhibited by these sequences are similar to those reported for the binding of unmodified TO1 to a range of DNA G4s126, again suggesting
that the binding mode of RNA Mango to TO1-Biotin is unusual. The fluorescence enhancements of these complexes at saturation were similar to that seen for RNA Mango (Table 2.2), implying that all of these G4 structures are ultimately able to rigidify the TO1-Biotin fluorophore at saturation. In comparison, DFHBI, the Spinach fluorophore, showed no detectable binding or fluorescence enhancement for any RNA other than RNA Spinach (Table 2.1, Table 2.2, Figure 2.2). We note however, that binding interactions 300-fold weaker than the nearly micromolar DFHBI:RNA Spinach binding interaction would have been hard to detect under our titration conditions (i.e. assuming a $K_D$ of 100 μM, at the 1 μM end point of our RNA titrations fluorescence signal would have only been 1% of that seen at full ligand binding).

### 2.2.2. Fluorophore ligand discrimination by RNA Mango

Since RNA Spinach bound TO1-Biotin and RNA Mango did not bind DFHBI detectably, we tested their ability to bind three other unrelated fluorophores. NMM and ThT are previously described G4 binders\textsuperscript{125,127}, while EtBr is an established double-stranded nucleic acid intercalator (Figure 2.1). Strikingly, RNA Mango could discriminate between these fluorophores, binding to them all with weaker affinity than to its selected ligand TO1-Biotin (Table 2.1). In distinct contrast, RNA Spinach bound NMM, ThT and EtBr with significantly higher affinity than to DFHBI (Table 2.1).

Crystal structures reveal that NMM binds selectively to parallel-stranded DNA G4s by stacking interactions between NMM and the G-tetrad face, and importantly, via its methyl group that fits into the center of the G4\textsuperscript{128}. NMM was therefore unsurprisingly the tightest binder to parallel-stranded biological and synthetic control G4 RNAs used in this study (Table 2.1). Nevertheless, binding of NMM to RNA Mango and RNA Spinach was stronger than these controls suggesting that artificial selection for fluorophore binding had also resulted in enhanced NMM binding and that the antiparallel G4 structure of RNA Spinach does not preclude NMM binding. Likewise, ThT bound worse to the control parallel-stranded quadruplexes compared to both artificially selected aptamers, with the four-stranded SynTet quadruplex showing the tightest fluorophore binding amongst the control quadruplexes (Table 2.1). While EtBr bound respectably to RNA Spinach, weaker binding to EtBr was measured for all other constructs.
2.2.3. Fluorescent efficiencies induced by RNA quadruplex binding are correlated with ligand discrimination

Consistent with the ability of RNA Mango to bind preferentially to TO1-Biotin, RNA Mango showed the lowest fluorescence enhancements of all the quadruplexes tested for fluorophores other than TO1-Biotin (Table 2.2). In marked contrast, all fluorophores tested had the highest fluorescence enhancement when bound to RNA Spinach. This indicated to us that the rigid G4 of RNA Spinach is well suited to bind a broad range of fluorophores and enhance their quantum yields, but the aptamer is unable to discriminate between these ligands. Further, a recent study\textsuperscript{129} indicates that DFHBI:RNA Spinach is strongly inhibited in 10 mM ATP, while TO1-Biotin:RNA Mango complex remains fully fluorescent in the presence of this concentration of ATP (data not shown). The ability of RNA Mango to distinguish between such ligands makes it a promising component for a system of orthogonal aptamers, each able to bind tightly and enhance the fluorescence of distinct fluorophore ligand partners.

<table>
<thead>
<tr>
<th>Table 2.4</th>
<th>Fluorescent efficiency of each fluorophore:RNA combination in units of (nM)\textsuperscript{-1} at room temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TO1-Biotin</td>
</tr>
<tr>
<td>RNA Mango</td>
<td>260</td>
</tr>
<tr>
<td>RNA Spinach</td>
<td>5.5</td>
</tr>
<tr>
<td>PITX1</td>
<td>0.52</td>
</tr>
<tr>
<td>hTR</td>
<td>0.78</td>
</tr>
<tr>
<td>SynTet</td>
<td>0.93</td>
</tr>
<tr>
<td>dsRNA</td>
<td>0.005*</td>
</tr>
</tbody>
</table>

\textsuperscript{1\textasciicircum{<}<1} less than the detection limit of the assay used. * value determined from linear increase in fluorescence observed during titration.

The overall signal-to-noise of a fluorescent RNA aptamer system is determined by both its fluorescence enhancement and binding affinity. Assuming that a fluorophore-aptamer system is optimally imaged with fluorophore concentration only several times higher than the \textit{K_D} of the complex, then the fluorescent efficiency (Equation 2.4)\textsuperscript{130,131} is a good metric to compare and contrast fluorophore:aptamer pairs (\textit{i.e.} the value of fluorescent efficiency represents the best combination of affinity and fluorescence enhancement of the fluorophore to overcome noise in an imaging experiment). TO1-Biotin:RNA Mango was by far the most fluorescently efficient complex (E = 260 nM\textsuperscript{-1}, Table 2.4) with TO1-Biotin:RNA Spinach being 50-fold worse having a fluorescent efficiency of 5.5 nM\textsuperscript{-1}, which in turn was just higher than the efficiency of the DFHBI:RNA.
Spinach complex (E = 3.8 nM⁻¹). Notably NMM, ThT and EtBr all have very poor fluorescent efficiencies owing mainly to the low Fₑ values observed upon binding. While TO1-Biotin did fluoresce strongly with a range of G4s (Table 2.2), RNA Mango imaging would be expected to use TO1-Biotin concentrations considerably lower than the $K_D$ values (Table 2.1) required to saturated fluorescent signal, giving the Mango system considerable theoretical contrast over both the naturally occurring biological G4s and to RNA Spinach. Thus, for example, in a given visual field we might expect that only when the local hTR RNA concentration is 300-fold higher than that of RNA Mango would the fluorescent signals from either the TO1-Biotin:Mango or the TO1-Biotin:hTR complexes be comparable (i.e. 260/0.78 = ~300 from Table 2.4). Improvements to RNA Mango’s binding affinity and fluorescence enhancement might reasonably enhance this contrast still further.

2.2.4. Temperature-dependent binding and fluorescent properties of RNA Mango and RNA Spinach

We decided to further characterize TO1-Biotin bound to either RNA Mango or RNA Spinach, given that these two complexes had the highest fluorescent efficiencies of any tested. First, we measured the effect of temperature on fluorescence. TO1-Biotin:RNA Spinach exhibited a sigmoidal-type fluorescence dependence (Figure 2.4A, blue curves) as seen previously for DFHBI:Spinach2⁶⁸. TO1-Biotin:RNA Mango, however, displayed a nearly linear decrease in fluorescence as temperature increased (Figure 2.4A, red curves). Fluorescence became undetectable for TO1-Biotin:RNA Spinach at ~68 °C and at ~79 °C for TO1-Biotin:RNA Mango, consistent with the higher binding affinity of the RNA Mango complex.
Figure 2.4 Temperature-dependent spectroscopy of TO1-Biotin complexed with RNA Mango or RNA Spinach.

(A) Normalized fluorescence of TO1-Biotin complexed with RNA Mango (red) or RNA Spinach (blue) as a function of temperature. (B) Simple derivatives of the RNA Mango fluorescence (red) and $A_{260}$ of unbound RNA Mango (purple) and RNA Mango with an excess of TO1-Biotin (green). (C) Derivatives of fluorescence (blue) and absorbance of unbound RNA Spinach (purple) and RNA Spinach with an excess of TO1-Biotin (green). For both (B) and (C): left Y-axis is the local slope of the $A_{260}$ temperature data (see supplement), right Y-axis is the local slope of the normalized fluorescence plot from panel A. Dots represent the local slope $(Y_{i+1}-Y_i)/(T_{i+1}-T_i)$ where $Y_i$ is either $A_{260}$ or the normalized RFU value and $T_i$ the corresponding temperature. Curves are interpolations of the data points. The darker colours represent decreasing temperature ramps of 1 °C/min, while the lighter colours represent increasing rates of 1 °C/min.
We next investigated the temperature-dependence of the TO1-Biotin complex by measuring absorbance at 260 nm using micromolar concentrations of RNA and ligand. The derivative of absorbance as a function of temperature was most revealing for both aptamers (Figure 2.4B, Figure 2.4C). As expected, based on the rigid global fold of RNA Spinach\textsuperscript{32,33}, dA\textsubscript{260}/dT was maximal at 68 °C in the presence or absence of TO1-Biotin, indicating that the folding of RNA Spinach is independent of fluorophore binding and that the disappearance of fluorescence is highly correlated with the melting of this RNA structure (Figure 2.4C). In distinct contrast, the dA\textsubscript{260}/dT curves for RNA Mango were strongly dependent on the presence or absence of fluorophore. When TO1-Biotin was present, dA\textsubscript{260}/dT was maximal at 77 °C and was highly correlated with the temperature at which fluorescence of the complex was observed to vanish (Figure 2.4B, Figure 2.4C). However, in the absence of TO1-Biotin, the maximum of the dA\textsubscript{260}/dT for RNA Mango shifted to 59 °C, indicating that the structure of RNA Mango is considerably stabilized by binding of the TO1-Biotin fluorophore.

Figure 2.5   Binding affinity curves of RNA Mango + TO1-Biotin as a function of temperature.
Curves were measured with fixed 20 nM TO1-Biotin. Error bars are the standard deviations of 3 replicates.
To explore the fluorescence dependence below the melting temperature of the RNA Mango complex we next measured TO1-Biotin:RNA Mango binding affinities and fluorescence enhancement as a function of temperature using 20 nM TO1-Biotin while simultaneously titrating RNA Mango concentrations (Figure 2.5). The $K_D$ was found to increase as the melting temperature of the complex was reached, moving from 3 nM at 25 °C to 26 nM at 60 °C (Figure 2.6, black symbols). Since the temperature dependence of fluorescence complex in Figure 2.4A was generated using 5 μM TO1-Biotin and 1 μM RNA Mango, such a small change in binding affinity would by itself be insufficient to explain the ~5-fold change in fluorescence observed over this temperature range. Rather, we infer that the fluorescence of the bound complex significantly attenuates due to either a change in the quantum yield, or extinction coefficient of the complex. Consistent with this hypothesis, the $F_E$ obtained from our temperature dependent $K_D$ titrations (Figure 2.6, purple dots) agreed well with the data obtained at a much higher complex concentration, and rather precisely reproduces the observed changes in fluorescence at the 50 fold higher concentration shown in Figure 2.4A (reproduced in Figure 2.6 by the solid red lines).

![Figure 2.6](image-url)

**Figure 2.6**  RNA Mango TO1-Biotin complex fluorescence and dissociation constant as a function of temperature. RNA Mango binding affinity to TO1-Biotin at intervals of 5 °C are represented by black points (error bars are the standard deviation of 3 replicates, see Figure 2.5). Normalized $F_{max}$ for each temperature is shown in purple, overlaid on top of the temperature dependence data for RNA Mango + TO1-Biotin from Figure 2.4(red solid lines). The two data sets have been arbitrarily normalized at the 25 °C temperature point for comparison.
2.2.5. Fluorescent lifetimes of TO1-Biotin:RNA Mango and TO1-Biotin:RNA Spinach are distinct

Since RNA Mango and RNA Spinach both bind by distinct mechanisms to TO1-Biotin, we wondered if the photophysics of these complexes differed in observables other than the fluorescence enhancement previously discussed. We measured the first-order fluorescence lifetimes (\(\tau\)) (Figure 2.7, Table 2.5) and Stokes shifts (Figure 2.8) of TO1-Biotin bound to either RNA Mango or RNA Spinach. For comparison, we also measured the fluorescence lifetimes of fluorescein (\(\tau = 4.1\) ns), as well as RNA Spinach bound to DFHBI (\(\tau = 4.3\) ns), both agreeing well with previously published values\(^{132,133}\). The TO1-Biotin:RNA Mango complex had a lifetime of 3.3 ns at room temperature and had a 25 nm Stokes shift comparable to the SynTet construct. This lifetime is similar to the lifetime of TO1 intercalated into dsDNA (2.6 ns\(^{118}\)), where quantum yields were also found to be comparable to that of RNA Mango. Interestingly, when TO1-Biotin was bound to RNA Spinach, the fluorescence lifetime nearly doubled to 6.0 ns consistent with the high brightness observed for this complex and with a ~15% decrease in emission peak width relative to the Mango TO1-Biotin emission peak width (Figure 2.8). The long lifetime observed with TO1-Biotin:RNA Spinach was also associated with a further 10 nm decrease in Stokes shift for this complex relative to the 25 nm shift observed for the TO1-Biotin:RNA Mango complex (Figure 2.8, Table 2.6). Based on these results, it appears highly likely that further selections to increase the quantum yield of RNA Mango will also result in changes in fluorescence lifetime.
Figure 2.7  Fluorescence lifetimes RNA aptamer fluorophore complexes.
Fluorescence lifetimes of fluorophore:RNA complexes were measured using time-correlated single photon counting (TCSPC). Colour-coding is as follows: red – TO1-Biotin:RNA Mango, blue – TO1-Biotin:RNA Spinach, dark green – DFHBI:RNA Spinach, bright green – fluorescein standard. Data was fitted to a first-order exponential decay between 13-25 ns and calculated lifetimes are listed in Table 2.5.

Table 2.5  Lifetimes of fluorescent decay experiments determined from single exponential fit of data shown in Figure 2.7.

<table>
<thead>
<tr>
<th>Complex</th>
<th>Lifetime (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA Mango + TO1-Biotin</td>
<td>3.3 ± 0.01</td>
</tr>
<tr>
<td>RNA Spinach + TO1-Biotin</td>
<td>6.1 ± 0.04</td>
</tr>
<tr>
<td>RNA Spinach + DFHBI</td>
<td>4.3 ± 0.02</td>
</tr>
<tr>
<td>Fluorescein</td>
<td>4.1 ± 0.01</td>
</tr>
</tbody>
</table>
Figure 2.8  Excitation and emission spectra for fluorophore/RNA pairs in WB buffer

Y-axes are normalized to 1, relative to the maximum value of each spectrum and fitted to a smoothen function (Kaleidagraph). Colours represent RNA construct used: RNA Mango – red, RNA Spinach – dark green, PITX1 – blue, hTR – purple, SynTet – light green, dsRNA – gray, no RNA – black.
Table 2.6  Visible excitation and emission peaks for each fluorophore/RNA complex used in this study (wavelengths in nm).

<table>
<thead>
<tr>
<th></th>
<th>TO1-Biotin</th>
<th>DFHBI</th>
<th>NMM</th>
<th>Thioflavin T</th>
<th>EtBr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ex</td>
<td>Em</td>
<td>Ex</td>
<td>Em</td>
<td>Ex</td>
</tr>
<tr>
<td>Unbound</td>
<td>500</td>
<td>525</td>
<td>418</td>
<td>500</td>
<td>398</td>
</tr>
<tr>
<td>RNA Mango</td>
<td>510</td>
<td>535</td>
<td>-</td>
<td>-</td>
<td>398</td>
</tr>
<tr>
<td>RNA Spinach</td>
<td>510</td>
<td>525</td>
<td>469</td>
<td>503</td>
<td>401</td>
</tr>
<tr>
<td>PITX1</td>
<td>510</td>
<td>533</td>
<td>-</td>
<td>-</td>
<td>401</td>
</tr>
<tr>
<td>hTR</td>
<td>510</td>
<td>533</td>
<td>-</td>
<td>-</td>
<td>401</td>
</tr>
<tr>
<td>SynTet</td>
<td>510</td>
<td>535</td>
<td>-</td>
<td>-</td>
<td>402</td>
</tr>
<tr>
<td>dsRNA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

2.3. Discussion

This study demonstrates that the discriminating binding observed for RNA Mango is not generally found in either RNA Spinach or in biologically derived G4 controls. How exactly the high affinity complex between RNA Mango and TO1-Biotin is formed from a kinetic perspective is currently unclear, but requires a PEG biotin side chain or similar derivatives\(^{108}\) to become fully fluorescent. We infer that a concerted folding process occurs due to the dramatic thermal stabilization of the RNA Mango structure observed upon TO1-Biotin binding together with previous data\(^{108}\) that indicates a large change in CD and T1 RNase protection upon addition of TO1-Biotin ligand. This is significant, as G4s are not generally thought to change structure upon ligand binding. While previous work has demonstrated that small molecules can either stabilize or destabilize G4s, these studies require significantly higher ligand concentrations to achieve such effects\(^{134,135}\). Likewise, the global structure of the promiscuous RNA Spinach aptamer including its G4 platform, is the same in both the bound and unbound states as shown by crystal structures\(^{32}\) and as corroborated by our UV melting data. In contrast, RNA Mango appears to resemble naturally occurring (but G4 independent) aptamers such as the purine riboswitches that use a concerted mechanism to achieve ligand specificity\(^{136–139}\). The adenine riboswitch discriminates by 30 fold between adenine and guanine, but binds 30 fold tighter to 2,6-diaminopurine than to adenine, while the guanine riboswitch does not bind to adenine detectably\(^{140,141}\). RNA Mango compares very favorably in this respect binding TO1-Biotin more than 100 times tighter than to DFHBI, ThT or EtBr. While the exact mechanism of this discrimination remains to be determined, a clear distinction between these natural aptamers and RNA Mango is the complete lack of
potential hydrogen bonding interactions found on the TO1 portion of the TO1-Biotin ligand, making the presumably hydrophobic binding mechanism of considerable interest to study further.

Improvements in RNA Mango appear likely given the data presented in this study. RNA Spinach, which we have demonstrated to ubiquitously bind and enhance the brightness of a broad range of fluorophores, dramatically enhanced the brightness of TO1-Biotin to levels never previously seen. Consistent with our findings, recent work has shown that RNA Spinach binds to novel fluorophores with higher affinity than to DFHBI. We speculate that the high quantum yield seen for the TO1-Biotin:RNA Spinach complex, results from the ability of RNA Spinach to effectively shield bound fluorophores from solvent, as all fluorophores studied showed maximal fluorescence enhancement when bound to RNA Spinach. This solvent shielding is also likely to explain the nearly two-fold change in fluorescent lifetime observed when TO1-Biotin was bound to RNA Spinach instead of RNA Mango and is partially consistent with the decreased fluorescence emission peak width observed for this aptamer. It remains to be seen if an RNA Mango aptamer can be evolved to greater brightness, but experience with RNA Spinach suggests that this should be possible. RNA Spinach has previously been engineered or re-selected to improve quantum yield, affinity and thermal stability resulting in the Spinach2, Broccoli and iSpinach aptamers. In particular, Broccoli and iSpinach were selected using fluorescence-activating sorting methods to directly enhance brightness. This methodology is extremely promising for the identification of new aptamers of superior brightness/fluorescence efficiency. Evolved RNA Mango aptamers could potentially be brighter than eGFP, which when combined with the already extremely high binding affinity of the Mango system, bodes well for the use of the RNA Mango system in the imaging and purification of RNA complexes in living systems.

2.4. Materials & Methods

2.4.1. RNA synthesis

RNA Mango, RNA Spinach and dsRNA were in vitro transcribed by run-off transcription using T7 RNA polymerase with 0.5 μM single stranded DNA template stand hybridized to a 20-nt T7 promoter top strand. dsRNA was generated by mixing equimolar
ratio of each complementary RNA strand. PITX1, hTR and SynTet RNA were purchased from Integrated DNA Technologies Inc. All RNAs were purified using denaturing polyacrylamide gels prior to use.

2.4.2. Fluorophore purification and characterization

TO1-Biotin was synthesized and purified as previously described\textsuperscript{108}. DFHBI was purchased from Lucerna, Inc. NMM was purchased from Frontier Scientific, Inc. Thioflavin T was purchased from Sigma-Aldrich Co. Ethidium bromide was purchased from Merck KGaA. Fluorophore concentrations were determined using the extinction coefficients listed in Table 2.7. Fluorophores were used with no further purification steps following purchase. Purities were measured by HPLC with a 50 mM TEAA and increasing acetonitrile gradient, which indicated close to >99% purity for all fluorophores with the exception of NMM. NMM has previously been described to be a racemic mixture of eight isoforms. Historically, NMM has been used as the mixture; hence our study follows the same convention.

Table 2.7  Extinction coefficients of unbound fluorophores used in this study in conditions as cited.

<table>
<thead>
<tr>
<th>Fluorophore</th>
<th>( \lambda_{\text{abs}} )</th>
<th>Extinction Coefficient (M(^{-1})cm(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>TO1-Biotin</td>
<td>500</td>
<td>63,000 \textsuperscript{108}</td>
</tr>
<tr>
<td>DFHBI</td>
<td>405</td>
<td>11,864 \textsuperscript{104}</td>
</tr>
<tr>
<td>NMM</td>
<td>379</td>
<td>145,000 \textsuperscript{143}</td>
</tr>
<tr>
<td>Thioflavin T</td>
<td>416</td>
<td>26,600 (Sigma-Aldrich)</td>
</tr>
<tr>
<td>EtBr</td>
<td>343</td>
<td>40,000 (Sigma-Aldrich)</td>
</tr>
</tbody>
</table>

2.4.3. Binding affinity and fluorescence enhancement determination

Fluorescence data was gathered using a Varian Cary Eclipse Spectrophotometer unless otherwise stated. Fluorescent titrations in the \textit{in vivo} mimicking buffer (WB: 140 mM KCl, 1 mM MgCl\(_2\), 10 mM NaH\(_2\)PO\(_4\) pH 7.2, 0.05% Tween-20) were performed to determine binding affinities for each fluorophore/RNA pair. Fluorescence was measured at the maximum excitation and emission wavelengths of each complex (see Table 2.6). Curves were fitted using least squares (Kaleidagraph 4.5) using the following equation if \( K_D \) was less than 100 nM:
\[ F = F_0 + \frac{F_{\text{max}}}{2} \left( K_D + [RNA] + [TO] - \sqrt{([RNA] - [TO])^2 + K_D (K_D + 2[RNA] + 2[TO])} \right) \]

*Equation 2.1*

where \( F \) is the fluorescence as a function of RNA concentration \([RNA]\), \( F_{\text{max}} \) is the maximal fluorescence of the bound complex and \( F_0 \) the fluorescence of the unbound fluorophore. When the fluorescence of \( F_0 \) was undetectable, it was set to zero.

Or to the following equation if \( K_D \) was greater than 100 nM:

\[ F = F_0 + \frac{F_{\text{max}} [RNA]}{K_D + [RNA]} \]

*Equation 2.2*

Fluorescence enhancement was determined using:

\[ F = \frac{F_{\text{max}}}{F_0} \]

*Equation 2.3*

\( F_{\text{max}} \) was determined using Equation 2.1 or Equation 2.2 as appropriate.

Fluorescent efficiency (E) is defined by:

\[ E = \frac{F}{K_D} \]

*Equation 2.4*

### 2.4.4. Temperature dependence of fluorescence

Temperature dependence measurements were started at 90 °C decreasing at a rate of 1 °C/minute until 20 °C, then returned at 1 °C/minute 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, 1 μM RNA either with or without 5 μM fluorophore. Temperature dependence of fluorescence and absorbance were measured using a Varian Cary Eclipse Fluorescence Spectrophotometer and a
Varian Cary 100 Bio UV-visible spectrophotometer monitoring at 260 nm. Data was fitted to a smoothen function to guide the eye.

2.4.5. Temperature dependence of $K_D$ and $F_{\text{max}}$

Temperature dependence of $K_D$ for TO1-Biotin:RNA Mango was measured using a SpectraMax M5 Multi-Mode Microplate Reader using Greiner black, clear-bottom 96-well microtiter plates. The fluorescence measurements were obtained at ex/em of 495/535 nm with a 530 nm cut-off filter using the bottom read function. Temperature measurements were stepped from 25 °C to 60 °C in increments of 5 °C. RNA titrations were done in triplicate with fixed amounts of TO1-Biotin (20 nM) in WB buffer. RNA Mango was titrated from 2048 nM to 1 nM by 2 fold dilutions. Data was fitted to Eq. 1, $K_D$ and $F_{\text{max}}$ (normalized to fluorescence at 25 °C) being plotted in Figure 2.6.

2.4.6. Time-resolved fluorescence lifetime measurements

Fluorescence lifetimes were measured using a Horiba-Jobin Yvon Fluorolog spectrophotometer. For TO1-Biotin measurements, the excitation source was a 494 nm NanoLED at a repetition rate of 1 MHz. Emission was passed through a 14.5 nm band pass filter centered at 525 nm for TO1-Biotin:RNA Spinach and 535 nm for TO1-Biotin:RNA Mango. For DFHBI:RNA Spinach, the excitation source was a 463 nm NanoLED at a repetition rate of 250 kHz. Emission was passed through a 14.5 nm band pass filter centered at 501 nm. For the fluorescein standard, the excitation source was a 494 nm NanoLED at a repetition rate of 1 MHz. Emission was passed with a 2 nm filter centered at 521 nm. Each measurement was taken with 100 nM fluorophore and 500 nM RNA.
Chapter 3.

Fluorogenic RNA Mango Aptamers for Imaging Small non-coding RNAs in Mammalian Cells

We studied traffic patterns and found that drivers move fastest through yellow lights, so now we just have the red and yellow lights.

Prof. J.I.Q.N.F. Jr.

This chapter is based entirely on the following manuscript:


*These authors contributed equally to the work.

I (S.C.Y.J.) planned and performed experiments biochemical and biophysical experiments to characterize of aptamers. A. Autour planned and performed microfluidics selection. A.C. planned and performed live and fixed cell fluorescence imaging experiments. A. Abdolahzadeh planned and performed sequence and structure characterization experiments of aptamers. A.G performed salt-dependence titrations. S.S.S.P. provided intellectual input for formaldehyde resistance assays. D.R., M.R. and P.J.U. supervised and helped plan experiments. All authors except for A.G. and S.S.S.P contributed to writing the manuscript.
3.1. Introduction

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\textsuperscript{104,113,116,144}. 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\textsuperscript{32,33,98,104}. While reported aptamer-fluorophore complexes make use of fluorophores with high extinction coefficients, notably RNA Mango\textsuperscript{108} and the cytotoxic Malachite Green binding aptamer\textsuperscript{98}, these systems suffer from low quantum yields. Conversely, systems with high quantum yields such as the GFP-mimic aptamers\textsuperscript{68,104,106,113} have intrinsically low extinction coefficients. As a consequence, such complexes are all significantly less bright than enhanced GFP\textsuperscript{119}, 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\textsuperscript{131}. 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\textsuperscript{104,113,145–147}. This suggests that using newly developed screening methodologies to select brighter fluorogenic RNA aptamers either by FACS\textsuperscript{68} or droplet-based microfluidics platforms\textsuperscript{106} 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\textsuperscript{108}. 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. Step-wise photobleaching in fixed cell images indicate that as few as 4-12 molecules can be detected in each foci, and photobleaching rates in live cells or \textit{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\textsuperscript{131}.

3.2. Results

3.2.1. 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\textsuperscript{108}. Its structure consists of a three-tiered G4 with mixed parallel and anti-parallel connectivity (Figure 3.1)\textsuperscript{36}. 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\textsuperscript{112}, 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 \textit{in vitro} compartmentalisation (µIVC, Figure 3.2A)\textsuperscript{106}. Interestingly, the initial screening shows that a significant fraction of molecules in the R12 library are brighter than Mango I (Figure 3.2B).
Figure 3.1  RNA Mango aptamers core sequences
Colour coded alignment of RNA Mango I, II, III, and IV. G residues in yellow are protected from 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 schematic 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), showing the T3 tier of the quadruplex and relevant propeller residues, colour coding matches the schematic and top panel.
Figure 3.2  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, analyzing 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, Table 3.5). Droplets containing no DNA yield 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 Figure 3.5.

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)$^{128}$ and TO3-Biotin$^{108}$ (Figure 3.3), which are both known to
interact with G4s. As expected, the NMM supplemented IVT mixture significantly reduces TO1-B/Mango I fluorescence (Figure 3.4).

**Figure 3.3** Chemical structures of TO1-Biotin, TO3-Biotin and NMM.

**Figure 3.4** Effect of N-methyl mesoporphyrin IX (NMM) concentration on TO1-biotin/Mango I fluorescence.

RNA aptamer at 1 µM was incubated with 100 nM of TO1-Biotin and a concentration of NMM ranging from 0 to 5 µM. Green fluorescence was recorded (ex: 492 nm/em: 516 nm) in a real-time thermocycler (Mx 3005P, Agilent). Values are normalized to that of the 0 µM NMM point and are the mean of two independent experiments and error bars correspond to ± 1 standard error.
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 TO1-B (Figure 3.2C). The stability of the RNA/TO1-B complex was further challenged by sorting the droplets at 45 °C as previously described\textsuperscript{106}. The relative fluorescence of the library increased 2.5-fold in the first round (~3 million variants analysed, Table 3.1), but it did not increase further over the later rounds (Figure 3.2C, Figure 3.5A).

### Table 3.1  Metrics of screenings in presence of NMM

<table>
<thead>
<tr>
<th>Round</th>
<th>Temperature (°C)</th>
<th>NMM (µM)</th>
<th>λ value</th>
<th>Fusion efficiency (%)</th>
<th>Number of analyzed droplets</th>
<th>Number of analyzed variants</th>
<th>Number of sorted droplets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>45</td>
<td>3</td>
<td>1.25</td>
<td>90</td>
<td>2,716,500</td>
<td>3,056,062</td>
<td>25,834</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>4</td>
<td>0.20</td>
<td>85</td>
<td>1,011,000</td>
<td>171,870</td>
<td>21,754</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>6</td>
<td>0.22</td>
<td>87</td>
<td>1,511,250</td>
<td>289,253</td>
<td>3,719</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>8</td>
<td>0.14</td>
<td>75</td>
<td>986,875</td>
<td>103,621</td>
<td>5,309</td>
</tr>
<tr>
<td>5</td>
<td>45</td>
<td>0</td>
<td>0.15</td>
<td>90</td>
<td>491,625</td>
<td>66,369</td>
<td>1,689</td>
</tr>
</tbody>
</table>

However, the ability of NMM to compete against TO1-B binding decreased progressively with each round (Figure 3.6A), 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 (Figure 3.2C). From the final enriched library, we cloned and sequenced 32 pool RNAs, and analysed their fluorogenic capacity (Figure 3.7A,B). While the brightest clone was R5-NMM-20, six of the thirteen brightest aptamers exhibited an almost identical sequence to clone R5-NMM-5 (Figure 3.7C).
Figure 3.5  Green-fluorescence profile of the screenings performed in the presence of NMM (left-panel) or TO3-Biotin (right panel).

The green fluorescence of 50,000 droplets was used to build each profile. Red bars indicate the limit over which droplets were gated and sorted as positive. The selection conditions (TO1-Biotin and NMM/TO3-Biotin concentrations as well as the temperature of the analysis device) are indicated. In TO3-Biotin screening, the two last experiments (J and K) correspond to the same experiment performed respectively at 45°C (J) and 25 °C (K).
Figure 3.6  Resistance or selectivity of TO1-Biotin binding variants in presence of competitors after each round of screening

(A) NMM resistance selection: Resistance of TO1-Biotin/RNA complexes to NMM. The fluorescence of the complex between TO1-Biotin and the RNAs from the libraries obtained after each round of screening was determined by mixing 2 µM RNA and 100 nM TO1-Biotin in the absence or in the presence of 3 µM NMM. The Resistance to NMM was calculated by normalizing the aptamer/TO1-Biotin fluorescence in the presence of NMM by the aptamer/TO1-Biotin fluorescence in the absence of NMM. (B) TO3-Biotin resistance selection: Resistance of TO1-Biotin/RNA complexes to TO3-Biotin. The fluorescence of the complex between TO1-Biotin and the RNAs from the libraries obtained after each round of screening was determined by mixing 300 nM RNA and 100 nM TO1-Biotin, in the absence or in the presence of 1.6 µM TO3-Biotin. The resistance to TO3-Biotin was calculated by normalizing the aptamer/TO1-Biotin fluorescence in the presence of TO3-Biotin by the aptamer/TO1-Biotin fluorescence in the absence of TO3-Biotin. (C) Selectivity of the libraries obtained after the different rounds of screening in the presence of TO3-Biotin. The green fluorescence of the TO1-Biotin/RNA complex was normalized to the red fluorescence of the TO3-Biotin/RNA complex to calculate the selectivity index. The blue bar indicates the concentration of competitor used during the screening step. Data was obtained at 25 °C, the values are the mean of three independent experiments and error bars correspond to ± 1 standard error.
Figure 3.7  Analysis of the clones obtained at the end of the screening process performed in the presence of NMM

(A) Brightness of the complexes formed between TO1-Biotin and individual variants isolated from the screenings in the presence of NMM. Aptamer-coding genes were PCR amplified, *in vitro* transcribed in the presence of TO1-Biotin and the fluorescence was monitored at 37 °C. The maximal fluorescence was normalized to that of Mango I. (B) Resistance of TO1-Biotin/RNA complex to NMM. TO1-Biotin fluorescence was monitored as in (A) in the absence or in the presence of 3 μM NMM. (C) Sequence analysis of the clones of interest. The sequences of the clones of interest (indicated by an asterisk in A and B) were aligned with Clustal X. The green boxes indicate Mango III (R5-NMM-20) and Mango IV (R5-NMM-5).

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 (Figure 3.3). 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 (Figure 3.5B). In subsequent rounds, TO3-Biotin concentration was gradually increased (Figure 3.2D and Table 3.2).

Table 3.2  Metrics of screenings in the presence of TO3-Biotin

<table>
<thead>
<tr>
<th>Round</th>
<th>Temperature (°C)</th>
<th>TO3-Biotin (nM)</th>
<th>TO1-Biotin (nM)</th>
<th>λ</th>
<th>Fusion efficiency (%)</th>
<th># of analyzed droplets</th>
<th># of analyzed variants</th>
<th># of sorted droplets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>45</td>
<td>0</td>
<td>100</td>
<td>1.25</td>
<td>95</td>
<td>2,328,630</td>
<td>2,765,248</td>
<td>4,607</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>50</td>
<td>100</td>
<td>0.20</td>
<td>80</td>
<td>1,106,750</td>
<td>177,080</td>
<td>8,328</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>75</td>
<td>100</td>
<td>0.15</td>
<td>94</td>
<td>1,019,990</td>
<td>143,818</td>
<td>6,446</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>100</td>
<td>100</td>
<td>0.20</td>
<td>95</td>
<td>1,002,500</td>
<td>190,475</td>
<td>11,478</td>
</tr>
<tr>
<td>5</td>
<td>45</td>
<td>200</td>
<td>100</td>
<td>0.17</td>
<td>90</td>
<td>1,044,500</td>
<td>159,808</td>
<td>12,205</td>
</tr>
<tr>
<td>6</td>
<td>45</td>
<td>400</td>
<td>100</td>
<td>0.25</td>
<td>85</td>
<td>1,015,880</td>
<td>215,874</td>
<td>8,779</td>
</tr>
<tr>
<td>7</td>
<td>25</td>
<td>800</td>
<td>100</td>
<td>0.20</td>
<td>79</td>
<td>1,087,500</td>
<td>195,750</td>
<td>8,463</td>
</tr>
<tr>
<td>8</td>
<td>45</td>
<td>1,600</td>
<td>100</td>
<td>0.26</td>
<td>88</td>
<td>768,250</td>
<td>175,775</td>
<td>4,729</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
<td>3,200</td>
<td>100</td>
<td>0.28</td>
<td>75</td>
<td>547,250</td>
<td>114,922</td>
<td>5,868</td>
</tr>
<tr>
<td>10</td>
<td>45</td>
<td>0</td>
<td>25</td>
<td>0.15</td>
<td>70</td>
<td>499,500</td>
<td>52,447</td>
<td>4,004</td>
</tr>
</tbody>
</table>

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 (Figure 3.2D). In later rounds, the relative fluorescence increased progressively until the TO3-Biotin concentration exceeded TO1-B by 32-fold (3.2 µM 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 (Figure 3.6B,C). Surprisingly, we found that, starting from round three; the libraries were dominated by a single cluster of sequences (cluster E, Figure 3.8) 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 (Figure 3.9A).

Figure 3.8  Sequence analysis of the variants isolated across the rounds of screening performed in the presence of TO3-Biotin
The first green box indicates Mango II (R2-1). For comparison, the sequence of Mango III (R5-NMM-20) was also added and boxed in green. The families of sequences clustering together are labelled (A to E).
Figure 3.9 Mutations and truncations of Mangos II, III, and IV.

(A) Mango II constructs. (B) Mango IV constructs. (C) Mango III constructs. $F_E$ is relative to the full-length construct, which was normalized to one. Constructs with binding affinities higher than the end-point of titration are labelled 'u.d.' (undeterminable). The closing stem regions are highlighted in purple. Guanine residues protected from DMS cleavage in the named Mango constructs of the study (Figure 3.10C) are highlighted in yellow.

3.2.2. 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 (Figure 3.7, Figure 3.8, Figure 3.9), we engineered the minimal reference constructs Mango II, Mango III and Mango IV (Figure 3.1), respectively, by truncation and sequence manipulation while maintaining the binding and fluorescent properties of the parental constructs (Figure 3.9). Mango II, III, and IV were found to be 1.5-, 4-, and
3-fold brighter than Mango I, respectively (Figure 3.10A). Mango II binds TO1-B with subnanomolar affinity, while Mango III and IV had slightly weaker affinities than Mango I (Figure 3.10A). 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 (Figure 3.10B). Notably, the brightness of the Mango III and Mango IV TO1-B bound complexes are 43,000 M⁻¹cm⁻¹ and 32,000 M⁻¹cm⁻¹, respectively, making Mango III 1.3 times brighter than enhanced-GFP (EGFP) a common benchmark for the characterization of improved fluorescent proteins¹¹⁹.

Mango I binds TO1-B by sandwiching it between the T3 layer of the G4 and A25 and A30 (Figure 3.1, yellow and red residues respectively)³⁶. This fluorophore binding core is isolated from an arbitrary RNA duplex (Figure 3.1, purple residues) by a GAAA tetraloop-like adapter¹⁴⁸ that inserts the TO1-B binding core between the third and fourth residues of the tetrancleotide motif (Figure 3.1, cyan residues)³⁶. Like Mango I, all of the new Mango aptamers are contained within an arbitrary closing stem (Figure 3.9). DMS probing, which correctly confirmed the three-tiered quadruplex structure of Mango I, indicates that Mango II also contains a three-tiered quadruplex structure (Figure 3.10C). In distinct contrast to Mango I, which has a contiguous run of protected three guanine residues forming G-stack 1 (Figure 3.1, 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 (Figure 3.10C, '<' symbol). In addition to this change, Mango II has a dinucleotide adenine in its third propeller loop region (Figure 3.10C), 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, Figure 3.1, Figure 3.10C, 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 (Figure 3.9A).
Figure 3.10  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 panel B but using 20 nM TO3-Biotin. Data for panels 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 (terminal $^{32}$PpCp 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$_2$O 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 Figure 3.1.

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 (Figure 3.10C, ‘<’ 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 (Figure 3.10C, * 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 (Figure 3.1, blue residues), that 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 (Figure 3.9B, Variant 20). This data, together with additional point mutational analysis, indicates 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 (Table 3.3, Figure 3.11) 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.

<table>
<thead>
<tr>
<th>Mango</th>
<th>Hill Coefficient</th>
<th>$K_0$ (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hill Coefficient</td>
<td>Na*</td>
</tr>
<tr>
<td>I</td>
<td>1.2 ± 0.1</td>
<td>N/A</td>
</tr>
<tr>
<td>II</td>
<td>1.4 ± 0.2</td>
<td>0.8 ± 0.1</td>
</tr>
<tr>
<td>III</td>
<td>1.0 ± 0.8*</td>
<td>0.5 ± 0.5*</td>
</tr>
<tr>
<td>IV</td>
<td>1.5 ± 0.2*</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Errors are the standard deviation of three independent measurements.

* Hill coefficients with an asterisk are estimated based on initial rise of fluorescence data.

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 (Figure 3.11). 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 (Figure 3.10C). 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$^{112}$ (Figure 3.12). Consistent with the formation of a G4 structure in Mango I, II, and IV the ligand bound CD spectra for each aptamer were quite similar (Figure 3.13).

**Figure 3.11  Salt dependence of new Mango variants**
Dependence of fluorescence of each Mango for (A) $K^+$, (B) $Na^+$ in place of $K^+$, and (C) $Mg^{2+}$ ions in a buffer containing 140 mM $K^+$. Each salt was titrated holding 25 nM RNA and 50 nM TO1-Biotin constant. 10 mM Tris buffer (pH 7.2) was used in place of phosphate of the WB buffer to avoid monovalent counter ions. Colour-coding is as follows: Black – Mango I, Blue – Mango II, Green – Mango III, Orange – Mango IV. When possible, data is fitted to the Hill equation and Hill coefficients are listed in Table 3.3. Hill coefficients are fitted to dark points. The detailed source of this inhibition in fluorescence has not been characterized. Points in lighter shade have been excluded from the fit. Error bars are standard deviations of three replicates.
Figure 3.12  Temperature-dependent fluorescence and UV absorbance spectroscopy of new Mango variants

Line plots of temperature-dependent spectroscopy for Mango I (data obtained from previous work), Mango II, Mango III, and Mango IV. Left panels: 1 µM RNA was incubated with 5 µM TO1-Biotin and subjected to temperature ramps while monitoring fluorescence (red shades). Right panels: A$_{260}$ for this sample with 5 µM TO1-Biotin (green shades) and without TO1-Biotin (blue shades) were collected and the simple derivative plotted together with the derivative of the fluorescence data. Starting at 90 °C, temperature was ramped down at a rate of 1 °C/min to 20 °C (darker shade) and returned to 90 °C at a rate of 1 °C/min (lighter shade).
Figure 3.13  Circular dichroism spectra
Circular dichroism spectra of Mango I, II, III, and IV. 5 µM RNA was measured in 10 mM Tris pH 7.5 buffer either alone, with 140 mM monovalent salt, and/or with 7 µM TO1-Biotin as indicated by the legend. Data is a line plot to guide the eye.

Mango III, the brightest of the three constructs, contains only nine guanines in its core, and is therefore unable to form a three-tiered G4. All nine core guanines are DMS protected (Figure 3.10C). Mutant analysis suggests that the helical region of Mango III is likely to extend an additional three bp into its core (Figure 3.1, Figure 3.10C, 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 (Figure 3.9C, 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% (Figure 3.9C, 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 (Figure 3.1) and has a ~100-fold higher affinity for potassium and sodium, while being only modestly inhibited by high levels of magnesium (Figure 3.11). Its sigmoidal fluorescent melting curve resembles the melting of RNA Spinach2^{113} and not the more linear melting curves observed for Mango I, II, and IV (Figure 3.12). Similarly, the CD spectrum of the bound Mango III complex is different in the 270-300 nm region from the other Mango constructs tested (Figure 3.13). 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 (Figure 3.14). 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.

**Figure 3.14** Excitation/emission of the new Mangos.
Excitation (dark curve) and emission (light curve) spectra of each Mango. All Mangos have $\lambda_{\text{ex}}$ max = 510 nm and $\lambda_{\text{em}}$ max = 535 nm.
3.2.3. Cellular imaging of Mango-tagged RNA

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 (Figure 3.15A) previously shown to improve cellular fluorescence and RNA stability\(^{104,149}\). Each Mango-tagged RNA, with or without the folding scaffold or terminator hairpin, exhibited comparable fluorescence intensities \textit{in vitro} (Figure 3.15B). 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 G4 mutant Mango sequence (Figure 3.15B and Table 3.4).

**Figure 3.15**  RNA scaffold diagrams.  
Diagrams of the 5S-Mango and U6-Mango constructs synthesized (sequences in Table 3.4) and their fluorescence emission as a function of wavelength after excitation at 505 nm, compared to unmodified Mango I-IV.
### Table 3.4 Constructs used in *in vivo* experiments

<table>
<thead>
<tr>
<th>Construct</th>
<th>Sequence (5' → 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>5S-30-Control</td>
<td>GUCUACGGCC AUACCAACCU GAACGGCCCU GAUCUGCUCU GAUCUGCUCU GAAUACGAGG GUGCAGCCCU UUGUAGACU</td>
</tr>
<tr>
<td></td>
<td>GAUCUGCAGA GCUCGCAACG GUCUGCAGCCCU GAUGAGUCU GAUCUGCAGG GUGCAGCCCU GUGUAGACU</td>
</tr>
<tr>
<td></td>
<td>GAAUGGGAAG ACCCUGGCAG GUGAGGCAGG GAUCUGCAGG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>CGACUCUGCA GUGUAGACU GUGCAGCAAG GACGUGCCAG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>UAGAUAUGAG GAUCUAACAU CAUGCUGAUCU CAUGCUGAUCU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td>5S-F30-Mango II</td>
<td>GUCUACGGCC AUACCAACCU GAACGGCCCU GAUCUGCUCU GAUCUGCAGG GUGCAGCCCU GAUGAGACU</td>
</tr>
<tr>
<td>mutant</td>
<td>GAUCUGCAGA GCUCGCAACG GUCUGCAGCCCU GAUGAGUCU GAUCUGCAGG GUGCAGCCCU GUGUAGACU</td>
</tr>
<tr>
<td></td>
<td>GAAUGGGAAG ACCCUGGCAG GUGAGGCAGG GAUCUGCAGG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>CGACUCUGCA GUGUAGACU GUGCAGCAAG GACGUGCCAG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>UAGAUAUGAG GAUCUAACAU CAUGCUGAUCU CAUGCUGAUCU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td>5S-F30-Mango I</td>
<td>GUCUACGGCC AUACCAACCU GAACGGCCCU GAUCUGCUCU GAUCUGCAGG GUGCAGCCCU GAUGAGACU</td>
</tr>
<tr>
<td></td>
<td>GAUCUGCAGA GCUCGCAACG GUCUGCAGCCCU GAUGAGUCU GAUCUGCAGG GUGCAGCCCU GUGUAGACU</td>
</tr>
<tr>
<td></td>
<td>GAAUGGGAAG ACCCUGGCAG GUGAGGCAGG GAUCUGCAGG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>CGACUCUGCA GUGUAGACU GUGCAGCAAG GACGUGCCAG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>UAGAUAUGAG GAUCUAACAU CAUGCUGAUCU CAUGCUGAUCU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td>5S-F30-Mango II</td>
<td>GUCUACGGCC AUACCAACCU GAACGGCCCU GAUCUGCUCU GAUCUGCAGG GUGCAGCCCU GAUGAGACU</td>
</tr>
<tr>
<td></td>
<td>GAUCUGCAGA GCUCGCAACG GUCUGCAGCCCU GAUGAGUCU GAUCUGCAGG GUGCAGCCCU GUGUAGACU</td>
</tr>
<tr>
<td></td>
<td>GAAUGGGAAG ACCCUGGCAG GUGAGGCAGG GAUCUGCAGG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>CGACUCUGCA GUGUAGACU GUGCAGCAAG GACGUGCCAG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>UAGAUAUGAG GAUCUAACAU CAUGCUGAUCU CAUGCUGAUCU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td>5S-F30-Mango III</td>
<td>GUCUACGGCC AUACCAACCU GAACGGCCCU GAUCUGCUCU GAUCUGCAGG GUGCAGCCCU GAUGAGACU</td>
</tr>
<tr>
<td></td>
<td>GAUCUGCAGA GCUCGCAACG GUCUGCAGCCCU GAUGAGUCU GAUCUGCAGG GUGCAGCCCU GUGUAGACU</td>
</tr>
<tr>
<td></td>
<td>GAAUGGGAAG ACCCUGGCAG GUGAGGCAGG GAUCUGCAGG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>CGACUCUGCA GUGUAGACU GUGCAGCAAG GACGUGCCAG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>UAGUAGAGAG AAGAUUAGAU AAGAUUAGAU AAGAUUAGAU AAGAUUAGAU AAGAUUAGAU AAGAUUAGAU</td>
</tr>
<tr>
<td>5S-F30-Mango IV</td>
<td>GUCUACGGCC AUACCAACCU GAACGGCCCU GAUCUGCUCU GAUCUGCAGG GUGCAGCCCU GAUGAGACU</td>
</tr>
<tr>
<td></td>
<td>GAUCUGCAGA GCUCGCAACG GUCUGCAGCCCU GAUGAGUCU GAUCUGCAGG GUGCAGCCCU GUGUAGACU</td>
</tr>
<tr>
<td></td>
<td>GAAUGGGAAG ACCCUGGCAG GUGAGGCAGG GAUCUGCAGG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>CGACUCUGCA GUGUAGACU GUGCAGCAAG GACGUGCCAG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>UAGAUAUGAG GAUCUAACAU CAUGCUGAUCU CAUGCUGAUCU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td>5S Δ78-98-F30-Mango IV</td>
<td>GUCUACGGCC AUACCAACCU GAACGGCCCU GAUCUGCUCU GAUCUGCAGG GUGCAGCCCU GAUGAGACU</td>
</tr>
<tr>
<td></td>
<td>GAUCUGCAGA GCUCGCAACG GUCUGCAGCCCU GAUGAGUCU GAUCUGCAGG GUGCAGCCCU GUGUAGACU</td>
</tr>
<tr>
<td></td>
<td>GAAUGGGAAG ACCCUGGCAG GUGAGGCAGG GAUCUGCAGG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>CGACUCUGCA GUGUAGACU GUGCAGCAAG GACGUGCCAG GUGUAGACU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td></td>
<td>UAGAUAUGAG GAUCUAACAU CAUGCUGAUCU CAUGCUGAUCU GAUCUGCAGG GUGCAGCCCU</td>
</tr>
<tr>
<td>U6-Control</td>
<td>GUGUCUGCAC GUGGAGACCA GAUCAGUCUU GAUCAGUCUU GAUCAGUCUU GAUCAGUCUU GAUCAGUCUU</td>
</tr>
<tr>
<td></td>
<td>ACAGAGAAGU UUGACGACG CAGCAGCAGC GAGAUGACG GAGAUGACG</td>
</tr>
<tr>
<td></td>
<td>CAUACUUGCG CAGGUCUCA CAGGUGUCA CAGGUGUCA CAGGUGUCA CAGGUGUCA CAGGUGUCA</td>
</tr>
<tr>
<td>U6-Mango IV</td>
<td>GUGUCUGCAC GUGGAGACCA GAUCAGUCUU GAUCAGUCUU GAUCAGUCUU GAUCAGUCUU GAUCAGUCUU</td>
</tr>
<tr>
<td></td>
<td>ACAGAGAAGU UUGACGACG CAGCAGCAGC GAGAUGACG GAGAUGACG</td>
</tr>
<tr>
<td></td>
<td>CAUACUUGCG CAGGUCUCA CAGGUGUCA CAGGUGUCA CAGGUGUCA CAGGUGUCA CAGGUGUCA</td>
</tr>
<tr>
<td></td>
<td>UCCUAUUUUU UCCUAUUUUU UCCUAUUUUU UCCUAUUUUU UCCUAUUUUU UCCUAUUUUU UCCUAUUUUU</td>
</tr>
</tbody>
</table>
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 (Figure 3.16A).

Figure 3.16  Effect of formaldehyde and pH on Mango fluorescence.
(A) Fluorescence of 50 nM RNA Mango I, II, III, and IV with 100 nM TO1-Biotin are measured at 30 second intervals at 25 °C. In i) RNA is first incubated with TO1-Biotin for one hour before addition of corresponding formaldehyde amounts and fluorescence is measured for another 100 minutes. In ii) RNA is incubated with the corresponding amounts of formaldehyde first for one hour, then TO1-Biotin was added. Data for each panel is normalized to RFU at the 0 % formaldehyde, 0 min point. (B) pH titration for 50 nM Mango (Mango I – black, Mango II – blue, Mango III – green, Mango IV – orange) and 100 nM TO1-Biotin. The fluorescence of each Mango with TO1-Biotin was measured as a function of pH. 50 mM Sodium Citrate buffer for pH 3-6 (open circles), 50 mM Sodium Phosphate for pH 6-8 (filled diamonds, 50 mM Tris for pH 8-9 (open
squares). RFU is normalized to 1 for Mango I at pH 7. Errors are standard deviations of 3 replicates.

Up to ~10 bright RNA Mango foci could be readily detected per cell with a fluorescence microscope, but not in control transfections (Figure 3.17A). A time course of this process (Figure 3.18A) 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 (Figure 3.16B). 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 (Figure 3.12). Consistent with 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 (Figure 3.10C). The mean intensity of the 5S-F30-Mango IV foci was two to three-fold higher than 5S-F30-Control background (Figure 3.17B). 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 sub-cellular 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, Figure 3.17C). In addition, we observe that the cytoplasmic 5S-Mango IV foci overlap with immunostained mitochondria (ATP5B, Figure 3.18C, Figure 3.19A), and not with constructs lacking Helix IV (Figure 3.20C,E), consistent with the observation that Helix IV of 5S rRNA mediates its import into mitochondria\textsuperscript{150}. Conversely, we do not observe overlap with other sub-cellular compartments, such as P-bodies, Endosomes, or Stress Granules, where the transfected RNA could be processed for degradation (Figure 3.19B).
Figure 3.17  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, * denotes p < 0.05 and ** means 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.
Figure 3.18  Formation of 5S-Mango foci and their dynamics in live cells.
(A) 5S-Mango IV imaged post cell fixation at 5, 15, 30 and 60 min after transfection using Lipofectamine based CRISPRMAX transfection reagent. (B) Live-cell imaging of 5S-Mango III 60 min post transfection and the RNA (125nM) was pre-incubated with TO1-Biotin (250 nM) prior to transfection. (C) Root mean-squared displacement (RMSD) of three observed foci with different diffusive behaviors, fast (Green), slow (Blue) and static (Red). Scale bars are 10 µm and arrows indicate foci of interest. (D) Normalized photobleaching traces of 5S-F30-Mango I and IV compared to 5S-F30-dBroccoli (20 µM DFHBI-1T) under constant illumination with an exposure time of 200 ms. A single exponential fit yields the photobleaching half-lives: 5.7 ± 0.2 s (Mango I), 3.7 ± 0.1 s (Mango IV) and 0.30 ± 0.01 s (dBroccoli).
Figure 3.19  Co-Immunofluorescence with direct RNA-Mango transfections.
(A) Cytoplasmic 5S-Mango IV foci are observed to localize with immunostained mitochondria, whereas cytoplasmic U6-Mango foci do not. Cells fixed and stained with TO1-Biotin (200nM – Green), Mitochondrial antibody (ATP5B – Red) and Hoechst 33258 (1µg/ml – Blue). Normalized fluorescence intensities as a function of distance along the inset white lines (bottom panels). (B) Localization of 5S-Mango IV and U6-Mango IV foci relative to immunostained P-Bodies (GW182), Endosomes (EEA-1) and Stress Granules (TIAR). Scale bars are 10 µm. All images are maximum projections except in A, which show a single focal plane.
Figure 3.20 2D immunostaining co-localization plots
Normalized pixel-by-pixel intensity plots of the TO1-Biotin signal vs. immunostaining signal for both Mitochondria (ATP5B, A-E) and snRNPs (LSm3, F-I). Background intensity values were determined either with a Mango IV-tagged construct in the absence of immunostaining or a control construct in the presence of both TO1-Biotin and the appropriate immunostain (A, B, F and G). All plots were normalized by subtracting the background signal in each channel and then normalizing to the highest significant pixel intensity from each of the channels. (C-E) Show the co-localization patterns of Mango tagged 5S, U6 and 5S Δ78-98 with Mitochondria. (H and I) Show co-localization patterns of Mango tagged U6 and 5S with snRNPs. The upper quadrants highlight the number of pixels contained within and therefore depict the co-localized (upper right) and distinct (upper left) Mango signal. Each plot contains multiple slices taken from ~five images with dimensions 1280x1280 to accurately determine co-localizing pixels. The number of cells for plots A, B, C, D, and E were 66, 26, 178, 217 and 156 respectively. The number of cells for plots F, G, H and I were 34, 108, 165, and 89, respectively.

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 (Figure 3.15, Table 3.4), known to be amenable to modification\textsuperscript{151-153}. The resulting construct exhibits comparable fluorescence intensity to Mango IV alone \textit{in vitro} (Figure 3.15). Direct transfection of U6-Mango IV snRNA yields fluorescent foci comparable to 5S-F30-Mango IV (Figure 3.17A), albeit with lower intensity (Figure 3.17B). As expected, the fraction of nuclear foci increased ~9-fold, while cytoplasmic foci decrease significantly (p < 0.01, Figure 3.17B). 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 (Figure 3.17C), as expected. Similar to 5S-F30-Mango IV, U6-Mango IV foci do not overlap with other sub-cellular compartments, such as P-bodies, Endosomes, or Stress Granules (Figure 3.19B). To further quantify the proportion of nuclear and cytoplasmic Mango tagged RNA, we created 2D intensity plots for each pixel from multiple images (Figure 3.21). 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 (Figure 3.20A,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-\textsymbol{\Delta}78-98-F30-Mango IV shown to be deficient in mitochondrial import\textsuperscript{150} (Figure 3.20C,D,E). Also, consistent with the images shown in Figure 3.17, 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 (Figure 3.20F,G,H,I). Taken together, these fixed cell data show that Mango IV can be used to label and image small cellular RNAs \textit{via} direct transfection \textit{of in vitro} transcribed RNAs, without affecting their expected sub-cellular localization.
Figure 3.21 2D nuclear co-localization plots
Pixel-by-pixel intensity plots of the TO1-Biotin signal (200 nM) vs Hoechst 33258 signal (1 µg/ml) for both 5S-Mango IV and U6-Mango IV compare to their respective controls in fixed cells. For the Mango specific signal, a threshold was set above the 5S-F30-Control background of 600 a.u. Whereas the threshold of the nuclear boundary is observed to be ~2000 a.u. based on Hoechst 33258 staining. The upper quadrants highlight the number of pixels contain within and therefore depict pixels observed outside (upper left) or inside the nucleus (upper right). Each plot contained five maximum projection images with dimensions 1280x1280. 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.

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 (Figure 3.19B). The foci observed were similar to those in Figure 3.17A, albeit with a lower background fluorescence in the nucleolus of fixed cells. The foci can be readily tracked revealing three distinct diffusive behaviors (fast, slow, and static) and their respective root mean square displacement coefficients could be quantified (Figure 3.18C). 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 (Figure 3.18D). These data are in good agreement with in vitro photobleaching analysis in aqueous droplets (Figure 3.22A). Under pulsed illumination (200 ms, 0.2 Hz), the lifetime of Mango RNA aptamers in fixed cells increases by >60-fold (from 11.7 s to >10 min, Figure 3.22B), as previously observed for the Spinach aptamer\textsuperscript{132}.

![Photobleaching curves](image)

**Figure 3.22 Photostability of new Mango variants in *in vitro* and in fixed cells.** (A) *In vitro* transcribed and purified 5S-F30-Broccoli or 5S-F30-Mango I to IV were incubated with their cognate dye (i.e. DFHBI-1T or TO1-Biotin) for an hour at room temperature prior to being encapsulated into an emulsion later loaded into a glass capillary and imaged. Each complex was exposed to a constant illumination their maximum excitation wavelength (i.e. 470 nm for Broccoli/DFHBI-1T and 508 nm for Mangos/TO1-Biotin) and their fluorescence emission was recorded respectively at 514 ± 24 nm and 540 ± 12 nm. All the values were background subtracted, relativized to the first measurement point and an exponential decay was fit to the data to compute the fluorescence half-life ($t_{1/2}$) of each complex (values indicated on the plot). A zoom on the first 10 seconds is shown in the inset. The values are the mean of three independent experiments and the error bars correspond to standard deviation. (B) Fluorescence stability of nucleoli signal in pSLQ-5S-F30-Mango IV expressing cells under constant illumination (dark orange) or pulsed excitation (200 ms, 0.2 Hz, light orange) with a 488 nm laser. Cells were fixed and stained with 200 nM TO1-Biotin. Constant illumination gave a $t_{1/2} = 11.7 ± 0.2$ s whereas pulsed illumination significantly increases the lifetime >60 fold.
To estimate the number of fluorescent 5S-F30-Mango IV molecules in each foci, we performed photobleaching-assisted microscopy on fixed cells (Figure 3.23A). A maximum likelihood estimate analysis of the photobleaching trajectories\textsuperscript{154,155}, 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 (Figure 3.23B). The number of observed steps correlates linearly with the initial foci intensity below 2,000 intensity units (Figure 3.23C). Altogether, these results indicate that each foci contains at least 4 to 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.

**Figure 3.23**  RNA Mango stoichiometry of individual foci.
(A) Representative fluorescent intensity trace (yellow) of a single 5S-F30-Mango IV foci from a fixed cell undergoing photobleaching under constant illumination (50 ms time resolution). Different intensity states (dashed line) are identified by maximum likelihood estimation, as described\textsuperscript{154,155}. Between 4 and 17 photobleaching steps were observed for eleven different foci from fixed cells, indicating the number of bright 5S-F30-Mango IV molecules in each foci. (B) Distribution of photobleaching step sizes reveals two main peaks at ~40 and ~90 fluorescence intensity units, corresponding to photobleaching of one or two 5S-F30-Mango IV molecules. (C) Number of photobleaching steps per foci as a function of the initial intensity. Below 2,000 fluorescence units, the number of steps is proportional to the initial foci intensity (dashed line). Above 2,000 fluorescence units, the number of steps is under estimated due to multiple simultaneous photobleaching steps in the initial decay. Therefore, the grey points were excluded from the linear fit (dashed line).
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 a mCherry reporter gene to identify successfully transfected cells (Figure 3.24A). 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 (Figure 3.24A, Figure 3.25A). 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 (Figure 3.24B).

Figure 3.24  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 Grayscale. 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 – log10 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 (Collin). Arrows depict significantly co-localized foci, scale bar = 10 µm.
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 (Figure 3.25A). Under the same conditions of fixation and staining, signal was not observed in untransfected cells or in cells expressing the 5S-F30-Broccoli construct (Figure 3.25B). The robust cytoplasmic signal observed after 12 h of pSLQ-5S-Mango IV expression enabled us to combine Mango imaging with immunofluorescence (Figure 3.24C). 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 (Figure 3.25C).

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\textsuperscript{156}. Upon expression, the Mango tagged mgU2-47 RNA formed well defined nuclear foci that were absent in the mgU2-47 Control RNA (Figure 3.24A,B). The nuclear foci also co-localized with immunostained Cajal-bodies (Figure 3.24C). Taken together these results demonstrate the ability of Mango tags to function as efficient genetically encoded reporters of RNA sub-cellular location.
Figure 3.25 Imaging of genetically encoded aptamer tagged 5S rRNA.
(A) Fixed time points of pSLQ-5S-F30-Mango IV expression, fixed and stained with 200 nM TO1-Biotin. TO1-Biotin – Yellow, mCherry – Red. (B) Fixed images of untransfected cells stained with 200 nM TO1-Biotin and pSLQ-5S-F30-Broccoli expressing cells stained with 10 µM DFHBI-1T. (C) Images of immunostained cells expressing 5S-F30-Mango IV and stained for stress granules (TIAR), P-bodies (GW182) and endosomes (EEA-1). TO1 signal – Green, Immunostain – Red and Hoechst 33258 (1 µg/ml) – Blue. All Scale bars = 10 µm.
3.3. 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 (Table 1.1, Figure 3.26)\textsuperscript{157,107,104,98,108,106}.

![Figure 3.26  Summary of contrast attributes of RNA aptamer-based fluorescent tracking systems.](image)

Fluorescence enhancement is plotted against binding affinity in nanomolar for RNA aptamer-based fluorescent tracking systems. Optimal contrast should have strong binding affinity (low dissociation constant) and high fluorescence enhancement. Amongst aptamer systems shown, the Mango family has the best contrast attributes.

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 \textasciitilde0.14. This
quantum yield is similar to that typically observed when thiazole orange is rigidified by intercalation into a duplex nucleic acid\textsuperscript{112,158}. Mango III and IV, when bound to TO1-B, have quantum yields of \( \sim 0.56 \) and 0.42 respectively. Such high quantum yields are comparable or exceed that of TO1-activating proteins\textsuperscript{159}. 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\textsuperscript{157}, the high brightness of the Mango aptamers combined with their extremely high binding affinity towards 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\textsuperscript{158} and RNA\textsuperscript{108,112}, 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\textsuperscript{37,107}. 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 sub-cellular 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 4 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 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. 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, combined with their unusual ability to withstand formaldehyde fixation all promise to be very useful for investigating biological systems in the future.

3.4. Materials & Methods

3.4.1. 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\(^{161}\) to have a final average number of DNA molecule per droplet (\(\lambda\)) of \(~0.2\) to \(1\) (Table 3.2, Table 3.5). 1 µL of this dilution was introduced in 100 µL of a PCR mixture containing 0.2 µM of Forward primer (5'-CTT TAC GAC TCA CTA TAG GAA CCC GCA AGC CAT C), 0.2 µM of Reverse primer (5'-CAG AAT CTC ACA CAG CC), 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 PTFE tubing (I.D. 0.75 mm tubing; Thermo Scientific) and infused into a droplet generator microfluidic device\(^{106}\) where it was dispersed into 2.5 pL droplets (production rate of \(~12,000\) droplets/second) carried by HFE 7500 fluorinated oil (3M) supplemented with 3% of a fluorosurfactant\(^{161}\). 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\(^{161}\) and subjected to an initial denaturation step of 2 min at 95 °C followed by 30 PCR cycles of: 30 sec at 95 °C, 30 sec at 55 °C, 1 min 30 at 72 °C.

<table>
<thead>
<tr>
<th>Round</th>
<th>Temperature (°C)</th>
<th>TO1-Biotin (nM)</th>
<th>(\lambda) value</th>
<th>Fusion efficiency (%)</th>
<th>Number of analyzed droplets</th>
<th>Number of analyzed variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>R12-library</td>
<td>25</td>
<td>100</td>
<td>1.25</td>
<td>95</td>
<td>168,441</td>
<td>200,023</td>
</tr>
<tr>
<td>Mango</td>
<td>25</td>
<td>100</td>
<td>0.54</td>
<td>95</td>
<td>63,200</td>
<td>1</td>
</tr>
</tbody>
</table>

ii. Droplet fusion: PCR droplets were then injected into a fusion device\(^{161}\) at a rate of \(~1,500\) droplets/second, spaced by a stream of HFE 7500 fluorinated oil supplemented with 2% fluorosurfactant. Each PCR droplet was synchronized with a 16 pL \textit{in vitro} transcription (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, 3,500 U T7 RNA polymerase (purified in the laboratory and estimated to have an activity around 2,500 U/µL by comparing it with commercial enzyme), 100 nM TO1-B\(^{108}\), 5 ng/µL inorganic pyrophosphatase (Roche) supplemented with the desired concentration of NMM. For the screenings performed in the presence of TO3-Biotin\(^{108}\), 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. 16 pL IVT droplets were produced at a rate of ~1,500 droplets per second 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)\textsuperscript{106}. 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\textsuperscript{106}. Droplets were re-injected at a frequency of ~200 droplets/second, 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 (3M) through the tubing. 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, the droplets broken by vortexing the mixture and DNA-containing aqueous phase was recovered.

3.4.2. 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 \textit{in vitro} transcription 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. 10 U of Baseline-Zero™ DNase (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 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 TimePCR Detection System (Bio-Rad). Finally, the cDNA was quantified using the calibration curve obtained with reactions performed with purified Mango II RNA.

3.4.3. 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 sec at 95°C, followed by 20 cycles of: 5 sec at 95 °C and 30 sec 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 MgCl2, 40 mM Tris-HCl pH 8.0, 5 mM DTT, 1 mM Spermidine and 70 µg/mL T7 RNA polymerase. After 4 hours 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, 1x 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 UV shadowing, sliced from the gel and transferred into dialysis tubing (MWCO = 3 500, 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 MgCl₂ 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 1.6 µM of TO3-Biotin. Both green (ex. 492 nm/em. 516 nm) and red (ex. 635 nm/em. 665 nm) fluorescence were measured.

### 3.4.4. 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. 2 µL of PCR product added to 18 µL of *in vitro* transcription mixture containing 2 mM of each NTP, 25 mM MgCl₂, 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 3005P, Agilent) for 2 hours at 37°C and the green fluorescence (ex. 492 nm/em. 516 nm) measured every minute.
3.4.5. DMS probing of Mango

DMS probing consists of four main steps i. DMS (Denaturing): Protocol is adapted from Lorsch and Szostak\textsuperscript{162}. 50 nM RNA was 3’ end labelled with \textsuperscript{32}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\textsubscript{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. 150 µL ice cold ethanol + 5 µL 3 M NaCl was then immediately added and pelleted as for the denaturing DMS protocol.

iii. Reduction: Pellets were resuspended 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\textsubscript{2}O, 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\textsubscript{2}O was added, the sample refrozen and lyophilized once again. The pellet was resuspended in a 50% formamide denaturing solution before being loading on a 15 % polyacrylamide gel (19:1 acrylamide:bis).

3.4.6. T1 RNase ladder and alkaline hydrolysis ladder

200 pmol 3’ end labelled RNA 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 0.17 M Tris-HCl.

### 3.4.7. Screening for minimal functional Mango motifs

To identify the minimal functional motif of each Mango, truncated constructs were designed as shown in Figure 3.9. 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.

### 3.4.8. Affinity measurements of Mango variants

Fluorescence data was 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$_2$PO$_4$ pH 7.2, 0.05% Tween-20) to determine binding affinities. Fluorescence was measured at the maximum excitation and emission wavelengths of each complex (Figure 3.14). Curves were fitted using least squares (Kaleidagraph 4.5) using the following equation for TO1-B:

$$ F = F_0 + \frac{F_{\text{max}}}{2} \left( K_D + [RNA] + [TO] - \sqrt{([RNA] - [TO])^2 + K_D (K_D + 2[RNA] + 2[TO])} \right) $$

*Equation 3.1*

Where $F$ is the fluorescence at a given $[RNA]$, $F_0$ is the unbound fluorescence and $F_{\text{max}}$ 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:

$$ F = F_0 + \frac{F_{\text{max}}[RNA]}{K_D + [RNA]} $$

*Equation 3.2*

$F_{\text{max}}$ was determined using Equation 3.1 or Equation 3.2, as appropriate.
3.4.9. Temperature-dependent fluorescence and UV melting curves

Temperature dependence measurements were started at 90 °C decreasing at a rate of 1 °C/min until 20 °C, then returned at 1 °C/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.

3.4.10. 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.).

3.4.11. Formaldehyde resistance assay

RNA Mango aptamers were incubated with TO1-B in WB buffer for at least one hour 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 minute using a Varian Cary Eclipse Fluorescence Spectrophotometer, ex/em = 510 ± 2.5/535 ± 5 nm.

3.4.12. Cell culture and maintenance

HEK293T cells (293T - ATCC® CRL-3216™) were grown in Dulbecco Modified Eagle’s Medium containing 10% Fetal 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% CO₂ in a humidified incubator. Cells used for imaging were cultured in Ibidi glass bottomed 8-well chamber slides (Ibidi GmbH).
3.4.13. **Plasmid and RNA synthesis**

DNA encoding the F30 folding scaffold\(^{149}\) was modified to incorporate the Mango RNA sequences (Table 2.1) and ordered from (Integrated DNA Technologies). The DNA was amplified by PCR to incorporate 5’ Sall and 3’ Xbal restriction sites (Table 3.6). PCR products were digested using Fast Digest enzymes (Thermo Fisher) and ligated into Sall/Xbal 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 (Table 3.6). 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 & 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 *E. 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 acrylamide: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 *in vitro* transcription and purification as described above.

**Table 3.6 Primers used for the construction of plasmids**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5′ Sall F30 primer</td>
<td>GGCCTCGACT TGCCATGTGT ATGTGGGTAC</td>
</tr>
<tr>
<td>3′ Xbal F30 primer</td>
<td>CGCTCTAGAT TGCCATGAAT GATCCCAGAG G</td>
</tr>
<tr>
<td>5′ BamHI T7-5S primer</td>
<td>GCCGGATCCCT AATACGACTC ACTATAGCT ACGGCCATAC CACCC</td>
</tr>
<tr>
<td>5′ BstXI 5S primer</td>
<td>TTGGAGAAACC ACCTTGTTGG GTCTACGCCC ATACACCC</td>
</tr>
<tr>
<td>3′ XhoI F30 Rev primer</td>
<td>CTACTCAGAG AAAAAATTGC CATGAATCAT CCCGAAGG</td>
</tr>
<tr>
<td>5′ BstXI mgU2-47</td>
<td>TTGGAGAAACC ACCTTGTTGG AGCAAGTGA TGAGTAATAC TGGCTG</td>
</tr>
</tbody>
</table>
3.4.14. **Plasmid and RNA transfection**

RNA was transfected directly into 8-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 MgCl₂ 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 hour in complete growth medium. FuGene 6 was used to transfec 400 ng of the pSLQ-5S-F30-Mango and control plasmids directly in the 8-well chamber slides following the manufacturer’s instructions. Plasmids were left to express between 12 – 48 h before fixation as described below.

3.4.15. **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 MgCl₂) 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 MgCl₂ 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 – 8µg/ml), Anti-Ribosomal Protein L7 (ab72550, Abcam – 1µg/ml), Anti-ATP5B (ab14730, Abcam – 1µg/ml), Anti-GW182 (ab7052, Abcam – 5µg/ml), Anti-EEA-1 (ab70521, Abcam – 1µg/ml), Anti-LSm3 (NBP2-14206, Novus Biologicals – 1µg/ml), Anti-TIAR (sc-398372, Santa Cruz – 4µg/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.
3.4.16. Fluorescence microscopy and live cell imaging

Live and fixed cell images were taken directly in the 8-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 8-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% CO₂ in a stage top incubator (Tokai Hit).

3.4.17. 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. 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\textsuperscript{154,155}. The step sizes were subsequently binned and the histogram was fit to a double Gaussian equation.

3.4.18. 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 ≥ 3x3 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 (Figure 3.17B). To create the 2D co-localization plots (Figure 3.21, Figure 3.20), 5-6 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 heat-map.

3.4.19. pH titrations

50 nM Mango was incubated with 100 nM TO1-B in the presence of 140 mM KCl, 1 mM MgCl\textsubscript{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 hour at room temperature. Mango fluorescence was measured with excitation and emission at 485 and 535 nm respectively.

3.4.20. 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, \textit{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\textsubscript{2}, 20 mM Na\textsubscript{2}H\textsubscript{2}PO\textsubscript{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 (ThermoFisher) and infused into a droplet generator.
microfluidic device where it was dispersed into 100 pL droplets carried by HFE 7500 fluorinated oil (3M) supplemented with 3% of a fluorosurfactant as described previously\textsuperscript{163}. 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 20x objective (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.
Chapter 4.

Structure studies of Mango I and III, and studying orientation-dependent FRET with Mango III

*The sum of the square roots of any two sides of an isosceles triangle is equal to the root of the remaining side!*

*H.J.S/Scarecrow*

This chapter is based on the following:


The sections on detailed FRET studies were taken from unpublished work by S.C.Y.J. and P.J.U. I performed all FRET work including data collection, theoretical structure modelling and scripting. This work was made possible by the aforementioned work on the crystal structures of Mango I and Mango III.
4.1. Introduction

Förster resonance energy transfer (FRET) is an effect of non-radiative energy transfer when transition dipoles of two fluorescent entities are coupled\textsuperscript{164}. FRET is frequently applied when studying biological interactions. The distance dependence of FRET lends itself to act as a spectroscopic ‘ruler’, to study distances and conformations of biomolecules such as molecular interactions in a cell\textsuperscript{165}, molecular conformation studies\textsuperscript{166}, and can be applied to the design of FRET biosensors\textsuperscript{167} to name a few.

In most applications, the orientation of the transition dipoles is ignored as the molecules in question are typically orientationally unconstrained. However, in principle, the efficiency of energy transfer (\(E_{\text{FRET}}\)) should depend on the distance as well as the orientation of the transition dipoles of the two fluorophores\textsuperscript{168–172}.

\(E_{\text{FRET}}\) is related to the sixth power of the distance between the two molecules, and is described as:

\[
E_{\text{FRET}} = \frac{1}{1 + (\frac{R}{R_0})^6}
\]

\textit{Equation 4.1}

Where \(R\) is the distance between the two fluorophores and the Förster distance \((R_0)\) is the distance where \(E_{\text{FRET}}\) is 50%. \(R_0\) is calculated by:

\[
R_0 = \frac{9000(\ln 10)\kappa^2 Q_D J}{128\pi N n^4}
\]

\textit{Equation 4.2}

where \(Q_D\) is the quantum yield of the donor fluorophore, \(N\) is Avogadro’s number, \(n\) is the index of refraction of the medium, \(J\) is the spectral overlap between the fluorophores and \(\kappa^2\) is the orientation factor. \(J\) is a constant that depends upon the shape of the emission spectra of the two fluorophores and the extinction coefficient of the acceptor.
\[ J = \int_{0}^{\infty} \varepsilon_A(\lambda) F_D(\lambda) d\lambda \]

Equation 4.3

where \( \varepsilon_A(\lambda) \) is the extinction coefficient excitation spectrum of the acceptor, \( F_D(\lambda) \) is the emission spectrum of the donor normalized to unity.

Of the above, all except for \( \kappa^2 \) are known constants, or can be calculated based on the properties of the two fluorophores. Thus, with a given FRET pair, \( E_{\text{FRET}} \) depends only on the distance \( R \), and the orientation factor \( \kappa^2 \). This factor is related to the orientation of the transition dipoles in the fluorophore partners. Qualitatively, FRET is most optimal when the dipoles are collinear, and least optimal when they are perpendicular. \( \kappa^2 \) takes on values between 0 (\( \uparrow, \rightarrow \)) and 4 (\( \rightarrow, \rightarrow \)) calculated by Equation 4.4. In a setting where fluorophore orientation is rotationally averaged, many assume a \( \kappa^2 \) of 2/3:

\[ \kappa^2 = (\hat{d} \cdot \hat{a} - 3(\hat{d} \cdot \hat{R})(\hat{a} \cdot \hat{R}))^2 \]

Equation 4.4

The vectors being defined in the following diagram:

**Figure 4.1  Vectors related to orientation factor \( \kappa^2 \)**

\( \hat{d} \) is the unit vector of the donor dipole and \( \hat{a} \) is the unit vector of the acceptor dipole, both of which can be oriented in any given direction. \( \hat{R} \) is the vector from the donor fluorophore (green circle) to the acceptor fluorophore (red circle).

FRET studies in nucleic acids have typically been with unconstrained fluorophores unless they are deliberately fixed covalently\(^{173}\). Orientation-dependence has been studied in double stranded nucleic acids (DNA and RNA) using Cy3 and Cy5.
fluorophores attached to the terminal phosphates via a 3-carbon tether\textsuperscript{174,175}, taking advantage of the partial end-stacking on double stranded nucleic acids with these fluorophores. These data showed FRET signals that decrease with distance as expected and that oscillate due to the rotation of the acceptor Cy5 transition dipole about the nucleic acid helix as it is extended away from the donor Cy3 dipole found at the opposite helical end. Here, the periodicity of FRET signals traced well with the periodicity of helical turning, but is nonetheless substantially rotationally averaged by the partial end-stacking\textsuperscript{176–178}.

In contrast, fluorogenic dye-binding aptamers bind fluorophores much more rigidly and, in principle, can be fully oriented. In addition, fluorescent signal is only seen when the ligand is bound. Thus, using aptamers rather than Cy-dyes in FRET studies should be much more sensitive to orientation. Furthermore, the covalent attachment of dyes to RNA in a biological setting is currently only available by introduction of exogenous pre-tagged RNA. Free fluorophore systems such as fluorogenic aptamers would be advantageous for this reason, particularly if the fluorophores are cell-permeable.

Mango and Spinach are families of fluorogenic dye-binding aptamers. Conveniently, both aptamers have cores that are enclosed by a standard A-form RNA duplex, making them easy to manipulate into arbitrary RNA structures. The Spinach core is flanked by two helices that stack rigidly against the core. Mango is closed by a single duplex and transitions into the core by either a GAAA tetraloop adaptor in Mango I and II, or a triplex stack in Mango III. Since Mango can bind red fluorescent analogs of its standard TO1-Biotin ligand, it can be used as a FRET partner with the green Spinach. My work in Chapter 2 suggests that these aptamers could be used together as a ‘semi-orthogonal’ fluorogenic aptamer pairs and has recently been demonstrated in a biological setting. The Andersen laboratory inserted Mango I and Spinach into origami structures that can be modulated either by changing the length of the A-form duplexes, or by introducing a ligand to alter the conformation of the origami structure\textsuperscript{179}. FRET signals from these structures were detectable in \textit{E. coli} cells. However, these structures showed little to no orientation effect, and displayed FRET signals indicative of an averaged $\kappa^2$. Here, I show that the global structure of Mango I does not allow orientation measurements due to the aptamer core’s flexible linkage to the A-form duplex. Further, I utilize Mango III, which unlike Mango I has a fixed core relative to the helix, so as to
enhance the orientation dependence of FRET between Mango III and Broccoli, a brighter, shorter variant of Spinach\textsuperscript{68}.

4.2. Results

4.2.1. Structure of Mango I

\textit{(excerpts from Trachman et al., Structural basis for high-affinity fluorophore binding and activation by RNA Mango)}

The structure of RNA Mango loosely resembles a kite, with an A-form-duplex ‘tail’ projecting away from a G4 ‘wing’ (Figure 4.2B,C). TO1–Biotin binds on one of the faces of the quadruplex. The RNA Mango G4 comprises three G-quartet tiers (T1, T2 and T3; Figs. 1 and 2). T1 and T2 are connected in parallel, and all guanine residues in these tiers are in the anti conformation. In T1, the four nucleotides have alternating 2′-endo and 3′-endo puckers, whereas in T2, all four guanines adopt the 2′-endo pucker. T3, the third quartet, is unusual in that three of its four guanines (G16, G21 and G26) are antiparallel to the adjacent guanines in T2 (these three residues all have anti glycosidic-bond angles and adopt the 3′-endo pucker), whereas the remaining guanine (G10) is parallel to T2 (and adopts a syn glycosidic angle and a 3′-endo pucker). Whereas the nucleobases of T1, T2 and the three antiparallel guanines of T3 are nearly coplanar within their corresponding tiers, the base of G10 is buckled, and its long axis is ~30° from the mean plane defined by the other guanines of T3. This unusual orientation of G10 allows its exocyclic amine to hydrogen-bond to its own pro-SP non-bridging phosphate oxygen and improves its packing against the biotin of the fluorophore (Figure 4.2C).
Figure 4.2  Overall structure of RNA Mango in complex with TO1–Biotin.
(A) Chemical structures of TO1–Biotin and TO3–Biotin. The two compounds differ in having one or three benzylidene (methine) carbons, respectively, connecting the two heterocycles of their TO moieties. PEG, linker. (B) Secondary structure of the RNA Mango–TO1–Biotin complex. Thin lines with arrowheads denote connectivity. Base pairs are represented with Leontis–Westhof symbols. The location of the fluorophore and two potassium ions (TO1, M_A and M_B, respectively) are indicated. Except where noted, this colour scheme is used throughout. (C) Cartoon representation of the RNA Mango–TO1 complex. Curved arrows indicate the direction of the chain, 5’ to 3’. Orange mesh depicts a simulated-annealing omit |Fo| – |Fc| map (TO1–Biotin was omitted from the calculation by using the final refined atomic coordinates) contoured at 1.5 σ. Purple and red spheres represent potassium ions and water molecules, respectively. Variable region, pairing element 1 (P1) is shown in gray.

Intramolecular hydrogen-bonding appears to stabilize the conformations of A20 and A25 of the ligand-binding site of RNA Mango. Thus, the phosphate of A25 hydrogen-bonds to the exocyclic amine of G26, the Hoogsteen face of A25 makes direct and water-mediated hydrogen bonds with the phosphate of G21, the phosphate of A20 hydrogen-bonds with the exocyclic amine of G21, and its Hoogsteen face makes water-mediated hydrogen-bonds with the phosphate of G16. Overall, in associating with RNA Mango, TO1–Biotin buries 645 Å² (71%) of its solvent-accessible surface area with the benzothiazole, and the biotin and the proximal part of the PEG linker are most deeply recessed into the RNA.

No tertiary interactions join the A-form duplex and the quadruplex of RNA Mango; they connect solely through a junction that resembles a GAAA tetraloop (Figure 4.3). Canonically, such tetraloops feature a sugar edge–Hoogsteen closing base pair between
the first and fourth residues, a chain-direction reversal between the guanine and the first adenine, and stacking of the three adenines on the 3’ side. In the case of RNA Mango, G5 and A27 make the closing base pair, and the chain reverses between G5 and A6. A6 then stacks on A7 on the 3’ side of the loop. However, rather than stacking underneath A7, the next residue of RNA Mango (G8) points $\sim120^\circ$ away and forms part of the G4. The fourth position in the tetraloop is instead occupied by A27, which lies immediately 3’ of the G4 (Figure 4.3A) rather than being adjacent in sequence to the second adenine of the GAAA motif. Thus, the G4 interrupts the GAAA tetraloop-like element of RNA Mango between the second and third adenines. As in a canonical GAAA tetraloop, the 2’-OH of the guanosine (G5) hydrogen-bonds with the N7 of the second adenine (A7), and the Watson–Crick face of this guanine hydrogen-bonds to the phosphate that follows the second adenine. In the junction, this is the phosphate of A27 rather than that of the next residue in the RNA chain (G8). Despite the interruption in the backbone, the GAAA motif of RNA Mango superimposes closely (r.m.s. deviation of 0.29 Å for all non-hydrogen atoms, excluding the phosphate of A27) on a conventional tetraloop$^{23}$ (Figure 4.3C). Unlike many GAAA tetraloops, which are involved in A-minor and stacking tertiary interactions$^{19,180–182}$, the junction of RNA Mango makes no such interactions (except crystal contacts).

Although it is devoid of tertiary interactions, the GAAA tetraloop-like junction appears to be functionally important, because variant sequences isolated from randomization-reselection experiments all contain the sequence (G/U)AAA at this location. Indeed, removing the GAAA stem from RNA Mango substantially disrupted function (Figure 4.3D and Table 4.1). We produced variant RNAs containing each of the eight GNRA tetraloops as well as the UAAA and UACG tetraloops. Dissociation constants for TO1–Biotin increased modestly, from 3 nM to 10 nM through the GNRA tetraloop series $N = A, G, U$ or $C$ with $R = A$, whereas the series $N = U, C, A$ or $G$ with $R = G$ had $K_D$ values increasing nearly linearly from $\sim15$ nM to $\sim35$ nM. The binding affinity of the UAAA construct was intermediate between that of the two GNRA series, whereas the UACG construct (a member of the thermodynamically stable$^{183}$ UNCG tetraloops) showed a significant decrease in binding affinity (Figure 4.3D).
Figure 4.3  The duplex–quadruplex junction of RNA Mango resembles a GAAA tetraloop.

(A) Cartoon representation of the junction with one flanking Watson–Crick base pair from the duplex (gray) and adjacent residues from the G4 (G8 and G26). (B) Hydrogen-bonding pattern within the junction. (C) Junction of RNA Mango superimposed on a canonical GAAA tetraloop (gray; PDB 4FNJ). (D) Effect on fluorophore binding affinity of mutations in the tetraloop-like junction motif.

Table 4.1  Binding affinities and relative fluorescence of Mango I tetraloop motifs

<table>
<thead>
<tr>
<th>Tetraloop</th>
<th>$K_D$ (nM)</th>
<th>Relative $F_E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAAA (wt)</td>
<td>3.6 ± 0.7</td>
<td>1.00 ± 0.01</td>
</tr>
<tr>
<td>GGAA</td>
<td>5.35 ± 2.48</td>
<td>1.02 ± 0.05</td>
</tr>
<tr>
<td>GUAA</td>
<td>6.66 ± 1.53</td>
<td>0.89 ± 0.03</td>
</tr>
<tr>
<td>GCAA</td>
<td>9.13 ± 1.23</td>
<td>0.85 ± 0.03</td>
</tr>
<tr>
<td>UAAA</td>
<td>9.29 ± 1.36</td>
<td>0.85 ± 0.04</td>
</tr>
<tr>
<td>GUGA</td>
<td>15.12 ± 3.67</td>
<td>0.93 ± 0.01</td>
</tr>
<tr>
<td>GCGA</td>
<td>17.93 ± 4.07</td>
<td>1.02 ± 0.04</td>
</tr>
<tr>
<td>GAGA</td>
<td>24.80 ± 9.26</td>
<td>0.94 ± 0.02</td>
</tr>
<tr>
<td>GGGA</td>
<td>34.70 ± 9.86</td>
<td>1.33 ± 0.10</td>
</tr>
<tr>
<td>UACG</td>
<td>105.00 ± 51.0</td>
<td>0.58 ± 0.06</td>
</tr>
</tbody>
</table>
4.2.2. Structure of Mango III
(excerpts from Trachman et al., Mango-III is a compact fluorogenic RNA aptamer for unusual structural complexity)

The structure of Mango III (Figure 4.4) is organized around a two-tiered G4 with all-parallel connectivity, except for G18 of the top tier (T2), which is antiparallel. This quadruplex, which coordinates a canonical axial K⁺ ion (MA), stacks on a base triple, which in turn stacks coaxially on an A-form duplex (paired element P1). Two nucleotides from the loop connecting G14 and G18 of T2, and four nucleotides from the loop that follows G20 (the last of the eight guanines of the G4) form the three base pairs of a second paired element (P2) that is juxtaposed with the G4. U17, which is extruded from the middle of P2, forms a trans Watson-Crick pair with A10 from the propeller loop that connects G9 to G13, on the diagonally opposite side of the T2 G-quartet. The benzothiazole and methylquinoline of the bound TO1-Biotin are sandwiched between the A10•U17 tertiary base pair and T2 of the G4.

Figure 4.4  Overall structure of the Mango III-TO1-Biotin complex
(A) Secondary structure of Mango-III. Thin lines with arrowheads and Leontis-Westhof symbols denote connectivity and base pairs, respectively. The TO1-Biotin ligand is represented by two orange squares for the two heterocycles. BzT - benzothiazole, MQ – methylquinoline (B) Cartoon representation of the three-dimensional structure of the complex coloured as in (A). Arrows indicate 5' to 3' chain direction, and purple spheres represent metal ions.

Mango I and the closely related Mango II are each comprised of a three-tiered G4 flexibly linked to an A-form helix through a GAA^A-tetraloop-like junction (^ denotes the site of insertion of the quadruplex)³⁶,³⁸. The residues of Mango III corresponding to those junctions adopt a completely different structure. Instead of folding like the first
three nucleotides of a GNRA tetraloop, G1, A2 and A3 of Mango III continue the duplex structure of P1 by Watson-Crick base-pairing with complementary nucleotides at the 3' end of the aptamer (Figure 4.4A). A25 of Mango III forms an expanded base triple with A7 (which is part of the propeller loop between G5 and G8 of the Mango III quadruplex) and U19 (which connects G18 an G20 of the Mango III quadruplex) (Figure 4.4A). The A7:U19:A25 triple is unusual because only two direct hydrogen bonds connect the three nucleobases, but each base also hydrogen bonds through its Watson-Crick face to a shared water molecule that lies approximately at the center of the triple.

The expanded A7:U19:A25 triple links the P1 duplex with the quadruplex of Mango III through three sets of interactions. First, the base triple stacks below the T1 of the quadruplex and above the closing A3:U26 base pair of P1 (Figure 4.4A). Second, the 2'-OH of U26 donates a hydrogen bond to the N1 of A7 of the triple, in effect expanding it to a pseudo-tetrad. Third, a hydrated cation (Mg<sup>2+</sup> in this crystal structure, but likely Mg<sup>2+</sup> under physiological conditions) and a network of ordered water molecules connects the phosphate of A7 to the minor groove faces of both, A3 and G4, thereby spanning the duplex-quadruplex junction (not shown). Overall, these interactions likely produce a rigid connection between P1 and the G4 moieties of Mango III.

**4.2.3. A rigid transition from ligand-binding core to external helix allows orientation FRET measurements between Mango III and Broccoli.**

Mango I and Mango III are structurally distinct in their transition from the ligand-binding core and external helix. The core of Mango I is essentially ‘hinged’ onto the external helix by two phosphodiester bonds between A7-G8 and G26-A27 (Figure 4.5A). Thus, relative to the helix, the position of the fluorophore is quite mobile. In contrast, the core of Mango III is stacked on to the external helix via an A7:U19:A25 triplex layer, hence the position of the core is expected to be rigid relative to the helix (Figure 4.5B). Similar to Mango III, the ligand-binding core of the Broccoli aptamer is in a fixed position relative to its external helices<sup>32,33</sup>. Since FRET is highly dependent on the orientation of the fluorophores (more precisely, the orientation of the transition dipoles of the fluorophores), FRET signals using Broccoli and Mango III as a partner should display predictable orientation dependence.
Figure 4.5 Comparison of connection between aptamer core and external helix between Mango I and Mango III.

(A) Mango I core (orange/light grey) is ‘hinged’ to the GAAA tetra loop (red) and helix (grey) via two phosphodiester bonds (black and yellow atoms). TO1-Biotin is shown in green. (B) Mango III core (orange/light grey) is rigidly stacked onto a triplex layer (red), followed by a duplex (grey). TO1-Biotin is shown in green with its cap shown in yellow.

To explore the effect on orientation dependence of the Mango I and Mango III necks on FRET, we built constructs that connect the helices of Mango and Broccoli via an A-form RNA duplex to form unimolecular constructs where distance and dipole orientation can be modulated by changing the length of the duplex (Figure 4.6, schematic).

Figure 4.6 Unimolecular Broccoli & Mango fusions
Schematic representation of Broccoli (green)/Mango III (red) aptamers in a unimolecular construct connected by an RNA duplex (gray).

We measured FRET using these constructs bound by green fluorescent DFHBI\textsuperscript{58,104} and red fluorescent TO3-Biotin\textsuperscript{54,108} and as a function of increasing RNA duplex length. Here, we use a simplified calculation of FRET (FRET\textsubscript{S}) which is the fluorescence of the acceptor at the emission peak (S\textsubscript{TO3}) divided by the sum of the
fluorescence of the donor emission peak ($S_{DFHBI}$) and acceptor emission peak (Equation 4.5).

$$FRET_s = \frac{S_{TO3}}{S_{TO3} + S_{DFHBI}}$$

*Equation 4.5*

Constructs with Mango I coupled to Broccoli as described (henceforth, M1-nbp-Broc) did not exhibit strong periodicity in FRET as a function of duplex lengths between 4-14 bp owing to the flexible junction between the Mango I core and helix connection (Figure 4.7). In contrast, Mango III coupled to Broccoli (M3-nbp-Broc) presented clear oscillation in FRET that traced well with the periodicity of an A-form duplex, demonstrating the suitability of Mango III in orientation-dependent FRET studies compared to Mango I. All subsequent studies were, therefore, done using these Mango III constructs.

![Graph](image)

**Figure 4.7**  
$FRET_s$ as a function of duplex length for M1-nbp-Broc/TO3-Biotin+DFHBI (blue) and M3-nbp-Broc/TO3-Biotin+DFHBI constructs (red)  
Comparison of duplex-length dependent FRET from Mango I compared to Mango III. For Mango I/TO3-Biotin (blue), concentrations are as follows: 500 nM RNA, 3 μM DFHBI, 3 μM TO3-Biotin. For Mango3/TO3-Biotin (red), 500 nM RNA, 3 μM DFHBI, 200 nM TO3-Biotin.

### 4.2.4. Optimal conditions for FRET with DFHBI-1T and YO3-Biotin

FRET requires sufficient spectral overlap to display a strong acceptor signal, while minimizing unwanted excitation of the acceptor by the excitation source (due to the
overlap being too close). In a previous study, Mango I was paired with Spinach to measure FRET for origami type structures using DFHBI-1T as the donor in the Spinach pocket and YO3-Biotin (benzooxazole variant of TO3-Biotin) as the acceptor in the Mango pocket (Figure 4.8A). This pair of fluorophores has greater spectral overlap (Figure 4.8B) relative to other options including DFHBI and TO3-Biotin and bind more specifically to Broccoli and Mango III respectively (Table 4.2).

Since Broccoli cannot fully distinguish DFHBI-1T and YO3-Biotin, concentrations needed to be optimized to maximize FRET signal by saturating the Spinach pocket with DFHBI-1T while limiting binding of Spinach to YO3-Biotin. In fixed concentrations of DFHBI-1T (1.5 μM) and M3-nbp-Broc RNA (30 nM), YO3-Biotin was titrated giving rise to FRET signal. The optimal condition was when FRET was maximal, but there was minimal loss in the donor or acceptor fluorescent readings (i.e. ~200 nM in Figure 4.8C). Henceforth, all FRET readings of M3-nbp-Broc experiments use these concentrations of RNA and fluorophores.

We were interested in determining biophysical values (eg. \( \kappa^2 \), and \( R_0 \)) of the Broccoli/DFHBI-1T to Mango III/YO3-Biotin FRET partner. Using our experimental data, and models we generated for M3-nbp-Broc constructs, we used a workflow to determine \( \kappa^2 \), \( R_0 \), and FRET. The workflow of this process is described in Figure 4.9.

**Table 4.2** Binding affinity and excitation/emission peaks of Broccoli and Mango III with DFHBI, DFHBI-1T, TO3-Biotin and YO3-Biotin.

<table>
<thead>
<tr>
<th></th>
<th>Broccoli</th>
<th>Mango</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( K_0 ) (nM)</td>
<td>ex/em max (nm)</td>
<td>( K_0 ) (nM)</td>
</tr>
<tr>
<td>DFHBI</td>
<td>420 ± 40(^a)</td>
<td>469/503(^a)</td>
<td>&gt;&gt;</td>
</tr>
<tr>
<td>TO3-Biotin</td>
<td>22 ± 10</td>
<td>637/658</td>
<td>25 ± 2</td>
</tr>
<tr>
<td>DFHBI-1T</td>
<td>270 ± 10</td>
<td>473/505</td>
<td>&gt;&gt;</td>
</tr>
<tr>
<td>YO3-Biotin</td>
<td>910 ± 40</td>
<td>592/615</td>
<td>10 ± 0.2</td>
</tr>
</tbody>
</table>

\(^a\)Values were measured with the Spinach aptamer rather than Broccoli and obtained from reference (112)

'>>' – \( K_0 \) was greater than detection limit of assay.
Figure 4.8  FRET compatibility of Broccoli/DFHBI-1T and Mango III/YO3-Biotin. (A) Chemical structures of DFHBI-1T and YO3-Biotin. (B) Spectral overlap of Broccoli/DFHBI-1T (green) and Mango III/YO3-Biotin (red). (C) Titration of YO3-Biotin with DFHBI-1T and RNA held at constant reveals optimal concentrations for specific binding of Broccoli/DFHBI-1T and Mango III/YO3-Biotin. Green = Broccoli donor emission, Red = Mango acceptor emission, Black = acceptor/(acceptor + donor).
Figure 4.9  Workflow of determining biophysical FRET values for Broccoli and Mango III partners
This flow chart describes the process of fitting experimental FRET data to theoretical model predictions of FRET and $\kappa^2$. The flowchart notation follows American National Standards Institute (ANSI) standards.

4.2.5. Construction of M3-nbp-Broc models

We expanded our M3-nbp-Broc construct series to extend 4-18 bp, or ~1.5 turns of an A-form RNA duplex and using the crystal structures of Mango III and Spinach $^{32,33}$ (assuming Spinach and Broccoli have similar structures which is likely based on their sequences) we built 3D structural models to visualize and quantify the position and orientation of each fluorophore for each construct (Figure 4.10). Briefly, the external
helices of Mango III and Spinach were used as adaptors to align, using the PyMol align function with an A-form RNA duplexes that range from 4-18 bp long.

![M3-nbp-Broc models](image)

**Figure 4.10  Structural models of M3-nbp-Broc constructs**
Fifteen structural models for the M3-nbp-Broc constructs were generated to calculate fluorophore positions and angles. The model uses crystal structures of Spinach/DFHBI and Mango III/TO1-Biotin as surrogates for the intended construct of Broccoli/DFHBI and Mango III/YO3-Biotin. Representative models of 6, 9, 12, and 15 bp lengths are shown here. Spinach is coloured in green and the helix region on the opposing side of Mango is hidden for clarity. Mango III is coloured in red and the adapter duplex is coloured in white. Sticks show the respective fluorophores. Left model is the side view, right model is the ‘front’ view looking from the Mango III side.

### 4.2.6. FRET oscillates as a function of connecting duplex length

We measured FRET as a function of increasing duplex length in the M3-nbp-Broc series of constructs using the conditions determined in Figure 4.8C. We expected similar effects on the signal as seen in Figure 4.7: 1) a decrease in FRET as a function
of distance and 2) an oscillation of the FRET as a function of helical turn due to the rigid positioning of the Mango III and Broccoli cores relative to the helix.

We analyzed our data using Gaussian fits for the donor and acceptor signals. These fits are shown in Figure 4.12. The Gaussian shape for the donor and acceptor signals were determined by the emission spectra of each aptamer/fluorophore individually (Equation 4.6). We fit our data simply by searching only for the amplitude (n1 and n2) of the acceptor or donor peak shapes together with possible small lateral shifts in wavelength (m1 and m2, Figure 4.11, Table 4.3). In all, we found that our acceptor and donor peak shapes consisted three Gaussians each. Using the known Gaussian shapes for either peak, we calculated FRET as the area underneath the acceptor divided by the sum of the area of the donor and acceptor curves together (Figure 4.12, Equation 4.7, Equation 4.8, Equation 4.9). As predicted, we saw FRET signal decrease over distance and oscillate with the helix twist. Our data tracked precisely with the 11 bp/helical turn of an A-form duplex with local maxima/minima at 5-6 bp intervals (Figure 4.12B) for connection lengths between 4 and 18 bp.

### Table 4.3  Fitted emission peak shapes in Gaussian analysis of M3-nbp-Broc constructs

<table>
<thead>
<tr>
<th>Duplex length (bp)</th>
<th>Broccoli/DFHBI-1T em peaks (nm)</th>
<th>Mango/YO3-Biotin em peaks (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500.5 520.2 544 m1 n1</td>
<td>616.1 634.3 663.5 m2 n2</td>
</tr>
<tr>
<td>4</td>
<td>499.8 519.5 543.3 -0.7 26.1</td>
<td>616.0 634.2 663.4 -0.1 10.3</td>
</tr>
<tr>
<td>5</td>
<td>500.1 519.8 543.6 -0.4 16.1</td>
<td>615.8 634.0 663.2 -0.3 22.7</td>
</tr>
<tr>
<td>6</td>
<td>500.2 519.9 543.7 -0.3 13.0</td>
<td>615.7 633.9 663.1 -0.4 28.5</td>
</tr>
<tr>
<td>7</td>
<td>500.5 520.2 544.0 0.0 8.9</td>
<td>615.7 633.9 663.1 -0.4 18.7</td>
</tr>
<tr>
<td>8</td>
<td>499.9 519.6 543.4 -0.6 18.9</td>
<td>615.8 634.0 663.2 -0.3 18.5</td>
</tr>
<tr>
<td>9</td>
<td>499.6 519.3 543.1 -0.9 35.2</td>
<td>616.0 634.2 663.4 -0.1 7.7</td>
</tr>
<tr>
<td>10</td>
<td>499.6 519.3 543.1 -0.9 39.8</td>
<td>616.1 634.3 663.5 0.0 10.0</td>
</tr>
<tr>
<td>11</td>
<td>499.7 519.4 543.2 -0.8 32.1</td>
<td>616.0 634.2 663.4 -0.1 13.8</td>
</tr>
<tr>
<td>12</td>
<td>499.8 519.5 543.3 -0.7 31.6</td>
<td>616.1 634.3 663.5 0.0 15.5</td>
</tr>
<tr>
<td>13</td>
<td>499.7 519.4 543.2 -0.8 36.8</td>
<td>615.9 634.1 663.3 -0.2 13.0</td>
</tr>
<tr>
<td>14</td>
<td>499.7 519.4 543.2 -0.8 45.2</td>
<td>616.1 634.3 663.5 0.0 8.4</td>
</tr>
<tr>
<td>15</td>
<td>499.4 519.1 542.9 -1.1 55.4</td>
<td>615.9 634.1 663.3 -0.2 5.4</td>
</tr>
<tr>
<td>16</td>
<td>499.5 519.2 543.0 -1.0 56.4</td>
<td>616.5 634.7 663.9 0.4 6.5</td>
</tr>
<tr>
<td>17</td>
<td>499.5 519.2 543.0 -1.0 49.9</td>
<td>616.1 634.3 663.5 0.0 9.8</td>
</tr>
<tr>
<td>18</td>
<td>499.6 519.3 543.1 -0.9 42.2</td>
<td>615.9 634.1 663.3 -0.2 9.5</td>
</tr>
</tbody>
</table>

Standard deviation

|                  | 0.3 0.3 0.3 | 0.2 0.2 0.2 |

See Equation 4.6 for details of the function. Briefly, m1 and m2 are deviations of the peak position on the x-axis and n1 and n2 are the amplitude terms for each peak shape.
Figure 4.11  Gaussian fits for M3-nbp-Broc constructs
For each construct, one replicate of raw FRET emission data is shown. Constructs are labelled in the top right corner of the graph by the length of the duplex (e.g. ‘4 bp’ is M3-4bp-Broc). Blue dots are data points and Gaussian fit is shown by a red line. Fit residuals are shown below each graph.
Figure 4.12  FRET efficiency as a function of duplex length.
(A) Representative data analysis. Emission curves are shown for Mango III - Broccoli constructs bound to optimal concentrations of DFHBI-1T and YO3-Biotin (see Figure 4.8), excited at 460 nm. Data points are shown in black. Data is fitted to two groups of three Gaussians that are tied together (Equation 4.6); three assigned to the donor/DFHBI-1T emission (green, Equation 4.7) and three assigned to the acceptor/YO3-Biotin emission (red, Equation 4.8). (B) FRET data as a function of connector duplex length. FRET is calculated by using the integrals of Gaussians shown in (A) (Equation 4.9). Open circles represent the data points. Error bars are standard deviations of n = 3 for duplex lengths: 4-10, 15-18 and n = 4 for duplex lengths: 11-14.

4.2.7. Calculation of FRET values using structure models

Our structure models (Figure 4.10) provided relative fluorophore positions and orientations. The two fluorophores are nearly planar in their respective structures; thus, we could define the plane and calculate the normal vectors of each fluorophore relative to the plane of the atoms. The angle between the normal vector of the donor and the acceptor fluorophores are never near zero (Table 4.4), indicating that the fluorophores are never completely parallel, a clear distinction from previous orientation studies that use partial, parallel, end-stacked cyanine dyes\textsuperscript{174,175}. The angle about the respective normal vectors and within the plane of the fluorophore will, henceforth, be defined as the ‘clock-face angles’. With our models, we could discern the potential $\kappa^2$ and related $R_0$ values by fitting the clock-face angles to our experimental data.

The orientation factor $\kappa^2$ is often taken as an average of 2/3 when orientation between FRET partners are not constrained. Since the M3-nbp-Broc constructs should highly constrain fluorophore positions, we expected a large variation in $\kappa^2$ as a function of construct length. It was anticipated that the transition dipoles of each fluorophore would
be in the plane of the compound, thus the relevant angles to consider are simple the clock-face angles of each fluorophore. As expected, $\kappa^2$ oscillated with maxima peaks at 5, 12 and 18 bp, and minima peaks at 9 and 15 bp duplex lengths (Figure 4.13A). Though $\kappa^2$ values can range from 0 to 4, the orientation factor ranges from 0 to 1 when the transition dipoles are parallel by the axis of its plane, and can only be greater than 1 when they approach a co-linear orientation. The M3_nbp_Broc had $\kappa^2$ values ranging between 0 to \~1.1. Thus, in these structures, FRET will never reach maximum efficiency (i.e. the transition dipoles are never collinear).

$R_0$ is the distance between the two fluorophores at which FRET is 50% efficient. This value is related to $\kappa^2$ to the 6th power. The $R_0$ values with respect to $\kappa^2$ of the Broccoli/DFHBI-1T and Mango III/YO3-Biotin FRET pair is shown in Figure 4.13B. If the fluorophores are not constrained and $\kappa^2$ is taken as an average 2/3, then $R_0$ for this FRET pair was found to be \~39 Å.

**Table 4.4**  Modelled relative angle and distance between normal vectors of the donor and acceptor fluorophores in M3-nbp-Broc structures

<table>
<thead>
<tr>
<th>Duplex length (bp)</th>
<th>Angle between normal vectors of donor and acceptor dipoles (rad)</th>
<th>Distance (Å)</th>
<th>Angle between normal vectors of donor dipole and distance R (rad)</th>
<th>Angle between normal vectors of acceptor dipole and distance R (rad)</th>
<th>Modelled $\kappa^2$ values</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1.50</td>
<td>35.6</td>
<td>1.30</td>
<td>1.71</td>
<td>0.03</td>
</tr>
<tr>
<td>5</td>
<td>2.04</td>
<td>36.4</td>
<td>1.21</td>
<td>1.58</td>
<td>0.20</td>
</tr>
<tr>
<td>6</td>
<td>2.53</td>
<td>38.0</td>
<td>1.26</td>
<td>1.44</td>
<td>0.88</td>
</tr>
<tr>
<td>7</td>
<td>2.70</td>
<td>40.7</td>
<td>1.41</td>
<td>1.34</td>
<td>1.04</td>
</tr>
<tr>
<td>8</td>
<td>2.33</td>
<td>44.2</td>
<td>1.59</td>
<td>1.30</td>
<td>0.46</td>
</tr>
<tr>
<td>9</td>
<td>1.82</td>
<td>48.1</td>
<td>1.74</td>
<td>1.33</td>
<td>0.02</td>
</tr>
<tr>
<td>10</td>
<td>1.27</td>
<td>51.9</td>
<td>1.81</td>
<td>1.39</td>
<td>0.18</td>
</tr>
<tr>
<td>11</td>
<td>0.72</td>
<td>55.4</td>
<td>1.80</td>
<td>1.47</td>
<td>0.67</td>
</tr>
<tr>
<td>12</td>
<td>0.21</td>
<td>58.2</td>
<td>1.72</td>
<td>1.54</td>
<td>0.99</td>
</tr>
<tr>
<td>13</td>
<td>0.44</td>
<td>60.4</td>
<td>1.59</td>
<td>1.57</td>
<td>0.82</td>
</tr>
<tr>
<td>14</td>
<td>0.98</td>
<td>62.1</td>
<td>1.46</td>
<td>1.56</td>
<td>0.30</td>
</tr>
<tr>
<td>15</td>
<td>1.53</td>
<td>63.6</td>
<td>1.35</td>
<td>1.52</td>
<td>0.00</td>
</tr>
<tr>
<td>16</td>
<td>2.07</td>
<td>65.3</td>
<td>1.31</td>
<td>1.44</td>
<td>0.34</td>
</tr>
<tr>
<td>17</td>
<td>2.55</td>
<td>67.4</td>
<td>1.34</td>
<td>1.37</td>
<td>0.94</td>
</tr>
<tr>
<td>18</td>
<td>2.69</td>
<td>70.1</td>
<td>1.42</td>
<td>1.31</td>
<td>1.03</td>
</tr>
</tbody>
</table>
Figure 4.13  Relationship between FRET, $\kappa^2$ and $R_0$

(A) Calculated $\kappa^2$ of M3-nbp-Broc constructs. (B) $\kappa^2$ is shown as a function of $R_0$ (in Å). Data is fit to a 6th power. Open circles show measured data points, fit is extrapolated to $\kappa^2 = 4$ to show expected $R_0$ for all potential $\kappa^2$ values. (C) Fitted FRET using calculated $\kappa^2$ from (A). Open circles are experimental values reproduced from Figure 4.12B. Red filled circles and dashed line show the fitted FRET values. The grey line depicts FRET if $\kappa^2$ was an averaged 2/3.

Together, we used our experimental data to fit FRET, $R_0$ and $\kappa^2$ with respect to our theoretical models (Figure 4.13C, red). Our structural model-based fits are reasonably within range of our experimental data; however, several aspects could be improved. The fit around the first maximum of our modelled values do not overlay perfectly with the experimental data, with the first fitted ‘peak’ being narrower than the data. This is coupled with the loss in amplitude of our modelled data relative to the experimental values in longer constructs. We suspect the current quality of our fits is partially limited as a consequence of non-optimal fluorophore concentrations (Figure
4.8C). Though we made efforts to optimize these conditions, it may still not be sufficient as we have sacrificed FRET ratio for a higher overall signal to ensure detectability in our assays (Figure 4.8). This may have resulted in inconsistencies in FRET ratio between different constructs. We also hypothesize that while both the Mango III and Broccoli cores are rigid relative to their helices, longer duplexes might be more rotationally flexible to adopt alternative orientations than our predicted models. Finally, background FRET signals could be a factor in our data quality. Despite our models predicting that $\kappa^2$ is close to 0 for the 9 and 15 bp constructs, our experimental data always displayed some signal with these constructs. We do not yet understand the source of this background and we were forced to include a constant background term ($m$, Equation 4.10) in our fits, perhaps in detriment to our models. It would be ideal to understand the origin of this background. These issues should be explored in future work to obtain improved FRET models for these constructs.

4.3. Discussion

To our knowledge, this is the first careful characterization of $\kappa^2$ with fluorogenic RNA aptamers and the first study of fluorophore orientation in RNA structures where the transition dipoles are not fully parallel. Previous studies of RNA duplex and three-way junctions utilized end-stacked cyanine dye FRET pairs, thus the fluorophores were assumed to be oriented completely parallel to each other\textsuperscript{174,175}. In addition, the transient end-stacking of cyanine dyes dampened the oscillation effect on FRET by the helical turn. The rigid binding of fluorophores to fluorogenic aptamers in our constructs did not have this effect, revealing steep peaks and troughs as a function of helix length. Still, our study yielded $\kappa^2$ of up to only $\sim$1.1, owing to the simple A-form duplex connectivity of our constructs. Other structures, such as previously described origami designs may position fluorophores in more optimal FRET positions.

For an averaged $\kappa^2$ of 2/3, we found that the Broccoli/DFHBI-1T and Mango III/YO3-Biotin FRET pair has $R_0$ of $\sim$39 Å and for more optimal $\kappa^2 = 4$, $R_0$ can reach as high as $\sim$50 Å, Figure 4.13B). This falls within the range of optimal Förster radii (10-100 Å) needed for studying biological macromolecules. This is comparable to the $R_0$ of most fluorescent compounds such as fluorescent protein partners\textsuperscript{184} and small molecule fluorophores\textsuperscript{185}. 
Further characterization of the Broccoli/DFHBI-1T and Mango III/YO3-Biotin FRET pair would be beneficial to understand its biophysical attributes. Fluorescence lifetime measurements of the donor as well as polarization studies could provide insight into the transition dipole angles. However, challenges remain for the polarization measurements of the M3-nbp-Broc constructs. Spinach bound to DFHBI have fluorescence decay rates \( \sim 4 \text{ ns}^{112,132} \), which is likely shorter than the diffusive tumbling rate about the length of rod-shaped molecules approximately the size of the M3-nbp-Broc constructs. Thus, larger structures that restrain tumbling will be required for polarization experiments.

The ligand promiscuity of Spinach does not allow it to be easily applied without tedious concentration optimizations as a FRET pair with other fluorogenic aptamers to date, and as discussed, could be detrimental to our \textit{in vitro} models as well. Thus, it remains to be seen if orthogonal fluorogenic aptamer pairs can be developed for efficient multicolour imaging, including FRET applications, which would greatly bolster the repertoire of fluorescence imaging techniques for RNA. My attempts at addressing this specific weakness are found in Chapter 5 of this thesis.

4.4. Materials & Methods

4.4.1. M3-nbp-Broc FRET measurements

All fluorescence data are steady state measurements using a Varian Cary Eclipse fluorescence spectrophotometer. Experiments were carried out in an \textit{in vivo} mimicking buffer: 140 mM KCl, 10 mM NaH\(_2\)PO\(_4\) pH 7.2, 1 mM MgCl\(_2\), 0.05% Tween-20. Samples containing 30 nM RNA, 1.5 \( \mu \text{M} \) DFHBI-1T and 200 nM YO3-Biotin were incubated for 1 hour at room temperature in aluminum foil, then fluorescence was measured by excited at 460 nm and emission scanned at a rate of 1 nm/second from 475 – 750 nm at 25 °C. PMT was set to 800 V and ex/em slit widths at 5/10 nm.

4.4.2. FRET spectrum fitting

Raw FRET spectrums were fit to pre-assigned Gaussians using the following equation:
\[
\begin{align*}
    n_1 \left( e^{ \left( \frac{x-500.5+m_1}{16.35} \right)^2 } + 0.4937 e^{ \left( \frac{x-520.2+m_1}{28.47} \right)^2 } + 0.2574 e^{ \left( \frac{x-544+m_1}{47.58} \right)^2 } \right) \\
    + n_2 \left( e^{ \left( \frac{x-616.1+m_2}{20.42} \right)^2 } + 0.3175 e^{ \left( \frac{x-634.3+m_2}{34.3} \right)^2 } + 0.0904 e^{ \left( \frac{x-663.5+m_2}{58.73} \right)^2 } \right)
\end{align*}
\]

*Equation 4.6*

Where \( n_1 \) and \( m_1 \) are the Y-scaling factor and X-shifting factor respectively of the Broccoli/DFHBI-1T Gaussian and \( n_2 \) and \( m_2 \) serve the same function for the Mango/YO3-Biotin Gaussian. Gaussian shapes were predetermined by the emission of Broccoli/DFHBI-1T without acceptor fluorophore and the Mango/YO3-Biotin emission spectrum. Emission peak shifts between each construct were insignificant (Table 4.3)

FRET was calculated by integration of the two Gaussian groups after determining value of \( n_1, n_2, m_1 \) and \( m_2 \) parameters.

Broccoli/DFHBI-1T integral:

\[
I_{\text{donor}} = \int_{0}^{\infty} n_1 \left( e^{ \left( \frac{x-500.5+m_1}{16.35} \right)^2 } + 0.8409 e^{ \left( \frac{x-520.2+m_1}{28.47} \right)^2 } + 0.2574 e^{ \left( \frac{x-544+m_1}{47.58} \right)^2 } \right) dx
\]

*Equation 4.7*

Mango/YO3-Biotin integral:

\[
I_{\text{acceptor}} = \int_{0}^{\infty} n_2 \left( e^{ \left( \frac{x-616.1+m_2}{20.42} \right)^2 } + 0.3175 e^{ \left( \frac{x-634.3+m_2}{34.3} \right)^2 } + 0.0904 e^{ \left( \frac{x-663.5+m_2}{58.73} \right)^2 } \right) dx
\]

*Equation 4.8*

\[
FRET = \frac{I_{\text{acceptor}}}{I_{\text{acceptor}} + I_{\text{donor}}}
\]

*Equation 4.9*

4.4.3. Generating theoretical structure models

Theoretical structural models of M3-nbp-Broc constructs were generated based on the crystal structures of Mango III, Spinach (PDB ID: 4TS2\textsuperscript{33}) and an A-form RNA
duplex. Models were built using a combination of UCSF Chimera, PyMol and python scripts.

4.4.4. Modelling of FRET parameters

Expected FRET was calculated using the theoretical structural models of the M3-nbp-Broc series. The models provided relative position coordinates of the Mango III and Broccoli fluorophores and a MatLab script was built to fit the direction of the dipoles using the least squares method and the experimental data.

\[
FRET = \frac{1}{1 + R_i \left(\frac{R_0^6}{\kappa^2}\right)} + m
\]

*Equation 4.10*

where \( R \) is the distance in Angstroms between the two fluorophores and \( \kappa^2 \) is the orientation factor and \( m \) is a background constant. \( R_i \) is a factor encompassing \( R_0 \) and \( \kappa^2 \). \( R_0 \) is calculated by:

\[
R_0 = \sqrt[6]{\kappa^2 \left(\frac{1}{R_i}\right)}
\]

*Equation 4.11*
Chapter 5.

Selection for an orthogonal red fluorescent aptamer

I just saw something incredibly cool! A big floating ball that lit up with every colour in the rainbow, plus some new ones that were so beautiful I fell to my knees and cried.

P.J.F

This chapter is based on unpublished work by myself (S.C.Y.J.), Batool Rayyan (B.R.) and Peter J. Unrau (P.J.U.)

I (S.C.Y.J.) planned and performed selection and most experiments. B.R. performed fluorescence characterization of individual clones after selection and P.J.U. provided intellectual input in planning experiments.
5.1. Introduction

To investigate the dynamic nature of RNA functions within cells require a robust fluorescence imaging systems. Recent solutions to this problem have used fluorogenic aptamers that bind small molecule fluorophores. Though the number of systems developed have grown over the last several years, an aspect of fluorescence imaging that has not been solved with these aptamers is the ability to multiplex (i.e. to use multiple aptamers in the same system). However, most of these systems are either available in a narrow fluorescence emission wavelength range (usually GFP-like channels), or are inefficient with poor contrast between bound vs unbound fluorophores. Further, there is also currently a lack of orthogonal fluorogenic aptamer systems that can be efficiently used in the same system, with minimal crosstalk.

Currently, multicolour imaging is possible, but has certain limitations. The most widely used system for this is the orthogonal MS2 and PP7 viral coat protein aptamers. Both are small, stem loop structures that orthogonally bind different viral coat proteins. This type of system has most recently been used to measure translation kinetics in cells. However, since these systems are not fluorogenic, the background signal is high, greatly reducing the signal-to-noise ratio (see section 1.4.2). In principle, fluorogenic aptamers could solve this problem.

We previously developed the Mango fluorogenic aptamer series. These aptamers bind with nanomolar affinity, and enhance the fluorescence of thiazole orange (TO)-based fluorophores by up to 4,000-fold. We developed two forms of the TO ligand, with the distinction in the length of the methine group between the two heterocycles of TO (Figure 1.13). Mango was selected for TO1-Biotin, a green fluorophore with excitation and emission at 510 and 535 nm respectively. Extension of the methine bridge from a one-carbon, to a three-carbon chain yielded TO3-Biotin, a red variant with excitation and emission at 637 and 658 nm respectively. The Mango aptamers bind both of these fluorophores with indiscriminate affinities. Thus, while we have fluorogenic dyes in two colours, there is no current way to distinguish them using an RNA tag. Other systems possess similar problems. For example, the Spinach series of aptamers is known to be highly promiscuous (Chapter 2). Recent systems that take advantage of quencher-binding aptamers to enhance fluorescence of quencher-fluorophore
conjugates also cannot efficiently distinguish colours by the RNA\textsuperscript{110,111}. As such, it is an urgent need to develop fluorogenic aptamer systems that are multicolour and orthogonal, such that they can be used in a single system to enable multicolour imaging. Here, I describe a selection for an aptamer for TO3-Biotin that is orthogonal to our Mango systems.

5.2. Results

5.2.1. Selection of an aptamer that discriminates TO3-Biotin from TO1-Biotin.

To isolate orthogonal RNA aptamers from an unbiased pool (87 random positions, diversity of $10^{14}$) that bind TO3-Biotin but not TO1-Biotin, we utilized a competition-based selection protocol using TO3-Biotin anchored to magnetic streptavidin beads and the opposing fluorophore (TO1-R) as the competitive wash. In the first five selection rounds, RNA was bound to beads and were washed using three competitor molecules: N-methyl mesoporphyrin IX (NMM), TO1-Acetate, TO3-Acetate (Figure 5.1). Since Mango and other G4 aptamers are known to bind NMM with near nanomolar affinity, NMM in this selection acts as a general G4 binder to remove non-specific G4 structures. The TO1 compound acts as a competitive wash while the purpose of TO3 is to remove fast off-rate binders. Beginning in round 6, the competitor molecule, TO1-Acetate in the previous rounds, was replaced with TO1-propyl and TO1-methoxyethane. ‘Blocked’ R groups such as propyl and methoxyethane replaced acetate to remove negatively charged competitors (the hypothesis being that negatively charged dyes may discourage high affinity nucleic acid binding via its own negative backbone). To further increase stringency, an additional pre-incubation step of the RNA with competitor molecules before binding to beads was included to eliminate RNA molecules with a slow off-rate from competitors. By round 10, NMM was eliminated to avoid removal of potentially interesting G4 RNA that have high affinity for TO3-Biotin. We also continued to increase wash times and reduced the duration of the RNA binding step to promote molecules with fast on-rate for TO3-Biotin. To promote point mutations which may improve already high affinity binders, in rounds 12 and 13, we included a mutagenic PCR step where residues were mutated at a rate of 0.66%. The detailed selection conditions for a total of 16 rounds are described in Figure 5.1.
5.2.2. Diagnostics reveal increase in TO3 preference in late selection rounds

In the later stages of the selection, we wondered how well the library as a bulk could bind TO3-Biotin over TO1-Biotin. Beginning in round 10, in parallel to bead-binding selection for TO3-Biotin, we also included an identical experiment but using TO1-Biotin in place of TO3-Biotin on the beads and applying the same RNA pool to both. This was used as a diagnostic to compare the ability of the pool to bind TO1-Biotin or TO3-Biotin under the same conditions. Interestingly, starting round 12, the ratio of RNA binding to TO3-Biotin vs. TO1-Biotin (henceforth, TO3:TO1) began to rise significantly. In rounds 11-14, TO3:TO1 increased from 0.1 to 2, 2.7 and 8.8 (Figure 5.2). The most dramatic change was between rounds 13 (TO3:TO1 = 2.7) and 14 (TO3:TO1 = 8.8), coinciding with the mutagenic PCR rounds. Thus, this improvement may be caused by the introduction of favourable point mutations. By round 15, TO3:TO1 began to plateau (and
in fact TO3:TO1 slightly dropped to 8) and by round 16, this ratio had collapsed to 3.7, suggesting that these selection rounds were detrimental and no longer enriching better binders.

**Figure 5.2**  **TO3-Biotin selection, bead-binding diagnostic.**  
TO3-Biotin vs TO1-Biotin binding profile for each round of selection. For each round, TO3:TO1 is the ratio of radioactive counts of RNA pool bound to beads with TO3-Biotin divided by RNA pool bounds to beads with TO1-Biotin in the same conditions.

Together with the bead-binding diagnostic, we also compared rounds 12-14 for their ligand preference when both TO3-Biotin and TO1-Biotin were present together, using the increase in fluorescence for each fluorophore as a read-out (Figure 5.3A). Conveniently, Mango III with an A10U mutation in its ligand cap (Figure 4.4) has an ~30-fold difference in binding affinity for TO1-Biotin (5.6 nM) than TO3-Biotin (170 nM) making it a useful negative control for this experiment (Trachman et al., unpublished data). Surprisingly, in contrast to the bead-binding diagnostic, there was no significant change in fluorescence profiles between rounds 12, 13 and 14. Still, there was a preference for TO3-Biotin over TO1-Biotin for all three rounds compared to the Mango III A10 U cap mutant. By replotting the ratio of TO3:TO1 fluorescence as a function of RNA
concentration from this data, a clear rise is seen for TO3 preference for R12, 13 and 14 at low concentrations of RNA (Figure 5.3B). As RNA concentration increases, the ratio decreases, perhaps due to saturation of the TO3-Biotin fluorophore, allowing the pool to bind TO1-Biotin, indicating a higher on-rate of the pool for TO3-Biotin compared to TO1-Biotin.

5.2.3. Studying individual clones from the TO3-Biotin selection

Since the library may consist of a combination of highly selective aptamers and still poorly binding sequences, we explored the affinities for either TO3-Biotin or TO1-Biotin of individual clones within the round 12 and round 15 pool. In total 42 clones (22 from R12 and 20 from R15) were measured for brightness when both TO3-Biotin and TO1-Biotin were present in excess. Promisingly, when fluorescence was monitored in both TO1 and TO3 channels simultaneously, with both dyes present in excess to the RNA, clones from the R12 pool had up to 10-fold TO3:TO1 ratio in fluorescence and clones from R15 was up to 9.5-fold. In comparison, the Mango III A10U control RNA, which is expected to prefer TO1>TO3 had a TO3:TO1 ratio of <0.1. On average, clones from the R12 and R15 pool had TO3:TO1 of 6 ± 1.7 and 5.9 ± 1.4 respectively indicating
no clear difference in ligand preference between each round, consistent with previous diagnostics.

![Graph showing TO3:TO1 binding ratio for ligand preference.](image)

**Figure 5.4**  TO3:TO1 binding ratio for ligand preference. Individual clones from R12 and R15 of the selection were measured for fluorescence in the presence of equimolar concentrations of TO1-Biotin and TO3-Biotin. The log of the TO3/TO1 fluorescence signal is shown. (A) Clones from the R12 library (red) compared to a Mango III A10U control. (B) Clones from the R15 library (green) compared to the same control as in (A).

### 5.3. Discussion

While fixed concentration measurements suggest that the isolated clones can potentially be an orthogonal binder for TO3-Biotin and not TO1-Biotin, further characterization such as affinity measurements will be needed to identify interesting sequences. It remains to be seen whether this selection has isolated aptamers that when paired with Mango, can be comparable to other orthogonal aptamer pairs. High resolution discrimination has previously been achieved with RNA aptamers. The well-known theophylline aptamer can distinguish between theophylline and caffeine by a 10,000-fold difference in binding affinity despite the ligand differing by only a methyl group\(^\text{189}\). Similarly, the MS2 and PP7 coat protein aptamers, often used for fluorescent RNA labelling, can distinguish from opposing ligands with \(\approx\)1,000 fold difference in binding affinity\(^\text{186–188}\). In practice, the goal of orthogonal fluorogenic aptamers would be multiplex imaging in a single system, thus these are comparable existing metrics for which we can aim for.
While binding affinity is the main parameter for ligand discrimination, in practice, fluorescence enhancement of one aptamer over the other could compensate for low binding orthogonality. In other words, one could aim to achieve orthogonality by fluorescent efficiency (discussed in Chapter 2)\textsuperscript{112} rather than just affinity. For example, at equilibrium, if an aptamer has near identical binding affinities to two ligands (say 1 nM), but enhances the fluorescence of one 1,000-fold brighter than the other (E of ligand A: $1,000/1 \text{ nM} = 1,000 \text{nM}^{-1}$, ligand B: $1/1 \text{ nM} = 1 \text{nM}^{-1}$: Equation 2.4). This is indistinguishable from if the fluorescence enhancement is equal at 1,000-fold, but binding affinity to ligand A is 1 nM vs 1,000 nM to ligand B (E of ligand A: $1,000/1 \text{ nM} = 1,000 \text{nM}^{-1}$, ligand B: $1,000/1,000 \text{ nM} = 1 \text{nM}^{-1}$). Thus, fluorescence enhancement could be a secondary, or equally important, parameter to optimize to develop orthogonal aptamers. One way to achieve this might be using microfluidics-based fluorescence selection, which selects purely by fluorescence without considering binding affinity.

Overall, the current stage of this \textit{in vitro} selection for an aptamer that can distinguish TO3-Biotin from TO1-Biotin shows promise, but future work will be needed to fully optimize and develop a practical orthogonal aptamer, whether that is by a difference in fluorescence enhancement or binding affinity.

\section*{5.4. Materials & Methods}

\subsection*{5.4.1. \textit{In vitro} selection}

An unbiased (N87) high diversity ($10^{14}$) DNA pool was used for the selection. 2 nMoles of RNA after \textit{in vitro} transcription from the DNA pool was used as the input to the first round of selection. Streptavidin-conjugated magnetic beads (Dynabeads M-270, Life technologies) were used as the solid support. Beads were derivatized with TO3-Biotin by incubating beads with dye at the specified concentrations in Table 5.1 for 15 minutes at 37 \textdegree C in an \textit{in vivo} mimicking buffer (140 mM KCl, 10 mM MgCl$_2$. To ensure dye had bound to the beads, absorbance of the input solution and the supernatant were read at 615 nm using a Nanodrop 2000 (ThermoFisher). Detailed conditions of the selection are described in Figure 5.1 & Table 5.1. RNA bound to beads were heat-denatured in 95\% formamide and 0.05 M EDTA to remove them from beads for RT-PCR. Concentrations of TO1-Biotin and TO3-Biotin are determined by absorbance in the \textit{in vivo} mimicking
buffer using extinction coefficients of 63,000 M$^{-1}$cm$^{-1}$ at 500 nm for TO1-Biotin, and 133,000 at 615 nm for TO3-Biotin.

Table 5.1  Concentration conditions for each round of TO3-Biotin selection.

<table>
<thead>
<tr>
<th>Round</th>
<th>Moles of RNA</th>
<th>Moles of TO3-Biotin</th>
<th>[RNA]</th>
<th>[TO3-Biotin]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 nMole</td>
<td>0.6 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>2</td>
<td>1 nMole</td>
<td>0.28 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>3</td>
<td>0.5 nMole</td>
<td>0.14 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>4</td>
<td>0.2 nMole</td>
<td>0.06 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>5</td>
<td>0.2 nMole</td>
<td>0.06 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>6</td>
<td>0.25 nMole</td>
<td>0.07 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>7</td>
<td>0.25 nMole</td>
<td>0.07 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>8</td>
<td>0.25 nMole</td>
<td>0.07 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>9</td>
<td>0.25 nMole</td>
<td>0.07 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>10</td>
<td>0.25 nMole</td>
<td>0.07 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>11</td>
<td>0.25 nMole</td>
<td>0.07 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>12</td>
<td>0.25 nMole</td>
<td>0.07 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>13</td>
<td>0.25 nMole</td>
<td>0.07 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>14</td>
<td>0.25 nMole</td>
<td>0.07 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>15</td>
<td>0.25 nMole</td>
<td>0.07 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
<tr>
<td>16</td>
<td>0.25 nMole</td>
<td>0.07 nMole</td>
<td>2.5 µM</td>
<td>0.7 µM</td>
</tr>
</tbody>
</table>

5.4.2. Reverse-transcription PCR

Eluted RNA molecules from each round of selection were regenerated using an RT-PCR protocol. One quarter volume of eluted RNA was reverse transcribed using Maxima RT enzyme (ThermoFisher) in the provided RT buffer, 0.5 mM dNTPs, 1 µM RT primer, 0.01 M DTT and 20 U/µL of Maxima enzyme. The RT-mix was pre-incubated at 65 °C before adding enzyme and incubating at 50 °C for one hour. The reaction was then base-treated in 0.1 M KOH at 0.5 µM of forward and reverse primer at 90 °C for 15 minutes, then neutralized with 0.5 M Tris-HCl to pH 7. A reaction was done in parallel with no RT enzyme added as a negative control. The RT reaction was amplified by PCR using Taq polymerase (NEB). A pilot reaction was done with + and − RT to determine the number of optimal PCR cycles.

5.4.3. Mutagenic PCR

To perform mutagenic PCR after reverse transcription reaction, we adapted a previously established protocol$^{190}$. Taq polymerase was used as in standard PCR, with
the addition of 5.5 mM MgCl₂ and 0.5 mM MnCl₂. The dNTP mixed was adjusted to 2 mM dGTP, 2 mM dATP, 10 mM dCTP, and 10 mM dTTP.

5.4.4. Fluorescence competition assay of libraries

Libraries by round were measured for their ability to bind TO3-Biotin over TO1-Biotin in a fluorescence competition assay. TO3-Biotin and TO1-Biotin are incubated at equimolar (100 nM) titrated with RNA in an in vivo mimicking buffer for 1 hour at room temperature, Cary Eclipse fluorescence spectrophotometer: excitation/emission of 510/535 (TO1-Biotin) and 637/658 nm (TO3-Biotin). PMT was set to 800 V, read time of 0.5 s, and ex/em monochromator slits were set to 5/10 nm.

5.4.5. Fluorescence competition assay of individual clones

Individually sequenced clones were measured for their ability to bind TO3-Biotin over TO1-Biotin by a similar assay as fluorescence competition with libraries, but RNA was incubated only at 10 nM instead of titrated.
References


105. Strack, R. L., Disney, M. D. & Jaffrey, S. R. A superfolding Spinach2 reveals the


118. Algar, W. R., Massey, M. & Krull, U. J. Fluorescence Resonance Energy Transfer


143. Ren, J. & Chaires, J. B. Sequence and Structural Selectivity of Nucleic Acid


170. Wu, P. & Brand, L. Orientation factor in steady-state and time-resolved resonance


