THE UNIVERSITY OF CHICAGO

CRYSTAL STRUCTURE-DRIVEN INVESTIGATIONS OF A FLUOROGENIC RNA APTAMER WITH DIVALENT CATION DEPENDENCE

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

HUW CONOR REES

CHICAGO, ILLINOIS

JUNE 2022

"Adversity is the first path to truth."

-Lord Byron

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2'OH	2' hydroxyl of the ribose ring in a nucleic acid
3'CITE	3'UTR-Cap-independent-translation enhancers
3'UTR	3' untranslated region
5'UTR	5' untranslated region
A	adenine
Å	angstrom, $1x10^{10}$ meters
ARES	atomic rotationally equivariant scorer
BL3	bartel ligase 3 ribozyme
bp	base pair
C	cytidine
CDR	complementarity determining region
cp	circularly permuted
Cryo-EM	cryogenic electron microscopy
Da	dalton, a unit of mass (1 Da = 1 g/mol
DFHBI	(5Z)-5-[(3,5-Difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2,3-dimethyl-4H-imidazol-4-one
DFHBI-1T	3,5-difluoro-4-hydroxybenzylidene 1-trifluoroethyl-imidazolinone
DFHO	3,5-difluoro-4-hydroxybenzylidene imidazolinone-2-oxime
$DMHBO^+$	[4-[(4Z)-4-[(4-hydroxy-3,5-dimethoxyphenyl)methylidene]-2-[(E)-hydroxyiminomethyl]-5-oxoimidazolidin-1-yl]phenyl]-trimethylazanium
DIR	dimethyl indole red
DNA	deoxyribose nucleic acid
dPAGE	denaturing polyacrylamide gel electrophoresis
dsDNA	double-stranded DNA
eGFP	enhanced GFP
eIF	eukaryotic initiation factor
EMSA	electromobility shift assay
Ex/Em	excitation and emission wavelengths of light used in a fluorescence assay
F	phenylalanine (proteins), Fluorescence (data)

Fabs	antibody fragments
FRET	fluorescence resonance energy transfer
G	guanidine (nucleic acids), or glycine (proteins)
GFP	green fluorescent protein
H-bond	hydrogen-bond
HBC530	(Z)-4-(1-cyano-2-(4-((2-hydroxyethyl)(methyl)amino)phenyl)vinyl)benzonitrile
HBC599	(Z)-4-(1-cyano-2-(6-((2-hydroxyethyl)(methyl)amino)benzo[b]thiophen-2-yl)vinyl)benzonitrile
HBI	4-(p-hydroxybenzylidene)imidazolidin-5-one
HDV	hepatitis delta virus
IRES	internal ribosome entry site
J	junction composed of ssRNA in an RNA tertiary structure
K	Kelvin
K^+	potassium cation
K_d	dissociation constant
kDa	kilo daltons
lncRNA	long non-coding RNA
MFE	minimum free energy
Mg^{2+}	Magnesium cation
mg	milligrams
mL	milliliters
mM	milli molar
Na^+	sodium cation
nM	nano molar
nmol	nano mol
nL	nanoliter
NMR	nuclear magnetic resonance
nPAGE	native polyacrylamide gel electrophoresis
NTP	nucleotide tri-phosphate
OTB	oxazole thiozole blue

P	base-paired region of a nucleic acid structure
Pb^{2+}	lead cation
PCR	polymerase chain reaction
pdb	protein data bank
rmsd	root mean squared deviation
RNA	ribonucleic acid
RT	room temperature
S	serine
SAM	S-adenosylmethionine
SELEX	systematic evolution of ligands by exponential enrichment
SHAPE	selective 2′-hydroxyl acylation analyzed by primer extension
Sr^{2+}	strontium cations
ssRNA	single-stranded RNA
G-Quadruplex	guanosine-rich base quadruple that sometimes forms in RNA tertiary structure
T	thymine (DNA)
TO	thiozole orange
tRNA	transfer RNA
U	uracil (nucleic acids), Units of a protein in a reactive mixture (methods sections)
UV	ultra-violet
WCF	Watson-Crick-Franklin
WT	wild type
ε	extinction coefficient
λ	wavelength of light
μg	micro gram
μL	microliter
μM	micro molar
μm	micrometer
ϕ	quantum yield

Acknowledgements

Finishing this Ph.D. has been a long and difficult road, and there are many people I would like to thank. First and foremost, I need to thank Dr. Joseph Piccirilli for being a great mentor and advisor. Taking a student to do research in a field in which they are unfamiliar in their 4th year is a brave decision, and one which I'm forever grateful for. Joe's rigorous attention to detail and integrity has left lasting lessons in how I will approach problems for the rest of my life, and his passion for intellectual pursuits is so admirably genuine it has been inspiring. Dr. Yossi Weizmann – my advisor for my first years here – was also incredibly thoughtful and kind, and I'm grateful for the time I got to spend in his laboratory, and his supportive attitude. I also would like to thank the two members of my thesis committee – Dr. Weixin Tang and Dr. Phoebe Rice, who provided support and essential critical feedback for this work. Dr. Phoebe Rice provided candid and useful feedback that aided the crystallographic aspects of this work, and her kindness and understanding were much appreciated. Dr. Weixin Tang demonstrated enthusiasm for the contents of this thesis that was invigorating, and discussions about the directions of future research were much appreciated.

I'd also like to extend thanks to all of the incredible people I got to work with throughout my Ph.D. From the Piccirilli Lab I would like to first thank Dr. Anna Lewicka who was instrumental in helping me get through the highs and lows of my Ph. D., and her technical mentorship was exceptional. In addition, Wojciech Gogacz came into the project at a late stage and made huge contributions for which I can't thank him enough, and I will enjoy working with him after I leave to ensure the project gets finished. I would also like to extend special thanks to Dr. Nan-Sheng Li, who synthetized many organic compounds and synthetic RNAs on behalf of my work over the years, and I owe the success of this project to him also. Thank you so much to

Michael Disare, Matthew Moore, Allison Brink, Agata Ryzko, and Daniel Krochmal for being my colleagues and friends along the way, and thank you to all alumni I knew as well. Special thanks also goes to members of the Weizmann lab and the Rice lab for support they provided over the years, and huge thanks goes out to the beamline-hosts at Argonne National Lab for their expertise at critical moments of my research.

Outside of the lab, I must thank my family and my partner Natalie for providing extensive support and encouragement. Nat provided incredible emotional support and reminded me every day that the hard times don't last forever, and that the future is bright. My parents Aideen and John have encouraged me at each step along the way from the very beginning, and without them nurturing my interest in science (even things not geological) I'd not be the person I am today. My sisters Ailbhe and Maeve have been kind and understanding, and are two amazing, brilliant people who I'm proud to be a brother to. The family dog Poppy reminds me that from her perspective a Ph.D. is meaningless and that at the end of the day I am good for walks, food, the occasional scratch, and that's enough. I'm very fortunate to have an amazing extended family who have been nothing but incredibly kind and supportive to me, and I thank them also for their support.

I'd like to thank all of the friends I made along the way, including my classmates, and especially Dr. Jon Keim, the ever patient roommate. I'd also like to thank the intramural softball team (Go Beer!), inner-tube water polo team, and other assorted groups of friends who made my time here amazing. Friends I made through Nat, including her ever-tolerant roommate Sam, are also deeply appreciated. I had the privilege of pursuing comedy as a hobby during my work, and that has made a lasting impact to my presentation style and communication. Primarily, I'd like to thank members of the Excited State improv troupe, with whom I had the amazing experience of founding and performing in a comedy show with some of the University of Chicago's amazing

faculty. Dr. Jamie Dolan (physicist), Dr. Yeva Shan (biologist), Prof. Randy Picker (law professor), Dr. Marshall Chin (physician), Dr. Naoko Muramatsu (public health), Dr. João Caldeira (physicist), Dr. Mike Chamberlain (physician), Stephanie Timpone (engineer), Kate Wiseman (educator), Jakob Schafer (engineer), Dr. Meredith Cenzer (entomologist), Tim Sozko (coach & director), and John Stoops (producer) for being fantastic to work with and instrumental in the Excited State show. I would also like to make a special mention for the late Dr. Takintayo Akinbiyi (statistician), who is greatly missed and will be one of the funniest men on stage I've ever had the fortune to improvise with. In addition, I'd like to thank Eric Henderson for the Science Riot comedy program, and Jonah Jurkens for helping me get started as a stand-up comic. I'm proud of the comedy I got to do, and will try to carry it on wherever I go from here!

Chapter 1: Introduction to RNA crystallography and Fluorogenic aptamers

1.1 The importance of RNA structural biology

Structural biology's explosive growth as a field has led to a transformation in how we think about the complex chemical functions in living organisms. In pursuing answers about protein, DNA and RNA function, investigators in the last century have discovered some of the most complicated chemical systems, processes and pathways known to mankind, and have applied a wide variety of powerful investigative tools to probe these questions. As proteins rose to dominate the structural map of biology, the structures of RNA began to be elucidated and scrutinized. Eventually, biologists realized that RNA contained a whole evolutionary history of chemically complicated and diverse functions that were derived from tertiary and quaternary structure, rather than primary sequence alone. It is now generally accepted that RNA may have been the first of the three main biomolecules (DNA, RNA, and proteins) to have developed critical biochemical processes for life billions of years ago such as self-replication, self-cleavage, and regulation of sequences [1]. Structural RNA biology has become an essential field in fundamental and applied biology, with modern medicine becoming more reliant on understanding RNA structure and function to fight auto-immune diseases and viral pathogens by targeting their structured RNA features [2], including one particular reverse-RNA genome coronavirus called SARS-CoV-2, which caused the Covid-19 pandemic [3].

RNA structures have been discovered and engineered that perform a wide variety of interesting functions, from methyltransferase catalysis [4] to ribosomal frame-shifting for viral replication [5], and riboswitches that regulate gene translation [6]. Naturally occurring riboswitches – allosteric RNAs that bind specific metabolite ligands – use conformational changes upon substrate binding to induce differentiation in biological pathways, including gene translation [7]. Ribozymes

are RNA structures that perform specific chemical reactions; mainly site-specific self-cleaving reactions of the phosphodiester bond that makes up the RNA backbone. These cleavages complement many post-translational protein driven processes in the transcriptome [8]. Internal ribosome entry sites (IRESs), are employed by viral 5' untranslated regions (5'UTRs) to cause translation to occur without the canonical translation signals such as eukaryotic initiation factor (eIF) 5'-capping, and therefore bypassing steps that allow for anti-viral proof-reading by the ribosomal machinery [2]. 3'UTR-Cap-independent-translation enhancers (3'CITEs)[9] such as the pea-enation-mosaic virus 3'CITE (PEMV2) [10] are structured RNAs that are able to perform long-range interactions over thousands of nucleotides to assist viral translation via structurally interfacing with the viral genome's 5'UTR, demonstrating that RNA structures can function across large stretches of genomic space and are not merely localized to short stretches of transcripts. Naturally occurring structured RNAs are not limited to well-defined regions of translated RNA, as long-non-coding RNAs (lncRNAs) have turned from being considered informational junk to an unexplored frontier for the structureome [11]. Non-pathogenic ncRNAs in the "dark genome" are continuously being identified to serve as modulators in fundamental processes such as the lncRNA Braveheart (Bvht), with its expression tied to cariogenic differentiation in stem cells [12,13], or the lncRNA HOTAIR which is identified as a cancer biomarker [14].

Engineering attempts have been extremely fruitful at mimicking and developing the potential of structured RNAs. Examples of synthetic ribozymes have been developed include an aptazyme [15] that uses the hammerhead ribozyme to cleave upon binding a tetracycline substrate [16], a riboswitch that uses the universal cofactor S-adenosylmethionine (SAM) as a methyl-group donor to methylate a site-specific adenosine [4], and a ribozyme that specifically bind the Pb^{2+} cation to induce RNA cleavage [17]. These examples demonstrate the additional diverse capabilities of

ribozymes to function as context- and sequence-specific genomic switches in these hypothetical biological scenarios. Another example of structured RNAs are fluorogenic RNA aptamers, which are synthetically engineered allosteric RNAs that bind a ligand to form a fluorescent complex [18], and have been widely applied to monitor cellular processes, particularly the RNA transcriptome *in vivo*.

1.2 Methods to investigate RNA structure

Investigating structured RNA and their biology can be mainly broken down into four components; sequencing and modification localization, computational structure prediction, chemical and enzymatic probing, and empirical 3D-structure elucidation, (Figure 1.1). Through revolutionary tools such as RNA-seq and bioinformatics, RNA sequences and their post-translational modifications have been studied and localized in cells on a tremendous scale via analysing differential gene expression [19]. Computational prediction of RNA structure is improving quickly, with powerful tools that are able to make more accurate RNA models in development, advancing beyond the aging but straight-forward minimum-free-energy (MFE) secondary structure prediction tools such as unafold [20]. These methods are improving at a dramatic rate as tools similar to the protein structure-predicting alphafold [21] are developed for RNA such as the Atomic Rotationally Equivariant Scorer (ARES) deep-learning model [22], but these software still need fundamental information on RNA structure from empirically derived structures and in-vitro assays [23].

Methods to empirically derive 3D models of RNA structure are similar to those used to study protein structure (Figure 1.1). The protein data bank (pdb; https://www.rcsb.org/) was

originally established to coalesce the models of proteins from researchers around the world, and became the home of thousands of RNA structures as the structural biology field grew. X-ray diffraction serves as the primary tool for detailed structural research, though this is now supplemented by neutron diffraction and the increasingly popular cryo-electron microscopy (Cryo-EM) [24]. Nuclear magnetic resonance (NMR) in both solution [25] and solid [26] formats have been very popular in the structural biology community to investigate protein and RNA structure and dynamics. NMR has been particularly desirable for structures that do not form crystal lattices very easily such as membrane proteins [27], which are notoriously difficult to express, solubilize and crystallize. For similar reasons, dynamic and homogenous RNAs that are recalcitrant to crystallography have also been studied via NMR. However, NMR datasets lack the level of structural detail that an X-ray structure can provide [28], particularly for larger proteins and RNAs.

Although the appeal of diffraction techniques, NMR, and electron microscopy is very high for their detail and accuracy, they can be very hard to use without prior empirical information about the RNA structure, and are not usually the first methods of investigation. *In vitro* chemical probing techniques such as SHAPE (Selective 2′-hydroxyl acylation analyzed by primer extension) [29] and in-line probing [30] are vital for deducing secondary structures of RNA under various buffer and ligand contexts by identifying nucleotides that are more dynamic or solvent-exposed. Fluorescence resonance energy transfer (FRET) has become a valuable technique for assessing proximity within an RNA structure, providing important clues about tertiary interactions and dynamics [31]. By taking advantage of the suite of naturally occurring and highly specific RNases that are available, such as ribonuclease T1 which cleaves the 3’ side of unpaired guanosines [32], or V1 which cleaves dsRNA [32], tertiary structure features can be assigned *in vitro* without a 3D-structure. Finally, reverse transcription can be vital for checking and understanding how the RNA

is translated from DNA [33], particularly in cases where structured or catalytic RNAs may alter the performance of polymerases such as T7 that are typically used for in vitro RNA synthesis [34,35].

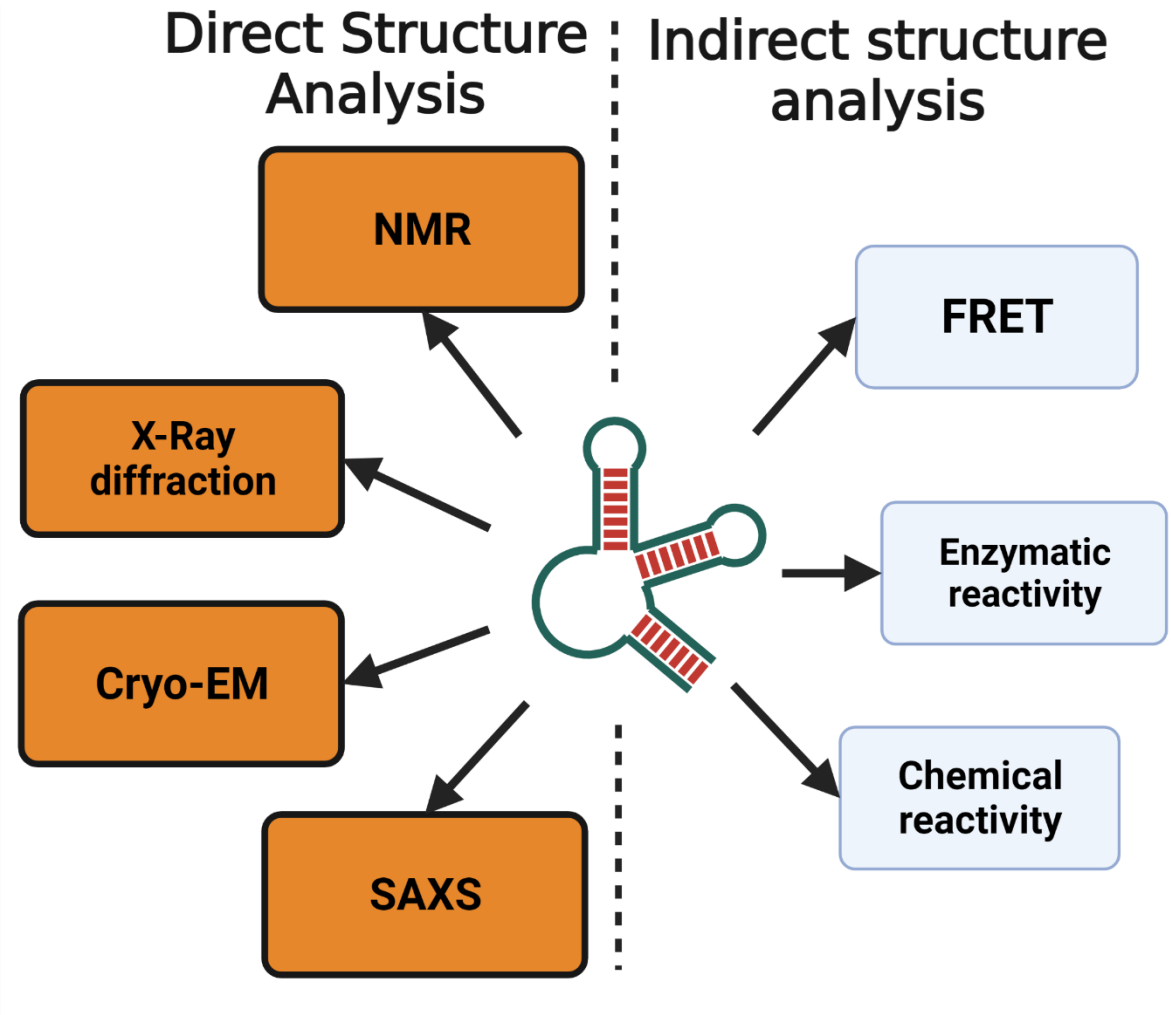


Figure 1.1: Direct and indirect methods of analysing RNA structure. Direct structure analysis gives insights into overall topology and tertiary structure and generating 3D models, potentially down to almost atomic-detail. Indirect structure analysis can indicate what parts of a sequence are reactive in various chemical or enzymatic contexts, the composition of the sequence, and relative proximity within a structure. Created using Biorender™.

1.3 RNA crystallography

The first published X-ray diffraction RNA structure in 1978 was that of a transfer RNA (tRNA); an examples of a fundamental structure that facilitates ribosomal protein synthesis through the triplet anticodon [36], followed by the first RNA helix structure published in 1989 [37]. By the mid-1990s, RNA structure gained traction with the publication of the hammerhead ribozyme structures, which gave a first detailed look at a naturally occurring catalytic RNA[38,39] after the initial proof of existence in the early 1980s [40]. From the beginning of the 2000s, RNA structures have been deposited to the pdb at an accelerating rate [41], including examples of riboswitches [42], more ribozymes [43], frame-shifting elements [44] and many RNAs in functional complexes with proteins [45].

Without the presence of other biomolecules, RNA has biophysical properties that make forming a crystal lattice particularly difficult, often seen as much greater than issues that make protein crystallography challenging [41]. Whereas proteins have 20 amino acids, and therefore a colossal range of potential sterics, charge interfaces and hydrophobic properties, RNA is composed of 4 types ribonucleotides that all exhibit similar chemical interfaces. The negatively charged phosphate backbone of all nucleic acids presents few opportunities for chemical discernment for creating specific crystal contacts, making lattice formation unfavourable. In addition, RNA lifetime in aqueous solutions is hampered by frequent cleavage when the 2'OH breaks the phosphodiester bond via nucleophilic attack. This cleavage is exacerbated by Mg^{2+} and other divalent ions that can catalyse the reaction by acting as a Lewis acid to make nucleophilic attack more favourable, and often cannot be excluded during structural studies as divalent ions can be critical to functional RNA structures [46]. RNA is also generally highly dynamic with many more rotatable bonds than proteins [47] inhibiting crystallization and also making it difficult to ensure the

correct fold of the RNA is being crystallized. Even when high-resolution diffraction data is collected for an RNA, the phase problem is a major obstacle to overcome [48] without either a substantial component of the dataset composed of previously solved structure (such as a distinct loop or pseudoknot), heavy metal soaking[48], or heavy atom labelling, all of which can be very detrimental to the RNA decrease the relevance of a derived 3D model.

When engineering an RNA construct for crystallography to circumvent these problems, making changes to the sequence can assist lattice-formation but at a risk of losing key structural details. Removing or replacing flexible domains and regions with well-structured RNA such as UUCG hairpins, and double-stranded regions can sometimes mitigate these problems [49]. For example, the Varkud-Satellite ribozyme structure was solved by using a well-structured RNA pentaloop for both structure and phasing assistance [50]. Modifying the ends of RNA to have overhanging bases or blunt ends can make a substantial difference in RNA crystallization. As a general rule, smaller RNAs have been successfully crystallized more often than larger RNAs [51], making minimal functional motifs a focal point for crystallization construct design.

1.4 Chaperone-assisted RNA Crystallography

Another way to overcome RNAs reluctance to crystallize is to use crystallography chaperones that assist with lattice formation, restrict dynamics, and aid structure solving once a diffraction dataset has been collected. An ideal crystallography chaperone binds to the RNA specifically with high affinity, ensuring that the equilibrium of the complex does not favor dissociation in solution and exists overwhelmingly as the complete complex. The chaperone should also be highly crystallisable and able to form substantial lattice contacts with other

chaperones or with RNA, and also large enough to provide an easy solution to the phase problem by giving the crystallographer a large target to identify in the initial dataset. An early example of such a chaperone, or crystallization module, was the RNA-binding U1A domain, which was used to characterize the 3D structure of the Hepatitis Delta Virus (HDV) ribozyme [52]. However, the U1A protein is relatively small (11 kDa), making it difficult to use for phase problems with all but very small RNAs. One alternative that has turned up are antibody fragments (Fabs) with a beta-sheet rich structure that bind tightly and specifically to a target via their complementarity determining regions (CDRs) [53]. Fabs can be easily expressed and purified, then complexed with the RNA to set up crystal trials [54]. When solving collected datasets, crystals containing Fabs can utilize their substantial mas (48 kDa) to make solving the phase problem much easier, as the Fab can be identified in the dataset via molecular replacement.

To use a Fab as a chaperone for an RNA of interest, there are two main approaches that can be considered (Figure 1.2). When an RNA is poorly characterized, *de-novo* phage display selections can be performed to identify Fab binders from a library of Fab conjugated to the pIII protein of M13 phage. This is advantageous as the RNA can be mostly unmodified from its native sequence for the selection [55]. When investigating RNAs with detailed pre-existing biochemical characterization such as SHAPE or in-line probing data, grafting an RNA hairpin with a loop that is already known to bind a Fab is a more straightforward option to apply to targets than phage display. For example Fab BL3-6 was mutated from the selection-identified Fab BL3 [54], and has been successfully deployed across a variety of RNA targets to find there structures [44,54,56–58]. In all of these cases, a previously identified hairpin proven to not be sequence specific for a particular RNA's function was changed to a BL3-6 binding AAACA loop closed by a GC pair to allow the formation of the Fab-RNA complex. Other examples of this approach can be seen with Fab HAVx

which was used to elucidate the structure of the hepatitis A IRES domain [59], and the synthetic self-alkylating ribozyme [60].

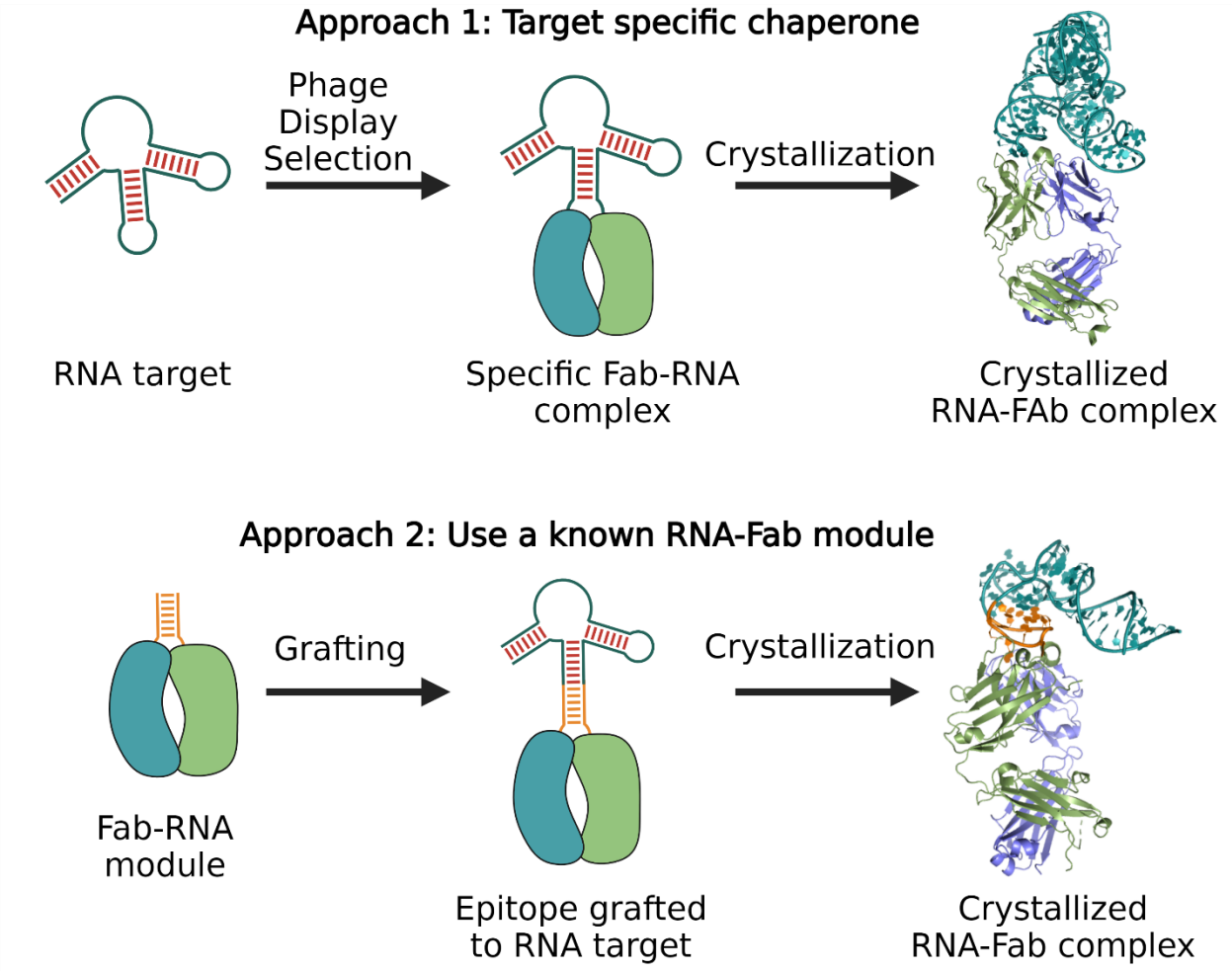


Figure 1.2: Approaches for using Fabs as crystallization chaperones. Using phage display, Fabs can be identified that bind to an RNA tertiary structure specifically and therefore used as chaperones with no further modifications to the RNA. The second approach involves replacing a non-essential helix or stem with a hairpin containing a Fab-binding loop, and using a pre-determined Fab as the crystallization chaperone. Created using Biorender™.

1.5 History of fluorescent visualization of RNA

For decades, visualizing RNA dynamics *in vivo* in real time had been a desire in many fields, and did not start to become a reality until the late 1990s [61]. The intrinsic problem of following RNAs in real time was that they were not specifically responsive to any radiation that could be used for microscopy, requiring that the RNA had an additional substrate or protein to be detected via fluorescence microscopy instead[62,63]. The RNA-phage derived MS2 coat protein had long been known to bind a particular RNA secondary structure with a tight low nM affinity[64–66], and once genetically inserted adjacent to an RNA of interest, an MS2-GFP (Green Fluorescent Protein) complex could bind to the RNA in real time for visualization via fluorescent microscopy, the first example being to follow ASH1 mRNA in yeast cells [61]. While this technique was a great step forward in real-time RNA analysis, it has two major drawbacks. The first of these is that multiple copies of MS2 need to be added to a transcript to be followed, which then needs to bind multiple MS2 and GFP proteins, making the overall complex gigantic compared to the RNA transcript it is visualizing. The second issue is that at least two extra proteins needed to be substantially expressed by a cell to visualize an RNA using the MS2-GFP system, which can be prohibitively demanding for certain cellular environments and organisms as a whole to synthesize in addition to the natural biological processes. The method has been updated over the years with ways to address these problems by using a split-GFP approach[67], or a brighter version of GFP called enhanced GFP (eGFP)[68], and to this day, the MS2-GFP systems are the benchmark against which other options for *in vivo* in fluorescence microscopy are judged [69]. However, more streamlined methods for directly visualizing RNA transcripts that would be compatible with the current fluorescence microscopes were in high demand from the mid-2000s onwards.

To devise alternatives, scientists looked to the GFP structure itself for inspiration. The crystal structures of GFP [70] allowed researchers to identify how the protein was able to stabilize the 4-(p-hydroxybenzylidene)imidazolidin-5-one (HBI) component that is responsible for fluorescence. The HBI motif is an organic structure that auto catalytically forms from three amino acid residues (S65, F54, and G67) within the core of GFP, resulting in an aromatic ring system that absorbs light in the visible spectrum[71]. As HBI is formed from amino acids in the peptide chain, HBI is intrinsically covalently linked to the GFP protein. This is important, as the HBI motif in aqueous solution does not emit absorbed light as, making the GFP structure key to HBI's fluorescent properties. GFP functions by excluding water from the HBI fluorophore's pocket and preventing the HBI from moving through covalent and steric interactions, and then preventing wavelengths absorbed by the photo-active HBI motif from energetically relaxing via non-radiative pathways [72]. In other words, the HBI motif absorbs the wavelengths of light, then is unable to release that energy through translation, rotation, chemical or quenching means easily, and so fluorescent emission occurs as a substantial means of energetic relaxation – increasing the quantum yield (φ) of the HBI in the protein relative to what it would be in aqueous solution. Mimicking this effect with an allosteric RNA that bound a similarly-photo responsive ligand to activate fluorescence was hypothesized to be possible (Figure 1.3).

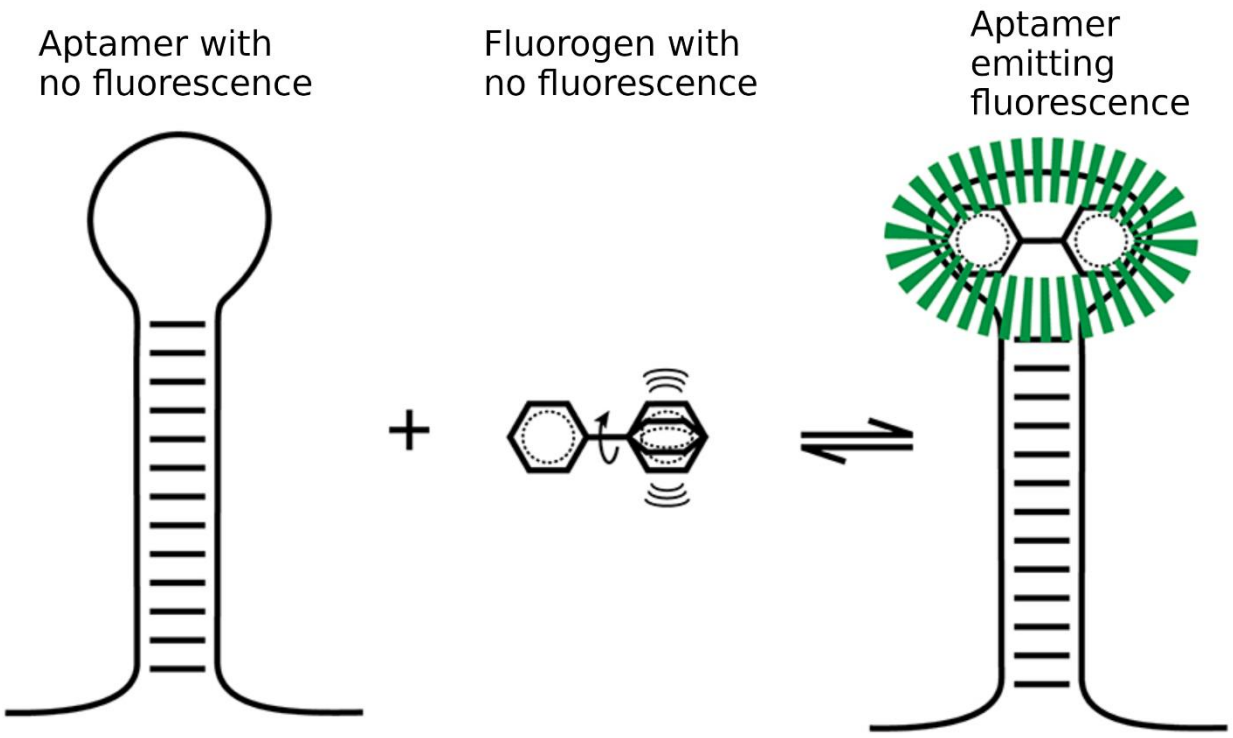


Figure 1.3. Fluorogenic aptamers emit light by binding a fluorogen that is not fluorescent in solution due to energy loss through non-radiative pathways. Once the ligand is bound by the aptamer, the radiative pathways become more favourable, quantum yield increases, and the absorbed light is now emitted as fluorescence. Figure adapted from Oullet. 2016[18].

The first fluorogenic aptamer with the potential to demonstrate this concept with an allosteric RNA was engineered through Systematic evolution of ligands by exponential enrichment (SELEX) [73] to bind the malachite green fluorogen (Figure 1.4) by Grate and Wilson [74], who originally intended to use the RNA aptamer for laser-mediated site-specific inactivation of RNA in cells. The pioneering malachite green aptamer (MGa) was hindered by the relatively high cytotoxicity [75] of the malachite green fluorogen, which produces free radicals upon excitation, and the fact that the RNA was liable to cleave upon radiation (as was the originally intended function). To improve upon these problems with MGa, less cytotoxic dyes were engineered form the fluorogenic cyanine dye thiazole-orange (TO) to produce dimethyl indole red (DIR) [76], which bound to the complementary aptamer with a high affinity (K_d = 87 nM). Another attempt to

improve upon MGa used the highly biorthogonal Hoechst derivatives (that were already well known to bind to the minor groove of AT-rich dsDNA helices [77]), and engineered the Hoechst ligand into an aptamer to detect transcription processes *in vivo* [78]. However, both of these new dye-aptamer complexes struggled to gain traction as biotools due to low fluorescent signals (DIR) or a lack of binding specificity (Hoechst), and neither rose to replace the RNA-binding-protein-GFP complexes that were dominant at the time.

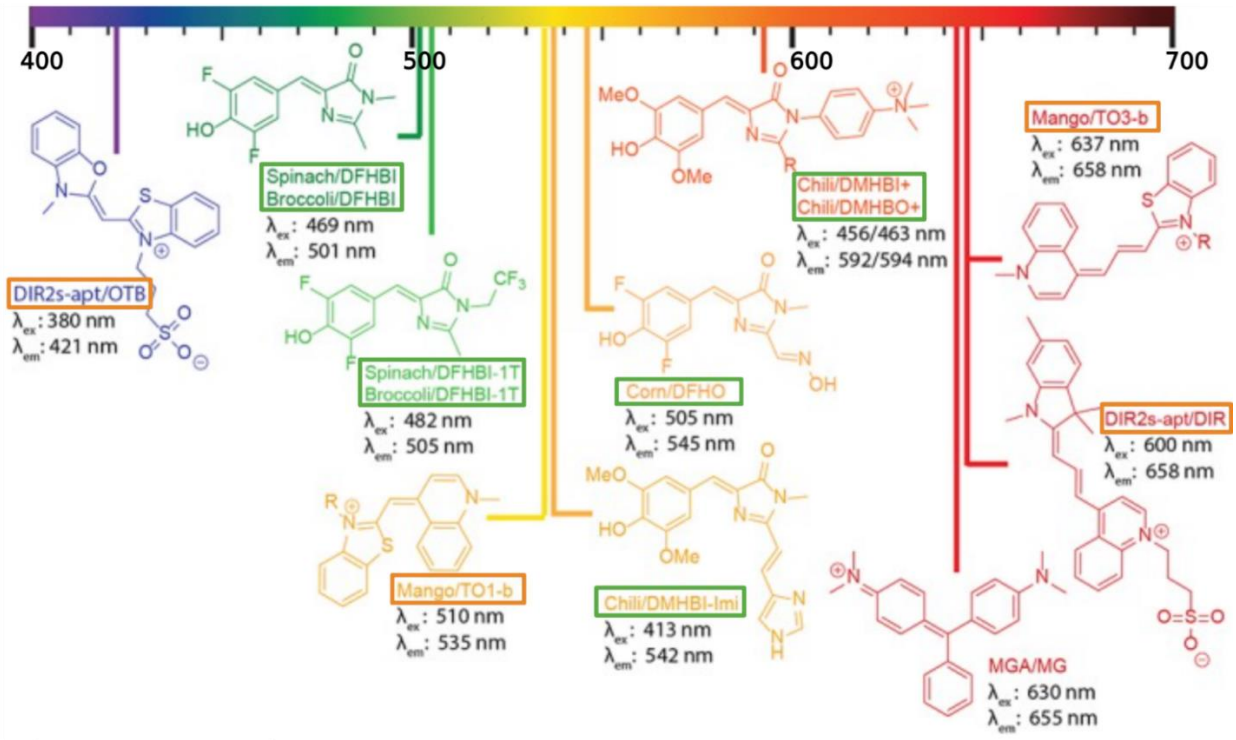


Figure 1.4. Pre-2019 fluorogens that have had RNA aptamers engineered to bind them. Derivates of GFP-binding fluorophores are boxed in green, and thiazole-orange derivatives are boxed in orange. Adapted from Truong *et al.*[79]

It was not until the Spinach fluorogenic aptamer was developed by Jaffrey and co-workers that the potential to directly observe in-vivo RNAs began to be truly realized (Figure 1.4). Through selecting RNA sequences to bind derivatives of the GFP active motif HBI, the Spinach aptamer was developed. Experimentation with the HBI motif lead to the creation of new fluorogens,

including (5Z)-5-[(3,5-Difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2,3-dimethyl-4H-imidazol-4-one (DFHBI) which exhibited enhanced properties to original HBI. With a great affinity for the bio-orthogonal ligand DFHBI in particular, Spinach quickly became a reliable tool for visualizing cellular dynamics without the need of a protein like the MS2-GFP system, as the low cytotoxicity of the GFP-derivatives and the brightness of Spinach were finally able to perform at the level the lumbering multimeric RNA-protein structures with simple and short RNAs that bound to a small fluorogen directly. Cell stress was decreased using Spinach, as even though multiple copies of spinach often needed to be cloned next to an RNA of interest, this would not result in the huge protein-RNA complexes of the MS2-GFP systems. Even with this breakthrough, Spinach still had two issues; poor folding in vivo, and a relatively poor affinity with its HBI-derivative ligands (around 500 nM) [80], the latter of which was in contrast to the tight binding of MS2-GFP complexes (10s nM).

In the last ten years, many new aptamers have been selected and optimized to overcome these initial problems and fulfil a wide range of needs (Table 1.1). The Jaffrey group engineered Spinach2 to have better folding properties and thermal stability which had been problems for the first iteration of Spinach [81], alongside improvements to the DFHBI fluorogen by replacing the methyl on the imidozolone ring with a tri-fluoroethyl moiety (DFHBI-1T: Figure 1.4) which when bound to Spinach2 exhibited a higher quantum yield. The same group were also responsible for the Broccoli aptamer, which was a substantial improvement over Spinach aptamers for in-vivo applications [80].

Many new aptamers have been developed that are fulfilling new needs, from stokes-shifting to even more compact 3D topologies. iSpinach was optimized for *in-vitro* applications [82]. An aptamer with very high affinity for its fluorogen came in the form of the Mango aptamer.

Presented by the Unrau group [83], Mango was selected to bind the from thiozole-orange family of fluorogen (TO-1, TO-3; Figure 1.4)) and exhibited a combination of tight affinity (3.2 nM) and brightness that Spinach aptamers had lacked (Table 1.1). These were further improved by structure-guided engineering to make MangoII-IV [84], and recently the Peach aptamer was engineered from Mango that has orthogonal ligand preference to TO3 over TO1, making dual-channel visualization with Mango (which preferentially binds the TO1 ligand) possible [85]. The DIR system was eventually improved with the development of the DIR2 aptamer which promiscuously bound two variants of cyanine dye, and allowed for dual-channel red/blue visualization [86]. Corn was an aptamer developed for even more precise quantitative reporting in cells by binding the less photolabile 3,5-difluoro-4-hydroxybenzylidene imidazolinone-2-oxime (DFHO) derivative of GFP ligands [87]. An aptamer with a very large stokes shift called Chilli was engineered from Spinach by Höbartner and co-workers, who also modified the GFP-derived ligand $DMHBO^+$ ([4-[(4Z)-4-[(4-hydroxy-3,5-dimethoxyphenyl)methylidene]-2-[(E)-hydroxyiminomethyl]-5-oxoimidazolidin-1-yl]phenyl]-trimethylazanium) to be positively charged and have the tightest affinity to its HBI-fluorophore at the time (12 nM) [87]. Although virtually every feature that could be desired of an aptamer has been engineered, from brightness to great folding *in vivo* to tight affinities, there are still challenges in engineering aptamers to have more and more of these attributes simultaneously[18,79].

One of the newest and most promising aptamers developed is the Pepper aptamer, which was evolved against an entirely new suite of ligands derived from HBC ((Z)-4-(1-cyano-2-(4-((2-hydroxyethyl)(methyl)amino)phenyl)vinyl)benzonitrile) [88]. With a low Kd (3.5 nM) to ligand HBC530 (Table 1.1) bright signal (3 times brighter than Broccoli), and a relatively long lifetime, Pepper has many properties that are highly promising as a biosensing RNA aptamer. In addition,

engineering of HBC ligands produced many different wavelength options for visualization, making Peppers potential versatility unparalleled.

Table 1.1: Tabulated data of selected fluorogenic aptamers. Extinction coefficients (ε) are for the RNA-ligand complex in ideal buffer conditions.

Fluorogen	Aptamer	λ ex (nm)	λ em (nM)	ε (ex. Coeff.)	ϕ (quantum yield)	Kd (nM)	References
Malachite Green	MGA	630	650	150000	0.19	117	[89]
DFHBI	Spinach	452	496	24271	0.72	537	[57,90]
TO1-Biotin	Mango	510	535	77500	0.14	3.2	[91]
TO3-Biotin	Mango	637	658	ND	ND	6-8	[91]
OTB	DIR2s-Apt	380	421	73000	0.51	662	[92]
DMHBO+	Chilli	456	592	22000	0.1	12	[92]
DMHBI+	Chilli	413	542	21000	0.4	63	[92]
DFHO	Corn	505	545	29000	0.25	70	[93]
DFHO	Squash	505	562	24600	0.25	54	[33,39]
HBC530	Pepper	485	530	65300	0.66	3.5	[88]

1.6 Structural findings of fluorogenic RNA aptamers

Since their inception, fluorogenic aptamers have attracted the attention of structural RNA scientists because of their functionality and relatively static structures that make them highly suitable for 3D structural investigations. As predicted from their intended designs and engineering, fluorogenic aptamers have been shown in crystal structures to work by binding a fluorophores that do not have significant fluorescence in solution and increasing the proportion of energetic relaxation via/through fluorescent emission pathways.

The first fluorogenic aptamer MGa was successfully crystallized with the malachite green ligand soon after its discovery [89]. The structure revealed that the malachite green ligand bound within a stacked formation between a base-quadruple and a base-pair backed up with two base-

triples on the other side of the malachite green ligand (Table 1.2). This first example showed many general features that would be exhibited by later structures of fluorogenic aptamers, with the ligand stabilized by a complex stacking arrangement within a bulge supported by an out of plane base system that assists the formation of a complete binding pocket. Interestingly, the malachite green ligand showed no specific hydrogen bonding to the RNA that was expected from the known structures of riboswitches at the time. Instead MGa entirely relied upon planar aromatic stacking between the similar profiles of aromatic structure in the ligand and the purine or pyrimidine rings of the RNA. Although the crystal structure was solved with Sr^{2+} atoms for solving the crystallographic phase problem, MGa exhibited no strict requirements for divalent cations, and the lack of Sr^{2+} in the core of the MGa structure agreed with this finding.

When the structure of Spinach was discovered[57,90], it was similar to MGa but also exhibited two new features for a fluorogenic aptamer (Table 1.2). Similar to MGa, Spinach binds DFHBI by forming two sets of stacking nucleic acids that sandwich the ligand, with an out-of-plane gateway nucleotide to complete the pocket. However, unlike MGa, Spinach's bulge unexpectedly formed a G-quadruplex; a structure that can form from G-rich strands of RNA that is reliant upon monovalent cations [94], particularly potassium. This explained Spinach's observed reliance on potassium cations for function, a feature that became a common feature in future fluorogenic aptamer structures that also relied on G-quadruplexes[79]. In addition to the G-quadruplex, Spinach was different to MGa because it bound DFHBI with a hydrogen-bond through a G nucleotide in the same plane to the ligand.

Following Spinach, multiple more GFP-derivative binding aptamers have had their structures solved including Corn [95], iSpinach [82], Chilli [92], and Squash [96]. With the notable exception of Squash, all of the aptamers bind their fluorophores using a G-quadruplex platform stack

sandwiched with another non-G-quadruplex stack on the other side. Corn is a dimer, and thus is the only aptamer to bind it's fluorogen with two unconnected G-quadruplex stacks. Most of the aptamers, except for Chilli, have bases out of plane with the ligand and stacking arrangement to complete their respective binding pockets (Table 1.2).

Mango aptamers, and presumably Peach, also use a G-quadruplex motif to bind the TO1 and TO3 ligands with very tight affinity [91]. Mango has been structurally engineered through a series of variants, in the form of Mango II-IV, which have all had their crystal structures solved. Mango-I - the original mango - binds the thiozole-orange ligands on top of a G-quadruplex and supported by a series of "propeller-like" nucleotides, making Mango the only aptamer to bind it's ligand without two sets of parallel planar stacks on either side of the ligand [96] and - to great surprise at the time - Mango I bound TO-ligands in multiple conformations. Mango-II, which has an A22U mutation from the original Mango I aptamer, and was found to discriminate better between the ligands with a five-fold better affinity for TO1 than TO3 [97] partly by removing the potential for multiple binding conformations that Mango-I exhibited. Mango-III was engineered with a brighter (18% increase) in signal and found to have an extra pseudoknot that supported the ligand binding from a distance [98]. Mango IV demonstrated great performance in cells, mainly due to its relatively flexible binding site which made it more resistant to perturbation in sometimes unpredictable cellular contexts [99], though lost some of the conformation specificity from earlier Mango generations.

Table 1.2: Descriptions of interactions around bound fluorogens for select aptamers with solved 3D structures.

pdb	Aptamer	H-bonding/other	Stacking 1.3	Stacking 1.2	Stacking 1.1	Fluorogen	Stacking 2.1	Stacking 2.2	Stacking 2.3	Side (out of plane)
1F1T	**MGA**	None	U11•A22•A26 triple	C10•G23•A27 triple	G8•C28 pair	Malachite Green	C7•G24•G29 A31 quadruple	C6•G33 pair	C5•G34 pair	U25, A30•A9
4KZD	**Spinach**	G28•H-bond keto	C25•G63 pair	G22•G26•G57•G61 QPx	G23•G27•G54•G59 QPx	DFHBI	U29•U50•A53 triple	C30•G52 pair	C31•G49 pair	A58
5V3F	**Mango**	None	G8•G13•G18•G23 QPx	G9•G14•G19 G24•QPx	G10•G16•G21•G26 QPx	TO1-Biotin	U15, A20, A25 (also side)	None	None	U15, A20, A25
6DB8	**DIR2s-Apt**	G39•SO3 moiety H-bond	G37, A13 stack individually	A14•U38•U42 triple	A15•G39•A41 triple	OTB	A40 stacking	None	None	None
7OAX	**Chili**	In plane G15 H-bond (C40 pair) (and K+)	G10•A11•C44 triple	G9•G12•G36•G42 QPx	G13•G32•G37•G41 QPx	DMHBO+	C16•G39, G14•G31	G17•U30 pair	C18•G29 pair	None
6E81	**Corn**	Dimerizes, A11 H-bonds to imine	A10•U17•A21•U27 QPx	G13•G16•G23•G26 QPx	G12•G15•G22•G25 QPx	DFHO	Dimer*	Dimer*	Dimer*	A11, A24
7KVV	**Squash**	A69, A70 H-bond phenolate	G42•G46•G71•G73 quadruple	A12•G41•G47•C72 quadruple	G13•A40•A70 triple	DFHO	U15•A69	G16•A49 pair	A17•U68 pair	G14, A48 (U39 pair)

Although researchers have made great efforts to acquire structural data on fluorogenic aptamers, not all have been successfully crystallized. The structure of Broccoli is presumably very similar to Spinach based on the mostly identical sequence, secondary fold, and metal dependencies, which has decreased interest in a crystal structure [100]. However, there have undoubtedly been attempts to crystallize Broccoli, potentially indicating a more dynamic structure that has difficulties forming crystal lattices. The very recent Peach remains uncrystallised with no other 3D-structural data, and is interesting as it differentiates the binding of the TO1 and TO3 ligands like Mango-II but this time favouring TO3, indicating Peach has structural features that the other Mango derivatives do not. Until very recently, the structure of a new fluorogenic aptamer named Pepper was unknown [88,101], with its novel ligand scope, tight affinity, and excellent performance *in vivo* making it an interesting target to pursue.

1.7 The aims of this thesis

In this thesis, the structure and function of the recently engineered Pepper aptamer are presented and investigated. By using the Fab-BL3-6 chaperone, crystal structures of the Pepper aptamer bound to the ligands HBC530 and HBC599 were solved to 2.3 Å and 2.7 Å resolution respectively. Mutational analysis confirmed observations of the structure by demonstrating the necessity of key interactions. By using the crystal structures to guide investigation, the unique Mg^{2+} occupancy and dependency have been more thoroughly understood, and the potential for future work with the Pepper aptamer has expanded to include potentially improving small molecule biosensors *in vivo*.

1.8 References for Chapter 1.

1. Robertson, M. P.; Joyce, G. F. Origins of the RNA world. *Cold Spring Harbor Perspectives in Biology* 1–22 (2012).

2. Weingarten-Gabbay, S.; Elias-Kirma, S.; Nir, R.; Gritsenko, A. A.; Stern-Ginossar, N.; Yakhini, Z.; Weinberger, A.; Segal, E. Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **351**, (2016).

3. Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N.; Bi, Y.; Ma, X.; Zhan, F.; Wang, L.; Hu, T.; Zhou, H.; Hu, Z.; Zhou, W.; Zhao, L.; Chen, J.; Meng, Y.; Wang, J.; Lin, Y.; Yuan, J.; Xie, Z.; Ma, J.; Liu, W. J.; Wang, D.; Xu, W.; Holmes, E. C.; Gao, G. F.; Wu, G.; Chen, W.; Shi, W.; Tan, W. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet (London, England)* **6736**, 1–10 (2020). [PMID:32007145]

4. Scheitl, C. P. M.; Ghaem Maghami, M.; Lenz, A. K.; Höbartner, C. Site-specific RNA methylation by a methyltransferase ribozyme. *Nature* **587**, 663–667 (2020). [PMID:33116304]

5. Marcheschi, R. J.; Staple, D. W.; Butcher, S. E. Programmed Ribosomal Frameshifting in SIV Is Induced by a Highly Structured RNA Stem-Loop. *Journal of Molecular Biology* **373**, 652–663 (2007). [PMID:17868691]

6. Breaker, R. R. Riboswitches and the RNAWorld. *Cold Spring Harbor Laboratory Press* 1–23 (2012).

7. Sherlock, M. E.; Breaker, R. R. Former orphan riboswitches reveal unexplored areas of bacterial metabolism, signaling, and gene control processes. *Rna* **26**, 675–693 (2020). [PMID:32165489]

8. Weinberg, C. E.; Weinberg, Z.; Hammann, C. Novel ribozymes: discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic acids research* **47**, 9480–9494 (2019). [PMID:31504786]

9. Gao, F.; Simon, A. E. Differential Use of 3′CITEs by the Subgenomic RNA of Pea Enation Mosaic Virus 2. *Virology* **519**, 194–204 (2017).

10. Truniger, V.; Miras, M.; Aranda, M. A. Structural and functional diversity of plant virus 3′-cap-independent translation enhancers (3′-CITEs). *Frontiers in Plant Science* **8**, 1–14 (2017).

11. Lu, Z.; Chang, H. Y. Decoding the RNA structurome. *Current Opinion in Structural Biology* **36**, 142–148 (2016). [PMID:26923056]

12. Hou, J.; Long, H.; Zhou, C.; Zheng, S.; Wu, H.; Guo, T.; Wu, Q.; Zhong, T.; Wang, T. Long noncoding RNA Braveheart promotes cardiogenic differentiation of mesenchymal stem cells in vitro. *Stem Cell Research and Therapy* **8**, 1–13 (2017). [PMID:28095922]

13. Klattenhoff, C. A.; Scheuermann, J. C.; Surface, L. E.; Bradley, R. K.; Fields, P. A.; Steinhauser, M. L.; Ding, H.; Butty, V. L.; Torrey, L.; Haas, S.; Abo, R.; Tabebordbar, M.; Lee, R. T.; Burge, C. B.; Boyer, L. A. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* **152**, 570–583 (2013). [PMID:23352431]

14. Gupta, R. A.; Shah, N.; Wang, K. C.; Kim, J.; Horlings, H. M.; Wong, D. J.; Tsai, M. C.; Hung, T.; Argani, P.; Rinn, J. L.; Wang, Y.; Brzoska, P.; Kong, B.; Li, R.; West, R. B.; Van De Vijver, M. J.; Sukumar, S.; Chang, H. Y. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010). [PMID:20393566]

15. Zhong, G.; Wang, H.; Bailey, C. C.; Gao, G.; Farzan, M. Rational design of aptazyme riboswitches for efficient control of gene expression in mammalian cells. *eLife* **5**, (2016). [PMID:27805569]

16. Wittmann, A.; Suess, B. Selection of tetracycline inducible self-cleaving ribozymes as synthetic devices for gene regulation in yeast. *Molecular BioSystems* **7**, 2419–2427 (2011). [PMID:21603688]

17. Wedekind, J. E.; Mckay, D. B. Crystal structure of a lead-dependent ribozyme revealing metal binding sites revelant to catalysis. *Nature Structural Biology* **6**, 261–268 (1999).

18. Ouellet, J. RNA fluorescence with light-Up aptamers. *Frontiers in Chemistry* **4**, 1–12 (2016).

19. Stark, R.; Grzelak, M.; Hadfield, J. RNA sequencing: the teenage years. *Nature Reviews Genetics* **20**, 631–656 (2019). [PMID:31341269]

20. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406–3415 (2003). [PMID:12824337]

21. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). [PMID:34265844]

22. Townshend, R. J. L.; Eismann, S.; Watkins, A. M.; Rangan, R.; Karelina, M.; Das, R.; Dror, R. O. Geometric deep learning of RNA structure. *Science* **373**, 1047–1051 (2021). [PMID:34446608]

23. Xu, X.; Chen, S. J. Topological constraints of RNA pseudoknotted and loop-kissing motifs: Applications to three-dimensional structure prediction. *Nucleic Acids Research* **48**, 6503–6512 (2020). [PMID:32491164]

24. Baker, M. Cryo-Electron Microscopy Shapes Up. *Nature* **561**, 565–567 (2018).

25. Markwick, P. R. L.; Malliavin, T.; Nilges, M. Structural biology by NMR: Structure, dynamics, and interactions. *PLoS Computational Biology* **4**, (2008). [PMID:18818721]

26. Van Der Wel, P. C. A. New applications of solid-state NMR in structural biology. *Emerging*

Topics in Life Sciences **2**, 57–67 (2018).

27. Puthenveetil, R.; Vinogradova, O. Solution NMR: A powerful tool for structural and functional studies of membrane proteins in reconstituted environments. *Journal of Biological Chemistry* **294**, 15914–15931 (2019). [PMID:31551353]

28. Barnwal, R. P.; Yang, F.; Varani, G. Applications of NMR to structure determination of RNAs large and small. *Archives of Biochemistry and Biophysics* **628**, 42–56 (2017).

29. Smola, M. J.; Rice, G. M.; Busan, S.; Siegfried, N. A.; Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nature Protocols* **10**, 1643–1669 (2015).

30. Mlýnský, V.; Bussi, G. Understanding in-line probing experiments by modeling cleavage of nonreactive RNA nucleotides. *Rna* **23**, 712–720 (2017). [PMID:28202709]

31. Stephenson, J. D.; Kenyon, J. C.; Symmons, M. F.; Lever, A. M. L. Characterizing 3D RNA structure by single molecule FRET. *Methods* **103**, 57–67 (2016). [PMID:26853327]

32. Nilsen, T. W. RNA structure determination using nuclease digestion. *Cold Spring Harbor Protocols* **8**, 379–382 (2013). [PMID:23547152]

33. Schwaber, J.; Andersen, S.; Nielsen, L. Shedding light: The importance of reverse transcription efficiency standards in data interpretation. *Biomolecular Detection and Quantification* **17**, 1–7 (2019).

34. Gholamalipour, Y.; Karunanayake Mudiyanselage, A.; Martin, C. T. NAR breakthrough article 3 end additions by T7 RNA polymerase are RNA self-templated, distributive and diverse in character—RNA-Seq analyses. *Nucleic Acids Research* **46**, 9253–9263 (2018).

[PMID:30219859]

35. Steitz, T. A. The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Current Opinion in Structural Biology* **19**, 683–690 (2009). [PMID:19811903]

36. Sussman, J. L.; Holbrook, S. R.; Warrent, R. W.; Church, G. M.; Kim, S.-H. Crystal Structure of Yeast Phenylalanine Transfer RNA. *Journal of Molecular Biology* **123**, 607–630 (1978).

37. Dock-Bregeon, A. C.; Chevrier, B.; Podjarny, A.; Johnson, J.; de Bear, J. S.; Gough, G. R.; Gilham, P. T.; Moras, D. Crystallographic structure of an RNA helix: [U(UA)6A]2. *Journal of Molecular Biology* **209**, 459–474 (1989). [PMID:2479753]

38. Leonard, G. A.; McAuley-Hecht, K. E.; Ebel, S.; Lough, D. M.; Brown, T.; Hunter, W. N. Crystal and molecular structure of r(CGCGAAUUAGCG): an RNA duplex containing two G(anti)· A(anti) base pairs. *Structure* **2**, 483–494 (1994). [PMID:7922026]

39. Pley, H. W.; Flaherty, K. M.; McKay, D. B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68–74 (1994). [PMID:7969422]

40. Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983). [PMID:6197186]

41. Felden, B. RNA structure: experimental analysis. *Current Opinion in Microbiology* **10**, 286–291 (2007). [PMID:17532253]

42. Garst, A. D.; Edwards, A. L.; Batey, R. T. Riboswitches: Structures and mechanisms. *Cold Spring Harbor Perspectives in Biology* **3**, 1–13 (2011). [PMID:20943759]

43. Müller, S.; Appel, B.; Balke, D.; Hieronymus, R.; Nübel, C. Thirty-five years of research into ribozymes and nucleic acid catalysis: where do we stand today? *F1000Research* **5**, 1511 (2016). [PMID:27408700]

44. Roman, C.; Lewicka, A.; Koirala, D.; Li, N.-S.; Piccirilli, J. A. The SARS-CoV-2 Programmed −1 Ribosomal Frameshifting Element Crystal Structure Solved to 2.09 Å Using Chaperone-Assisted RNA Crystallography. *ACS Chemical Biology* **16**, 1469–1481 (2021). [PMID:34328734]

45. Ippolito, J. A.; Kanyo, Z. F.; Wang, D.; Franceschi, F. J.; Moore, P. B.; Steitz, T. A.; Duffy, E. M. Crystal structure of the oxazolidinone antibiotic linezolid bound to the 50S ribosomal subunit. *Journal of Medicinal Chemistry* **51**, 3353–3356 (2008). [PMID:18494460]

46. Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. Cations in charge: Magnesium ions in RNA folding and catalysis. *Current Opinion in Structural Biology* **22**, 262–272 (2012). [PMID:22595008]

47. Murray, L. J. W.; Arendall, W. B.; Richardson, D. C.; Richardson, J. S. RNA backbone is rotameric. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 13904–13909 (2003). [PMID:14612579]

48. Golden, B. L.; Kundrot, C. E. RNA crystalization. *Journal of Structural Biology* **142**, 98–107 (2003).

49. Hendrix, D. K.; Brenner, S. E.; Holbrook, S. R. RNA structural motifs: Building blocks of a modular biomolecule. *Quarterly Reviews of Biophysics* **38**, 221–243 (2005). [PMID:16817983]

50. Suslov, N. B. The Crystal Structure of the Varkud Satellite Ribozyme: Implications for RNA Catalysis, Biology and Evolution, 2012

51. Holbrook, S. R. Structural principles from large RNAs. *Annual Review of Biophysics* **37**, 445–464 (2008). [PMID:18573090]

52. Ferré-D'Amaré, A. R.; Doudna, J. A. Crystallization and structure determination of a hepatitis delta virus ribozyme: Use of the RNA-binding protein U1A as a crystallization module. *Journal of Molecular Biology* **295**, 541–556 (2000). [PMID:10623545]

53. Carpenter, B.; Hemsworth, G. R.; Wu, Z.; Maamra, M.; Strasburger, C. J.; Ross, R. J.; Artymiuk, P. J. Structure of the human obesity receptor leptin-binding domain reveals the mechanism of leptin antagonism by a monoclonal antibody. *Structure* **20**, 487–497 (2012). [PMID:22405007]

54. Koldobskaya, Y.; Duguid, E. M.; Shechner, D. M.; Suslov, N. B.; Ye, J.; Sidhu, S. S.; Bartel, D. P.; Koide, S.; Kossiakoff, A. A.; Piccirilli, J. A. A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination. *Nature Structural and Molecular Biology* **18**, 100–107 (2011). [PMID:21151117]

55. Koldobskaya, Y.; Duguid, E. M.; Shechner, D. M.; Suslov, N. B.; Ye, J.; Sidhu, S. S.; Bartel, D. P.; Koide, S.; Kossiakoff, A. A.; Piccirilli, J. A. A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination. *Nature Structural and Molecular Biology* **18**, 100–107 (2011). [PMID:21151117]

56. Shelke, S. A.; Shao, Y.; Laski, A.; Koirala, D.; Weissman, B. P.; Fuller, J. R.; Tan, X.; Constantin, T. P.; Waggoner, A. S.; Bruchez, M. P.; Armitage, B. A.; Piccirilli, J. A. Structural basis for activation of fluorogenic dyes by an RNA aptamer lacking a G-

quadruplex motif. *Nature Communications* **9**, 1–10 (2018).

57. Huang, H.; Suslov, N. B.; Li, N. S.; Shelke, S. A.; Evans, M. E.; Koldobskaya, Y.; Rice, P. A.; Piccirilli, J. A. A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nature Chemical Biology* **10**, 686–691 (2014). [PMID:24952597]

58. Koirala, D.; Shelke, S. A.; Dupont, M.; Ruiz, S.; DasGupta, S.; Bailey, L. J.; Benner, S. A.; Piccirilli, J. A. Affinity maturation of a portable Fab–RNA module for chaperone-assisted RNA crystallography. *Nucleic Acids Research* **46**, 2624–2635 (2018). [PMID:29309709]

59. Koirala, D.; Shao, Y.; Koldobskaya, Y.; Fuller, J. R.; Watkins, A. M.; Shelke, S. A.; Pilipenko, E. V.; Das, R.; Rice, P. A.; Piccirilli, J. A. A conserved RNA structural motif for organizing topology within picornaviral internal ribosome entry sites. *Nature Communications* **10**, 1–13 (2019). [PMID:31399592]

60. Krochmal, D.; Shao, Y.; Li, N. S.; DasGupta, S.; Shelke, S. A.; Koirala, D.; Piccirilli, J. A. Structural basis for substrate binding and catalysis by a self-alkylating ribozyme. *Nature Chemical Biology* (2022).

61. Bertrand, E.; Chartrand, P.; Schaefer, M.; Shenoy, S. M.; Singer, R. H.; Long, R. M. Localization of ASH1 mRNA particles in living yeast. *Molecular Cell* **2**, 437–445 (1998). [PMID:9809065]

62. Mendez, M. A.; Szalai, V. A. Fluorescence of unmodified oligonucleotides: A tool to probe G-quadruplex DNA structure. *Biopolymers* **91**, 841–850 (2009). [PMID:19548317]

63. Kwok, C. K.; Sherlock, M. E.; Bevilacqua, P. C. Effect of loop sequence and loop length on the intrinsic fluorescence of G-quadruplexes. *Biochemistry* **52**, 3019–3021 (2013).

[PMID:23621657]

64. Lim, F.; Spingola, M.; Peabody, D. S. The RNA-binding site of bacteriophage Qβ coat protein. *Journal of Biological Chemistry* **271**, 31839–31845 (1996). [PMID:8943226]

65. Bernardi, A.; Spahr, P. F. Nucleotide sequence at the binding site for coat protein on RNA of bacteriophage R17. *Proceedings of the National Academy of Sciences of the United States of America* **69**, 3033–3037 (1972). [PMID:4507620]

66. Fouts, D. E.; True, H. L.; Celander, D. W. Functional recognition of fragmented operator sites by R17/MS2 coat protein, a translational repressor. *Nucleic Acids Research* **25**, 4464–4473 (1997). [PMID:9358153]

67. Valencia-Burton, M.; McCullough, R. M.; Cantor, C. R.; Broude, N. E. RNA visualization in live bacterial cells using fluorescent protein complementation. *Nature Methods* **4**, 421–427 (2007). [PMID:17401371]

68. Cormack, B. P.; Valdivia, R. H.; Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996). [PMID:8707053]

69. Garcia, H. G.; Tikhonov, M.; Lin, A.; Gregor, T. Quantitative Imaging of Transcription in Living Drosophila Embryos Links Polymerase Activity to Patterning. *Current Biology* **23**, 2140–2145 (2013). [PMID:24139738]

70. Ormo, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y. Crystal Structure of the Aequorea victoria Green Fluorescent Protein. *Science* **273**, 1392–1395 (1996).

71. Chalfie, M. Green fluorescent protein. *Photochemical & Photobiological Sciences* **62**, 651–656 (1995).

72. Prasher, D. C.; Eckenrode, V. K.; Ward, W. W.; Prendergast, F. G.; Cormier, M. J. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* **111**, 229–233 (1992). [PMID:1347277]

73. Ellington, A. D.; Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990). [PMID:1697402]

74. Grate, D.; Wilson, C. Laser-mediated, site-specific inactivation of RNA transcripts. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 6131–6136 (1999). [PMID:10339553]

75. Kraus, G. A.; Jeon, I.; Nilsen-Hamilton, M.; Awad, A. M.; Banerjee, J.; Parvin, B. Fluorinated analogs of malachite green: Synthesis and toxicity. *Molecules* **13**, 986–994 (2008). [PMID:18463600]

76. Constantin, T. P.; Silva, G. L.; Robertson, K. L.; Hamilton, T. P.; Fague, K.; Waggoner, A. S.; Armitage, B. A. Synthesis of new fluorogenic cyanine dyes and incorporation into RNA fluoromodules. *Organic Letters* **10**, 1561–1564 (2008). [PMID:18338898]

77. Sando, S.; Narita, A.; Aoyama, Y. Light-up Hoechst-DNA aptamer pair: Generation of an aptamer-selective fluorophore from a conventional DNA-staining dye. *ChemBioChem* **8**, 1795–1803 (2007). [PMID:17806095]

78. Sando, S.; Narita, A.; Hayami, M.; Aoyama, Y. Transcription monitoring using fused RNA with a dye-binding light-up aptamer as a tag: A blue fluorescent RNA. *Chemical Communications* No. 33 3858–3860 (2008). [PMID:18726014]

79. Truong, L.; Ferré-D'Amaré, A. R. From fluorescent proteins to fluorogenic RNAs: Tools

for imaging cellular macromolecules. *Protein Science* **28**, 1374–1386 (2019). [PMID:31017335]

80. Paige, J. . S.; Wu, K. Y.; Jaffrey, S. R. RNA Mimics of Green Fluorescent Protein. *Science* **333**, 642–646 (2011). [PMID:21798953]

81. Ts--, S.; Shaaban, K. A.; Singh, S.; Elshahawi, S. I.; Wang, X.; Larissa, V.; Sunkara, M.; Copley, G. C.; Hower, J. C.; Andrew, J.; Kharel, M. K.; Thorson, J. S. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat RNA. **67**, 223–230 (2014).

82. Autour, A.; Westhof, E.; Ryckelynck, M. iSpinach: A fluorogenic RNA aptamer optimized for in vitro applications. *Nucleic Acids Research* **44**, 2491–2500 (2016). [PMID:26932363]

83. Dolgosheina, E. V.; Jeng, S. C. Y.; Panchapakesan, S. S. S.; Cojocaru, R.; Chen, P. S. K.; Wilson, P. D.; Hawkins, N.; Wiggins, P. A.; Unrau, P. J. RNA Mango aptamer-fluorophore: A bright, high-affinity complex for RNA labeling and tracking. *ACS Chemical Biology* **9**, 2412–2420 (2014). [PMID:25101481]

84. Autour, A.; Jeng, S. C. Y.; Cawte, A. D.; Abdolahzadeh, A.; Galli, A.; Panchapakesan, S. S. S.; Rueda, D.; Ryckelynck, M.; Unrau, P. J. Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells. *Nature Communications* **9**, (2018). [PMID:29440634]

85. Kong, K. Y. S.; Jeng, S. C. Y.; Rayyan, B.; Unrau, P. J. RNA Peach and Mango: Orthogonal two-color fluorogenic aptamers distinguish nearly identical ligands. *Rna* **27**, 604–615 (2021). [PMID:33674421]

86. Tan, X.; Constantin, T. P.; Sloane, K. L.; Waggoner, A. S.; Bruchez, M. P.; Armitage, B.

A. Fluoromodules Consisting of a Promiscuous RNA Aptamer and Red or Blue Fluorogenic Cyanine Dyes: Selection, Characterization, and Bioimaging. *Journal of the American Chemical Society* **139**, 9001–9009 (2017). [PMID:28644615]

87. Steinmetzger, C.; Palanisamy, N.; Gore, K. R.; Höbartner, C. A Multicolor Large Stokes Shift Fluorogen-Activating RNA Aptamer with Cationic Chromophores. *Chemistry - A European Journal* **25**, 1931–1935 (2019). [PMID:30485561]

88. Chen, X.; Zhang, D.; Su, N.; Bao, B.; Xie, X.; Zuo, F.; Yang, L.; Wang, H.; Jiang, L.; Lin, Q.; Fang, M.; Li, N.; Hua, X.; Chen, Z.; Bao, C.; Xu, J.; Du, W.; Zhang, L.; Zhao, Y.; Zhu, L.; Loscalzo, J.; Yang, Y. *Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs*; 2019; Vol. 37

89. Baugh, C.; Grate, D.; Wilson, C. 2.8 Å Crystal structure of the malachite green aptamer. *Journal of Molecular Biology* **301**, 117–128 (2000). [PMID:10926496]

90. Warner, K. D.; Chen, M. C.; Song, W.; Strack, R. L.; Thorn, A.; Jaffrey, S. R.; Ferré-D'Amaré, A. R. Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nature Structural & Molecular Biology* **21**, 658–663 (2014). [PMID:25026079]

91. Trachman III, R. J.; Demeshkina, N. A.; Lau, M. W. L.; Panchapakesan, S. S. S.; Unrau, P. J.; Ferré-D'Amaré, A. R. Structural basis for high-affinity fluorophore binding and activation by RNA Mango. *Physiology & behavior* **13**, 139–148 (2017).

92. Mieczkowski, M.; Steinmetzger, C.; Bessi, I.; Lenz, A. K.; Schmiedel, A.; Holzapfel, M.; Lambert, C.; Pena, V.; Höbartner, C. Large Stokes shift fluorescence activation in an RNA aptamer by intermolecular proton transfer to guanine. *Nature Communications* **12**, 1–11

(2021). [PMID:34112799]

93. Song, W.; Filonov, G. S.; Kim, H.; Hirsch, M.; Li, X.; Moon, J. D.; Jaffrey, S. R. Imaging RNA polymerase III transcription using a photostable RNA-fluorophore complex. *Nature Chemical Biology* **13**, 1187–1194 (2017). [PMID:28945233]

94. Millevoi, S.; Moine, H.; Vagner, S. G-quadruplexes in RNA biology. *Wiley Interdisciplinary Reviews: RNA* **3**, 495–507 (2012).

95. Warner, K. D.; Sjekloa, L.; Song, W.; Filonov, G. S.; Jaffrey, S. R.; Ferré-D'Amaré, A. R. A homodimer interface without base pairs in an RNA mimic of red fluorescent protein. *Nature Chemical Biology* **13**, 1195–1201 (2017). [PMID:28945234]

96. Dey, S. K.; Filonov, G. S.; Olarerin-George, A. O.; Jackson, B. T.; Finley, L. W. S.; Jaffrey, S. R. Repurposing an adenine riboswitch into a fluorogenic imaging and sensing tag. *Nature Chemical Biology* **18**, (2021). [PMID:34937909]

97. Trachman, R. J.; Abdolahzadeh, A.; Andreoni, A.; Cojocaru, R.; Knutson, J. R.; Ryckelynck, M.; Unrau, P. J.; Ferré-D'Amaré, A. R. Crystal Structures of the Mango-II RNA Aptamer Reveal Heterogeneous Fluorophore Binding and Guide Engineering of Variants with Improved Selectivity and Brightness. *Biochemistry* **57**, 3544–3548 (2018). [PMID:29768001]

98. Trachman, R. J. I.; Autour, A.; Jeng, S. C. Y.; Abdolahzadeh, A.; Andreoni, A.; Cojocaru, R.; Garipov, R.; Dolgosheina, E. V.; Knutson, J. R.; Ryckelynck, M.; Unrau, P. J.; Ferré-D'Amaré, A. R. Structure and functional reselection of the Mango-III fluorogenic RNA aptamer. *Nature Chemical Biology* **(Accepted)**, (2019).

99. Trachman, R. J.; Cojocaru, R.; Wu, D.; Piszczek, G.; Ryckelynck, M.; Unrau, P. J.; Ferré-D'Amaré, A. R. Structure-Guided Engineering of the Homodimeric Mango-IV Fluorescence Turn-on Aptamer Yields an RNA FRET Pair. *Structure* **28**, 776-785.e3 (2020). [PMID:32386573]

100. Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R. Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *Journal of the American Chemical Society* **136**, 16299–16308 (2014). [PMID:25337688]

101. Huang, K.; Chen, X.; Li, C.; Song, Q.; Li, H.; Zhu, L.; Yang, Y.; Ren, A. Structure-based investigation of fluorogenic Pepper aptamer. *Nature Chemical Biology* **17**, 1289–1295 (2021). [PMID:34725509]

Chapter 2: Crystal structures of Pepper aptamer in complex with HCB530 and HBC599

2.1 Introduction

Pepper is one of the latest additions to the existing suite of RNAs that can bind small molecules to form fluorescent complexes (Figure 2.1). Developed by Chen *et al.* [1] via SELEX, Pepper was engineered to find a new aptamer that had fewer of the weaknesses of current GFP-derivative binding aptamers. Before Pepper, there was a notable lack of aptamers that had the combination of tight affinity (sub 100 nM), excellent brightness in vivo, and a range of excitation and emission wavelengths to choose from to allow for visualization throughout cellular environments [2]. Each of these qualities had been previously engineered into allosteric RNAs, such as Mango with tight affinity to TO1 [3] (3 nM), or the high-brightness of DIR2 [4], but these examples were lacking in the other aforementioned properties [2]. Instead of engineering from the HBI-derivative or TO binding families of aptamers, Pepper was developed by Chen *et al* [1] via SELEX, to bind variants of the HBC fluorophore, chosen as the fluorogenic ligand for its capacity to accommodate electron-withdrawing groups that can promote red-shifted spectral properties (Figure 2.2). With extremely low K_d (3.5 nM for HBC530), and intense brightness both *in vitro* and *in vivo*, Pepper showed great promise as a replacement for previous aptamers.

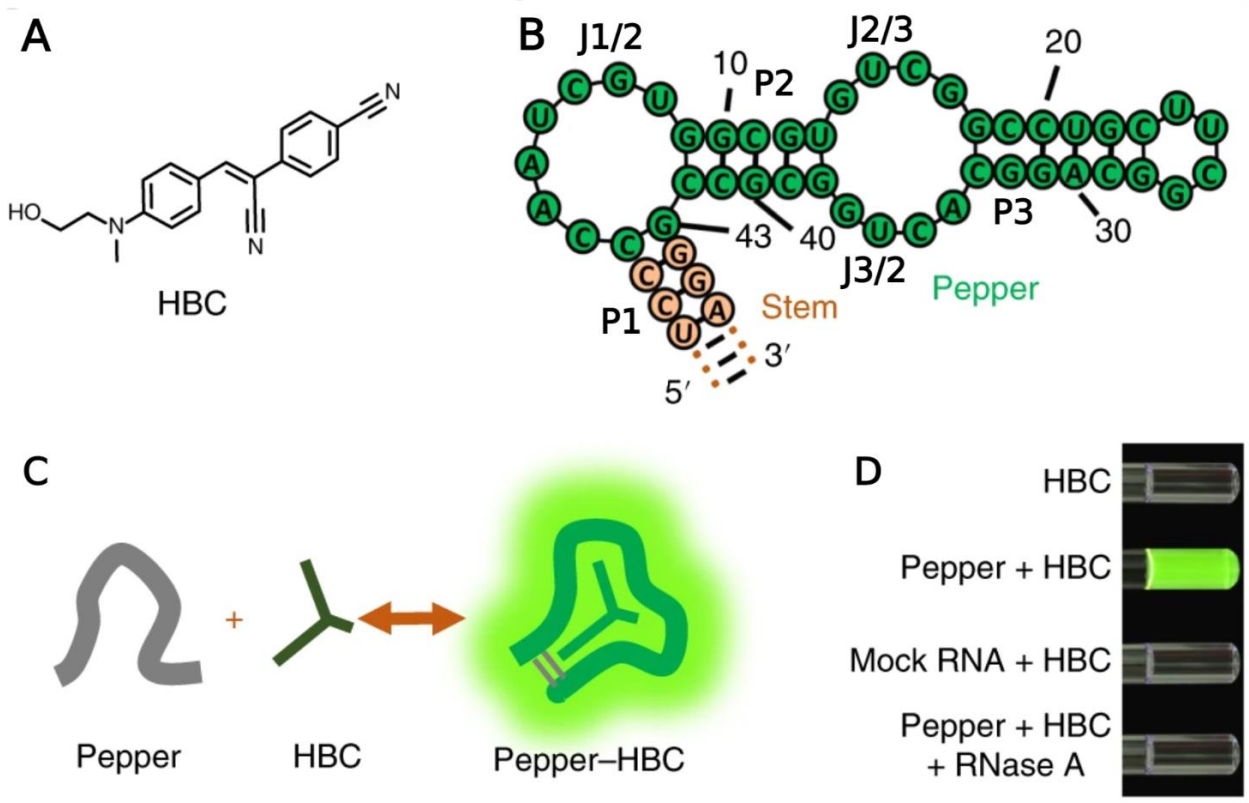


Figure 2.1: The Pepper aptamer was engineered to bind HBC derivatives for visualizing cell dynamics in real time. A) HBC (or HBC530) that is the core compound against which the Pepper RNA was evolved. B) The minimal motif for the Pepper aptamer after initial engineering. The predicted secondary structure contains two bulges and a hairpin. C,D) The Pepper aptamer binds HBC to induce fluorescence of various absorbance and emission wavelengths. Controls show that the Pepper RNA is specifically required to induce fluorescence. Adapted from Chen *et al*.[1]

In addition to better K_d and brightness measurements compared to previous aptamers, Pepper appeared to have a potentially simple structure with no G-quadruplex. The predicted secondary structure of Pepper's minimal motif showed three helical regions (P1-P3) with an asymmetric bulge (J1/2) separating P1 and P2 and a symmetric bulge (J2/3, J3/2) separating P2 and P3 (Figure 2.1). The bulges both lack the guanosine content to assemble a G-quadruplex [5], making this an unlikely feature of Pepper's structure. In addition, the metal-dependency of Pepper fluorescence activation has a requirement for Mg^{2+} ions but not K^+ ions, further indicating the

absence of a K^+-dependent G-quadruplex in Pepper [6]. Mutational analysis of the Pepper construct indicated that the symmetric bulge in the secondary structure contributes an essential function to fluorogenic activity [1]. Additionally, mutations to the base of P2 abolished activity, whereas exchange of the terminal loop of P3 with an MS2-binding module had no effect on fluorescence. While the HBC was expected to bind in the symmetric bulge, how the asymmetric bulge J1/2 contributed to binding was a mystery.

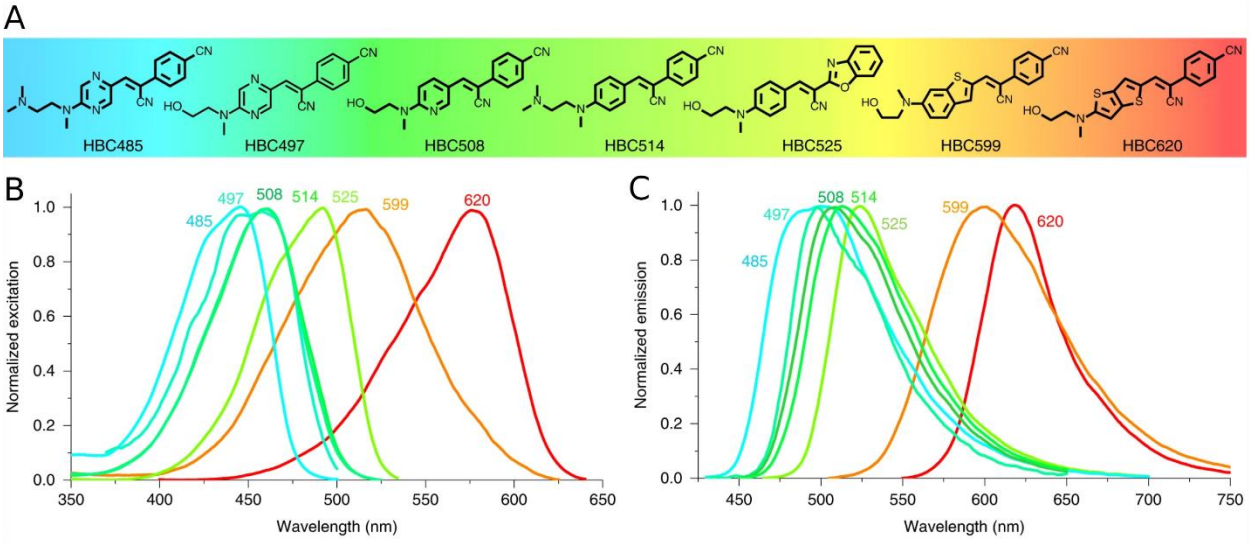


Figure 2.2: HBC derivatives with various excitation and emission wavelengths A) The initial suite of HBC derivatives reported, with nomenclature referring to their respective emission wavelengths. B) Excitation spectra of Pepper bound to the various HBC derivatives. C) Emission spectra of Pepper bound to the various HBC derivatives. Adapted from Chen *et al.*[1]

To help understand this new member of the fluorogenic aptamer family, we deployed chaperone assisted RNA crystallography using Fab BL3-6 to obtain two crystal structures of Pepper in complex with Fab chaperone bound to the ligands HBC530 and HBC599 to 2.3 and 2.7 Å resolution, respectively. These structures showed that Pepper binds its HBC ligands in a way that is similar to previous aptamers, and agreed with the pre-existing data that indicated no presence of G-quadruplexes. Key tertiary interactions are visible in the crystal between J1/2 and

P2, J1/2 and J3/2, and the novel base-quadruple stack formed by J2/3 and J3/2. Pepper has a relatively simple tertiary structure that utilizes a few critical tertiary interactions, and appears to be dependent on divalent ions due to specific binding sites that we assigned as Mg^{2+} in the crystal structures.

2.2 Methods

Pepper RNA synthesis and purification

The dsDNA template (Table 2.1) used to transcribe the RNA was prepared via PCR of a single-stranded DNA template purchased from integrated DNA technologies (IDT DNA). The first nucleotide of the reverse primer was modified with 2'-OMe to reduce transcriptional heterogeneity at the 3' end [7]. RNA was prepared via *in vitro* transcription for 3.5 hours at 37ºC in buffer composed of 40 mM Tris-HCl pH 7.9, 10 mM NaCl, 25 mM $MgCl_2$, 2 mM Spermidine, 10 mM DTT, 30 U/mL Rnase inhibitor, 2.5 U/mL TIPPase, 4 mM of each nucleotide tri-phosphate (NTP), 8 ng/uL of DNA template, and 40 μg/mL of in-house-prepared T7 RNA polymerase. To remove DNA template from the reaction, 10 U/mL RNase-free DNase I (www.promega.com) was added and the reaction incubated for a further 30 mins at 37ºC. After the Phenol/Chloroform pH 4.3 extraction to remove proteins, the RNA was purified via denaturing polyacrylamide gel electrophoresis. The RNA band of interest was quickly visualized via UV shadowing and excised from the gel. Using the mash-and-soak method, the RNA was collected, aliquoted into small fractions, and stored at –80 °C until further use.

Crystallization of Pepper-BL3-6-Fab complex

The RNA (Table 2.1) sample was refolded in a buffer containing 10 mM Tris pH 7.5, 100 mM KCl, 5 mM $MgCl_2$ (supplied as 10×). For refolding, RNA was heated at 90 °C for 1 min in deionized water and then cooled on ice for 2 min before folding buffer and dye was added. This was then followed by incubation at room temperature for 30 min. The refolded RNA was then introduced to 1.1 equivalents of the BL3-6 Fab (expressed as soluble protein in phagemid as an expression vector and, purified by affinity and ion exchange chromatography using protein A, G, and Heparin columns (GE Healthcare), respectively [8]) at room temperature for 30 min and concentrated to 3 mg/mL using 10 kDa cut-off, Amicon Ultra-15 column. The formation of Fab–RNA complex was confirmed by native poly-acrylamide gel electrophoresis. To prevent excess nucleation events, RNA was then passed through the 0.2 μm cut-off, Millipore centrifugal filter units. A Mosquito liquid handling robot (STP Labtech) was used to set up high-throughput hanging drop vapor diffusion crystallization screens at room temperature using commercially available screening kits from Hampton Research, Sigma, and Jena Bioscience. After additive screening with diffracting conditions, the best-diffracting crystals of the Pepper-BL3-6-Fab-HBC530 complex were obtained in a condition from the Natrix screen: 0.02 M Magnesium sulfate hydrate, 0.002 M Cobalt(II) chloride hexahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.0, 0.0005 M Spermine, 4% v/v 2,5-Hexanediol. For the complex with the HBC599 fluorophore, the optimal condition was from the Index screen: 0.2 M Ammonium Acetate, 0.1 M BIS-TRIS pH 6.5, and 45% v/v (+/-)-2-Methyl-2,4-pentanediol. Crystals appeared and grew to full size within 2 days in 100 nl + 100 nl hanging drops. The crystals were looped without cryo-protectant before being flash-frozen in liquid nitrogen. Diffraction datasets were collected at 80 K at the Advanced Photon Source NE-CAT section beamlines 24-ID-C and 24-ID-E at Argonne National Lab.

Fluorescence assay experiments

Florescence measurements were conducted using a Neo2 Synergy BioTek plate reader. RNA heated to 95°C for 1 minute in water, cooled on ice for 2 minutes, before buffer (1x; HEPES 40 mM (pH 7.5), 100 mM KCl, 5 mM $MgCl_2$) and dye (1 μM) were added were incubated for 30 minutes at 25 °C before measurements were taken (Ex/Em 485/530 for HBC530, 515/599 for HBC599, gain 100, ±20) at 25°C, and the reported results were the average of three measurements. Measurements under variable pH were made with separately prepared HEPES buffers (1x; HEPES 40 mM (pH variable), 100 mM KCl, and 5 mM $MgCl_2$.

Synthesis of HBC derivatives

HBC530 and **HBC599** were prepared by Nan-Sheng Li according to the reported procedure as shown in Schemes 1-2.[1] **HBC530** was prepared according to the procedure reported as shown in Figure 2.3. To a stirring solution of *N*-methyl-*N*-(2-hydroxyethyl)-4-aminobenzaldehyde (180 mg, 1.0 mmol) and 4-cyano-benzeneacetonitrile (310 mg, 2.2 mmol) in dry methanol (20 mL), 2 drops of pyrrolidine were added. After stirring at rt overnight, the solvent was removed, the residue was isolated by silica gel chromatography, eluting with 50% ethyl acetate in hexane to give **HBC530** as an orange solid: 233 mg, 77% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.89 (d, *J* = 8.8 Hz, 2H), 7.73 (d, *J* = 8.4 Hz, 2H), 7.68 (d, *J* = 8.8 Hz, 2H), 7.49 (s, 1H), 6.80 (d, *J* = 9.2 Hz, 2H), 3.89 (m, 2H), 3.63 (t, *J* = 5.6 Hz, 2H), 3.13 (s, 3H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 151.83, 145.67, 139.86, 133.15, 132.20, 125.65, 120.20, 119.15, 111.69, 109.80, 99.91, 58.41, 53.95, 38.88.

Scheme 1

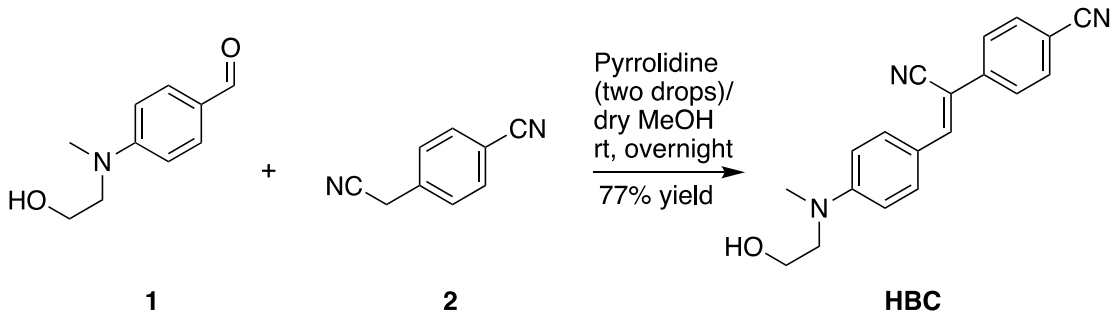


Figure 2.3: Scheme 1: Synthesis of HBC530 (HBC), by Nan-Sheng Li

HBC599 was synthesized according to the procedure reported as shown in Figure 2.4. We obtained the correct product **5** as a minor product. The details for the step from **4** to **5** and the NMRs of the product and intermediates are shown as follows.

Compound **4**: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.67 (dd, *J* = 8.9, 1.2 Hz, 1H), 7.22 – 7.17 (m, 2H), 7.14 (dd, *J* = 5.5, 1.0 Hz, 1H), 6.93 (ddd, *J* = 8.9, 2.6, 1.1 Hz, 1H), 4.31 (td, *J* = 6.1, 0.9 Hz, 2H), 3.68 (t, *J* = 6.1 Hz, 2H), 3.06 (s, 3H), 2.03 (d, *J* = 1.0 Hz, 3H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 171.03, 146.89, 142.04, 130.95, 123.84, 123.30, 121.76, 112.13, 104.46, 61.59, 51.80, 39.05, 20.90.

Compounds **5** and **5a**: Phosphorous oxychloride (0.11 mL, 1.16 mmol) was added dropwise to a stirred 0 ^{o}C solution of compound **4** (240 g, 0.96 mmol) in dry CH_2Cl_2(10 mL) and dry DMF (1 mL) under the protection of Ar. The resulting mixture was warmed to room temperature and stirred for an additional 5 hr. The mixture was quenched with a saturated solution of sodium carbonate and extracted with CH_2Cl_2. The organic layer was dried over Na_2SO_4, then the solvent was removed under reduced pressure to give the crude product which was purified by silica gel column chromatography, eluting with 15-25% ethyl acetate in hexane to yield compound **5a** (0.171 g,

64%) (top spot on TLC) and compound **5** (19 mg, 7.1%) (low spot on TLC). Compound **5** is more visible on TLC with long wave UV than with short wave UV.

Compound **5**: ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.88 (s, 1H), 7.78 (s, 1H), 7.67 (d, *J* = 9.0 Hz, 1H), 7.19 (s, 1H), 6.98 (d, *J* = 2.4 Hz, 1H), 6.87 (dd, *J* = 9.0, 2.4 Hz, 1H), 4.23 (t, *J* = 6.0 Hz, 2H), 3.64 (t, *J* = 6.0 Hz, 2H), 3.03 (s, 3H), 1.94 (s, 3H).

Compound **5a**: ^{1}H NMR (400 MHz, $CDCl_3$) δ 10.50 (s, 1H), 7.83 (d, *J* = 8.6 Hz, 1H), 7.35 (d, *J* = 5.5 Hz, 1H), 7.17 (d, *J* = 5.5 Hz, 1H), 7.14 (d, *J* = 8.6 Hz, 1H), 4.15 (t, *J* = 5.7 Hz, 2H), 3.37 (t, *J* = 5.7 Hz, 2H), 2.90 (s, 3H), 1.85 (s, 3H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 190.25, 170.77, 155.36, 139.06, 136.28, 129.89, 128.69, 123.05, 122.45, 117.68, 61.44, 57.58, 43.70, 20.79.

Compound **6**: ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.82 (s, 1H), 7.73 (d, *J* = 1.4 Hz, 1H), 7.63 (dd, *J* = 9.0, 1.4 Hz, 1H), 6.98 (d, *J* = 2.2 Hz, 1H), 6.88 (d, *J* = 9.0 Hz, 1H), 3.87 – 3.74 (m, 2H), 3.60 – 3.49 (m, 2H), 3.02 (s, 3H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 183.91, 150.25, 146.30, 138.51, 135.45, 129.38, 126.97, 113.13, 103.40, 60.22, 54.99, 39.24.

HBC599: ^{1}H NMR (400 MHz, DMSO) δ 8.45 (s, 1H), 7.98 – 7.91 (m, 2H), 7.88 (d, *J* = 8.6 Hz, 3H), 7.76 (d, *J* = 9.2 Hz, 1H), 7.23 (d, *J* = 2.2 Hz, 1H), 6.99 (dd, *J* = 9.0, 2.4 Hz, 1H), 4.82 (t, *J* = 5.4 Hz, 1H), 3.06 (s, 3H). 129.89, 128.69, 123.05, 122.45, 117.68, 61.44, 57.58, 43.70, 20.79.

Scheme 2

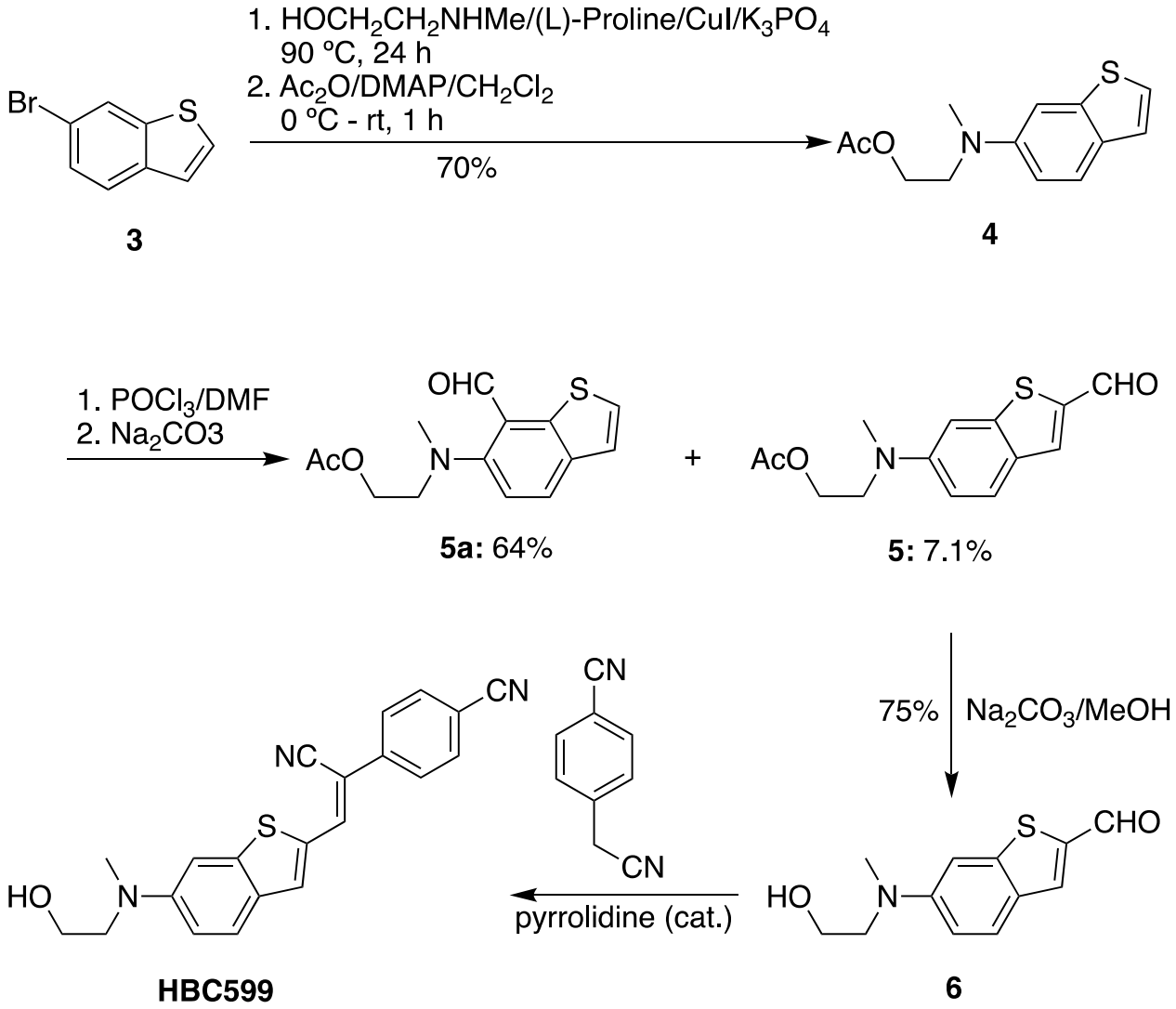


Figure 2.4: Scheme 2: Synthesis of HBC599, by Nan-Sheng Li

Table 2.1: Pepper-BL3-6 and Pepper WT sequences

ID	Type	Sequence 5'-3'	Description
Pepper-WT	DNA	GCG TAA TAC GAC TCA CTA TA GGTA CCT ACC AAT CGT AGC GTG TCG ACC TGC TTC GGC AGG CAC TGG CGC TGT AGG TAC	dsDNA template with T7 promoter. Purchased from IDT (see methods)
Pepper-WT	RNA	GGU ACC UAC CAA UCG UAG CGU GUC GAC CUG CUU CGG CAG GCA CUG GCG CUG UAG GUA C	RNA sequence used for Pepper-WT (T1 construct from original authors) generated using T7 RNA polymerase (see methods).
Pepper-BL3-6	DNA	GCG TAA TAC GAC TCA CTA TA GGTA CCT ACC AAT CGT AGC GTG TCG ACC AGC TGC **GAAACAC** GCA GCT GG CAC TGG CGC TGT AGG TAC	dsDNA template with T7 promoter and sequence encodes DNA for **BL3-6 binding loop**. Purchased from IDT (see methods)
Pepper-BL3-6	RNA	GGU ACC UAC CAA UCG UAG CGU GUC GAC CAG CUG C**GA AAC AC**G CAG CUG GCA CUG GCG CUG UAG GUA C	RNA sequence generated using T7 RNA polymerase (see methods). Contains **BL3-6 binding loop**. Construct used for crystallization

Figures, Schemes and Crystallography

Figures were created using Biorender™. Images of 3D models were generated using pymol, and organic structures were drawn using Chemdraw. All crystallography was performed using Phenix and WinCoot software packages, with feedback from the pdb during deposition.

2.3 Results

Construct and Crystallization

The Pepper crystallization construct design started from the minimal T1 motif [1] (Figure 2.5) and modifications were made that would assist crystallization without significantly changing Pepper's fluorogenic activity. We replaced the original stem-loop of P3 with a cognate pentaloop hairpin (GAAACAC) to make a binding site for the Fab BL3-6 crystallization chaperone [8]. The

other main changes included extension of P1, addition of two G residues to the 5' end to assist transcription, and extension of P3 to increase the distance between the Fab and the aptamer core. Fab binding to this construct was confirmed via electrophoretic-mobility-shift assays (EMSA; Figure 2.6). To ensure the resultant crystal structure would have relevance for the mechanism of fluorescence activation, we confirmed via fluorescence spectroscopy that the construct activates the fluorescence of HBC ligands in the presence of Fab (Figure 2.6).

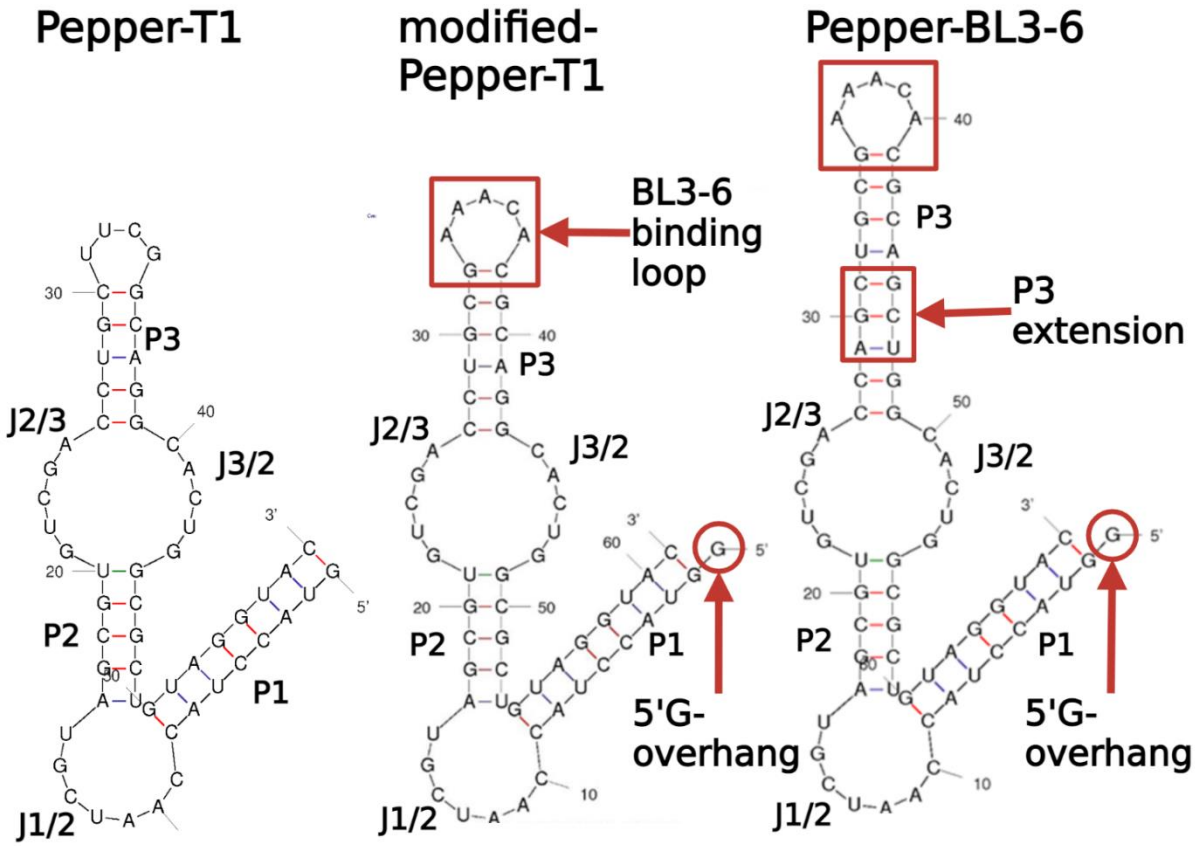


Figure 2.5: Generation of the Pepper-BL3-6 crystallization construct from original T1 motif[1]. RNA secondary structure predictions and images were generated using unafold[9]. The original T1 motif was designed by Chen et al[1], and used as a starting point for designing a crystallization construct. The first iteration of constructs involved grafting the BL3-6 binding loop into the structure by replacing the UUCG hairpin, and adding a 5'G overhang to assist with transcription. The final construct that crystallized had an extension to P3.

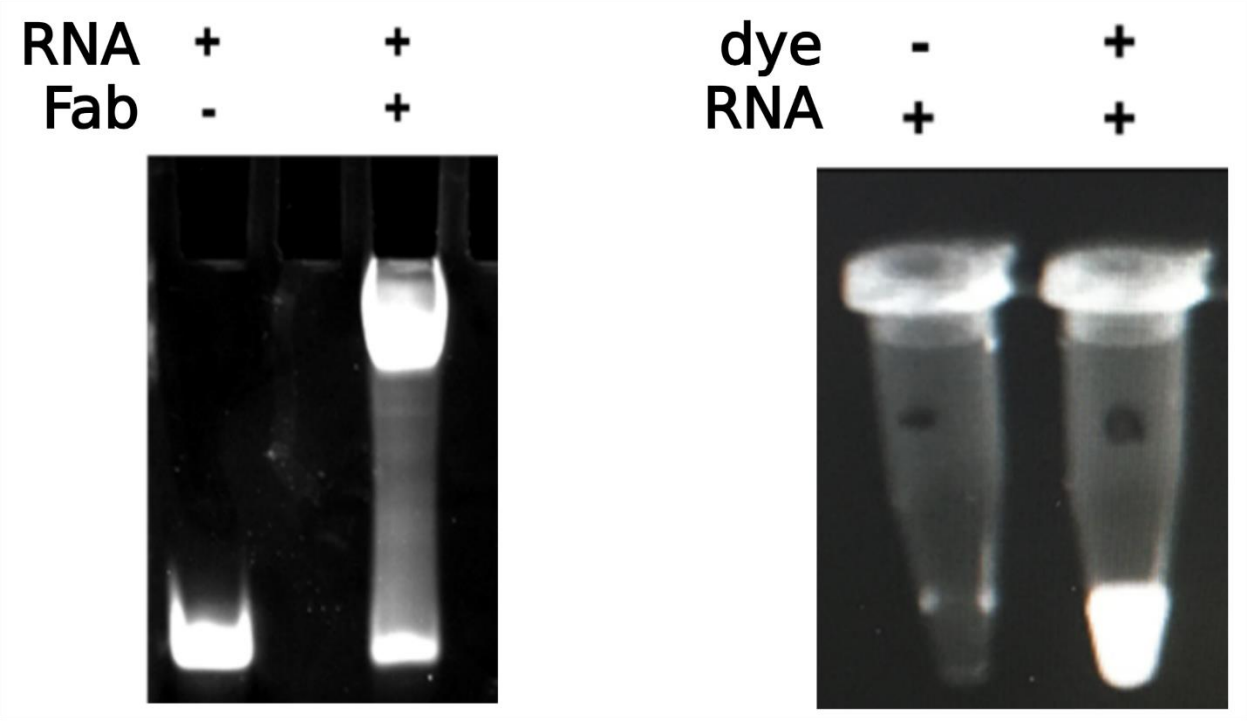


Figure 2.6: Evidence for Fab-RNA complex and Pepper-BL3-6 binding to HBC ligands. Left) 10% nPAGE Electromobility-shift-assay (EMSA) of Pepper-BL3-6 RNA with and without Fab-BL3-6. Upon binding the Fab, the RNA (visualized with Ethidium bromide staining) becomes much less mobile, and appears higher up on the gel. Right) To confirm that the RNA still bound the dye, a RNA was exposed to UV-light with and without the dye. The RNA was folded identically in each case.

Overall, four constructs were designed that had variable P1 and P3 lengths; they were able to bind Fab BL3-6 and activate HBC to produce a strong fluorescent signal, and one construct, referred to hereafter as Pepper-BL3-6 (Figure 2.5), formed diffracting crystals (Figure 2.7) when in complex with ligand (Table 2.2). The Pepper-BL3-6-HBC530 crystals were solved to 2.3 Å resolution via molecular replacement with Fab BL3-6 variable and constant regions (4kzd)[6] followed by rounds of building and refinement of the RNA using the Phenix and Wincoot software packages. Due to the high quality of the data, the ligand density became unambiguous as the RNA was built. Ligand orientations were decided by best fit with the electron density difference maps. Metal ions were assigned by density, coordination, occupancy, and temperature factor relative to

adjacent RNA. The Pepper-BL3-6-HBC599 crystals were solved to 2.7 Å resolution via molecular replacement to place the RNA and Fab, then inserting the HBC599 ligand into the revealed density (Table 2.2).

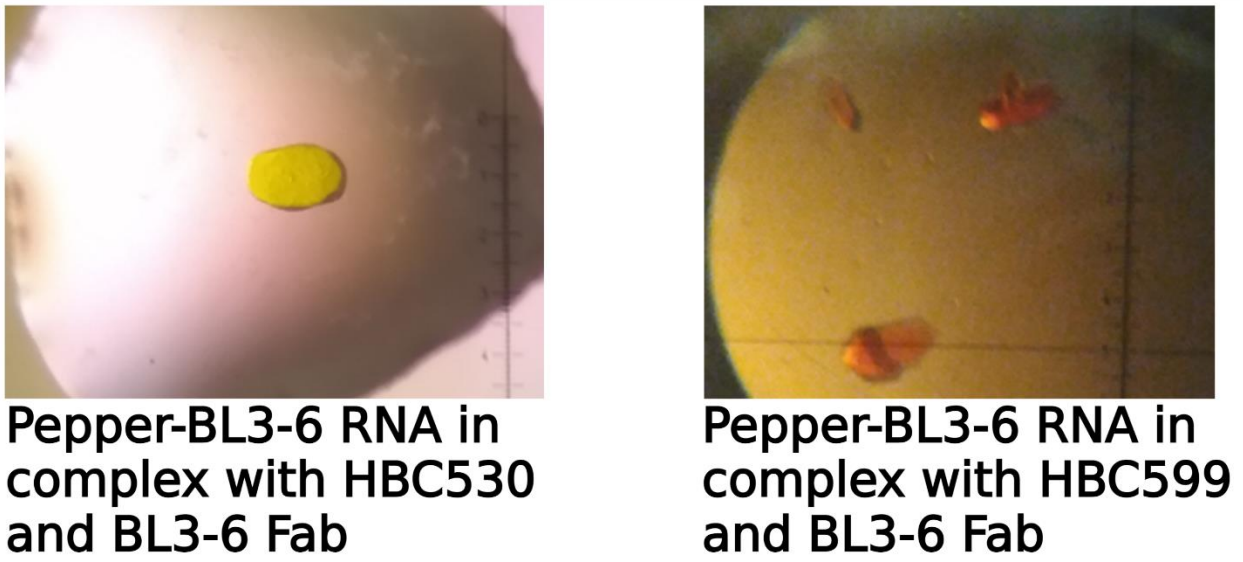


Figure 2.7: Photographs of crystals grown from Pepper-BL3-6 RNA in complex with HBC ligands and Fab BL3-6. The colour of the complexes are clearly visible with the naked eye. These crystals are taken from similar well-conditions that later diffracted, but are not necessarily the crystals that produced the structure datasets.

Table 2.2: Refinement statistics of the Pepper-aptamer structures bound to HBC530 and HBC599 respectively. Generated using Phenix.

	Pepper-BL3-6-HBC530	**Pepper-BL3-6-HBC599**
Wavelength (Å)	0.9792	0.9792
Resolution range (Å)	58.9 - 2.24 (2.32 - 2.24)	58.67 - 2.66 (2.755 - 2.66)
Space group	P 21 21 21	P 21 21 21
Unit cell	61.609 96.992 148.255 90 90 90	61.247 96.071 148.165 90 90 90
Total reflections	297064 (29908)	83083 (5232)
Unique reflections	43547 (4280)	25100 (2083)
Multiplicity	6.8 (7.0)	3.3 (2.5)
Completeness (%)	99.86 (99.58)	96.79 (83.02)
Mean I/sigma(I)	18.17 (1.22)	15.61 (1.27)
Wilson B-factor	62.32	84.72
R-merge	0.06286 (1.283)	0.04627 (0.5635)
R-meas	0.06827 (1.385)	0.05491 (0.6941)
R-pim	0.02627 (0.5184)	0.02909 (0.3967)
CC1/2	0.998 (0.56)	0.997 (0.723)
CC*	0.999 (0.847)	0.999 (0.916)
Reflections used in refinement	43498 (4281)	24995 (2083)
Reflections used for R-free	1997 (197)	1996 (166)
R-work	0.2198 (0.3446)	0.2182 (0.3895)
R-free	0.2634 (0.4234)	0.2548 (0.4200)
CC(work)	0.869 (0.716)	0.906 (0.736)
CC(free)	0.903 (0.653)	0.941 (0.551)
RMS(bonds)	0.002	0.005
RMS(angles)	0.58	0.83
Ramachandran favored (%)	96.77	93.5
Ramachandran allowed (%)	2.77	5.34
Ramachandran outliers (%)	0.46	1.16
Rotamer outliers (%)	0.8	3.49
Clashscore	3.79	7.37
Average B-factor	70.29	81.8
macromolecules	67.39	83.05
ligands	87.28	97.19
solvent	76.45	78.57

Overall Structure

The Pepper-BL3-6 structures with bound ligand HBC derivatives (Figure 2.8.) mostly conform to secondary structure predictions from programs such as mfold[9], containing three A-form helical stems and two bulges. P1 and P2 stack directly with each other; a joining region (J1/2) connects them on the 5'-side, but they share a continuous 3' strand. The symmetric bulge between P2 and P3 harbours the ligand binding pocket and causes P3 to tilt into the ligand pocket and sit at a roughly 30^{o} angle from the orientation of P1 and P2. Both structures contain site-bound Mg^{2+} cations, one located in J1/2 distal to the ligand binding site referred to herein as M_A and one located in J3/2 proximal to the ligand binding site referred to herein as M_B. The HBC530 structure contains a third site-bound Mg^{2+} also distal to the ligand binding site located in J2/3, referred to herein as M_C (Figure 2.8).

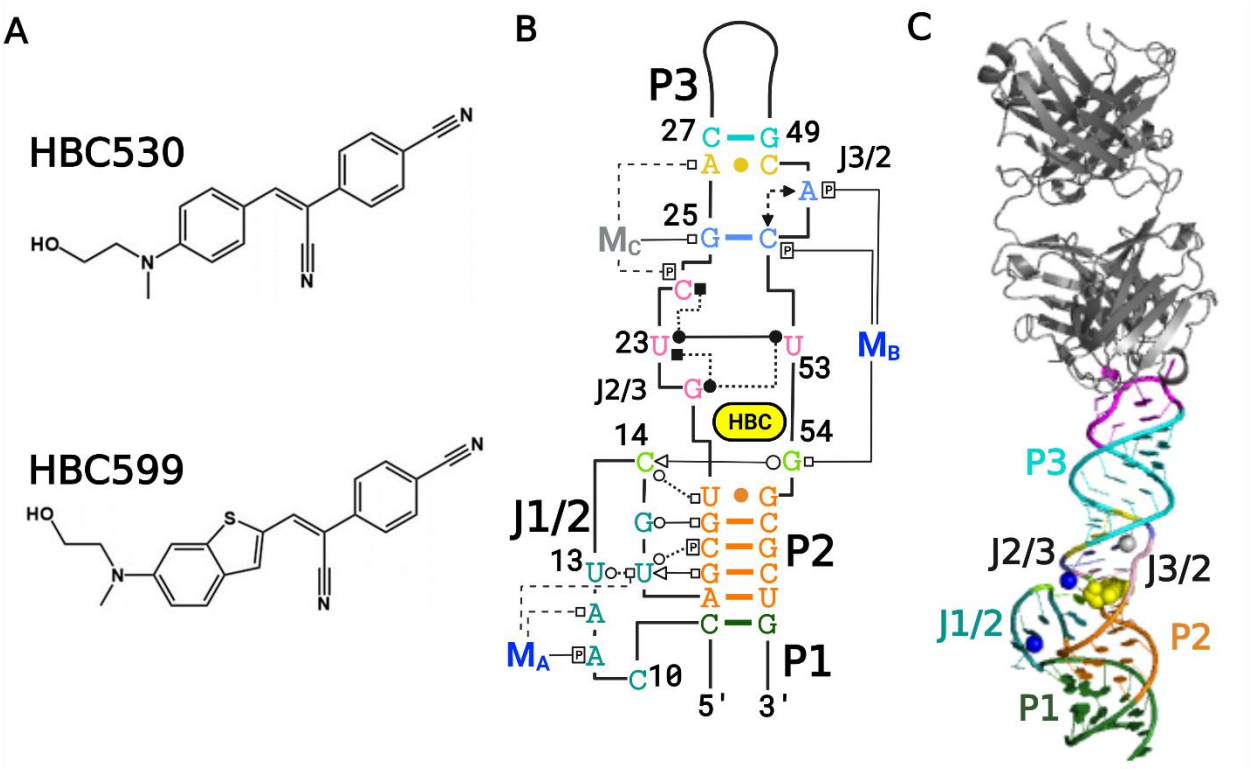


Figure 2.8: Overall structure of the Pepper-BL3-6 aptamer. A) Chemical structures of the HBC ligands, HBC530 and HBC599 activated by Pepper RNA. B) Secondary structure of the Pepper aptamer bound to HBC530, with color coded regions, and the three Mg^{2+} cations that are visible in both structures C) Fab BL3-6 bound the Pepper with the HBC530 ligand.

J1/2 contains seven nucleotides, C10 to U16; this loop, which would be predicted to be structurally flexible based on MFE calculations (Figure 2.9), adopts a well-defined structure in the crystal, with J1/2 calculated to have average temperature-factors similar to the paired regions of the aptamer after structure solving (Figure 2.10). The J1/2 bulge adopts a conformation that is similar to the letter "Q" in cursive writing, with the backbone forming a tight turn at C14 and U16, with A17 re-joining the P2 helical stack above C9. This elaborate fold allows J1/2 the bulge to reach the ligand binding site and cooperatively interact with J3/2 to form essential interactions for ligand binding. In the crystal lattice, the first nucleotide of J1/2 C10 forms an intermolecular crystal contact with the 5'-most nucleotide, G1 – an unintentional but beneficial interaction that assists crystallization (Figure 2.11), though is unlikely to occur in solution. J1/2 is organized by M_A that

is present in both structures, which forms an inner sphere interaction with the 5'PO of A11, and outer sphere interactions with the 5'PO of C10 and the Hoogsteen edges of A11 and A12 (Figure 2.8). At the apex of J1/2, C14 forms a tertiary interaction with G54 adjacent to the ligand binding pocket, and its amino group (C14-N4) lies within range to form a tertiary interaction with U21's keto group (U21-O4) (Figure 2.12). In addition, J1/2 forms tertiary interactions with the minor groove of P2 through G15 and U16 without structurally perturbing the P2 A-form helix (Figure 2.12).

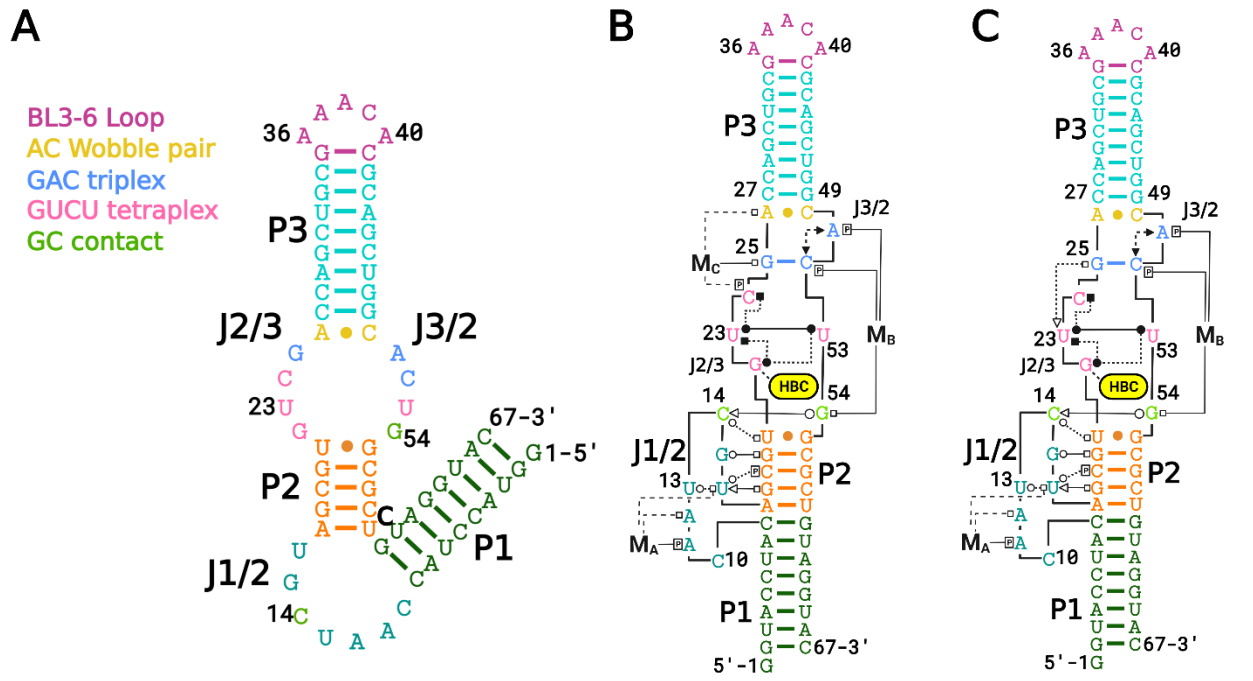


Figure 2.9: Secondary and tertiary structures of the Pepper-BL3-6-HBC complexes. A) Pepper-BL3-6 RNA secondary structure prediction forms P1, P2 and P3 helices, but the junctions are modelled as ssRNA. B) The full tertiary structure of Pepper-BL3-6 bound to HBC530 with the three Mg^{2+} ions assigned. C) The full tertiary structure of Pepper-BL3-6 bound to HBC599 with the three Mg^{2+} ions assigned.

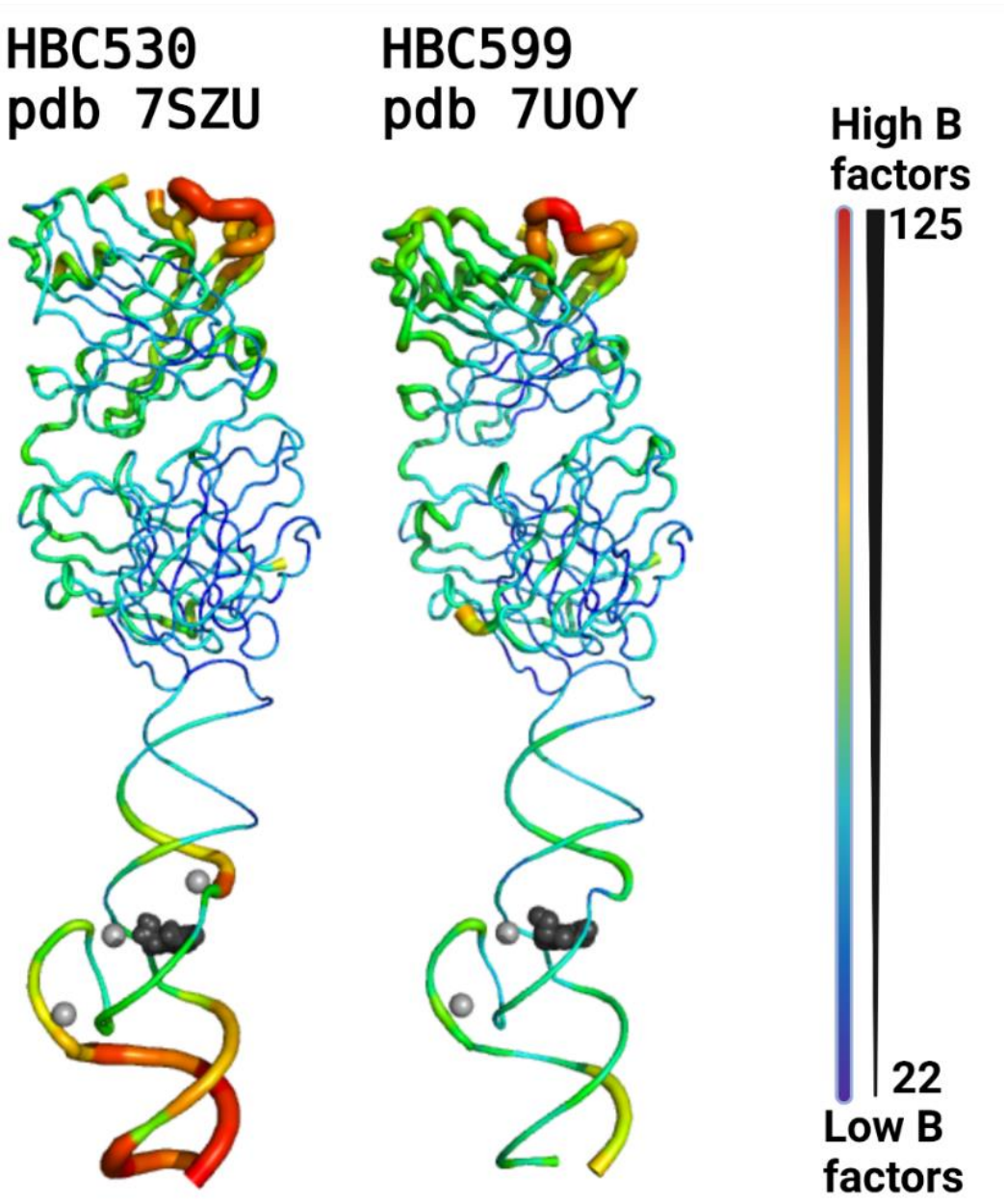


Figure 2.10: Pymol-generated temperature putty cartoons of the Pepper-BL3-6 structures. Thickness and color both correspond to B-factors. Assigned Mg^{2+} and Ligands shown in light and dark gray respectively.

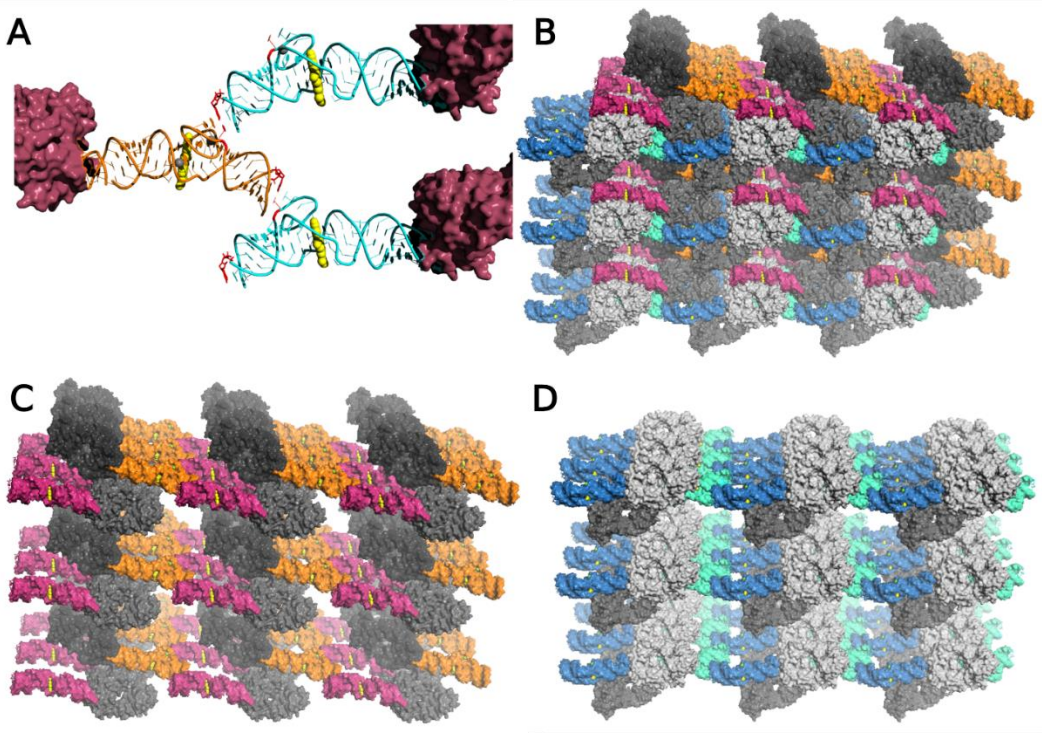


Figure 2.11: Crystal packing of Pepper-BL3-6 structures. A) The intermolecular base pair (G1, C10 shown in red) facilitates the formation of "sheets" of Pepper-BL3-6-Fab complexes that can stack together (B, C, D). The Fab relieves the necessity of RNA-RNA contacts in the structure, and Fab-RNA contacts dominate the lattice formation. Per RNA-Fab complex with 19021 $Å^2$ of available surface area for crystal lattice formation, RNA-RNA interactions accounted for 269 $Å^2$ (1.4 %), Fab-Fab interactions accounted for 36 $Å^2$ (0.1%), and Fab-RNA contacts accounted for 1020 $Å^2$ (5.4%), with the rest contacting solvent (93.1%)

J1/2 leads into P2, a helix composed of four canonical Watson-Crick-Franklin (WCF) base pairs followed by a U21•G55 wobble pair. This pair is the base of the asymmetric bulge that forms the ligand binding site. On the 5' side of this J2/3 bulge starting with G22, the backbone of the next two nucleotides U23 and C24 runs perpendicular to the helical stack to allow G22, U23 and C24 to all lie in the same plane and form a quadruple-base interaction that includes U53 on the 3'-side of the bulge (Figure 2.12). In this quadruple layer, G22, U23, and U53 interact with each other, but the position of the fourth base, C24, differs slightly between the two structures. In the HBC599 complex structure, the exocyclic amine of C24 resides within hydrogen bonding distance of U23's O2 keto group, but the two groups are 4Å away in the Pepper-HBC530 complex (Figure

2.13)). The Mg^{2+} (M_C) present in the HBC530 structure but not in the HBC599 structure may contribute to this difference.

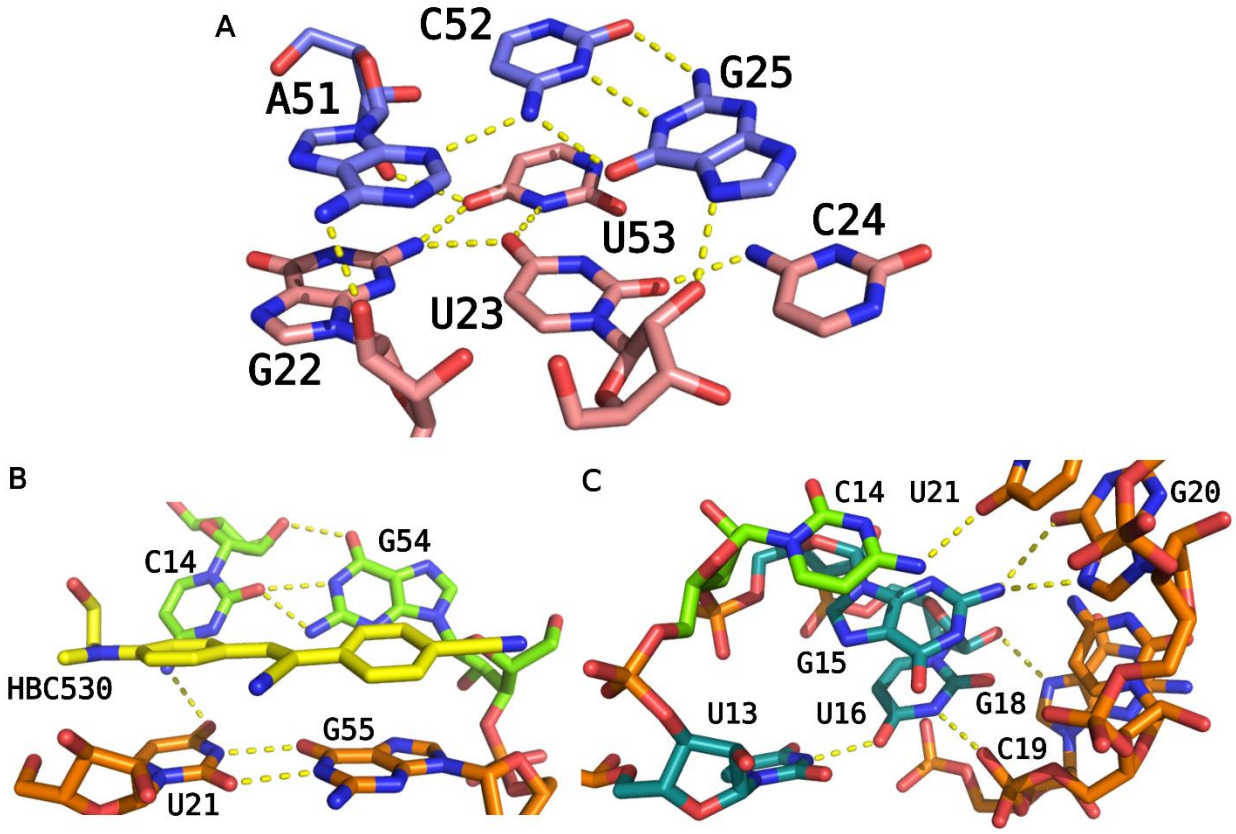


Figure 2.12: Tertiary structures in the Pepper-BL3-6 structure. A) Above the binding site, a base quadruple and base triple stacking arrangement is formed. In HBC599, G25-N7 forms a hydrogen bond to the 2'OH of the U53 ribose ring. B) The ligand stacks on the U21G55 wobble pair, and is adjacent to an out-of-plane tertiary interaction between C14 and G54. C) J1/2 forms tertiary interactions with the P2 helix minor groove.

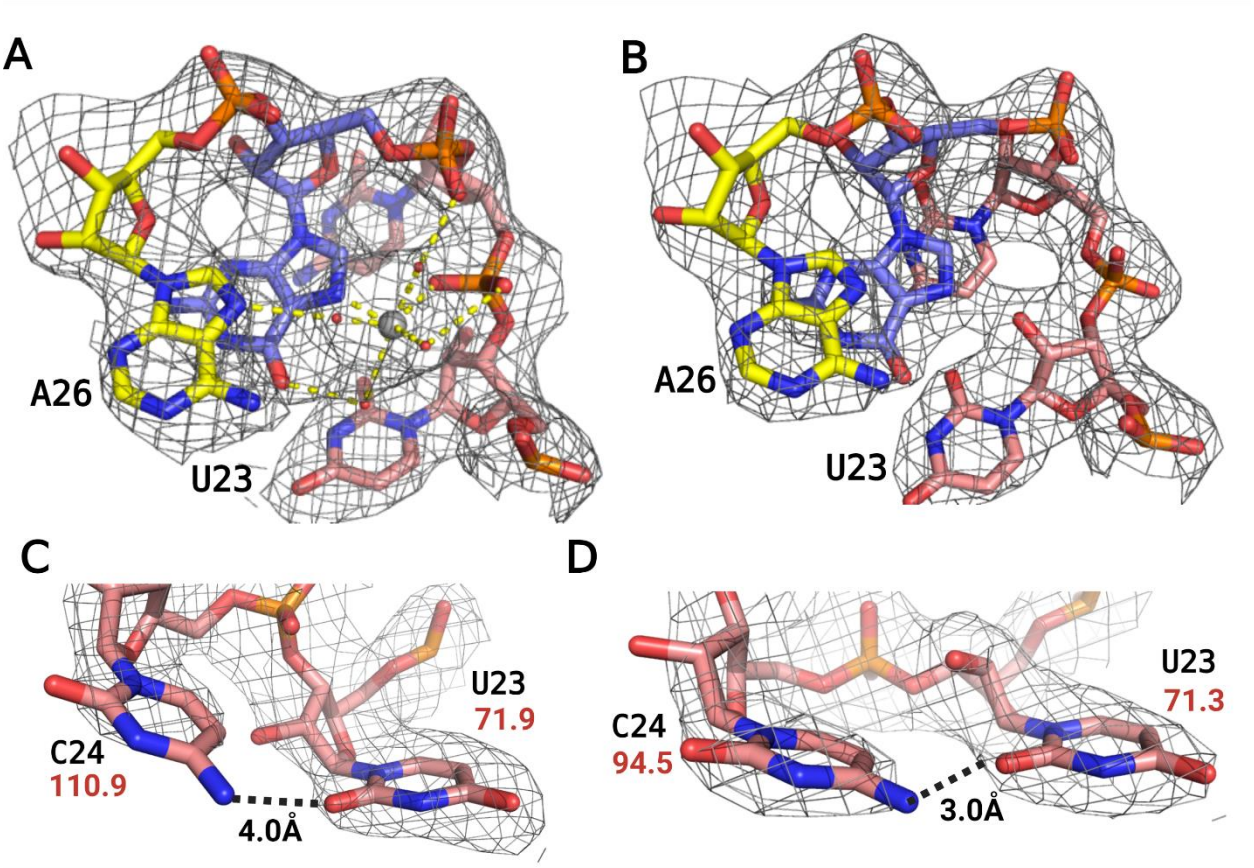


Figure 2.13: Effect of Mg^{2+} (M_C) seen in Pepper-HBC530 on the C24 orientation. A, B) Density map overlay to show the presence of the Mg^{2+} in Pepper-HBC530 and its absence in Pepper-HBC599. C, D) Density comparison of C24 to U25 in Pepper-530 and Pepper-599 respectively, with average temperature factors (red) derived from the crystal structures. The distance from C24 to the U25-O2 is higher in the Pepper-530 structure. The temperature factors show that C24 is more dynamic than neighbouring nucleotide U23 in the crystal lattice.

The base quadruple layer sits on top of the HBC ligand, sandwiching it with the UG wobble pair below (Figure 2.14). With the U21-G22 backbone forming one side of the binding pocket, the aforementioned C14-G54 tertiary interaction closes the other side via a WCF/sugar-edge (SE) base pair. In this tertiary interaction, the sugar-edge of C14 from J1/2 interacts with the WCF face of G54, which adopts a C2'-endo pucker that facilitates its orientation towards C14.

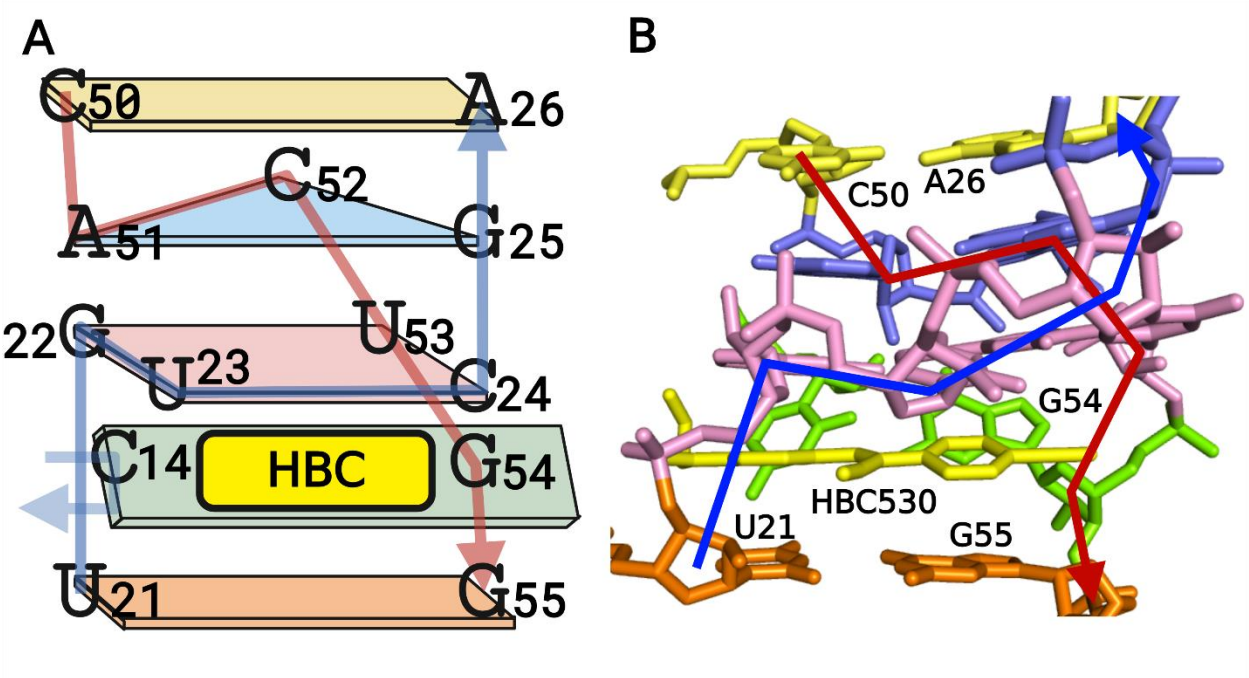


Figure 2.14: Stacking schematic of the Pepper HBC binding site. A) Schematic of the J2/3 and J3/2 stacking arrangement B) 3D cartoon view of stacking with 5' (blue) and 3'(red) backbones traced

Stacked upon the quadruple base layer is a base triple involving nucleotides G25, A51, and C52 (Figure 2.14). G25 and C52 form a WCF base pair, and A51 completes the triple using its N3 endocyclic nitrogen to accept a hydrogen bond from the N4 exocyclic amino group of C52. The ribose of A51 adopts a C2'-endo pucker, which allows A51 and C52 to remain in the same plane. In the HBC530 structure, G25 uses its N7 nitrogen to coordinate directly to M_C, whereas, in the HBC599 structure, which lacks M_C, N7 accepts a hydrogen bond from the U23-2'OH below (Figure 2.13). Above the base triple, an A26-C50 wobble base pair closes the symmetric bulge with a distance of 2.8 Å between A26-N1 and C50-O2. This distance implies that that A26-N1 is protonated, consistent with the crystal growth occurring at pH 6.5 or lower. P3 follows this asymmetric bulge and contains the AAACA BL3-6 pentaloop, which binds Fab BL3-6 in a similar way to previous Fab-BL3-6-RNA complexes[10–13].

Ligand binding pocket

The HBC ligands sit in the stacking pocket (Figure 2.14) formed by the symmetric bulge between P2 and P3. Both ligands stack with one ring sandwiched between U21 below and G22 above (Figure 2.15), with the G22 ribose adopting a C2'-endo pucker, possibly to accommodate the increased phosphate linker distance from U21. The other ring of the ligand stacks between G55 below and the U23/U53 of the quadruple-base layer. The pi-pi distances between the bases and ligand are around 3.4 Å, which is within the expected range for such stacking interactions. Finally, both HBC ligands orient the α,β unsaturated nitrile in the middle of the molecule away from the C14-G54 pair, probably due to steric reasons.

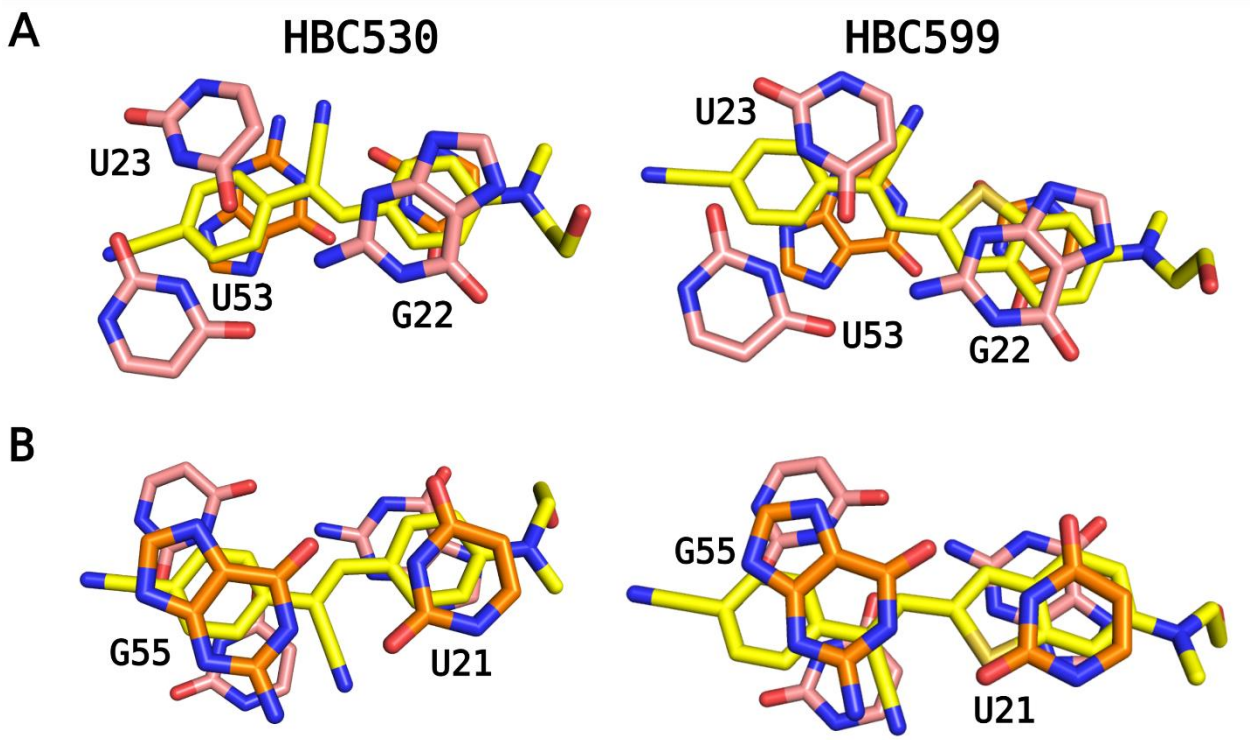


Figure 2.15: HBC ligand stacking from above and below. A) View from above shows that one ring sits sandwiched between U21 and G22 – possibly allowing the tail of the HBC ligands to form a H-bond with G22-N7. The tetrad (C24 not shown) has U23 and U53 making a wobble pair over the other ring of HBC. B) U21-G55 wobble pair below the ligand.

Like other fluorogenic RNA aptamers, Pepper binds its ligand via a platform generated by a co-planar arrangement of multiple bases, though not using G-quadraplexes as observed in several other aptamers before[6,14–17]. Chen *et al.* reported that changing the U21•G55 wobble pair on the underside of the ligand to a GC pair results in weaker fluorescence[1]; possibly a more stable G•C pair attenuates dynamics needed for ligand binding or alters the base pair geometry in a way that affects the structure or stacking preferences of the binding pocket.

Although the electron density clearly defines the location of the HBC ligands within our structures, there is ambiguity with respect to ligand orientation. We modeled the ligand in the orientation that allows the N-hydroxyethyl tail to form a H-bond to G22-N7 (Figure 2.16); nevertheless, the density for HBC599 in particular does not exhibit crystallographic evidence that this interaction is static.

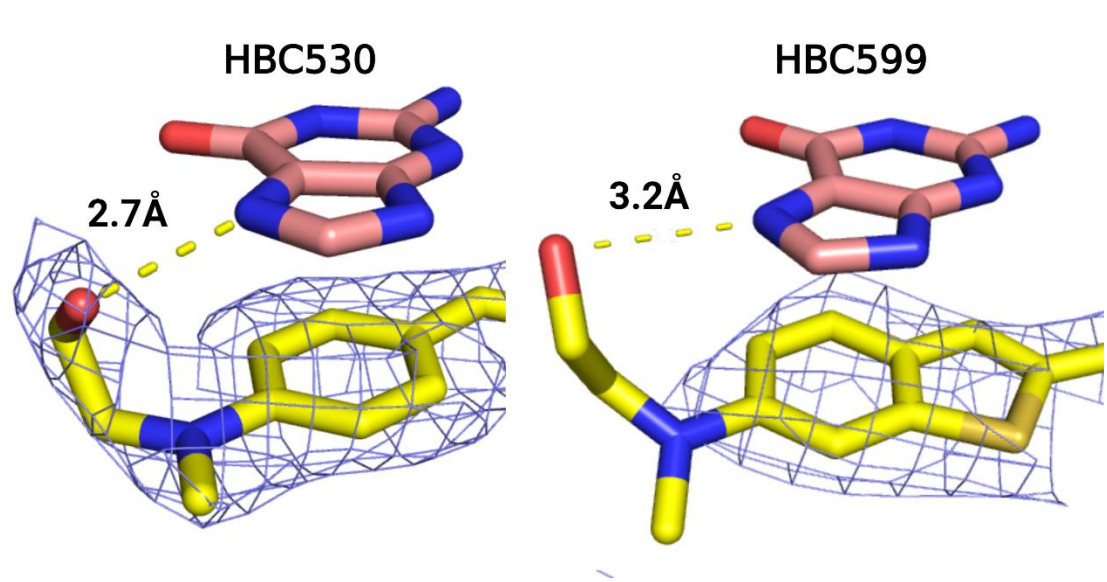


Figure 2.16: Distances between G22-N7 and the tail of the HBC ligands, with ligand density shown

Metal binding pockets

Pepper's fluorogenic activity exhibits a strict dependence on divalent cations, consistent with the site-bound Mg^{2+} cations observed in our structures. The M_A in J1/2 has an inner-sphere contact to the pro-R_P oxygen of A11, supported by three outer-sphere contacts that coordinate to N7 of A11, N7 of A12 and N6 of A12 (Figure 2.17). A second metal ion, M_B can be assigned in both structures due to substantial electron densities above G54 and adjacent to the binding pocket. G54 forms an inner-sphere interaction with this Mg^{2+} through N7, in addition to forming the tertiary interaction to C14 (Figure 2.17). This M_B forms two additional inner-sphere interactions with the phosphates of A51 and C52 and may make the A51 C2'-endo pucker more favourable. Only in the structure of Pepper bound to HBC530 is a third metal ion visible, M_C, which causes C24 to shift relative to the HBC structure (Figure 2.13, 2.17). Compared to previous studies of Mg^{2+} interactions with RNA structures in the pdb, the contacts made by the Mg^{2+} cations in Pepper are not unusual, with the inner sphere pro-Rp oxygen being the most likely atom in RNA to form inner sphere interactions with Mg^{2+}. Previous work has also shown that the most likely nucleobase atom to make inner-sphere contacts with cations are purine N7s [18], in agreement with findings here. The only uncommon interaction is the outer sphere N6 of A12 [19].

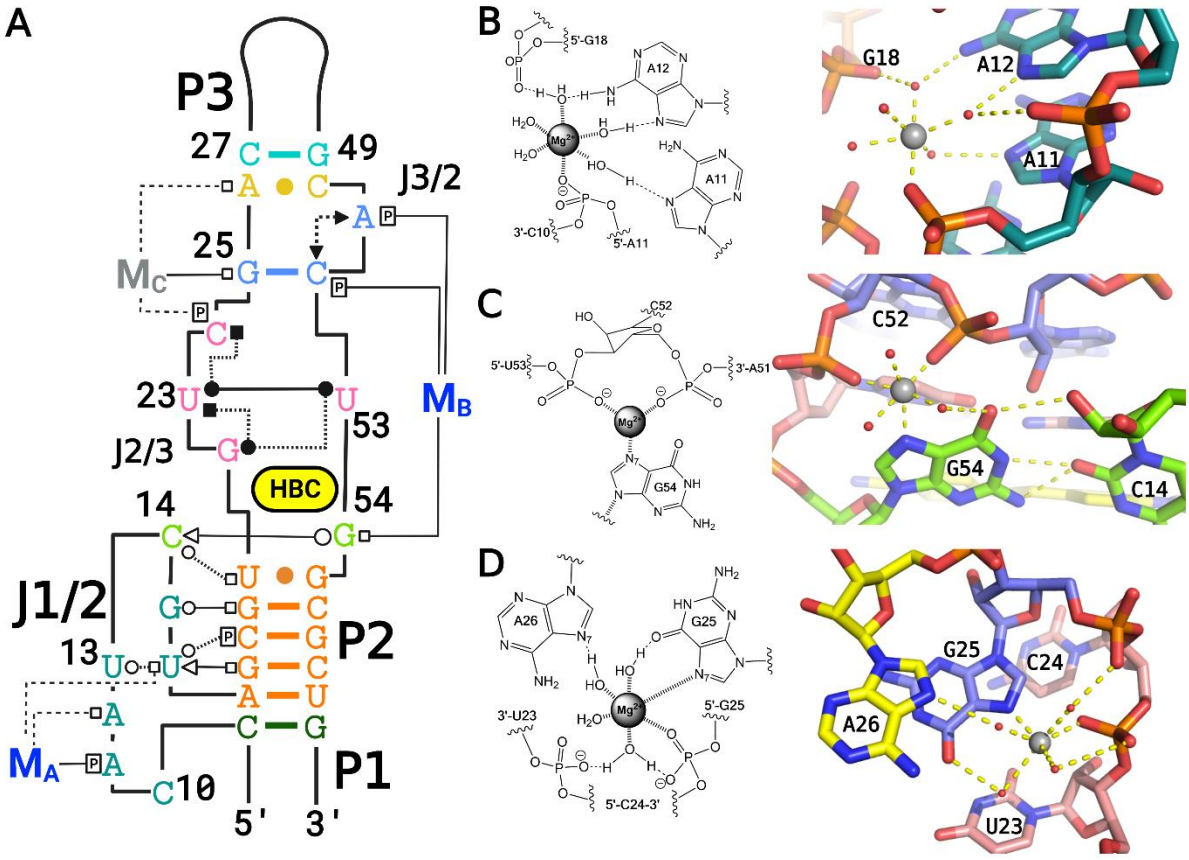


Figure 2.17: Mg^{2+} sites identified in the crystallized Pepper-BL3-6 aptamers. A) Tertiary structure diagram of Pepper bound to HBC530, with the Magnesium ions labeled. B) Diagram and cartoon of the J1/2 Mg^{2+} (M_A) with one inner sphere contact to A11-PO. C) Diagram and cartoon of the binding site Mg^{2+} (M_B) with three inner sphere contacts to G54-N7 and the C52 and U53 phosphates. The G54-C14 interaction is also highlighted D) Diagram and cartoon of the triple base Mg^{2+} (M_C) that is present only in the HBC530 structure, with one inner-sphere contact to G25-N7, with a distance in-between inner and outer sphere to C24-PO. This Mg^{2+} induces a change in the position of C24 of the base quadruple.

2.4 Discussion

Through Fab-chaperone assisted crystallography, the 3D structures of Pepper RNA in complex with ligands HBC530 and HBC599 were successfully derived. The structures show that Pepper adopts a base-quadruple stack fold between J2/3 and J3/2 that sandwiches the ligand against the G•U wobble pair at the top of P2, complemented by a tertiary interaction between C14

and G54 that completes the binding pocket. The structures identified binding sites for Mg^{2+} that appear to be specific and essential to the folding of Pepper, explaining the previously identified divalent dependence of the aptamer. HBC appears to bind overwhelmingly through pi-pi aromatic stacking, but there remains the possibility that a hydrogen bond between G22-N7 and the tail of the HBC ligands is a key stabilizer for binding.

Pepper can be compared to the other non-G-quadruplex aptamers, namely Squash[20], DIR2[21] and MGa[22]. All four of these aptamers utilize a quadruple and triple base system, which is most likely the key stabilizer of the stacking interactions with the aromatic faces of the ligands. Of the non G-quadruplex aptamers, Pepper exhibits the tightest affinity to its respective ligand, and the lowest affinity aptamers (MGa, DIR2) lack a hydrogen-bond between the ligand and RNA and out-of-plane supporting bases respectively, indicating that these features could be essential to forming a tight affinity binding pocket. Pepper is comparable to Squash in affinity, and both aptamers appear to have specific Mg^{2+} binding in their crystal structures, indicating that this support could be contributing to thermodynamic stability. Even compared to G-quadruplex aptamers, Pepper's conditions for binding are not usual, with K^+ dependence seemingly substituted with Mg^{2+} dependence.

The crystal chaperone was critical for lattice formation and structure solving (Figure 2.18). By using the Fab for molecular replacement, the density for the RNA and the ligand became clear quickly without the need for heavy metals. The Fab facilitated crystal contacts primarily by contacting the RNA, rather than other Fabs (77% of all crystal contacts). The RNA-RNA intermolecular base-pairing interaction was serendipitous, particularly as it did not appear to affect the structure in a way that distorts its ligand binding. Previous examples of Fabs as crystallization chaperones have also exhibited this sort of assistance, with Fabs dominating the crystal contacts

for both Spinach (99%)[6] and DIRs (87%)[21], making Pepper consistent in how the Fab facilitates crystallization.

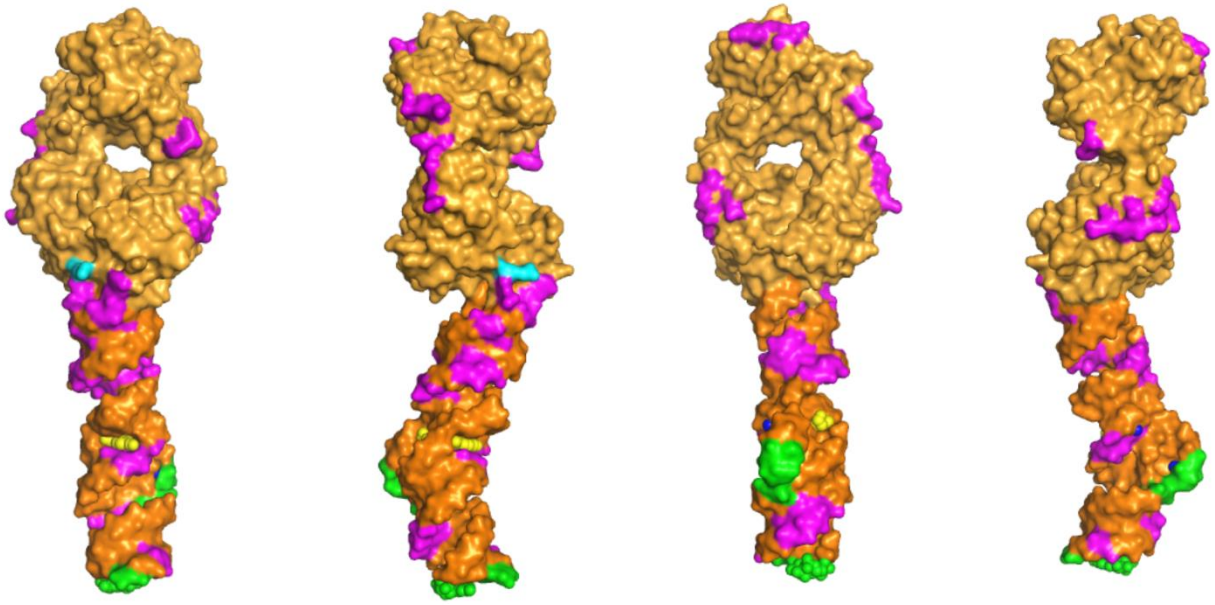


Figure 2.18: Pepper-BL3-6 RNA in complex with Fab and HBC530 crystal contacts. The structure turned 90^{o} anti-clockwise from left to right. RNA-RNA contacting residues (green) dominate the RNA-only crystal lattice. Fab-RNA contacts (magenta) dominate the Fab-RNA crystal lattice, and there are minimal Fab-Fab (cyan) contacts.

Overall, the crystal structures of Pepper were key for identifying three features of HBC binding that were not previously known. Firstly the stacking that was suspected to be formed by the symmetric bulge to hold the ligand is now known to be a unique base-quadruple, though the contribution of C24 of could be minimal, and a base-triple supports the formation. Secondly, the tertiary interactions of J1/2 have been found to be mainly interacting with P2 and J3/2 through the keystone C14-G54 interaction. Finally, the Pepper's Mg^{2+} dependence appears to be derived from specific, consistent binding sites that stabilize the structure and allow HBC to bind. Validating the crystal structure to identify just how important these features are will require mutational analysis and variants that can test the metal sites in great detail.

2.5 References for Chapter 2

1. Chen, X.; Zhang, D.; Su, N.; Bao, B.; Xie, X.; Zuo, F.; Yang, L.; Wang, H.; Jiang, L.; Lin, Q.; Fang, M.; Li, N.; Hua, X.; Chen, Z.; Bao, C.; Xu, J.; Du, W.; Zhang, L.; Zhao, Y.; Zhu, L.; Loscalzo, J.; Yang, Y. Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs. *Nature Biotechnology* **37**, 1287–1293 (2019). [PMID:31548726]

2. Truong, L.; Ferré-D'Amaré, A. R. From fluorescent proteins to fluorogenic RNAs: Tools for imaging cellular macromolecules. *Protein Science* **28**, 1374–1386 (2019). [PMID:31017335]

3. Dolgosheina, E. V.; Jeng, S. C. Y.; Panchapakesan, S. S. S.; Cojocaru, R.; Chen, P. S. K.; Wilson, P. D.; Hawkins, N.; Wiggins, P. A.; Unrau, P. J. RNA Mango aptamer-fluorophore: A bright, high-affinity complex for RNA labeling and tracking. *ACS Chemical Biology* **9**, 2412–2420 (2014). [PMID:25101481]

4. Tan, X.; Constantin, T. P.; Sloane, K. L.; Waggoner, A. S.; Bruchez, M. P.; Armitage, B. A. Fluoromodules Consisting of a Promiscuous RNA Aptamer and Red or Blue Fluorogenic Cyanine Dyes: Selection, Characterization, and Bioimaging. *Journal of the American Chemical Society* **139**, 9001–9009 (2017). [PMID:28644615]

5. Agarwala, P.; Pandey, S.; Maiti, S. The tale of RNA G-quadruplex. *Organic & biomolecular chemistry* **13**, 5570–5585 (2015). [PMID:25879384]

6. Huang, H.; Suslov, N. B.; Li, N. S.; Shelke, S. A.; Evans, M. E.; Koldobskaya, Y.; Rice, P. A.; Piccirilli, J. A. A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nature Chemical Biology* **10**, 686–691 (2014). [PMID:24952597]

7. Kao, C.; Rüdisser, S.; Zheng, M. A simple and efficient method to transcribe RNAs with reduced 3′ heterogeneity. *Methods* **23**, 201–205 (2001). [PMID:11243833]

8. Koldobskaya, Y.; Duguid, E. M.; Shechner, D. M.; Suslov, N. B.; Ye, J.; Sidhu, S. S.; Bartel, D. P.; Koide, S.; Kossiakoff, A. A.; Piccirilli, J. A. A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination. *Nature Structural and Molecular Biology* **18**, 100–107 (2011). [PMID:21151117]

9. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406–3415 (2003). [PMID:12824337]

10. Koldobskaya, Y.; Duguid, E. M.; Shechner, D. M.; Suslov, N. B.; Ye, J.; Sidhu, S. S.; Bartel, D. P.; Koide, S.; Kossiakoff, A. A.; Piccirilli, J. A. A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination. *Nature Structural and Molecular Biology* **18**, 100–107 (2011). [PMID:21151117]

11. Shelke, S. A.; Shao, Y.; Laski, A.; Koirala, D.; Weissman, B. P.; Fuller, J. R.; Tan, X.; Constantin, T. P.; Waggoner, A. S.; Bruchez, M. P.; Armitage, B. A.; Piccirilli, J. A. Structural basis for activation of fluorogenic dyes by an RNA aptamer lacking a G-quadruplex motif. *Nature Communications* **9**, 1–10 (2018).

12. Koirala, D.; Shelke, S. A.; Dupont, M.; Ruiz, S.; DasGupta, S.; Bailey, L. J.; Benner, S. A.; Piccirilli, J. A. Affinity maturation of a portable Fab-RNA module for chaperone-assisted RNA crystallography. *Nucleic Acids Research* **46**, 2624–2635 (2018). [PMID:29309709]

13. Roman, C.; Lewicka, A.; Koirala, D.; Li, N.-S.; Piccirilli, J. A. The SARS-CoV-2 Programmed −1 Ribosomal Frameshifting Element Crystal Structure Solved to 2.09 Å Using Chaperone-Assisted RNA Crystallography. *ACS Chemical Biology* **16**, 1469–1481

(2021). [PMID:34328734]

14. Warner, K. D.; Chen, M. C.; Song, W.; Strack, R. L.; Thorn, A.; Jaffrey, S. R.; Ferré-D'Amaré, A. R. Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nature Structural & Molecular Biology* **21**, 658–663 (2014). [PMID:25026079]

15. Trachman III, R. J.; Demeshkina, N. A.; Lau, M. W. L.; Panchapakesan, S. S. S.; Unrau, P. J.; Ferré-D'Amaré, A. R. Structural basis for high-affinity fluorophore binding and activation by RNA Mango. *Physiology & behavior* **13**, 139–148 (2017).

16. Warner, K. D.; Sjekloa, L.; Song, W.; Filonov, G. S.; Jaffrey, S. R.; Ferré-D'Amaré, A. R. A homodimer interface without base pairs in an RNA mimic of red fluorescent protein. *Nature Chemical Biology* **13**, 1195–1201 (2017). [PMID:28945234]

17. Mieczkowski, M.; Steinmetzger, C.; Bessi, I.; Lenz, A. K.; Schmiedel, A.; Holzapfel, M.; Lambert, C.; Pena, V.; Höbartner, C. Large Stokes shift fluorescence activation in an RNA aptamer by intermolecular proton transfer to guanine. *Nature Communications* **12**, 1–11 (2021). [PMID:34112799]

18. Leonarski, F.; D'Ascenzo, L.; Auffinger, P. Mg2+ ions: Do they bind to nucleobase nitrogens? *Nucleic Acids Research* **45**, 987–1004 (2017). [PMID:27923930]

19. Zheng, H.; Shabalin, I. G.; Handing, K. B.; Bujnicki, J. M.; Minor, W. Magnesium-binding architectures in RNA crystal structures: Validation, binding preferences, classification and motif detection. *Nucleic Acids Research* **43**, 3789–3801 (2015). [PMID:25800744]

20. Dey, S. K.; Filonov, G. S.; Olarerin-George, A. O.; Jackson, B. T.; Finley, L. W. S.; Jaffrey,

S. R. Repurposing an adenine riboswitch into a fluorogenic imaging and sensing tag. *Nature Chemical Biology* **18**, (2021). [PMID:34937909]

21. Shelke, S. A.; Shao, Y.; Laski, A.; Koirala, D.; Weissman, B. P.; Fuller, J. R.; Tan, X.; Constantin, T. P.; Waggoner, A. S.; Bruchez, M. P.; Armitage, B. A.; Piccirilli, J. A. Structural basis for activation of fluorogenic dyes by an RNA aptamer lacking a G-quadruplex motif. *Nature Communications* **9**, 1–10 (2018).

22. Baugh, C.; Grate, D.; Wilson, C. 2.8 Å Crystal structure of the malachite green aptamer. *Journal of Molecular Biology* **301**, 117–128 (2000). [PMID:10926496]

Chapter 3: Structure guided investigation of Pepper-BL3-6 structures

3.1 Introduction

In the previous chapter, structures of the Pepper aptamer have shown that the ligand is bound between two helical stacks supplemented by out-of-plane bases that close up the binding pocket. Supporting this formation are approximately two key divalent binding sites that appear to support the structure and could potentially explain Pepper's dependence on divalent cations. However, there are three main questions that remain to be answered. Firstly is understanding how the crystal structure relates to the functional structure in solution. Although precautions were made to ensure that the crystal would produce a relevant crystal dataset to aqueous conditions, mutational and salt dependence assays can be used to confirm that the crystal structure is providing a model that is realistic for solution based understanding of the aptamer when it is fluorescing in cells. Secondly, although the crystal structures identified two consistent Mg^{2+} binding sites, these cations in the crystal lattice may not be the same sites that are adopted in solution when Pepper binds HBC. Thirdly, the HBC tail is almost certainly forming a hydrogen bond with G22-N7 in the crystal lattice (see chapter 2), but how much this contributes to specificity and HBC binding remains unknown.

Salt dependencies of fluorogenic aptamers are important as this determines the range of environments in which an aptamer can be expected to function. For example, average Mg^{2+} concentrations across eukaryotic cells are approximately 0.5 mM [1], but mitochondrial environments tend to be much higher (above 10 mM) [2]. Pepper's reliance on Mg^{2+} is not unique among fluorogenic RNAs; Squash has exhibited Mg^{2+} dependency[3], and even solely K^{+} reliant G-quadruplex-containing aptamers perform significantly better with divalent ions present, though often this effect can be also achieved by very high monovalent concentrations (1-2M) to create the

atmospheric charge saturation that RNA folding requires [4]. This ionic atmosphere saturation at high monovalent concentrations can be exploited to identify divalent specific effects, as Na^+ and K^+ generally cannot substitute divalents that chelate with RNA structures specifically (Figure 3.1)[5]. By understanding in greater detail how metals are stabilizing and supporting functional aspects of RNA structures, engineering them for improved performance or for bi-functional synthetic RNA aptazymes [6] becomes substantially more likely to succeed.

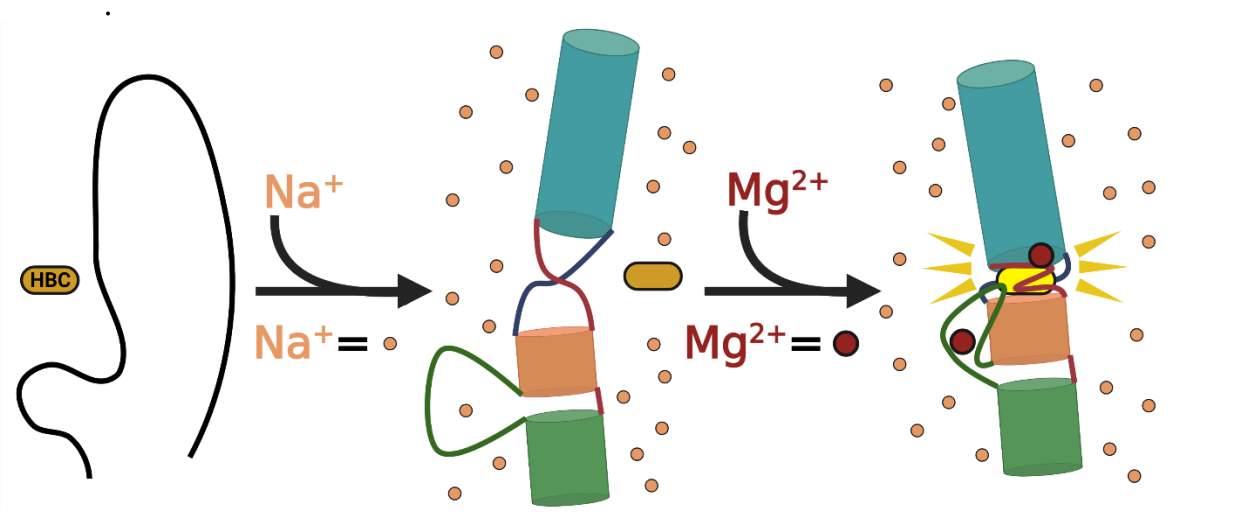


Figure 3.1: Ionic atmosphere saturation using monovalents can be used to ensure that divalents are occupying binding sites and minimally partaking in the atmospheric charge neutralization that RNA folding requires. This allows divalent dependence to be studied more precisely as divalents will be more likely to fulfil structural roles than atmospheric saturation. The HBC only fluoresces (right) when the Mg^{2+} ions bind, making stoichiometric measurement of Mg^{2+} via fluorescence possible [5].

Even though metals can often be assigned in crystal structures, establishing if they correlate to function when the RNA is in solution can be challenging for a variety of reasons. RNA crystals are by nature extremely dense in negative charge, and metal ions are highly attracted to crystal lattice sites that are not electron rich enough to bind them when in solution [7]. From crystal structure data of the four non-G-quadruplex aptamers, the Malachite Green and DIR2 aptamers are able to function without divalent ions in solution [8,9], which correlates with these datasets lacking divalents

in sites close to the aptamer core, though they both accommodate metal ions in their crystal structures. For Pepper, conducting tests to assess how important the crystal-assigned Mg^{2+} are for function is important for describing the specific nature of Pepper's divalent cation dependence. One way of assessing the metal binding in solution is to use α-thio phosphates, where a non-bridging oxygen on a phosphate is substituted to a sulfur and potentially perturbing interactions that involved the oxygen (Figure 3.2). This type of technique has been widely applied to study the kinetics of ribozymes [10].

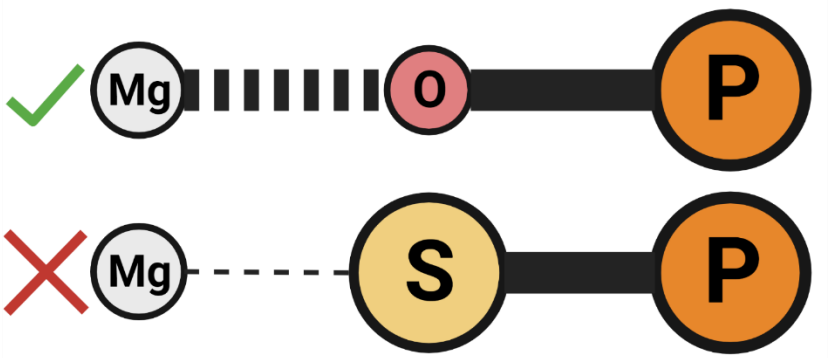


Figure 3.2: Thio-effect for testing Mg^{2+}-oxygen inner sphere chelation in RNA. Mg^{2+} and oxygen have a strong interaction on which RNA structures can depend for function. Mg^{2+} and sulfur have a much weaker interaction, which can seriously perturb RNA functions which depended on that particular Mg^{2+}-oxygen interaction.

Structures of fluorogenic aptamers in complex with their ligands have shown that hydrogen-bonding can be a very common feature of aptamer-ligand affinity and specificity, particularly for HBI-derived aptamers [11]. In the case of Spinach, DFHBI was found to form several hydrogen-bonds to its respective aptamer, and perturbation of these sites using mutational analysis and modifications showed that they did contribute to Spinach's K_d (537 nM) [7], as the K_d was increased by 100s nM. Squash [12], Corn [11], and Chilli [13] have all also exhibited hydrogen-bonding between RNA and ligand in their crystal structures, though the investigations into how much these contribute to binding affinity have varied in depth. As Pepper has only one visible hydrogen-bond

between the ligand and RNA, it is an opportunity to see how much it contributes to Pepper binding though atomic perturbations, and gauge how much the HBC relies on this interaction in comparison to hydrophobic stacking.

To understand how the Pepper structure is able to bind HBC so tightly, we probed the structure through mutational assays, metal dependence studies, and other modification that tested the contributions of other structural features. Through this work we deduced that Pepper only appears to need around two divalent ions per RNA to function thermodynamically, and began to probe specifically where these ions bind. We also confirmed that the crystal structure is in strong agreement with the most likely structure in solution through mutational assays that had impacts that were understandable from the crystal structure, and that the hydrogen-bond between HBC and the Pepper RNA is not decisive component of HBC binding. We also opened the door to future engineering of Pepper, by making sure its requirements are precisely understood.

3.2 Methods

Pepper RNA synthesis and purification

Pepper-BL3-6 RNA was prepared from dsDNA and T7 RNA polymerase as described in earlier (see chapter 2).

For the split-Pepper assays, RNA oligos (Table 3.1) were synthesized in-house by Nan-Sheng Li using solid-phase synthesis on a 1-µmol scale using an Expedite Nucleic Acid Synthesis System (8900) by following standard RNA synthesis protocols. The oligonucleotides were released from solid support with 3:1 NH4OH/EtOH at 55 °C for 8 h, desilylated with 300 µL 6:3:4 N-methylpyrrolidinone/triethylamine/triethylamine-3HF at 65 °C for 2 h and precipitated by n-

BuOH. The oligomers were further purified by dPAGE, collected in pure water and stored at −80 °C until further use.

HBC syntheses

HBC530 and HBC599 were synthesized as described earlier (see chapter 2) by Nan-Sheng Li.

Fluorescence assay experiments

Florescence measurements were conducted in the same manner as previously described (see chapter 2), with HBC at 1 μM unless otherwise stated.

Dissociation constant identification

To measure the K_d of HBC ligands with Pepper constructs, constant 1 nM RNA over varying HBC concentration were measured via fluorescence assay, and the measurements fitted to a Hill equation with one binding site (n=1, Figure 3.3) using OriginPro. To ensure that the values obtained were accurate for the binding constant (K_d), several data points must be made with both components below the dissociation constant concentration, a measurement is made with the variable not present (0 nM), and several measurements made at the saturation point (plateau).

Hill Mg^{2+} binding isotherm

To assess Mg^{2+} stoichiometry in the Pepper aptamer, high monovalent background concentrations were used [5]. Florescence measurements were conducted using a Neo2 Synergy BioTek plate reader. RNA heated to 95°C for 1 minute in water, cooled on ice for 2 minutes, before buffer (1x; HEPES 40 mM (pH 7.5), 2M NaCl, supplied as 2x) and dye (1 μM) were added. $MgCl_2$ was then added as a variable to the appropriate concentration. The solutions were incubated for 30

minutes at 25 ºC before measurements were taken (Ex/Em 485/530 for HBC530, 515/599 for HBC599, gain 100, ±20) at 25ºC, and the reported results were the average of three measurements. Date was then fit to a hill equation (Figure 3.3).

$$F = A + B\frac{\left(\frac{[M^{2+}]}{[M^{2+}]_{1/2}}\right)^{n}}{1+\left(\frac{[M^{2+}]}{[M^{2+}]_{1/2}}\right)^{n}}$$

Figure 3.3: Equation used for fitting fluorescence data over changing Mg^{2+} concentrations with a 2M NaCl background. Using Originpro, the following 4 parameters were returned; the unfolded Pepper fluorescence A, the maximal change in fluorescence B, the midpoint $[M^{2+}]_{1/2}$, and the apparent hill coefficient n. 3 independent measurements were taken for the input to yield the final values.

Table 3.1: Split-Pepper RNA sequences

ID	Type	Sequence 5'-3'	Description
Split-Pepper-Strand1	RNA	CC UAC CAA UCG UAG CGU GUC GAC CAG C	RNA sequence synthesized using solid-phase synthesis (see methods)
Split-Pepper-Strand2	RNA	G CUG GCA CUG GCG CUG UAG G	RNA sequence synthesized using solid-phase synthesis (see methods)

Figures and Schemes

Figures were generated in Biorender™. 3D models were generated using pymol. Organic structures were drawn using Chemdraw.

3.3 Results

Mutational Analysis

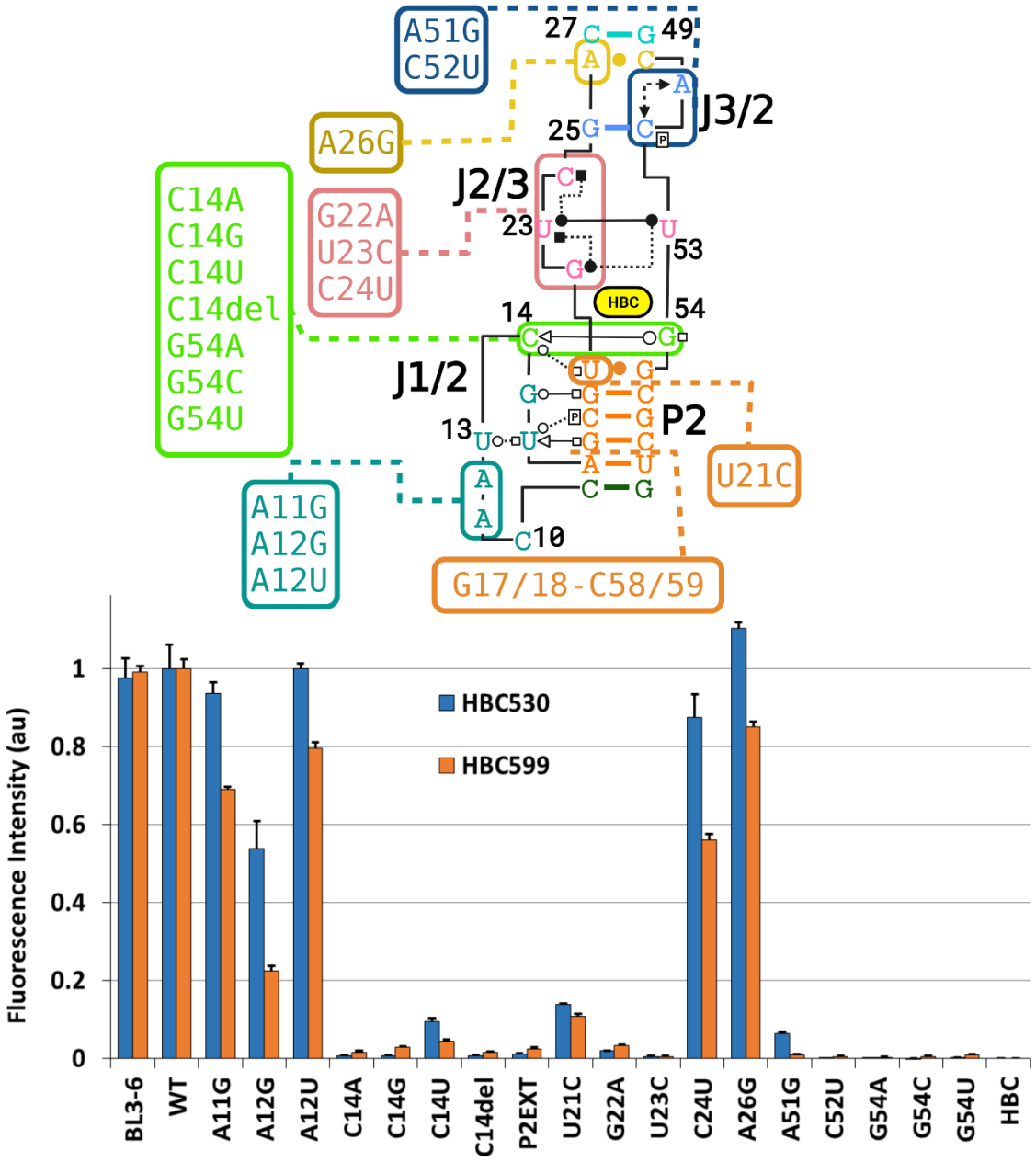


Figure 3.4: Assessment of base contributions to Peppers fluorogenic activity via mutational analysis with HBC530 (blue) and HBC599 (orange). Pepper-BL3-6 was incubated with a HBC ligand for 30 minutes at 25 °C. Measurements were made with 100 nM RNA, 1 μM HBC, 40 mM HEPES pH 7.5, 100 mM KCl, and 5 mM $MgCl_2$.

To probe the functional relevance of our Pepper structures for fluorogenic activity, mutational analysis was conducted (Figure 3.4). First, we tested mutations in several layers stacked above or below the ligand. The quadruple base layer was probed by separate mutations at G22 and U23. A G22A mutation reduced fluorescence to barely above background with both HBC530 and HBC599, consistent with the importance of the hydrogen-bonds mediated by the G22-exocyclic amine. U23 has proximity to all other members of the tetrad and may serve as the keystone nucleotide. Mutating U23C reduced fluorescence to almost background levels, potentially also due to disrupting the quadruple base platform for ligand binding. A C24U mutation mostly retained fluorescence, consistent with the structures implicating C24 to be more dynamic than the rest of the tetrad (see chapter 2). In G25-A51-C52 base triple above the tetrad, a C52U mutation, which breaks the canonical G25-C52 base pair at the heart of the triplex, abolishes fluorescence, supporting the importance of WCF base pair in this triplex. A51G, which loses fluorescence indicates that the interaction between A51 and C52 is also critical. The non-canonical A-C base pair at the transition between the bulge and P3 tolerates conversion to a canonical base pair through an A26G mutation with no change in signal. In the layer below the ligand, U21C causes a serious drop in signal. Presumably this mutation converts the U21-G5 wobble pair to a C•G WCF base pair, which could disrupt the hydrogen bond between the C14-2'-OH and U21-O4.

Next we tested the key nucleotides that sit out of plane to the ligand and form the back of the binding site – C14, and G54. G54 mutations (G54A, G54C, and G54U) all caused a complete loss of fluorescent signal, presumably by disrupting the interactions with C14 or M_B, indicating that this residue makes a vital contribution to Pepper's fluorogenic activity. Although G54A should be able to maintain the inner-sphere contact with the M_B through N7 and a favorable interaction with C14's 2'-OH via N6, the interactions with C14's O2 keto group would be disrupted. Deletion of

C14 (C14del) and transversion mutations (C14A, C14G) abolished signal. However, the transition mutation, C14U maintained around 30% WT fluorescence from HBC530 but abolished signal from HBC599. P2ext – where an extra base pair was inserted into P2 – increased the distance helical turn between J1/2 and the binding pocket, also reduced signal to background.

Finally, we constructed mutations to probe the Mg^{2+} binding site in J1/2. M_A forms one inner sphere contact through the pro-R_P oxygen of A11 and several outer sphere interactions to the Hoogsteen faces of A11 and A12. The first mutation – A11G – maintained fluorescence with HBC but led to a 30% loss with HBC599. A12U also maintained high fluorescent signal, despite most likely changing the outer-sphere arrangements of M_A significantly. A12G maintained a weaker signal, indicating that the structure of J1/2 was affected but not completely abolished.

Metal Dependence analysis

The dependence of Pepper on Mg^{2+} for folding and function can reflect contributions to the folded RNA relative to the unfolded RNA from both site-bound metal ions and diffusely-associated metal ions that form the ion atmosphere. To saturate the ion atmosphere and thereby isolate the contribution of site-bound Mg^{2+} to Pepper function, we conducted fluorescence measurements of samples containing 100 nM M RNA and 5 uM ligand in a 2M NaCl background [15] (Figure 3.5.). Fits of fluorescence versus Mg^{2+} yielded Hill coefficients (n) of 1.82 and 1.86 for the HBC530 and HBC599 complexes respectively, suggesting that the aptamer requires two-site bound metal ions, most likely M_A and M_B (see chapter 2) . Although both ligands exhibit an apparent requirement for two Mg^{2+} ions, the different apparent $[Mg]_{1/2}$ values for HBC599 ($[Mg^{2+}]_{1/2}$ = 7.8 mM) compared to HBC530 ($[Mg^{2+}]_{1/2}$ = 4.0 mM) suggests energetic coupling between Mg^{2+} and ligand binding.

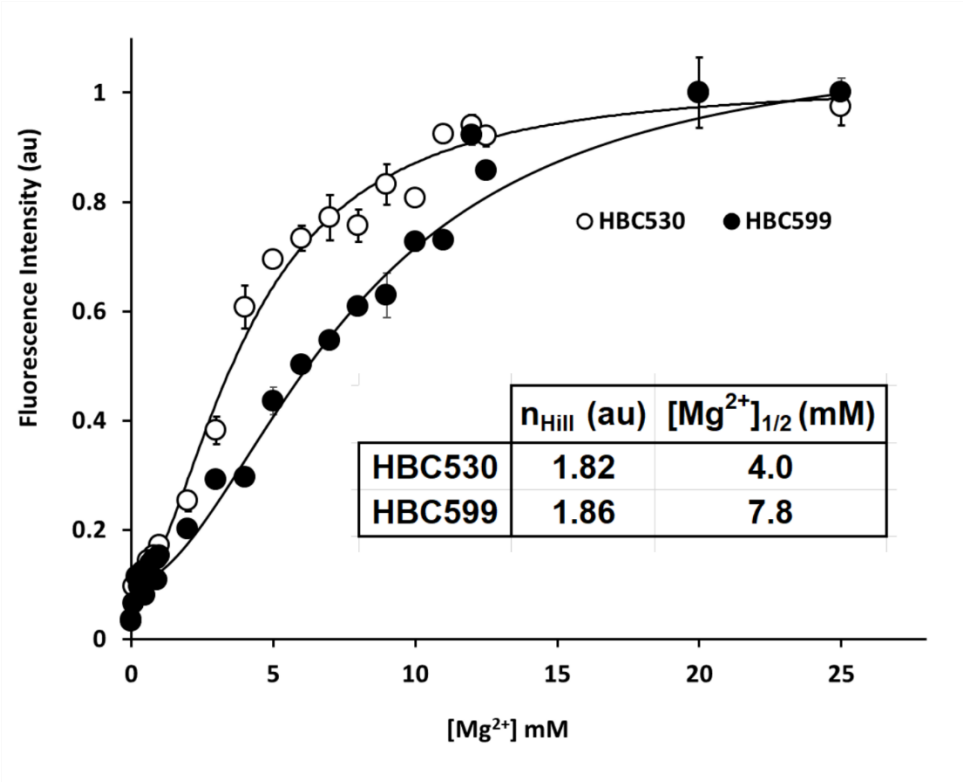


Figure 3.5: Hill binding isotherm showing fluorescence over varying concentrations of Mg^{2+} with Pepper-BL3-6 bound to HBC530 and HBC599. Three replicate fluorescence measurements were averaged, with appropriate weighting and assuming normally distributed errors, to yield the final values of $[M^{2+}]_{1/2}$ and n for each Pepper complex. 2M NaCl-background fluorescence isotherm showing how fluorescence of Pepper-BL3-6 changes over $[Mg^{2+}]$, with n_{Hill} = 1.82 for binding HBC530 and n_{Hill} = 1.86 for binding HBC599 respectively. With saturated ion-atmosphere from 2M NaCl, Mg^{2+} binding is limited to non-monovalent binding sites on the aptamer.

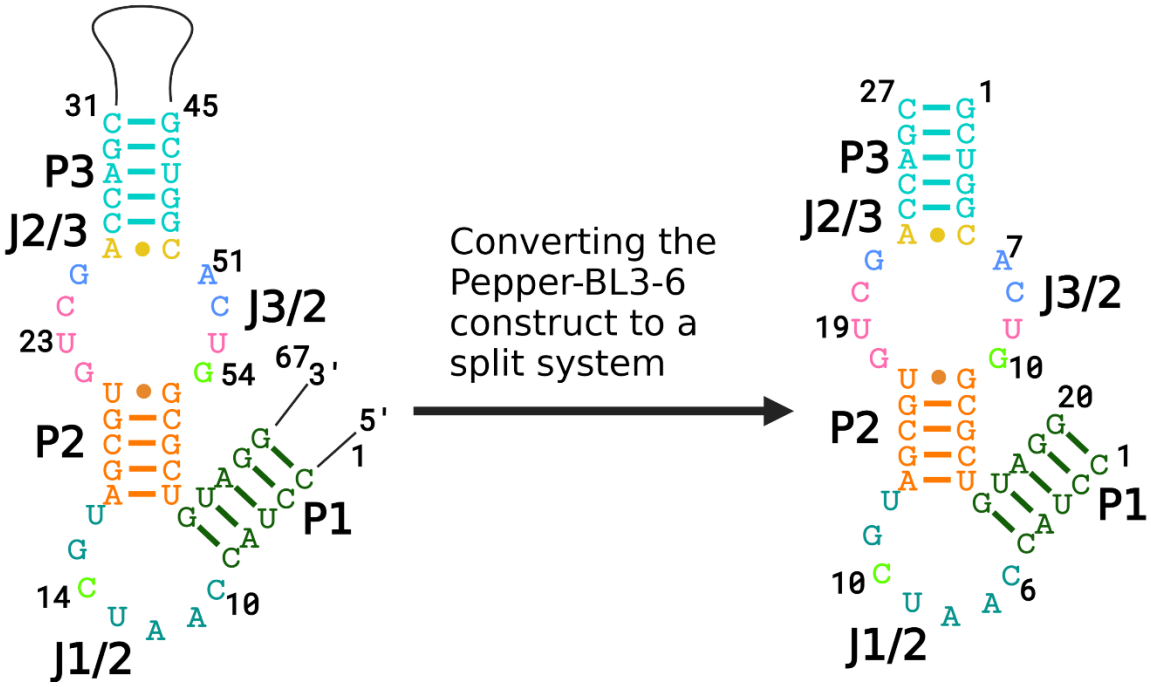


Figure 3.6: Split-Pepper construct designed from the crystal structure of Pepper-BL3-6. The RNA was made using solid-phase syntheses to incorporate α-thio phosphates and 7-deaza-G in positions to test the Pepper structure in solution. Synthesized by Nan-Sheng Li

To precisely assess Mg^{2+} binding visible in the Pepper crystal structures, a minimal-split-Pepper motif was designed from the crystal structures, and was demonstrated to have similar fluorescence to Pepper BL3-6 (Figure 3.6). Phosphates making inner-sphere interactions with M_A and M_B in the crystal structures were substituted with α-thio-phosphates, to disrupt the strong Mg-O bond. These thio-modifications changed R and S oxygens on the phosphate in a 1:1 ratio, meaning that only 50% of each modification would be affected as they were not separated after synthesis. This translates to a 50% signal decrease indicating that the RNA with this thio modification completely loses fluorescence activity. Modifications that targeted M_A (A11-PS) had only slightly reduced fluorescence, indicating that no cations need to interact with the A11-phosphate for fluorescence (Figure 3.7). A modification to nearby phosphates (A12-PS, U13-PS) had a similar effect despite not being assigned to have an inner-sphere chelation to Mg^{2+} in the diffraction dataset. In contrast, substituting oxygens that coordinated with M_B with sulfurs had a

stronger effect (A51-PS and C52-PS), which accumulated when both modifications were introduced at the same time, serving as evidence that M_B is the key source of divalent dependency in Pepper.

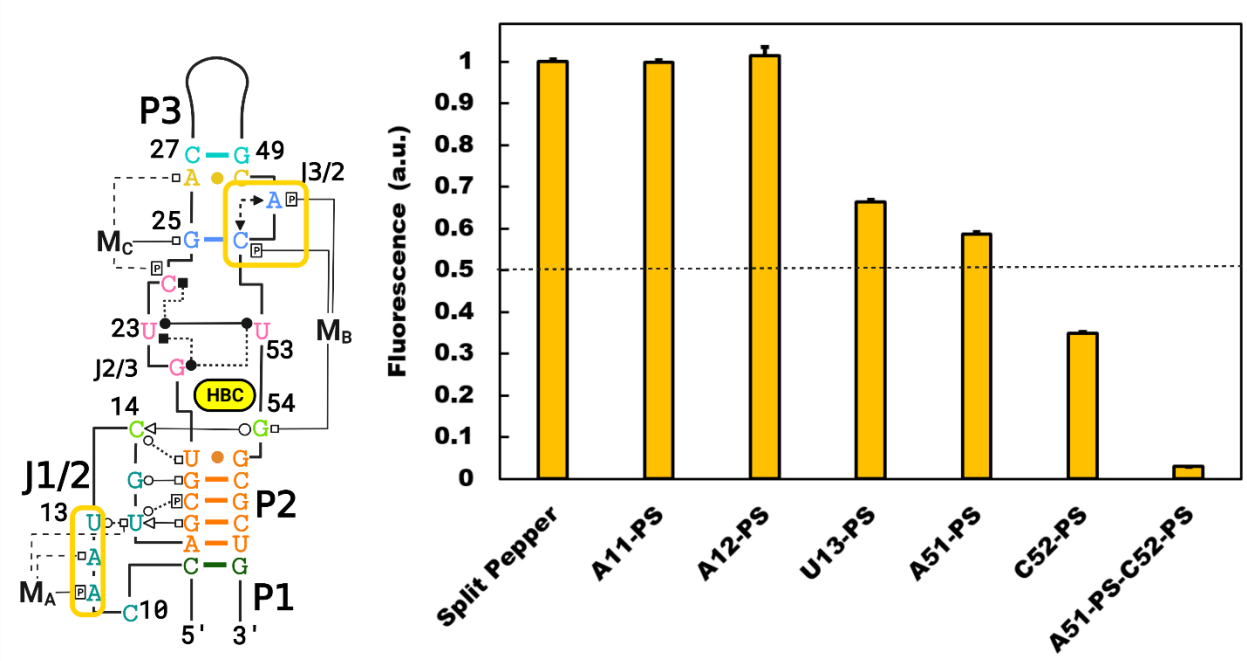


Figure 3.7: Split-Pepper thio-modifications study to assess Mg^{2+} binding sites. A minimal Pepper construct was designed based on the crystal structure. R-α-thio-phosphate modifications were introduced to key sites that showed inner sphere Mg^{2+} coordination in the Pepper-BL3-6 structure (yellow bars) to inhibit Mg^{2+}-oxygen interaction and disrupt function. The 50% mark indicates where activity is substantially low enough that Pepper function has been essentially wiped out by that modification.

To complement the thio-substitutions, 7-deaza-guanasines (N_7→C-H) were introduced to test the interactions seen with G54 and M_B, and G25 and M_C (HBC530 only) (Figure 3.8). In agreement with the data, the 7-deaza-G54 showed a decreased signal, whereas the 7-deaza-G25 had a no significant decrease, reflecting the implied importance of the M_B and M_C to structure. The decrease for the 7-deaza-G54 modification is surprisingly small, however, as this would completely remove the ability for Mg^{2+} to bridge J3/2. This indicates that the majority of the

structural contribution of M_B may be from the phosphate contacts instead, which is in agreement with Mg^{2+} generally forming tighter interactions with oxygen than nitrogen [16].

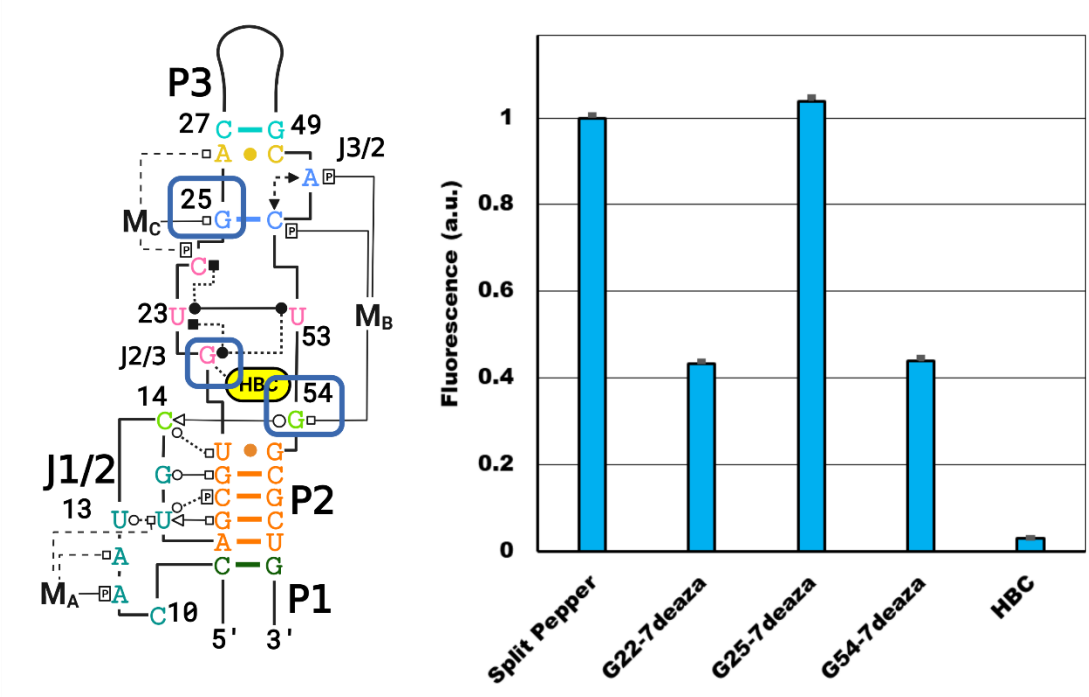
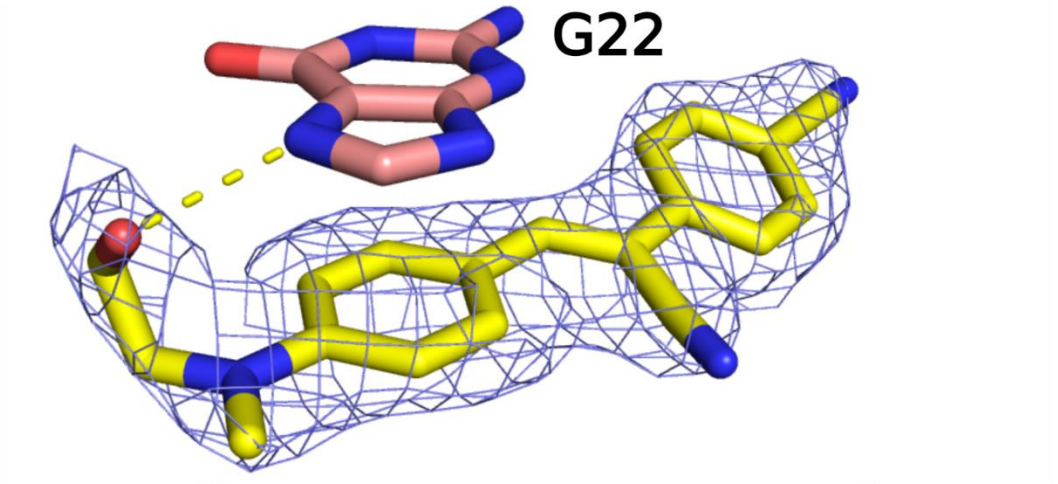
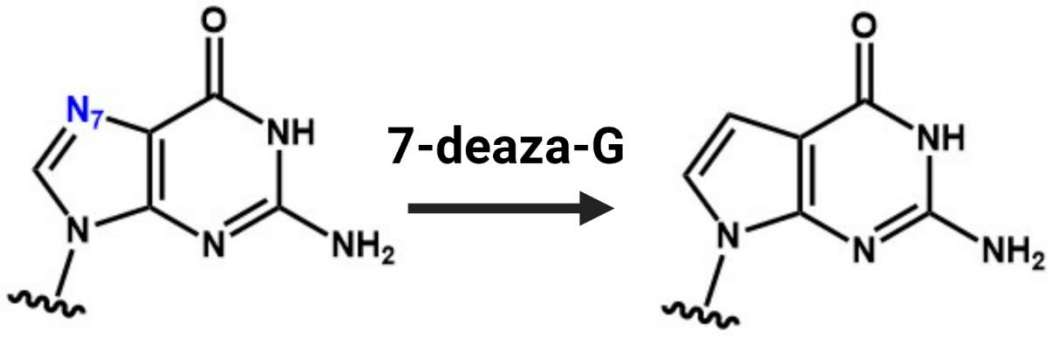


Figure 3.8: Split-Pepper modifications study to assess Mg^{2+} binding sites and the ligand-RNA hydrogen bond.

Hydrogen-bonding between HBC and Pepper RNA

In the previous chapter, the HBC ligand was shown to potentially form hydrogen-bond with G22-N7 In the HBC530 structure density indicated that this is indeed where the ligand tail (the ethyl alcohol motif on the tertiary amine) is positioned in the crystal lattice, but in HBC599 such density is not visible, indicating that the tail is dynamic in solution. This brings into question how much the hydrogen-bond contributes energetically to the binding of HBC in solution. It is also notable that during the SELEX procedure that generated Pepper, the HBC ligand was most likely attached to a surface via this tail motif [17], which would imply that the binding of HBC should not be reliant on this H-bond.





HBC530 K_d	Pepper-BL3-6	Split-Pepper	7-deaza-G22
K_d (nM)	8.5	15.9	36.1

Figure 3.9: 7-deaza-G modification can test hydrogen-bond seen in Pepper crystals. Measurement of K_d with Pepper RNAs with HBC530

Using a similar split-Pepper system described earlier, RNA with the 7-deaza-G22 was synthesized, which would remove the ability for the HBC ligand to form a hydrogen-bond with the N7 (Figure 3.9). When 7-deaza-G22 was tested with HBC530, the fluorescence of the split-Pepper system dropped below half of the original signal (Figure 3.8). To further investigate this, K_d of HBC530 was measured with this 7-deaza-G, and the K_d was shown to increase two-fold between the split-Pepper and the atomically modified system. The 7-deaza-G modification

approximately doubled the Kd, meaning a two-fold loss in binding affinity, though the affinity is still very high for a fluorogenic aptamer overall (36.1 nM) [18].

3.4 Discussion

Mutational analysis of Pepper revealed that the crystal structure appears to be representative of the Pepper structure in solution, as mutations had the expected effects on binding. C14 and G54 are crucial to the binding of HBC, and the base-quadruple and base-triple also sequence specific, losing function if mutated. A key exception to this is C24 of the base-quadruple which, as the crystal structures alluded to, is more dynamic and less essential to binding, with a mutation to a uracil only slightly affecting fluorescence. J1/2 appeared to tolerate some mutations close to the M_A binding site, but overall appears to be in an optimized state.

The Mg^{2+} stoichiometry, via hill analysis, revealed that around 2 divalent ions bind for HBC activation to occur, though this does not mean two cations must be bound statically to the RNA at once. Follow-up experiments showed that the M_B close to the binding site is critical to Pepper's function and most likely occupies that site during fluorescence, but M_A, though in both crystal structures, is unlikely to be specifically bound to that phosphate during or after HBC binding to the phosphate of A11. As we have no data on the kinetics of HBC binding, it's likely that divalents are involved in interactions that are not represented in the final static structure, as the J1/2 loop may have Mg^{2+} dynamics during binding or in solution that are not represented in the crystal structure. The higher $[M^{2+}]_{1/2}$ of HBC599 to HBC530 correlates with HBC599's higher K_d, which may allude to a cooperative binding of the ligand and Mg, though without future testing there are numerous possibilities.

HBC binding appears to be not be heavily dependent on the hydrogen-bonding to G22, HBC530 was shown to still bind tightly to a Pepper modified with 7-deaza-G22. In the case of Spinach, which has several hydrogen bonds to DFHBI in the bound state, removing these hydrogen bonds resulted in K_d increases on the order of 100s nM [19], which demonstrates a more significant thermodynamic reliance on these interactions. As the increase seen in Pepper is very small, this does not provide evidence for the hydrogen-bond's impact on binding. It is important to consider that 7-deaza-G modifications do also affect the electronic pi-system profile of the guanosine interfacing with HBC ligands, meaning that the observed changes in K_d are not necessarily due to hydrogen bonds alone, though the data safely rules the hydrogen-bond out as the key stabilizer of binding.

Overall, the investigations have shown that the crystal structures of Pepper represent the solution structure fairly well, though also raises questions about the dynamics of HBC binding. We currently don't know if the binding pocket is pre-organized like it is in Spinach[19], and in what temporal sequence HBC and Mg^{2+} bind to the RNA. However, the RNA sequence of Pepper does appear to be highly optimized, with the most effective sequence apparently identified for in vitro testing. There are avenues for Pepper's performance in vivo to be improved through lowering the divalent dependence of the aptamer, which was a problem during initial engineering [13], and if dynamics of the binding are studied, potentially Pepper's performance can be further enhanced.

3.5 References for Chapter 3

1. Goytain, A.; Quamme, G. A. Identification and characterization of a novel family of membrane magnesium transporters, MMgT1 and MMgT2. *American Journal of Physiology*

- *Cell Physiology* **294**, 495–502 (2008). [PMID:18057121]

2. Pilchova, I.; Klacanova, K.; Tatarkova, Z.; Kaplan, P.; Racay, P. The Involvement of Mg2+ in Regulation of Cellular and Mitochondrial Functions. *Oxidative Medicine and Cellular Longevity* **2017**, (2017). [PMID:28757913]

3. Dey, S. K.; Filonov, G. S.; Olarerin-George, A. O.; Jackson, B. T.; Finley, L. W. S.; Jaffrey, S. R. Repurposing an adenine riboswitch into a fluorogenic imaging and sensing tag. *Nature Chemical Biology* **18**, (2021). [PMID:34937909]

4. Warner, K. D.; Chen, M. C.; Song, W.; Strack, R. L.; Thorn, A.; Jaffrey, S. R.; Ferré-D'Amaré, A. R. Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nature Structural & Molecular Biology* **21**, 658–663 (2014). [PMID:25026079]

5. Frederiksen, J. K.; Li, N. S.; Das, R.; Herschlag, D.; Piccirilli, J. A. Metal-ion rescue revisited: Biochemical detection of site-bound metal ions important for RNA folding. *Rna* **18**, 1123–1141 (2012). [PMID:22539523]

6. Su, Y.; Hammond, M. C. RNA-based fluorescent biosensors for live cell imaging of small molecules and RNAs. *Current Opinion in Biotechnology* **63**, 157–166 (2020). [PMID:32086101]

7. Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. Cations in charge: Magnesium ions in RNA folding and catalysis. *Current Opinion in Structural Biology* **22**, 262–272 (2012). [PMID:22595008]

8. Baugh, C.; Grate, D.; Wilson, C. 2.8 Å Crystal structure of the malachite green aptamer.

Journal of Molecular Biology **301**, 117–128 (2000). [PMID:10926496]

9. Tan, X.; Constantin, T. P.; Sloane, K. L.; Waggoner, A. S.; Bruchez, M. P.; Armitage, B. A. Fluoromodules Consisting of a Promiscuous RNA Aptamer and Red or Blue Fluorogenic Cyanine Dyes: Selection, Characterization, and Bioimaging. *Journal of the American Chemical Society* **139**, 9001–9009 (2017). [PMID:28644615]

10. Ren, A.; Micura, R.; Patel, D. J. Structure-based mechanistic insights into catalysis by small self-cleaving ribozymes. *Current Opinion in Chemical Biology* **41**, 71–83 (2017). [PMID:29107885]

11. Warner, K. D.; Sjekloa, L.; Song, W.; Filonov, G. S.; Jaffrey, S. R.; Ferré-D'Amaré, A. R. A homodimer interface without base pairs in an RNA mimic of red fluorescent protein. *Nature Chemical Biology* **13**, 1195–1201 (2017). [PMID:28945234]

12. Truong, L.; Kooshapur, H.; Dey, S. K.; Li, X.; Tjandra, N.; Jaffrey, S. R.; Ferré-D'Amaré, A. R. The fluorescent aptamer Squash extensively repurposes the adenine riboswitch fold. *Nature Chemical Biology* **18**, (2021).

13. Mieczkowski, M.; Steinmetzger, C.; Bessi, I.; Lenz, A. K.; Schmiedel, A.; Holzapfel, M.; Lambert, C.; Pena, V.; Höbartner, C. Large Stokes shift fluorescence activation in an RNA aptamer by intermolecular proton transfer to guanine. *Nature Communications* **12**, 1–11 (2021). [PMID:34112799]

14. Kao, C.; Rüdisser, S.; Zheng, M. A simple and efficient method to transcribe RNAs with reduced 3′ heterogeneity. *Methods* **23**, 201–205 (2001). [PMID:11243833]

15. Das, R.; Travers, K. J.; Bai, Y.; Herschlag, D. Determining the Mg2+ stoichiometry for

folding an RNA metal ion core. *Journal of the American Chemical Society* **127**, 8272–8273 (2005). [PMID:15941246]

16. Dolgosheina, E. V.; Jeng, S. C. Y.; Panchapakesan, S. S. S.; Cojocaru, R.; Chen, P. S. K.; Wilson, P. D.; Hawkins, N.; Wiggins, P. A.; Unrau, P. J. RNA Mango aptamer-fluorophore: A bright, high-affinity complex for RNA labeling and tracking. *ACS Chemical Biology* **9**, 2412–2420 (2014). [PMID:25101481]

17. Chen, X.; Zhang, D.; Su, N.; Bao, B.; Xie, X.; Zuo, F.; Yang, L.; Wang, H.; Jiang, L.; Lin, Q.; Fang, M.; Li, N.; Hua, X.; Chen, Z.; Bao, C.; Xu, J.; Du, W.; Zhang, L.; Zhao, Y.; Zhu, L.; Loscalzo, J.; Yang, Y. Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs. *Nature Biotechnology* **37**, 1287–1293 (2019). [PMID:31548726]

18. Ouellet, J. RNA fluorescence with light-Up aptamers. *Frontiers in Chemistry* **4**, 1–12 (2016).

19. Huang, H.; Suslov, N. B.; Li, N. S.; Shelke, S. A.; Evans, M. E.; Koldobskaya, Y.; Rice, P. A.; Piccirilli, J. A. A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nature Chemical Biology* **10**, 686–691 (2014). [PMID:24952597]

Chapter 4: Conclusions – Consolidation of Pepper structures

4.1 Introduction

In the previous chapter, the Pepper-BL3-6 structures were probed to confirm that the crystal structure observations were consistent with Pepper binding HBC in solution. The mutational data showed that the crystal-derived 3D-models appears to be an accurate representation of how the RNA is positioned when Pepper binds HBC in solution, with mutations that would by hypothesized to cause disruption in the crystal having deleterious effects in solution. In addition, the metal assignments – namely M_A and M_B – appear to be in agreement with the stoichiometric data, with approximately two Mg^{2+} ions being essential to Pepper function. Modifications targeting key phosphate oxygens and Guanosine-N7s showed that the Mg^{2+} close to the binding site, M_B, is the key contributor to divalent dependence in Pepper, with M_A having no observable influence outside of the crystals. Finally, the hydrogen bond that HBC ligands can form with the Pepper RNA appears to not be a critical component of binding thermodynamically, though this is inconclusive and needs further investigation,

During writing of this thesis, and while a publication of thesis material was being written, a publication was released which had also solved the structures of the Pepper RNA aptamer, using heavy-metal soaking instead of chaperone assisted crystallography [1]. Since the Fab-RNA structures were solved before these RNA-only structures were released, they did not inform the RNA structure and are therefore being treated as contemporary for the writing of this thesis. A comparison is included for scholarship purposes.

In this chapter, the consolidated knowledge gleaned from the Pepper-BL3-6 structure will now be compared to an RNA-only crystal structure of the Pepper aptamer that was published

during the closing stages of this work. In this comparison we will see that the Pepper structures appear to tell a consistent story of HBC binding, and that our structures are overall in great agreement. Finally, future work for the Pepper aptamer will be proposed – specifically how Pepper can be integrated into aptazymes, and the thesis conclusions will be drawn and summarized.

4.2 Comparison to a circularly permuted Pepper construct.

After we solved the structures of Pepper bound to ligands HBC530 and HBC599 respectively, Huang *et al.* reported the X-ray crystal structure of a circularly permuted (cp) Pepper construct bound to HBC530 and several other ligands [1], though not with HBC599. The overall structure reported is highly similar to our structures (Figure 4.1) despite different space groups and lattice packing interactions, supporting the functional relevance of the observed 3D architecture. Overlapping of the comparable regions between the structures gave excellent convergence (Figure 4.1), with an rmsd (root mean squared deviation) value below 1 $Å^2$ showing that the structures have bases in virtually identical positions relative to each other.

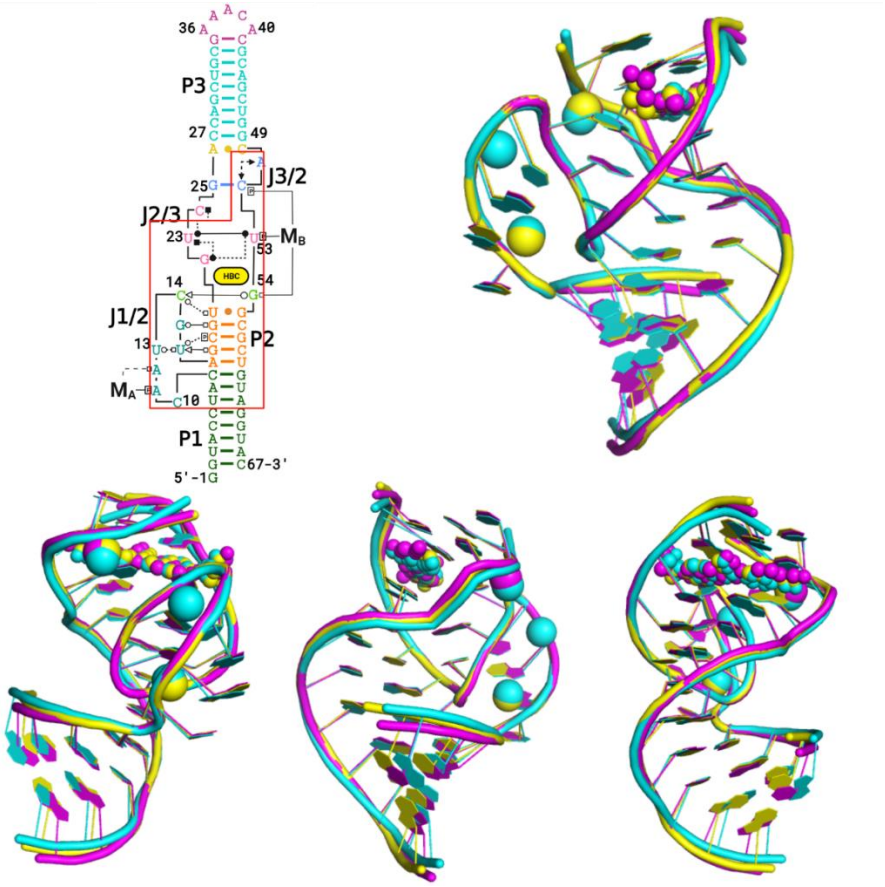


Figure 4.1: RMSD comparison of Pepper structures. Partial overlay Fab-RNA Pepper from this work (7SZU yellow, 7U0Y magenta) and the RNA-only crystal structure by Huang *et. al.*[1] (7EOH, cyan), with the structure turned 90° anti-clockwise from left to right. Tertiary structure diagram with red outline showing the included nucleotides that best match between structures. RMSD between 7SZU and 7U0Y; 0.274 Å^2. 7SZU and 7EOH; 0.370 Å^2. 7U0Y and 7EOH; 0.428 Å^2. Rmsd calculations and 3D models generated in pymol. Created in Biorender™.

The main differences include the angle subtended by P2 and P3, which is a few degrees more acute in cp structure, and the number of metal ions visible. This different angle causes the rmsd comparison to increase sharply as the Pepper structures deviate from each other, but does imply that this angle between helices is not fixed and may be flexible in a functional Pepper-ligand complex. The cp structure contains five Mg ions (M1-M5). In contrast, our HBC530 and HBC599 structures contain two metal ions, M_A and M_B that correspond directly to M1 and M3 in the cp structure (Figure 4.2). A third metal ion unique to HBC530 does not correspond to any metal ion

in the cp structure. M4 and M5 in the cp structure reside distal to the ligand binding site within the P1 helix (P3 in our construct) and likely are unnecessary for function as shown by the hill coefficients (see chapter 3). Our structures appear to lack these two metal ions but the corresponding helix in our construct (P3) contains an entirely different sequence identical to the original aptamer [2], and also potentially due to the extra charge stabilization requirements of RNA-only crystal packing.

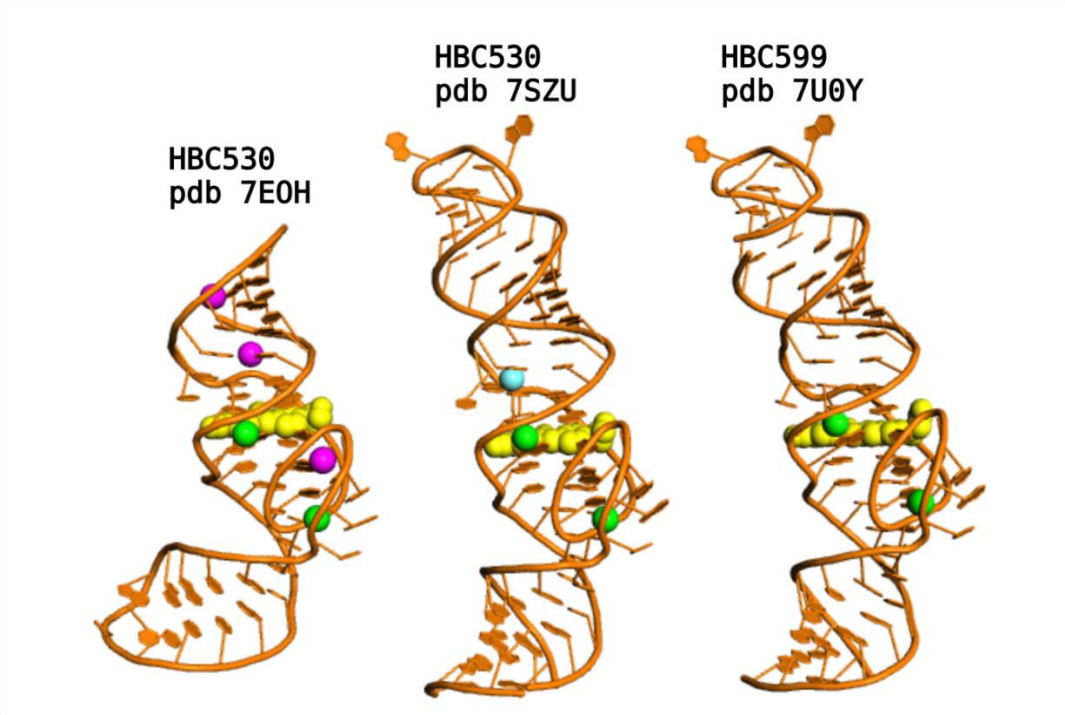


Figure 4.2: Mg^{2+} assignment comparison of Pepper structures. Comparison of the Mg^{2+} assigned in this work (7SZU, 7U0Y) to the currently existing Pepper HBC530 structure from Huang *et. al.* (pdb 7EOH)[1]. The pre-existing Pepper structure has five assigned Mg^{2+} (magenta), two of which correlate to assigned Mg^{2+} visible in the Pepper-BL3-6 with HBC599 structures (green). Mg1 and Mg3 in their previous work correlate to Mg1 and Mg2 in this publication. The third Mg^{2+} (cyan) seen in Pepper-BL3-6-HBC530 is not equivalent to any of the Mg^{2+} seen in the previous Pepper structures. 3D models generated in pymol.

Most likely these modest differences between the structures reflect the distinct lattice interactions (Figure 4.3). The cp RNA crystal lattice buries significantly more of the total RNA

surface area compared to the Fab-RNA crystal lattice (8.6% vs 2.6%). Lattice contacts in the cpRNA crystals (732.9 Å^2) involve exclusively interactions with other RNA molecules that can strongly influence the electrostatic environment due to RNAs negative charge In contrast, among the lattice contacts in the Fab-RNA crystals, only 18 % (286.6 Å^2) involve interactions with other RNA molecules.

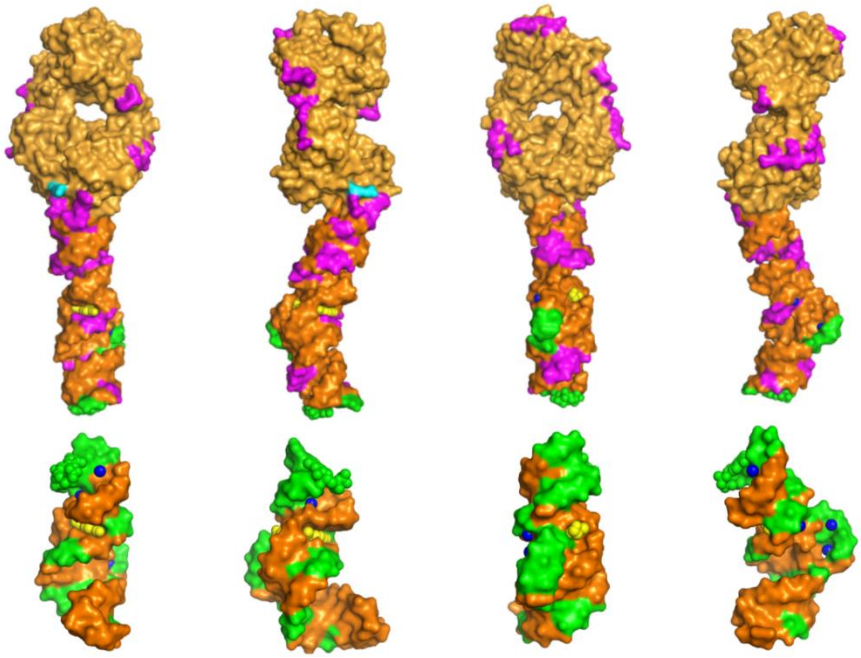


Figure 4.3: Crystal packing comparison of Pepper structures. Crystal packing contacts of Fab-RNA Pepper from this thesis (top row, 7SZU) and the RNA-only crystal structure by Huang *et. al.*[1] (bottom row, 7EOH), with the structure turned 90^o anti-clockwise from left to right. RNA-RNA contacting residues (green) dominate the RNA-only crystal lattice. Fab-RNA contacts (magenta) dominate the Fab-RNA crystal lattice, and there are minimal Fab-Fab (cyan) contacts. Models generated in pymol.

The Fab-RNA crystal is clearer on which parts of the RNA are flexible which may be more useful for making comparisons to the solution structure. While the visible differences between the RNA-only crystals and the Fab-RNA crystals do not appear to substantially impact the mode of ligand binding, the differences in uncertainty and dynamics in the crystals is informative. The high-resolution of the RNA-only crystals decisively answers questions of ligand orientation, which was not absolute in the RNA-Fab structures. Looking at temperature-factor diagrams of the Fab-RNA

crystals and RNA-only crystals also shows agreement in the dynamics of the RNA, with P1 in a highly dynamic state compared to P3 even when it is not harbouring the ends of the RNA (Figure 4.4), and the relative dynamism of C24. This difference in flexibility could be due to the crystal packing, where fewer RNA-RNA contacts, and the higher number of divalent ions needed for RNA-RNA crystals, leading to a less restricted environment which better represented RNA flexibility.

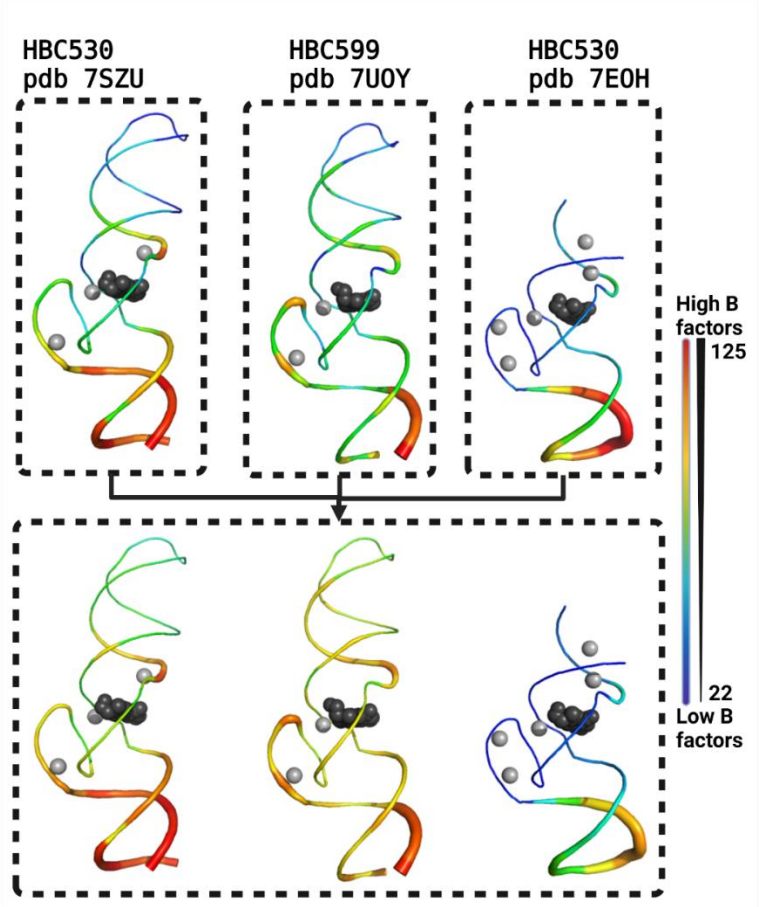


Figure 4.4: Temperature-factor comparison of Pepper structures. Pymol-generated B-factor putty temperature cartoons of structures from this work compared to (7EOH)[1]. Thickness and color both correspond to B-factors. Assigned Mg^{2+} and Ligands shown in gray. Top: temperature scale relative to the individual structure. Bottom: Temperature scale relative to all three structures. Figure generated in Biorender™.

4.3 Proposed utilization for Pepper in aptazymes frameworks

Because the fluorescence of Pepper appears to be coupled to the structuring of J1/2 aided by Mg^{2+} binding and the bulge caused by J2/3 and J3/2, we suggest that Pepper has potential to serve as a framework for development of fluorogenic sensors for a variety of substrates. Pepper has the potential to be coupled to allosteric RNAs through its P1 or P3 stems as communication modules [3], similar to how the SAM riboswitch was engineered into the Corn aptamer to couple DFHO fluorescence to SAM binding [4], or Spinach to detect thiamine 5′-pyrophosphate binding [5], along with many others [6,7]. As Pepper is not reliant on a G-quadruplex and has tight K_ds, Pepper could perform significantly better than Spinach and other HBI-derivative aptamers, with a compact and simple tertiary structure that is more tolerated in cells than the G-quadruplexes [8]. With the wide variety of excitation wavelengths available to Pepper through its HBC ligands, multiple reporting systems can be set up in close proximity that sense different RNAs, allowing for more complex and detailed cell visualization.

As described in chapters 2 and 3, J1/2 is structured seemingly by one Mg^{2+} ion and has a unique contribution to the binding of HBC which opens up more difficult but interesting engineering pathways for Pepper. Because J1/2's purpose appears to be to be able to contribute a cysteine to the HBC binding site, modifications to J1/2's structure that still allow this tertiary interaction to form could be tolerated. Engineering J1/2 to bind other metals or a small molecule to become structured to could open up more complex aptazyme communication modules that don't rely on helix formation.

Finally, Pepper is able to bind a wide variety of derivatives with a tight Kd, and crystal structures have demonstrated that this is predominantly through stacking with the aromatic ring systems [1]. Ligand promiscuity has been seen in Mango[9] and DIR2 [10] aptamers, but the structure of

Pepper with HBC has shown opportunities at both ends of the HBC ligand to engineer linkers to larger components such as proteins or larger organic molecules. Therefore there may be potential in engineering Pepper to bind HBC derivatives with even tighter affinities – even at the cost of photophysical properties – to design a biosensor that will bind and report an HBC tag that can be connected to other moieties. While this has been potentially viable for previous fluorogenic aptamers, HBC appears to be much more ideally suited now that the Pepper-HBC structure has shown the tails to be sterically unhindered.

4.4 Conclusions and implications of the Pepper structure

In this thesis, the structure of the Pepper aptamer has been elucidated via chaperone assisted crystallography. The Pepper structure is reliant on a base-quadruple and base-triple stack that sandwiches the HBC ligand with another helical stack, with supporting out-of-plane tertiary interactions forming a binding pocket, and a hydrogen bond between the ligand and RNA. It is most likely that the structure is thermodynamically facilitated by specific divalent binding sites that are essential to HBC binding, and may contribute to the tight affinity of the aptamer. Chaperone-assisted crystallography has been a crucial tool for 3D RNA structure studies, and continues to be so. The Pepper structure with Fab BL3-6 serves as a great example of how the Fabs can stabilize a crystal lattice, and reduce the need for excessive cations for charge neutralization. The structures of Pepper have shown how much improvement has been made to fluorogenic aptamers over the years, and serve as a new benchmark for allosteric RNA engineers.

Table 4.1: Tabulated data of fluorogenic aptamers with pdb entries with descriptions of the binding sites. Pepper does not utilize a G-quadruplex, and one of the stacks is a WCF helix topped by a G•U wobble pair, which is simpler than other aptamers shown here.

pdb	Aptamer	H-bonding/ other	Stacking 1.3	Stacking 1.2	Stacking 1.1	Fluorogen	Stacking 2.1	Stacking 2.2	Stacking 2.3	Side (out of plane)
1F1T	**MGA**	None	U11•A22•A26 triple	C10•G23•A27 triple	G8•C28 pair	Malachite Green	C7•G24•G29 A31 quadruple	C6•G33 pair	C5•G34 pair	U25, A30•A9
4KZD	**Spinach**	G28•H-bond keto	C25•G63 pair	G22•G26•G57•G61 QPx	G23•G27•G54•G59 QPx	DFHBI	U29•U50•A53 triple	C30•G52 pair	C31•G49 pair	A58
5V3F	**Mango**	None	G8•G13•G18•G23 QPx	G9•G14•G19 G24•QPx	G10•G16•G21•G26 QPx	TO1-Biotin	U15, A20, A25 (also side)	None	None	U15, A20, A25
6DB8	**DIR2s-Apt**	G39•SO3 moiety H-bond	G37, A13 stack individually	A14•U38•U42 triple	A15•G39•A41 triple	OTB	A40 stacking	None	None	None
7OAX	**Chilli**	In plane G15 H-bond (C40 pair) (and K+)	G10•A11•C44 triple	G9•G12•G36•G42 QPx	G13•G32•G37•G41 QPx	DMHBO+	C16•G39, G14•G31	G17•U30 pair	C18•G29 pair	None
6E81	**Corn**	Dimerizes, A11 H-bonds to imine	A10•U17•A21•U27 QPx	G13•G16•G23•G26 QPx	G12•G15•G22•G25 QPx	DFHO	Dimer	Dimer	Dimer	A11, A24
7KVV	**Squash**	A69, A70 H-bond phenolate	G42•G46•G71•G73 quadruple	A12•G41•G47•C72 quadruple	G13•A40•A70 triple	DFHO	U15•A69	G16•A49 pair	A17•U68 pair	G14, A48 (U39 pair)
7SZU	**Pepper**	G22 H-bond ethyl alcohol	A26•C50 pair	G25•A51•C52 triple	G22•U23•C24•U53 quadruple	HBC530	U21•G55 pair	G20•C56 pair	C19•G57 pair	C14•G54

Pepper is unusual amongst fluorogenic aptamers not only for its unique ligand scope, but also its stacking arrangement and specific Mg^{2+} binding. All known fluorogenic aptamers function by folding to create a binding pocket for organic chromophores bearing planar aromatic pi-systems. The binding pockets consist of planar nucleobase layers that sandwich the ligand through stacking interactions and in most cases, one or more out-of-plane nucleobases that wall off one side of the binding pocket (Table 4.1). The out-of-plane system can involve up to three nucleotides, as seen in the malachite green aptamer[11] and Spinach[12,13]. Various combinations of

nucleobases comprise the two stacking layers and fluorogenic aptamers fall broadly into two categories. Those that use a G-quadruplex fold to form one of the stacking layers, including Spinach[12,13], Mango[9,14,15], Chili[16], and Corn[16], and those that do not, including the Malachite Green aptamer[11], DIR2[11], Squash[11], and Pepper[11]. The latter category of aptamers contain a base triple or base quadruple. The crystal structures of the G-quadruplex class all contain bound K^+ to support the quadruplex fold. Within the other aptamer class, the crystal structures of Squash and Pepper contain site-bound Mg^{2+}. This leaves DIR2 and Malachite Green as the only aptamers with crystal structures to not exhibit a visible metal that could be critical for binding.

Pepper's unique set of properties is in agreement with its unique structure and requirements, and its simple tertiary fold in comparison to other aptamers is an advantage for in vivo utilization of the aptamer. Pepper is distinct in its lack of a G-quadruplex and is reliant on Mg^{2+}, making it only comparable to Squash [17]. Squash was evolved from an adenosine riboswitch to access an In vivo compatible fold, but at the cost of a lower K_d (54 nM for DFHO-Squash vs 3.5 nM HBC-Pepper) and a much lower brightness (see chapter 1, Table 1.1). Pepper's structural features contribute to the aptamer with the most desirable set of parameters seen so far, and Engineering future aptamers to favor structural features seen in Pepper, such as the out-of-plane nucleotides, would be ideal for those seeking to make new ligand-RNA complexes for biosensing. For future aptamers, evolution in low K^+ and Mg^{2+} could select for aptamers with similar properties to Pepper without the salt dependence. The main downside of Pepper is that it still has a high Mg^{2+} requirement for function – indeed this was a problem in the original selection [18] – and finding aptamers with similar performance with a lower reliance on divalents could lead to even better fluorogenic RNAs to be engineered in the future.

4.5 References for Chapter 4

1. Huang, K.; Chen, X.; Li, C.; Song, Q.; Li, H.; Zhu, L.; Yang, Y.; Ren, A. Structure-based investigation of fluorogenic Pepper aptamer. *Nature Chemical Biology* **17**, 1289–1295 (2021). [PMID:34725509]

2. Chen, X.; Zhang, D.; Su, N.; Bao, B.; Xie, X.; Zuo, F.; Yang, L.; Wang, H.; Jiang, L.; Lin, Q.; Fang, M.; Li, N.; Hua, X.; Chen, Z.; Bao, C.; Xu, J.; Du, W.; Zhang, L.; Zhao, Y.; Zhu, L.; Loscalzo, J.; Yang, Y. Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs. *Nature Biotechnology* **37**, 1287–1293 (2019). [PMID:31548726]

3. Zhong, G.; Wang, H.; Bailey, C. C.; Gao, G.; Farzan, M. Rational design of aptazyme riboswitches for efficient control of gene expression in mammalian cells. *eLife* **5**, (2016). [PMID:27805569]

4. Kim, H.; Jaffrey, S. R. A Fluorogenic RNA-Based Sensor Activated by Metabolite-Induced RNA Dimerization. *Cell Chemical Biology* **26**, 1725-1731.e6 (2019). [PMID:31631009]

5. You, M.; Litke, J. L.; Jaffrey, S. R. Imaging metabolite dynamics in living cells using a Spinach-based riboswitch. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E2756–E2765 (2015). [PMID:25964329]

6. Porter, E. B.; Polaski, J. T.; Morck, M. M.; Batey, R. T. Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nature Chemical Biology* **13**, 295–301 (2017). [PMID:28092358]

7. Su, Y.; Hammond, M. C. RNA-based fluorescent biosensors for live cell imaging of small molecules and RNAs. *Current Opinion in Biotechnology* **63**, 157–166 (2020).

[PMID:32086101]

8. Guo, J. U.; Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353**, 1–23 (2016).

9. Trachman III, R. J.; Demeshkina, N. A.; Lau, M. W. L.; Panchapakesan, S. S. S.; Unrau, P. J.; Ferré-D'Amaré, A. R. Structural basis for high-affinity fluorophore binding and activation by RNA Mango. *Physiology & behavior* **13**, 139–148 (2017).

10. Shelke, S. A.; Shao, Y.; Laski, A.; Koirala, D.; Weissman, B. P.; Fuller, J. R.; Tan, X.; Constantin, T. P.; Waggoner, A. S.; Bruchez, M. P.; Armitage, B. A.; Piccirilli, J. A. Structural basis for activation of fluorogenic dyes by an RNA aptamer lacking a G-quadruplex motif. *Nature Communications* **9**, 1–10 (2018).

11. Baugh, C.; Grate, D.; Wilson, C. 2.8 Å Crystal structure of the malachite green aptamer. *Journal of Molecular Biology* **301**, 117–128 (2000). [PMID:10926496]

12. Huang, H.; Suslov, N. B.; Li, N. S.; Shelke, S. A.; Evans, M. E.; Koldobskaya, Y.; Rice, P. A.; Piccirilli, J. A. A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nature Chemical Biology* **10**, 686–691 (2014). [PMID:24952597]

13. Warner, K. D.; Chen, M. C.; Song, W.; Strack, R. L.; Thorn, A.; Jaffrey, S. R.; Ferré-D'Amaré, A. R. Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nature Structural & Molecular Biology* **21**, 658–663 (2014). [PMID:25026079]

14. Trachman, R. J.; Cojocaru, R.; Wu, D.; Piszczek, G.; Ryckelynck, M.; Unrau, P. J.; Ferré-D'Amaré, A. R. Structure-Guided Engineering of the Homodimeric Mango-IV

Fluorescence Turn-on Aptamer Yields an RNA FRET Pair. *Structure* **28**, 776-785.e3 (2020). [PMID:32386573]

15. Trachman, R. J. I.; Autour, A.; Jeng, S. C. Y.; Abdolahzadeh, A.; Andreoni, A.; Cojocaru, R.; Garipov, R.; Dolgosheina, E. V.; Knutson, J. R.; Ryckelynck, M.; Unrau, P. J.; Ferré-D'Amaré, A. R. Structure and functional reselection of the Mango-III fluorogenic RNA aptamer. *Nature Chemical Biology* **(Accepted)**, (2019).

16. Mieczkowski, M.; Steinmetzger, C.; Bessi, I.; Lenz, A. K.; Schmiedel, A.; Holzapfel, M.; Lambert, C.; Pena, V.; Höbartner, C. Large Stokes shift fluorescence activation in an RNA aptamer by intermolecular proton transfer to guanine. *Nature Communications* **12**, 1–11 (2021). [PMID:34112799]

17. Dey, S. K.; Filonov, G. S.; Olarerin-George, A. O.; Jackson, B. T.; Finley, L. W. S.; Jaffrey, S. R. Repurposing an adenine riboswitch into a fluorogenic imaging and sensing tag. *Nature Chemical Biology* **18**, (2021). [PMID:34937909]

18. Tan, X.; Constantin, T. P.; Sloane, K. L.; Waggoner, A. S.; Bruchez, M. P.; Armitage, B. A. Fluoromodules Consisting of a Promiscuous RNA Aptamer and Red or Blue Fluorogenic Cyanine Dyes: Selection, Characterization, and Bioimaging. *Journal of the American Chemical Society* **139**, 9001–9009 (2017). [PMID:28644615]