THE UNIVERSITY OF CHICAGO

RNA CATALYSIS:

STRUCTURE, FUNCTION AND EVOLUTION

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

SAURJA DASGUPTA

CHICAGO, ILLINOIS

AUGUST 2018

To

Reason

and

Imperfection

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2AP	2-aminopurine
A	Adenine
AMP	Adenosine monophosphate
AMPA	α-Amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
Asp	Aspartate
ATP	Adenosine triphosphate
bp	Base-pair
bya	Billion years ago
C	Cytidine
cDNA	Complementary DNA
CLEC2	C-type lectin family 2
CMCT	1-cycohexyl-3-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate
CPEB3	Cytoplasmic polyadenylation element binding protein 3
Cryo-EM	Cryo-electron microscopy
DEPC	Diethyl pyrocarbonate
DMS	Dimethyl sulfate
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
DNAzyme	DNA enzyme
dPAGE	Denaturing polyacrylamide gel electrophoresis
DTT	Dithiothreitol
EF-Tu	Elongation factor-thermo-unstable

EMSA	Electrophoretic mobility shift assay
Fab	Fragment of antigen binding
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
FPLC	Fast protein liquid chromatography
FRET	Förster resonance energy transfer
G	Guanosine
GlcN6P	Glucosamine-6-phosphate
GMP	Guanosine monophosphate
GTP	Guanosine triphosphate
H-bond	Hydrogen bond
HDV	Hepatitis delta virus
HH	Hammerhead
His	Histidine
HIV	Human immuno-deficiency virus
HP	Hairpin
HPLC	High performance liquid chromatography
HTS	High-throughput sequencing
KE	Kethoxal
LUCA	Last universal common ancestor
Lys	Lysine
M.W.	Molecular weight
mfe	Most favorable energy

miRNA	Micro RNA
mRNA	Messenger RNA
NAIM	Nucleotide analog interference mapping
nbO	Non-bridging oxygen
NMR	Nuclear magnetic resonance
nSEC	Native size-exclusion chromatography
nt	Nucleotides
NTP	Nucleoside triphosphate
P/C/I	Phenol/chloroform/isoamyl alcohol
PCR	Polymerase chain reaction
PDB	Protein data bank
PNK	Polynucleotide kinase
PRE	Paramagnetic relaxation enhancement
pRNA	Promoter-associated RNA
pyrroloC	Pyrrolo-cytosine
QM/MM	Quantum mechanics/Molecular mechanics
RNA	Ribonucleic acid
RNase	Ribonuclease
rRNA	Ribososmal RNA
RT	Reverse transcription/ Reverse transcriptase
SAD	Single anomalous dispersion
SASBDB	Small angle scattering biological databank
SAXS	Small-angle x-ray scattering

Ser	Serine
siRNA	Small-interfering RNA
S_N2	Substitution nucleophilic type-2
snRNA	Small nucleolar RNA
TEN	Tris-EDTA-NaCl
TiPPase	Thermostable inorganic pyrophosphatase
TPP	Thiamine pyrophosphate
tRNA	Transfer RNA
TS	Twister-sister
Tyr	Tyrosine
U	Uridine
VS	Varkud satellite
WC	Watson-Crick

ACKNOWLEDGEMENTS

My PhD research involved daily interactions with inanimate systems cleverly designed to do my bidding. But in that process, I have realized how deeply human the scientific endeavor really is. While on countless occasions I have felt fortunate (without knowing what that means) to get an experiment to work, I feel I should take this opportunity to thank some of the people who I feel fortunate to have crossed paths with during my tenure as a PhD student.

I begin by thanking my advisor Dr. Joseph Piccirilli for giving me the opportunity to work in a dynamic and friendly research group and creating a pleasant and accepting environment. Working closely with him has taught me to look at scientific results with nuance and respect the gaps in our understanding of any scientific problem. Joe has also taught me the importance of clear and precise communication, and I treasure the training I have received while writing papers with him. Joe has allowed complete intellectual independence and has been extremely patient with me on multiple occasions when I was lost amidst the unsightly debris of failed projects. It is due to his encouragement to pursue what is interesting and take full ownership of it, did I succeed in exploring different areas of research that sometimes led me to uncharted territories. I also thank Joe for affording me the opportunity to collaborate with leaders in their fields and expand my technical and intellectual abilities. Scientific contributions aside, Joe has often revealed a caring and personable side that has made working with him an even more pleasant experience. I will take this sentiment with me to any position of responsibility I might encounter in future.

I thank my candidacy committee (Dr. Stephen Kent, Dr. Chuan He and Dr. Sergey Kozmin) and thesis committee (Dr. Bryan Dickinson and Dr. Phoebe Rice) for their time and useful suggestions. I wish to thank Dr. Phoebe Rice for being a fantastic collaborator, ever-ready to pass on her expertise to fledgling structural biologists like myself. Over the years she has been the go-to person when stuck with any issue related to structure phasing or refinement. I also thank her for her valuable suggestions, which greatly improved my dissertation. Every conversation with her has imbued me with a little more confidence and a lot more excitement toward science.

I express my earnest gratitude to Melinda Moore (Chemistry) for taking care of the administrative side of things so that I could write my papers instead of worrying about additional paperwork. I also wish to thank Julien Owens, Theresa San Juan (BMB) and Jessica Krupinski (Fisher Gordon stockroom) for assisting me in ordering and billing-related issues.

The Piccirilli lab has been a wonderful place to work in. I thank members of the Piccirilli lab, past and present, whom I have had the opportunity to interact with. I consider Dr. Nikolai Suslov and Dr. Nicole Tuttle to be my mentors in a lot of ways, even if we overlapped for only a few months. Nikolai took a hapless first year graduate student under his wing and introduced him to the VS ribozyme and the wonderful world of RNA catalysis. The contents of this thesis would likely be very different if it weren't for your tiny act of kindness. I remember the first chat with Nicole at the bookstore, where she entertained my naiveté about scientific research with utmost concern. I'm glad I have been able to keep in touch with both of you and shared much more than science over the years. I remember late-night chats with Dr. Armando Hernendez when we discussed religion, politics, music and of course science. I thank Dr. Karthik Sundaram for always reminding me that "science never sleeps" (and my rebuttal- "but scientists do"), Dr. Hao Huang for introducing me to crystallography software suites and for solving the crystal structure of the Spinach RNA aptamer, which inspired the work discussed in the appendix of this thesis (and sharing the love for the band 'Muse'). I wish to thank Dr. Sandip Shelke for our reflective chats over coffee and for showing me the operations of the fluorimeter, Dr. Deepak Koirala for training me to use the circular dichroism spectrometer, Benjamin Weissman for his advice on running ribozyme kinetics experiments, and being the only person, other than Joe, who shared my excitement about RNA catalysis in the lab. I thank Dr. Nan-sheng Li for catering to the synthetic needs of my projects and Dr. Yaming Shao for sharing technical knowledge in structure solving. I am grateful to have known Dr. Shabana Shaik as a co-worker and as a friend. You are a shining embodiment of perseverance and fortitude. I cherish our candid conversations and thank you for making sure I ate well during the busiest months near the end.

I greatly appreciate help from Dr. Carter Abney (ICP-MS), Dr. Andrew Sandstorm (crystallography theory) and Dr. James Fuller (SAXS of the VS ribozyme). I am indebted to the generosity of the members

of the Ruthenberg, Perozo, Adams, Kossiakoff, Kent and Krishnan labs for sharing instruments and reagents in times of urgent need. I must acknowledge the contributions of the NE-CAT beamline staff at the Advanced Photon Source (APS), Argonne National Laboratory, for their relentless support during data collection and especially Dr. Surajit Banerjee for his help with data processing and accompanying me in my strolls around the APS, adding much needed levity to the otherwise long and tiring hours in front of the beamline controls.

I thank the custodial staff for keeping our lab space as pristine as possible and especially Gayle Smith for making sure that we have clean glassware at the start of every work day and Josephina Luna for watering my plant when I forgot to do so. I will always remember Josephina's smile when she saw me leave earlier than usual and her caring – "take it easy, okay?"

Going back sixteen years, I remember being a teenager who was 'thinking about getting a PhD in Chemistry' without really knowing what that meant. Achintya Kumar Biswas, the eight-grade chemistry teacher had fed the flames instead of dismissing my wide-eyed fantasy. To you, I express my deepest gratitude. I would also like to thank Dr. Dipankar Sen for introducing me to the wonders of nucleic acids way back in 2009 and Dr. Pradeepkumar P. I., for supervising my Master's thesis at IIT Bombay, further cultivating my interest in nucleic acids. My PhD work owes a lot to these experiences. I thank my acquaintances from the 'Science and Faith' discussion group for creating a healthy atmosphere of lively debate on some of the most fundamental questions of existence over wonderful tea and snacks.

I am extremely fortunate to have friends who have made my PhD years, the best years of my life. Together we have spent hours chatting incessantly, explored the cultural scene of Chicago, visited over twenty states, staged three plays, jammed all night, wrote new music and shared bits of our personal lives with each other. A huge shout out to you. I cannot forget my students; you were my first friends in Chicago and in the US. I wish you the best.

My fiancée, soon-to-be Dr. Urmimala Basu, has been a part of my academic journey for over a decade, and has witnessed me mature socially and intellectually. We have tried to interpret each other's experimental data, proofread each other's writing, refined each other's arguments and supported each other

during occasional bursts of despondency. It has been quite an experience surfing the waves with you, and I am looking forward to more of it.

It is impossible to appreciate the role of your family in shaping who you are, till you start living away from them. My parents have been my biggest supporters and given me complete autonomy to choose my adventures. A lot of our future is determined by what we decide to imitate during the early years of our development. I think watching my father surrounded by law books and my mother's penchant for literature had primed me for this attempt at an academic career. Your sacrifices have shaped my life in ways I cannot begin to imagine. Thank you is all I have got.

Finally, I'd like to acknowledge the influence of two individuals who I have never met but are a big part of my intellectual life: Richard Dawkins for being a gladiator of reason and his lifelong fight to raise scientific consciousness among the masses, and the late Leonard Cohen for opening my eyes to the limitations and imperfections that define everything we do. Well perhaps…

The reason for imperfection…

is the imperfection of reason.

ABSTRACT

Catalytic RNA or ribozymes present excellent systems to study the foundational principles of biological catalysis. In addition, they serve as models for investigating RNA structure and its relation to function. We report the first crystal structures of the Varkud satellite (VS) ribozyme, the largest among endonucleolytic ribozymes. The structures reveal a modular organization, in which independently folding domains are assembled into the functional conformation of the ribozyme by three-way junctions. The catalytic domain of the ribozyme recognizes and binds the substrate through tertiary interactions, and substrate docking is accompanied by a remodeling of the substrate structure that results in the formation of a catalytically-relevant active site. The catalytic strategies employed by the VS bear resemblance to that used by protein and DNAzyme ribonucleases, highlighting the robustness of the chemistry behind catalytic RNA cleavage. In addition, the active site of the VS ribozyme is strikingly similar to that of the hairpin and hammerhead ribozymes, although the three endonucleolytic ribozymes have distinct sequences and structures. The presence of these architectural features in the context of what appears to be distinct mechanisms of catalysis underscores their functional importance and bolsters the case for convergent evolution.

However, our understanding of the possible mechanisms for the emergence of distinct endonucleolytic ribozyme function during evolution and the ease of access to distinct catalytic motifs involved in RNA cleavage is limited. We have explored the mutational connections between the VS, hairpin and hammerhead ribozymes and delineated plausible pathways by which these distinct ribonuclease motifs can be accessed via intersection of their neutral networks. Intersections between neutral networks are possible due to the existence of bifunctional sequences that exhibit catalytic functions corresponding to both networks. We have identified two such

bifunctional sequences that can support hairpin and VS, and hammerhead and hairpin dual functions. Bifunctional sequences present plausible evolutionary nodes toward increasing complexity in functional RNA as illustrated by the hammerhead, hairpin and VS ribozymes in our study. Our results provide a framework to investigate the evolutionary origins of distinct catalytic function in RNA.

Chapter - 1

INTRODUCTION

1.1 Enzymes: A brief introduction

Life is a struggle for energy and time. Life's uphill battle with energy manifests itself in a multitude of biochemical reactions that are thermodynamically unfeasible in the absence of additional help. Arguably, the most important reaction in the cell –formation of peptide bonds by stitching together amino acids to create the primary functional currency of life, proteins, does not occur spontaneously (Watson et al., 1987). To overcome this struggle, biology couples exergonic reactions usually involving hydrolysis of high energy bonds, with endergonic ones like peptide bond formation, to drive these unfavorable reactions forward. For biochemical transformations that are thermodynamically feasible, there is still no free lunch. Most reactions are so sluggish that it might appear that they don't take place at all – a situation hardly distinguishable from an energetically untenable reaction. The reverse reaction of peptide formation, peptide hydrolysis is energetically allowed; however, its half-life extends to hundreds of years (Radzicka and Wolfenden, 1996). This might seem useful as we need stable proteins to keep most of life together but certain circumstances demand rapid breakdown of proteins (e.g. during digestion) therefore, the fight is primarily against time.

For this purpose, we need biological catalysts or enzymes, which speed up biochemical reactions by leading them through alternate paths with lower energy barriers (Figure 1.1) (Stryer et al., 2002).

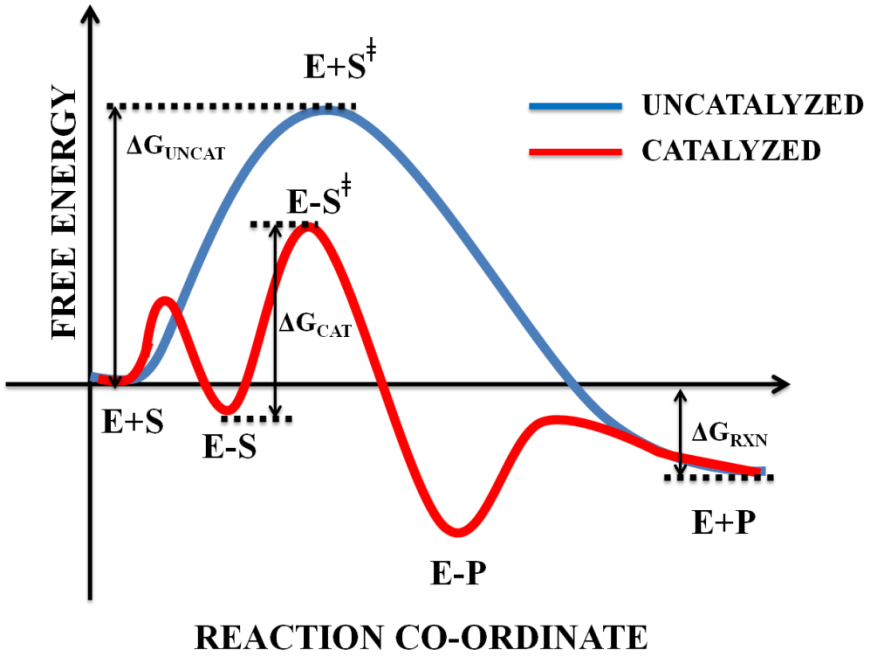

Figure 1.1 Enzyme catalysis. Enzymes lower the reaction activation energy ($\Delta G_{UNCAT} \rightarrow \Delta G_{CAT}$) by changing the reaction pathway to one with a low energy transition state.

Enzymes catalyze more than 3500 different types of biochemical transformations (Lodish et al., 2003) and an extremely overwhelming majority of those enzymes are proteins. Protein enzymes are versatile molecules that support phenomenal rate accelerations (e.g. 1.4 X 10^{17}-fold by OMP decarboxylase) (Radzicka and Wolfenden, 1995) and excellent turnover numbers (600,000 per second by carbonic anhydrase) (Lindskog, 1987). One of the pioneers of enzymology, William Jencks pointed out quite accurately, "One of the revolutions of this century is the development of knowledge of catalysis by enzymes to the point where it can be examined in ordinary chemical terms. There is optimism that chemistry can explain enzymatic catalysis without requiring vitalistic or mystical hypotheses." (Jencks, 1987). We have come a long way in the thirty years since this pronouncement, in terms of understanding the molecular mechanisms, energetic implications and biological significance of enzyme function. In very simplistic form, enzyme function involves four kinetically distinguishable steps (Stryer *et al.*, 2002) (Figure 1.2):

1. *Folding*. Macromolecules are generally flexible. A single molecule can adopt multiple conformations in three dimensions, most of which are non-functional. Therefore, structural features intrinsic to enzyme molecules stabilize the catalytically active fold, creating energy minima where the active molecules can rest at. Most of these structural features involve weak interactions like hydrogen bonds and hydrophobic interactions (van der Waals interactions, stacking, etc.). The presence of these non-covalent forces in enzyme catalysis distinguishes it from its chemical counterpart (Jencks, 1987).

2. *Substrate recognition/binding*. Complementarity in structural features between a catalytically competent enzyme and its substrate(s) enables tight binding with strong specificity to form the enzyme-substrate complex. Binding between enzyme and substrate is facilitated by the aforementioned weak interactions, which supply the entropic cost of intimate association (Jencks, 1987). It has been suggested that enzymes have a greater affinity/complementarity to a configuration that is intermediate to the reactants and products of that reaction pathway – the 'activated complex', occupying the reaction transition state. This increased affinity lowers the activation barrier of the reaction leading to catalysis.

3. *Chemical step of catalysis*. Once bound to the enzyme, the substrate is housed in the active site, a three-dimensional cleft generally occupying a small segment of the enzyme molecule where the chemical step of catalysis takes place. Neighboring residues within the active site proximal to the substrate participate in chemistry, which generally involves proton transfer, underscoring the importance of acid-base properties of catalytic residues in the active site. The catalytic cleft also brings into close proximity, two or more substrates under specific reaction conditions.

4. *Product release.* Following catalysis, the affinity of the enzyme toward the 'processed substrate' or product decreases as a consequence of an altered field of interactions between them, resulting in product release. This enables the enzyme to carry out further rounds of catalysis (i.e. multi-turnover catalysis).

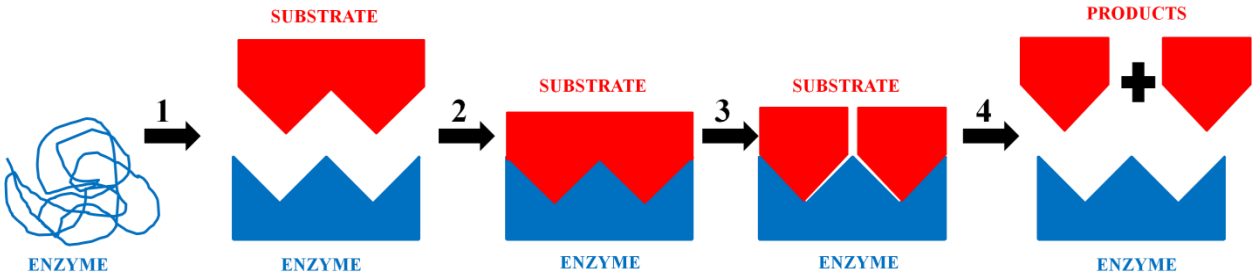

Figure 1.2 Steps in enzyme catalysis. Enzyme catalysis involves four fundamental steps: (1) enzyme folding, (2) substrate binding, (3) catalysis and (4) product release.

We have achieved an in-depth understanding about each of the four steps by studying the global structures of enzymes using a wide spectrum of biophysical techniques, of which nuclear magnetic resonance (NMR), x-ray crystallography and cryo-electron microscopy (cryo-EM) have had the largest impact (rcsb.org) These coupled with *in vitro* biochemistry have unraveled the chemical processes occurring in the active sites of these enzymes. Investigations into protein enzymes elicited a powerful message central to biocatalysis – enzymes possess well-defined, roughly globular shapes that employ a chemically diverse toolkit of residues with varying pH sensitivities, electrostatics and hydrophobicity/hydrophilicity. This led to a rather self-fulfilling prophecy: proteins are the only biomolecule that appeared to satisfy all these criteria. It is interesting to note that proteins were put on a similar pedestal of being the only biomolecule complex enough to carry genetic information in the early 1900s. However, a rather pedestrian molecule found in the nucleus, DNA staked its claim as the hereditary entity in biology through

sets of ingenious experiments by Fred Griffith in 1928 (Griffith, 1928) and Avery-MacLeod-McCarty in 1944 (Avery et al., 1944). Decades later, in a historical déjà-vu, another simple molecule considered fairly uninteresting at that time and structurally similar to DNA was found to possess the capacity to fold into compact particles - a prerequisite for enzyme function; this molecule was RNA. It was demonstrated by numerous biophysical techniques that ribosomal RNA (rRNA), a type of RNA instrumental in protein synthesis, assumed a distinct three-dimensional structure (Vasiliev et al., 1978; Vasiliev et al., 1986). This was not obvious as it may seem today, as the uniformly negative charge on the phosphodiester backbone of RNA would appear to hinder the formation of compact folds. The ground-breaking, first ever high-resolution structure of RNA, that of transfer RNA (tRNA), was solved by Alex Rich and his group in 1973 (Kim et al., 1973). This high-resolution structure firmly established that RNA could adopt a compact folded structure with clefts and cavities much like proteins. RNA was not just an elongated tape of nucleotides, but a 'tape with a shape' (Yarus, 2010). The implications of this discovery were captured in a portentous statement by Francis Crick – "tRNA looks like Nature's attempt to make RNA do the job of a protein" (Crick, 1966). However, even before this structural revelation, with the discovery of mRNA and its role in transferring genetic messages (Brenner et al., 1961; Gros et al., 1961; Jacob and Monod, 1961) to the message-decoding center in the ribosome (consisting of rRNA and proteins), via the involvement of an adaptor molecule, tRNA (Crick, 1958), an obvious challenge to the central dogma: 'genetic message residing within DNA gets copied to RNA which is translated by the ribosome to make proteins' (Figure 1.3) appeared in the form of a 'chicken or egg' paradox. Protein enzymes called polymerases are required to make DNA and RNA from parent DNA, (or RNA) but RNA in the form of mRNA, tRNA and rRNA is used to make proteins. Which biopolymer held primacy?

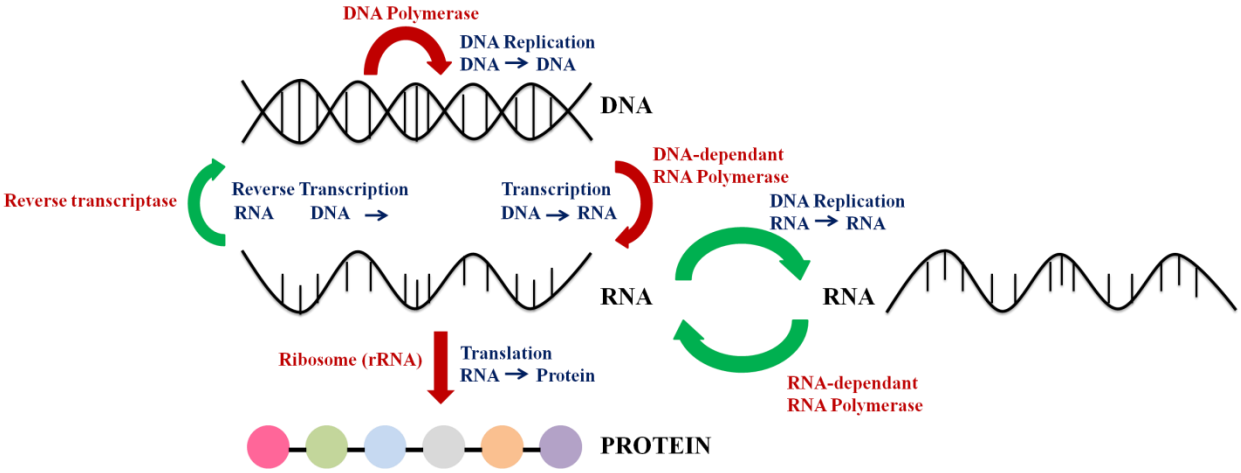

Figure 1.3 Central dogma of molecular biology. The central flow of information in life has been outlined as follows: DNA replication is followed by transcription into RNA. Both these processes require protein enzymes (polymerases). RNA (mRNA) is read by the ribosome (rRNA) with the help of tRNA to produce proteins. RNA can be reverse-transcribed to DNA or be transcribed from RNA by protein enzymes. Steps in the flow of genetic information are shown in blue and enzymes catalyzing these steps are shown in red. Red arrows indicate steps in the traditional (and simplified) central dogma, and green arrows indicate alternate direction of information flow.

From a biosynthetic viewpoint (ribonucleotides produce deoxyribonucleotides, the building blocks of DNA through the action of ribonucleotide reductase) and most likely an evolutionary perspective, DNA is a modification of RNA (Elledge et al., 1992). Therefore, the top contenders for the ancestral biopolymer are protein and RNA. Francis Crick's optimistic assertions that a primitive translational apparatus for making proteins "had no protein at all and consisted entirely of RNA" and "the first enzyme was an RNA molecule with replicase property" (Crick, F. H., 1968), as well as Leslie Orgel's arguments (Woese, 1967) for the seniority of RNA over proteins in the earth's biochemical history, bolstered by Harold White III's observation that half of all proteins relied on cofactors that are nucleotides or their derivatives (White III, 1976) (section 1.2.1) set the stage for the discoveries that altered the way we think about enzymes.

1.2 Ribozymes and primitive biocatalysis

1.2.1 RNA World resurfaces

The solution to the aforementioned conundrum came from two accidental discoveries in the 1980s. The groups of Tom Cech and Sidney Altman demonstrated for the first time that RNA was capable of catalyzing biochemical reactions on its own. Cech's discovery of the self-splicing group I intron from the prokaryotic *Tetrahymena* (Kruger et al., 1982) and Altman's discovery that the 5'-end maturation of tRNAs was catalyzed solely by the RNA portion of the ribonucleoprotein complex, RNase P (Guerrier-Takada et al., 1983) expanded enzyme catalysis to RNA, a molecule previously relegated to the singular purpose of managing the flow of genetic information (Figure 1.3). Catalytic RNAs, only a novelty at the time, were christened 'ribozymes', a contraction of ribonucleic acid enzymes (Kruger et al., 1982). These discoveries culminated in the coinage of the term 'RNA World', by Walter Gilbert in 1986 (Gilbert, 1986) (followed by a Nobel Prize shared jointly by Cech and Altman in 1989), a term that has stuck ever since and continues to capture scientific and public fascination alike.

Catalytic RNA that broke and forged bonds in RNA (group I intron catalyzes RNA cleavage followed by ligation and RNase P catalyzes RNA cleavage) was an elegant way out of the 'chicken-egg' paradox; a scenario could be imagined where RNA would perform enzyme function, almost exclusively handled by proteins in modern biology, as well as carry genetic message, currently encoded in DNA. This constitutes the crux of the RNA World hypothesis: the primitive scenario dated to approximately 3.6-3.8 billion years ago, where life passed through a period of biochemical function and evolution using RNA as its primary functional biomolecule. RNA would be the sole molecule carrying genetic information and supporting enzyme function. Retroviruses like HIV use an RNA genome and are often referred to as living fossils of the RNA

World. Replicases made entirely of RNA would trigger Darwinian evolution in hypothetical ribo-life (or 'ribosaurs', a term lovingly coined by Peter Moore) (Yarus, 2010), by copying itself using Watson-Crick (WC) base-pairing to ensure information transfer (the absence of such copying mechanism in proteins suggest primitive, self-sustaining life must have been devoid of proteins). Ribozyme polymerases would create copies of a variety of RNA templates and some of these functional RNAs would regulate gene expression, cut or join RNA sequences to increase functional complexity and support primitive metabolism. Peptidyl transfer leading to the synthesis of proteins through the development of a universal genetic code was a later innovation in RNA catalysis that possibly had roots in specific interactions between certain RNA sequences and amino acids, and a rudimentary form of a peptidyl transferase ribozyme (Yarus, 2010). Contemporary peptidyl transfer (i.e. protein synthesis) is carried out by the ribosome, which is a ribonucleoprotein particle; however, early claims by Harry Noller that the catalytic machinery of the ribosome was constructed entirely with RNA (Noller, 1992), was triumphantly supported by crystal structures of the ribosome that appeared at the dawn of the new millennium, from work done by the Steitz, Ramakrishnan and Yonath groups (Ban et al., 2000; Wimberly et al., 2000; Schluenzen et al. 2000). The observation that the active site for peptide bond formation is buried within the core ribosomal RNA, and no amino acid chain is present within 18 Å of it, firmly established the ribosome as a ribozyme; the most important enzyme in the cell was made of RNA. The spliceosome, another ribonucleoprotein machine, larger and more dynamic than the ribosome, carries out a reaction similar to the group I intron, but in eukaryotic cells. If the ribosome is important for accurate translation of the message created with mRNA, the spliceosome is responsible for editing it. By selectively removing non-coding 'introns' and joining the flanking 'exons', the spliceosome creates usable messages for translation and is also able to create multiple

messages from the same string of ribonucleotides, contributing to information economy (Collins and Guthrie, 2000). Much like rRNA in the ribosome, certain spliceosomal small nucleolar RNAs (snRNAs) were implicated in catalysis (Collins and Guthrie, 2000), a suggestion that received strong functional support from the Piccirilli and Staley groups (Fica et al. 2013). Paralleling the ribosome story, once again high-resolution structures from three groups, of Shi, Nagai and Lührmann, this time using cryo-EM (probably reflecting the changing trends), laid to rest all speculations about RNAs role in catalysis and demonstrated that the spliceosome is indeed a ribozyme at heart (Yan et al., 2015; Galej et al., 2016; Rauhut et al., 2016).

In addition to these behemoths, smaller functional RNAs have been and are consistently being discovered in extant biology. Riboswitches are small RNA motifs that bind specifically to ligands that range from ions (divalent metal ions and even an anion, F^-) to amino acids and a host of metabolites (e.g. coenzymes like TPP, FMN, nucleotides and their derivatives like guanine, adenine and prequeuosine and 2'-deoxyguanosine, AND signaling small molecules like cyclic di-AMP) that are likely derived from RNA (Breaker, 2010; Nelson and Breaker 2017). These motifs, found in the 5'-untranslated regions of prokaryotic mRNA, regulate gene expression via conformational changes triggered by ligand binding. It is not improbable that ribozymes worked in tandem with riboswitches to regulate gene expression and metabolism. Not surprisingly, one such RNA - the *glmS* riboswitch-ribozyme was discovered in gram-positive bacteria (Winkler et al., 2004) that embodies both ribozyme and riboswitch functions and regulates the expression of a gene that codes for L-glutamine: fructose 6-phosphate amidotransferase (Milewski, 2002), by autocatalytic cleavage of the message. In the last thirty years since the term ‘RNA World’ was coined, ribozymes have expanded to include a large family of small self-cleaving RNA (see section 1.2.4 for more detail) of which the *glmS* is a unique member. The discovery, as well as the

structural and mechanistic investigations of another self-splicing intron (group II intron) has added yet another large ribozyme to the list (Zhao and Pyle, 2017). Despite variations in size, structure and catalytic mechanisms, all contemporary ribozymes are involved in (with the exception of the peptidyl transferase center of the ribosome) catalyzing phosphoryl transfer reactions which might highlight its importance especially in the context of the historical requirement for RNA-based replicases or polymerases.

In the later part of the RNA World, dated to about 3.6 billion years in the past (oldest fossil remains dated to 3.5 billion years resemble modern organism with a DNA-RNA-protein biology) (Schopf, 2017), in addition to being the bearer (such as modern-day DNA) and propagator (e.g. modern day protein polymerases) of genetic information, ribozymes likely had to catalyze a wide plethora of biochemical reactions, most of which passed onto the three domains of modern life via our last universal common ancestor (LUCA) (Figure 1.4). Therefore, it may be expected that the ribo-life preceding LUCA had the same metabolic reactions operating in it; however, they were ribozyme-catalyzed (Hirao and Ellington, 1995). Excursions into laboratory RNA evolution have expanded catalytic capabilities of ribozymes beyond their natural functions. *In vitro* selection has identified ribozymes that catalyze RNA alkylation, RNA cleavage, ligation, polymerization, phosphorylation, aminoacyl tRNA synthesis and peptide bond formation, all of which are important for an RNA-based biology and some contain seeds to a transition to the modern Protein-World from the primitive RNA World (Chen et al., 2007). Ribozymes capable of forging C-C bonds classically in the domain of synthetic organic chemistry have further underscored the versatility of RNA catalysis (Chen et al., 2007). Experiments recreating the biochemical diversity of ribozyme-catalyzed reactions that might have been prevalent in the RNA World, coupled with possible artifacts from that time uncovered by ‘molecular archeology’ make a strong case for the

existence of an RNA World. These 'RNA World artifacts' include, but perhaps are not limited to cofactors and signaling small molecules of ribonucleotide origins, riboswitches that binds them, ubiquitous roles of RNA in the process of information transfer in forms of mRNA, tRNA, rRNA, and numerous small RNAs like siRNA and miRNA that regulate the information, and the requirement for a short sequence of RNA to initiate DNA replication (Hirao and Ellington, 1995; Benner et al., 1989; Breaker, 2017). Although, the RNA World hypothesis is a widely accepted one, there are some valid objections to it.

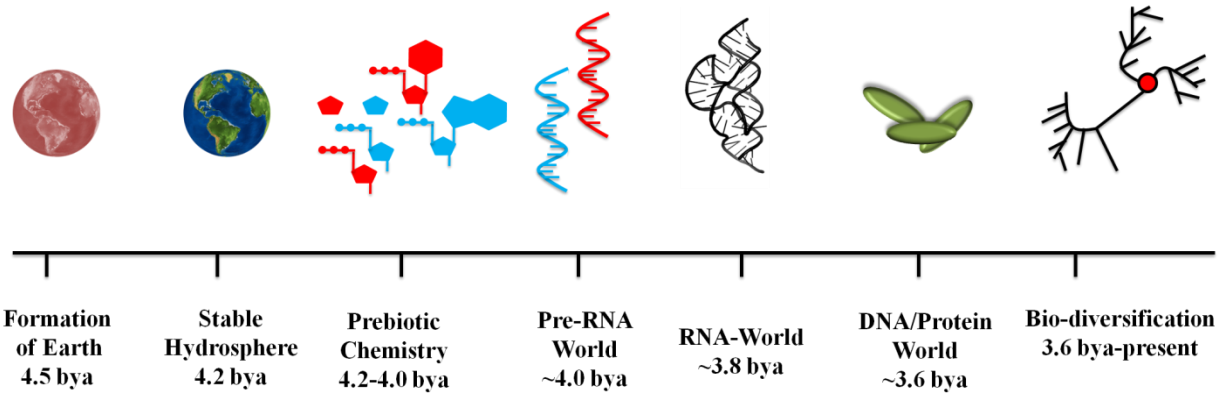

Figure 1.4 Timeline depicting the birth and journey of life on earth. Numbers stated are tentative and subject to modifications with the advent of new data. Adapted from Joyce, 2002.

These objections begin with messy prebiotic chemistry (described as 'a prebiotic chemist's nightmare') in a pre-RNA World, where the synthesis of ribonucleotides has been demonstrated to be far from favorable (Robertson and Joyce, 2012). Syntheses of nucleobases and sugars have been fraught with difficulties with the formation of a motley collection of undesirable side-products and the nucleobase-sugar coupling reaction has been largely impossible. Template-directed synthesis of RNA (Figure 1.5) before the emergence of the first RNA polymerase ribozyme is slow and has been historically considered to be inefficient in creating RNA sequences

of sufficient length and homogeneity (RNA backbones with a very few exceptions have 3'-5' linkages) to be catalytically active (Kawamura, 2012). The molecule that is central to the 'RNA-first' perspective to the origin of life and its evolution – a self-replicating ribozyme (colorfully referred to as 'a molecular biologist's dream') (Figure 1.5) has not been observed in its true sense (Robertson and Joyce, 2012). Further, the RNA World hypothesis appears to be in direct opposition with another popular theory of the origin of life – in which life is assumed to have started in submarine hydrothermal vents (Kawamura, 2012) at high temperatures detrimental to the integrity and function of RNA. Several of these concerns are being gradually alleviated by research led by Szostak (template-directed RNA synthesis), Joyce and Holliger (self-replicating ribozyme), and Benner and Sutherland (prebiotic synthesis of ribonucleotides) among others (Adamala and Szostak, 2013; Engelhart et al., 2015; Attwater et al. 2013; Horning and Joyce, 2016; Powner and Sutherland, 2011), although unanswered questions abound. Perhaps the following summarizes the status of the RNA World the best: "The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)" (Bernhardt, 2012). Notwithstanding the controversies regarding the role of RNA catalysis in the evolution of life, contemporary ribozymes present interesting systems to investigate the structural and functional diversity of RNA and study biomolecular catalysis in general.

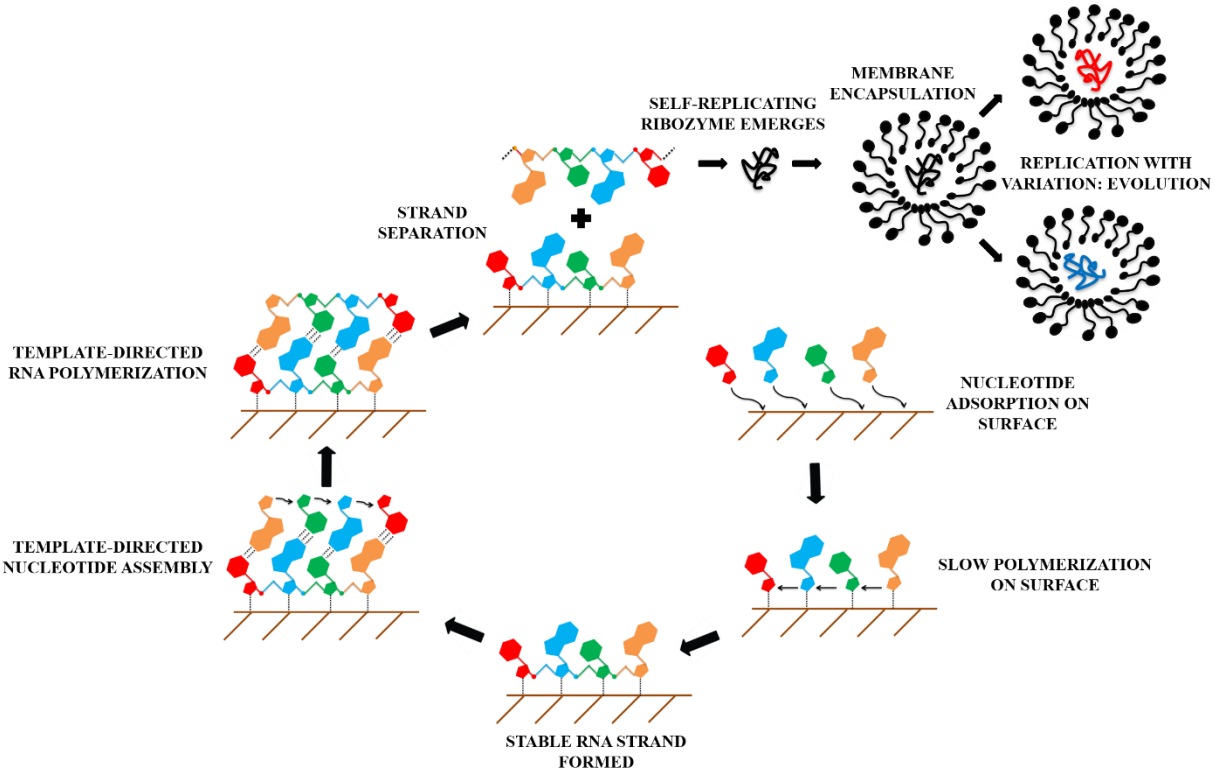

Figure 1.5 Simplified schematic for the origin of life via the RNA World. The prebiotically synthesized ribonucleotides are first adsorbed on a solid surface (clay is an exciting prospect). Due to proximity effects these units polymerize into short RNA strands on the surface. Template-directed assembly of nucleotides with WC complementarity allows polymerization of a second strand. After strand separation due to environmental perturbation, the newly assembled strand can break free from the template strand. Template-directed uncatalyzed RNA synthesis probably led to the emergence of the first catalytic RNA capable of self-replication. Membrane encapsulation of these self-replicating ribozymes provided closed systems that could interact selectively with their environments. Replication errors would invariably create sequence variants with differential activities and functions ushering in Darwinian evolution by natural selection. These self-sustaining systems capable of self-replication and Darwinian evolution could be considered proto-life.

1.2.2 Structural resources of catalytic RNA

Since the revelation that RNA could adopt well-defined three-dimensional structures (Kim et al., 1973), there has been substantial progress in understanding the relationship between RNA structure and function. The resurgence of RNA structure determination that started with two ribozymes – the hammerhead (Pley et al., 1994) and the P4-P6 domain of the group I intron (Cate et al., 1996) in the 1990s, has expanded to include high-resolution structures of two large self-splicing intron ribozymes (group I and group II introns), three ribonucleoprotein ribozymes (RNase P, ribosome and spliceosome), and eight classes of self-cleaving ribozymes (section 1.2.4) in addition to numerous classes of riboswitches and other non-coding RNA. Cumulatively, these structures have provided a wealth of information regarding the construction of functional RNA. Structure-function analyses of protein enzymes benefit from a large collection of functional structures available in the database that enables a 'first-guess assignment' of function to a new protein, based on the presence of motifs that are already annotated to established functions (Bork and Koonin, 1998). In the absence of a similar database for RNA, function is not easily related to recurring RNA motifs. The contributions of some of these motifs to overall folding and eventual function are, however, well documented. Peter Moore's suggestion that "structures of all RNAs would eventually prove to be nothing more than concatenations of RNA motifs—of which it was surmised there would not be very many—and runs of A-form helix" (Moore, 1999), might be overly simplistic given the paucity of RNA structures, leading to a high probability that many such 'motifs' remain undiscovered, but points to a logical and modular approach to RNA structural biology. RNA structure is roughly made of stretches of A-form helices that are either interrupted by 'internal loop motifs' or end in 'terminal loop motifs' (Moore, 1999).

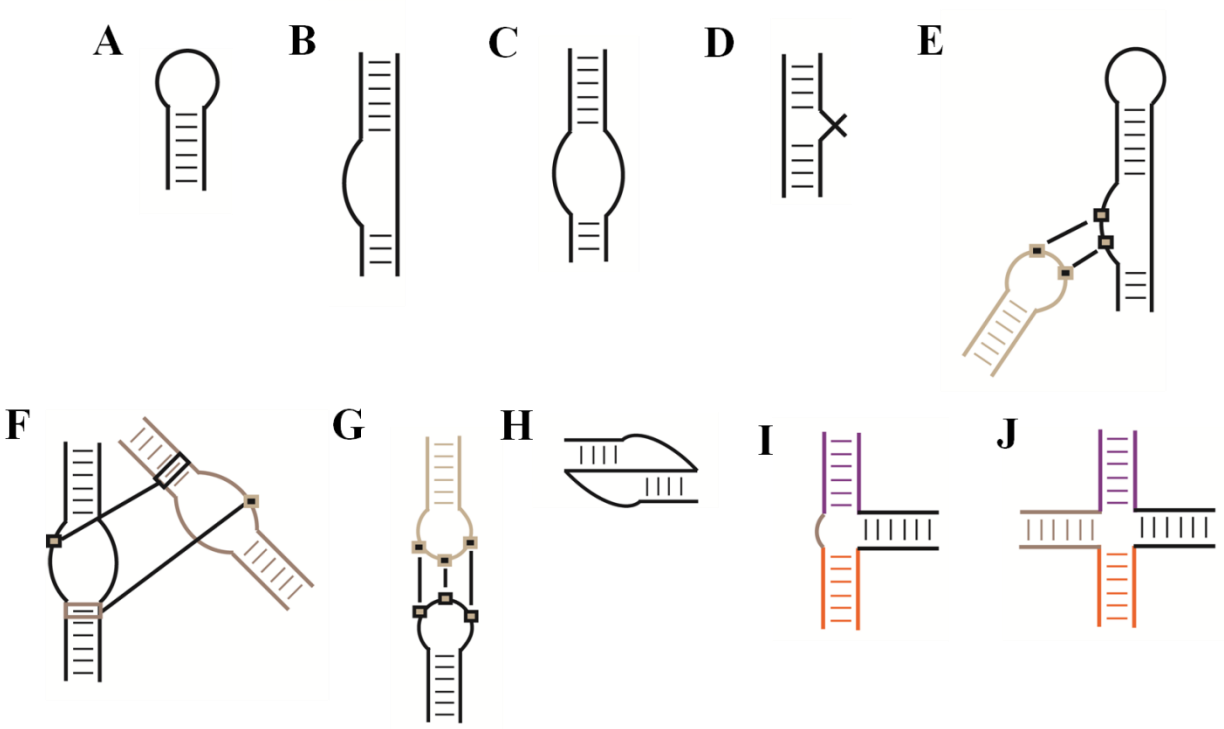

Figure 1.6 Recurring structural motifs in RNA. A. Hairpin loop. **B.** Asymmetric internal bulge. **C.** Symmetric internal bulge. **D.** Cross-strand stack. **E.** Internal bulge-tetraloop receptor. **F.** Ribose-zipper. **G.** Kissing-loop. **H.** Pseudoknot**. I.** Three-way junction. **J.** Four-way junction.

Internal loop motifs (Figure 1.6B-F) include cross-strand purine stack, A platform, bulged – G-motif, and bulge-helix-bulge, among others that often organize catalytic sites in ribozymes or act as acceptors for hairpin loops or metal ions (Moore, 1999).

Terminal loop motifs (Figure 1.6A, E, G, H) are often instrumental in organizing the global geometry of RNA by mediating intramolecular or intermolecular tertiary contacts. These include 'UNRN' U-turns and tetraloops (UNGC, GNRA, and CUYG), where N, R and Y stand for any nucleotide, purines and pyrimidines, respectively (Figure 1.6A) (Moore, 1999). Association of motifs in three-dimensional space that are spatially separated in the secondary structure of the RNA, results in the formation of tertiary motifs. Terminal loops associate with each other via the formation of WC base-pairs to form kissing-loop interactions (Figure 1.6G) (Paillart et al., 1996) that are directly involved in biological processes like the regulation of ColE1 plasmid replication

(Eguschi and Tomizawa, 1991), initiation of HIV1 dimerization (Paillart et al., 1996), substrate/ligand recognition and binding (Zhang and Ferré-D'Amaré, 2015; Suslov et al. 2015; DasGupta et al., 2017), and organization of RNA quaternary structure (Suslov et al. 2015, Jones and Ferré-D'Amaré, 2015). Terminal loops interact with internal bulges to form tetraloop-receptor structures (Figure 1.6E) (Pyle, 2010), or other unstructured single stranded region via WC base-pairing to form pseudoknots (Figure 1.6H) (Ferré-D'Amaré et al., 1998; Ren et al., 2016; Theimer and Feigon, 2006). Ribose zippers are formed by the interactions between unpaired and antiparallel strands of two separate internal loops (Figure 1.6F) (Moore, 1999; Rupert and Ferré-D'Amaré, 2001) via hydrogen bonds between the 2'-hydroxyl groups from both strands. To facilitate these interactions, the individual motifs must be brought in close proximity to each other by favorable coaxial stacking of the helices harbouring them. This global organization is performed by helical junctions (Figure 1.6I, J). These junctions often utilize non-canonical interactions between nucleotides that constitute these structural devices to modulate their spatial orientation. This is particularly important in larger RNAs that exhibit hierarchical folding and organization (Pyle, 2010, Suslov et al. 2015). In summary, these structural features support the catalytically competent fold of ribozymes and bring functional nucleotides, including ones that are close to the active site, setting up chemical step of catalysis.

1.2.3 Chemical resources of catalytic RNA

Proteins are clearly Nature's favorite biocatalysts. Since the advent of translation about 3.6 billion years ago, natural selection has been active in selecting proteins over RNA for enzymatic roles. In addition to the intrinsic instability of RNA, (a half-life of a few hundred days compared to half-lives of a few hundred years for the same linkage in DNA and the peptide bond in protein), other factors such as unfavorable electrostatics (a negatively charged backbone in RNA hinders compact folding; a problem obviated by a neutral amide linkage in proteins) and the lack of chemical versatility presented by its residues (four chemically similar aromatic heterocycles versus at least twenty functionally diverse amino acids, see section 1.1) make RNA inferior to proteins in terms of their enzymatic prowess (e.g. endonucleolytic ribozymes accelerate RNA cleavage by 10^6-fold in contrast to the ~10^{11}-10^{13} enhancement by RNase A, a protein enzyme) (Thompson et al., 1995, Emilsson et al., 2003). These limitations notwithstanding, ribozymes appear to make the best of their evolutionary predicament, utilizing every chemical moiety to varying degrees in catalysis (Figure 1.7).

Nucleobases are the obvious choices for catalysis, as they can directly participate in proton transfer, thereby acting as general acids or bases in the catalytic step. Proteins can deploy several amino acid residues for the same purpose; however, histidine is the most extensively used in general acid/base catalysis (Stryer et al., 2002). The imidazole moiety of histidine has a pKa ~6, which makes it slightly acidic but not so much that it's conjugate base is completely unreactive. More importantly, this means at the near-physiological pH of 7, 1 in 10 histidine molecules will be protonated, thereby capable of functioning as a general acid, and their deprotonated form as a general base. In sharp contrast, none of the pKas of RNA nucleobases are close to 7 (pKas of imino nitrogen atoms are ~3.5 for adenine, ~4.2 for cytosine, ~9.2 for guanine and uracil) (Figure 1.7)

(Lilley and Eckstein, 2008). At a pKa of ~4.2 cytosine is a fairly strong acid; however, at neutral pH, only 1 in 1000 molecules would be in the protonated state required for general acid catalysis. The remaining 999 deprotonated molecules are too weak to act as general bases. Cytosines with pKas shifted from 4.2 to 7.7-9.6 have been observed in pseudoknotted RNA underscoring the direct impact of RNA folding on the chemistry of the ribozyme active site (Cornish and Giedroc, 2006). Similarly, pKa shifts from 4.2 to near neutrality is possible in the presence of proximal phosphate groups, whereas a decrease in pKa is expected in guanine if it's located close to a metal ion, thereby enhancing its basic capabilities (Lilley, 2011).

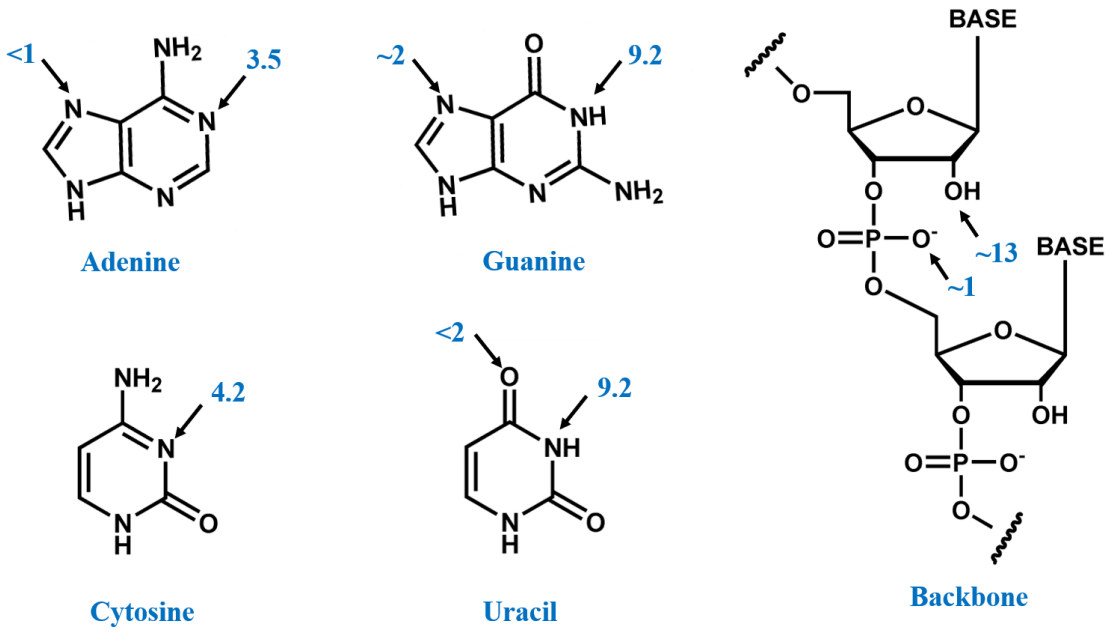

Figure 1.7 Unperturbed pK_a values for RNA. The four nucleobases of RNA with their unperturbed pK_a values for various atoms (Izatt et al., 1971; Narlikar and Herschlag, 1997; Simorre et al., 1997). pK_a values of ionizable groups in the phosphodiester backbone (Acharya et al., 2003; Li and Breaker, 1999; Lyne and Karplus, 2000). Adapted from Lilley and Eckstein, 2008.

Ribose sugars are not functionally as diverse as nucleobases but harbor the all-important 2'- hydroxyl group that contributes to the instability of RNA. Structural and computational studies (Martick and Scott, 2006; Chen et al., 2017) suggest that the 2'-hydroxyl group of a catalytic

guanine (G8) forms a crucial hydrogen bond with the 5'-oxygen leaving group during nucleolysis in the hammerhead ribozyme. The newly discovered pistol self-cleaving ribozyme is speculated to use a 2'-hydroxyl moiety for a similar purpose (Ren et al., 2016). The 2'-hydroxyl group of the exogenous guanosine cofactor in the group I intron splicing reaction forms a hydrogen bond with the bridging 3'-oxygen leaving group during the branching step of splicing (Herschlag et al., 1993; Houghland et al., 2008).

Divalent metal ions like Mg^{2+} are essential for RNA folding; however, they can also interact directly with the substrate or active site residues to mediate or assist catalysis. Metal binding might modulate pKas of catalytic nucleobases, act as Lewis acids, interact with both ribozymes and substrates to stabilize reaction transition states, or leaving groups in reactions that involve RNA cleavage (Lilley and Eckstein, 2008). Structural and biochemical studies implicate catalytic metal ions in self-splicing by the group I (Shan et al. 2001; Stahley and Strobel, 2005) and group II introns (Sontheimer et al., 1999; Gordon et al., 2000), and more recently in the spliceosome (Fica et al., 2013). Smaller self-cleaving ribozymes like HDV, hatchet and twister-sister likely utilize hydrated metal ions as general bases for site-specific cleavage in addition to nucleobases acting as putative general acids (Ren, et al., 2017). Previously unforeseen functions of metal ions in the hammerhead ribozyme that include modulating pKa of general base G12 and mediating direct interactions with the 5'-oxygen leaving group, suggest possible catalytic roles (Ren, et al., 2017). Biochemical and computational data suggest additional roles of divalent metal ions in stabilizing the transition state of the VS ribozyme (Weissman et al., unpublished results).

In the light of protein's late arrival on the catalytic scene, at least 200 million years after ribozymes, one can easily invert Crick's exclamation about the 'protein-like' structure of tRNA to: proteins are nature's attempt to do the (catalytic) job of RNA. Endonucleolytic ribozymes

provide excellent systems to investigate the roles of various players in RNA catalysis discussed above and more importantly reveal fundamental principles underlying the construction of efficient catalytic systems with RNA.

1.2.4 Catalyzing RNA cleavage

RNA strands spontaneously fall apart in a matter of months – probably the reason why evolution chose the more stable DNA to store hereditary information. The chemical reaction responsible for RNA degradation involves a nucleophilic (S_N2-type) attack by the 2'-oxygen atom (nucleophile) on the adjacent phosphate (electrophilic P) resulting in the formation of a 2'-3'-cyclic phosphate with the release of the 5'-oxygen (leaving group). Stereochemical inversion at the phosphorous center suggests a concerted reaction; however, the extents of bond making and breaking in the transition state under different biochemical environments dictate its associative/dissociative nature (Lilley and Eckstein, 2008). This is a topic of active research, where kinetic isotope effects have been used successfully to interrogate the transition state to high precision (Harris et al., 2015; Weissman et al., unpublished results). Investigations into alkaline hydrolysis of RNA in the 1950s, unraveled details about the reaction mechanism, and provided clues about possible strategies enzymes could employ to lower the activation energy of this reaction (Fedor, 2009). Contrary to evolutionary history, protein enzymes have been studied over a longer period of time and in greater detail. Therefore, mechanistic understanding of protein ribonucleases, provide a framework within which to study ribozymes that cleave RNA (see chapter 3 for a detailed comparison). Nucleolysis catalyzed by proteins exhibit much greater rate accelerations compared to those catalyzed by RNA, however, unlike protein ribonucleases, an endonucleolytic ribozyme is not only substrate specific but also largely specific toward a certain

dinucleotide linkage at a particular location in the substrate. This is primarily due to base-pairing capabilities of RNA that define the hot-spot for cleavage. Such specificity in RNA cleavage might have been essential to ribozyme-based gene regulation in the RNA World, but lost its necessity in modern biology, where protein nucleases were merely required to degrade unused or foreign RNA. Irrespective of the biochemical nature of the catalyst, cleavage of an RNA strand requires the following basic strategies (Figure 1.8) (Cochrane and Strobel, 2008; Emilsson et al., 2003):

1. *Facilitating an in-line attack.* This involves organizing the cleavage site nucleotides in a configuration such that the 2'-O nucleophile, P of the scissile phosphodiester and the 5'-O leaving group form a dihedral angle that is close to 180° to allow the lowest energy path for intramolecular S_N2-type attack. This is a substantial deviation from the usual 70° angle maintained by these groups when stacked in an A-form helix, requiring additional stabilization of a splayed conformation of the nucleotides flanking the scissile phosphate.
2. *Stabilization of the transition state.* Favorable interactions between an enzyme and the activated complex at the transition state of the reaction play a vital role in lowering the activation barrier. The pentacoordinated transition state (more precisely, activated complex but we will use the former for consistency in literature), is dianionic therefore, moieties able to alleviate its negative charge would contribute to catalysis. Such stabilization could come from the enzyme molecule itself or through tightly-bound metal ions.
3. *Activation of the nucleophile.* The alkoxide is a superior nucleophile than its weak conjugate acid, the hydroxyl. Consequently, groups that mediate proton transfer from the 2'-OH moiety thereby acting as bases, provide significant catalytic assistance. Metal-bound water molecules or enzyme residues can act as general bases by deprotonating the 2'-OH group, thereby activating the nucleophile.

4. *Stabilization of the leaving group.* Complimentary to relative destabilization of the 2'-oxygen nucleophile, neutralization of the negative charge building up on the 5'-oxygen leaving group in the course of the reaction lowers the overall energy of the reaction trajectory.

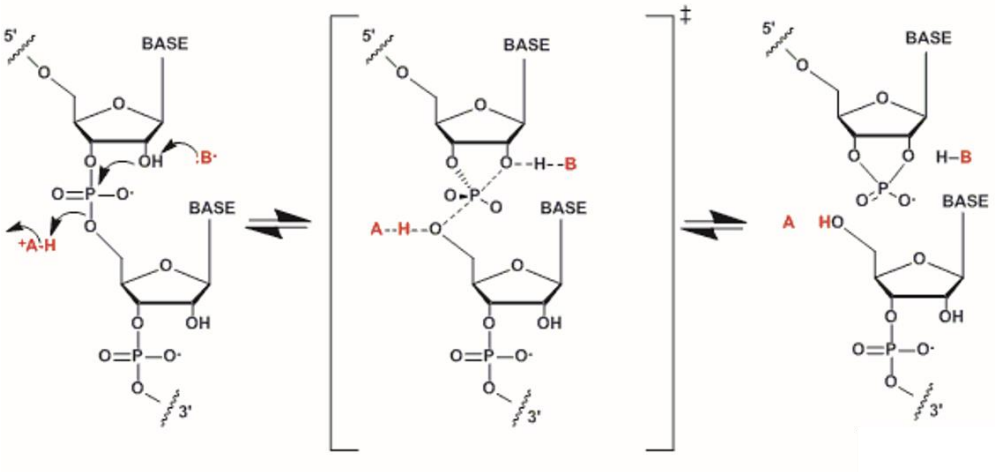

Figure 1.8 Endonucleolytic cleavage reaction. The cleavage reaction may proceed in steps or *via* a phosphorane intermediate, or in a concerted fashion via a single transition state. General acids ($^{+}$A-H, red) and bases (:B^{-}, red) or specific acids and bases catalyze the reaction *via* proton transfer to stabilize the leaving group and activate the nucleophile, respectively.

It is important to note that while the limits for rate enhancements for each strategy can be estimated independent of others, each type of enhancement often comes from the same moiety, and in reality, it might be problematic to decouple these strategies (Emilsson et al., 2003; Lilley, 2011). However, treating them separately provides a clear mechanistic picture of nucleolytic catalysis and guides further investigation.

1.2.5 The increasing diversity of endonucleolytic ribozymes

Protein enzymes use metal ions (Tesmer et al., 1999), amino acid residues (Raines, 1998) or a combination of both (Drum et al., 2002) to catalyze RNA cleavage employing the strategies discussed in the previous section. Ribozymes, having a negatively-charged backbone unsurprisingly use metal ions to catalyze RNA cleavage, as part of splicing in group I (Piccirilli et al., 1993; Shan et al. 2001; Stahley and Strobel, 2005) and group II introns (Sontheimer et al., 1999; Gordon et al., 2000), and the spliceosome (Fica et al., 2013). The simplest endonucleolytic ribozyme, the hammerhead (HH) was considered to be an obligate metalloenzymes by default (Pyle, 1993). The discovery that RNA could use their nucleobases that were once thought to be poorly suited for catalysis, for site-specific cleavage in the Hepatitis Delta virus (HDV) endonucleolytic ribozyme (Nakano et al., 2001; Shih and Been, 2001; Das and Piccirilli, 2005) created a mild paradigm shift in RNA catalysis. Using both hydrated metal ion and nucleobase as catalytic moieties, the HDV served as a historical intermediate for the next step in understanding ribozyme mechanism. A landmark study by Scott and co-workers (Murray et al., 1998) demonstrated catalytic RNA cleavage in the hammerhead (HH) and two other endonucleolytic ribozymes, the hairpin (HP) and Varkud satellite (VS) in the absence of divalent metal ions (but in high concentrations of monovalent ions). This suggested structural as opposed to catalytic roles of metal ions in these ribozymes, which put catalytic responsibilities exclusively on moieties within the RNA molecule. A dual nucleobase model for catalysis was proposed for the hairpin ribozyme that found strong support from crystal structures that captured both the precursor as well as the transition state mimic of the cleavage reaction (Rupert and Ferré-D'Amaré, 2001, Liberman et al., 2012) and from computational studies (Heldebrand et al., 2014). A similar two nucleobase model was proposed for the VS ribozyme in light of biochemical data (Wilson and Lilley, 2011)

which finds structural validation in the results presented in chapter 3. Mechanistic details of hammerhead catalysis have been refined in recent years in the light of new structures (Martick and Scott, 2006; Mir and Golden, 2016), biochemical data (Mir et al., 2015) and computation (Chen et al., 2017) These results have implicated two nucleotides (one nucleobase and one ribose) in catalysis, in addition to catalytic metal ions.

The HDV, hammerhead and hairpin ribozymes are all found in satellites of viruses (HDV ribozymes were discovered in both + and – strands of the HDV RNA found in a satellite of the Hepatitis B virus, hammerhead ribozymes are found in plant satellite viruses like the Tobacco Ringspot virus, with its complementary strand harboring the hairpin ribozyme motif) (Kuo et al., 1988; Haselhoff and Gerlach, 1988; Buzayan et al., 1986). VS ribozymes occur within satellite RNA in the mitochondria of fungus *Neurospora* (Saville and Collins, 1990). The primary function of these ribozymes appears to be resolution of concatemeric intermediates in the rolling-circle replication of their respective plasmids (Jimenez et al., 2015) (discussed in Chapter 2 and 3 in relation to biological functions of the VS ribozyme). Recent bioinformatic explorations have identified hammerhead and HDV-like motifs in the non-coding transcriptome of mammals. Hammerhead ribozymes are, in fact, present in all domains of life (Jimenez et al., 2015) ranging from viroids, plants, amphibians, insects and mammals (rodents). Highly active hammerhead motifs are embedded in the 3' UTRs of mammalian C-type lectin type II (CLEC2) genes. In these locations, the motif is fragmented into substrate and catalytic sequences. CLEC2-associated hammerhead motifs have been shown to cleave both *in vitro* and *in vivo*, lowering proteins expression in mouse cells (Jimenez et al., 2015; Garcia-Robles et al., 2016). A role in controlling polyadenylation sites in mRNA has been suggested for mammalian hammerhead; however, a more in-depth understanding of these motifs is needed before any definite function can be attributed to

them (Jimenez et al., 2015; Garcia-Robles et al., 2016). Another self-cleaving motif, CPEB3 found in mammals (including humans) has been implicated in similar roles. This motif maps to the second intron of the mammalian cytoplasmic polyadenylation element binding protein 3 locus (Salehi-Ashtiani et al., 2006). Interestingly, biochemical characterization of the CPEB3 ribozyme revealed parallels to the HDV ribozyme. Although the two RNAs have distinct primary structures (sequences), the CPEB3 sequence could be easily threaded through an HDV secondary structure, which makes a case for a common evolutionary history between the two (Webb, 2011). Like the hammerhead, variants of the HDV ribozyme have been found scattered throughout phylogeny, from viruses, bacteria, fungi and plants to members of the flatworms, insects, fish and mammals (Jimenez et al., 2015). While the functions of these ubiquitous self-cleaving motifs are still obscure, their presence might represent an additional layer of post-transcriptional gene regulation hitherto unknown.

The discovery of ribozyme-mediated gene-regulation predates the more recent discoveries of CLEC2 and CPEB3 ribozymes. The *glmS* ribozyme was discovered through a bioinformatic search designed to scour the bacterial genome for riboswitches. The *glmS* ribozyme occurs as a conserved self-cleaving motif in the 5' UTR of the mRNA corresponding to the *glmS* gene that codes for L-glutamine: fructose 6-phosphate amidotransferase, an enzyme that produces glucosamine-6-phosphate (GlcN6P) (Winkler et al., 2004). The *glmS* ribozyme binds GlcN6P to catalyze cleavage of the mRNA it is part of, thereby degrading the instruction to make more GlcN6P-synthesizing protein. This sets up a negative feedback mechanism of gene regulation for modulating intracellular concentrations of GlcN6P. *GlmS* is the only ribozyme to utilize a small molecule cofactor (extending parallels to cofactor-dependent protein enzymes) and the only riboswitch to use RNA cleavage without modulation of secondary structure to regulate gene

expression (other riboswitches use either transcription termination or translation initiation by 'switching' secondary structures), thus serving as a conceptual link between ribozymes and riboswitches, two supposed 'relics' from the RNA World (Gesteland et al., 2006). Mechanistically, the *glmS* extend the recurrent theme of nucleobase-mediated catalysis to include a cofactor that appears to have co-opted the role of general acid by stabilizing the 5'-oxygen leaving group (Klein and Ferré-D'Amaré, 2006). *In vitro* selection experiments have 'evolved' a cofactor-independent *glmS* motif that retains the global fold but utilizes a distinct catalytic mechanism (Lau and Ferré-D'Amaré, 2013; Bingaman et al., 2017). Interestingly, this ribozyme-only sequence is just a few point mutations away from the wild-type *glmS* ribozyme-riboswitch sequence, leading to speculations about evolutionary acquisition of cofactor-dependence, and broader implications regarding the nature of cooperation between ribozymes and riboswitches in ancient gene regulatory circuits.

Bioinformatic searches have become more robust in recent years leading to the discovery of four new classes of self-cleaving ribozymes by Breaker and co-workers (Roth et al., 2014; Weinberg et al., 2015). These new ribozymes namely, twister, twister-sister (TS), pistol and hatchet present variations in the major themes of RNA catalysis by utilizing nucleobases exclusively or in conjunction with RNA-bound Mg^{2+} to catalyze site-specific RNA cleavage, and organizing their catalytic fold using multi-helical junctions or pseudoknots (Ren et al., 2017). Despite overall similarities, these ribozymes present subtle novelties in their catalytic strategies. Structural and biochemical data, coupled with computational results have implicated the N3 atom of adenine in stabilizing the 5'-oxygen leaving group (δ catalysis) in the twister and pistol (adenine or guanine) classes (Lilley, 2017; Ren et al., 2017). This is surprising because pKa of the N3 atom is lower than that of the N1 atom of adenine (a more common general acid), pushing it farther from

neutrality. However, in twister, the pKa of this nucleobase is raised to near-neutrality in the environment of the active site possibly due to its proximity to negatively charged phosphates (Lilley, 2017). A similar interaction is expected in pistol; however, it is to be noted that more functional data is required to make confident assignments (Lilley, 2017). Twister-sister and hatchet ribozymes are thought to be metallo-ribozymes like the HDV; however, in the absence of a high-resolution structure of the latter, further similarities or differences remain obscure (Lilley, 2017; Ren et al., 2017; Lee and Lee, 2017).

Within the framework of data available at this time, we can tentatively classify endonucleolytic ribozymes according to structural principles of organization, and catalytic moieties used (Figure 1.9; Table 1.1) (Lilley, 2017). The hammerhead, hairpin, VS and twister-sister ribozymes organize their active folds primarily using multi-helical junctions, whereas HDV, *glmS*, twister and pistol use pseudoknots. VS, hairpin and twister are thought to use nucleobases exclusively for general acid-base catalysis, while HDV, twister-sister, and speculatively hatchet use hydrated divalent metal ions in similar roles. Hammerhead and possibly pistol use 2'-hydroxyl groups as general acids and a guanine nucleobase as the general base and could be using divalent metals for transition state stabilization. *GlmS*, using a small molecule cofactor as a general acid with a guanine acting as the general base belongs to a separate category.

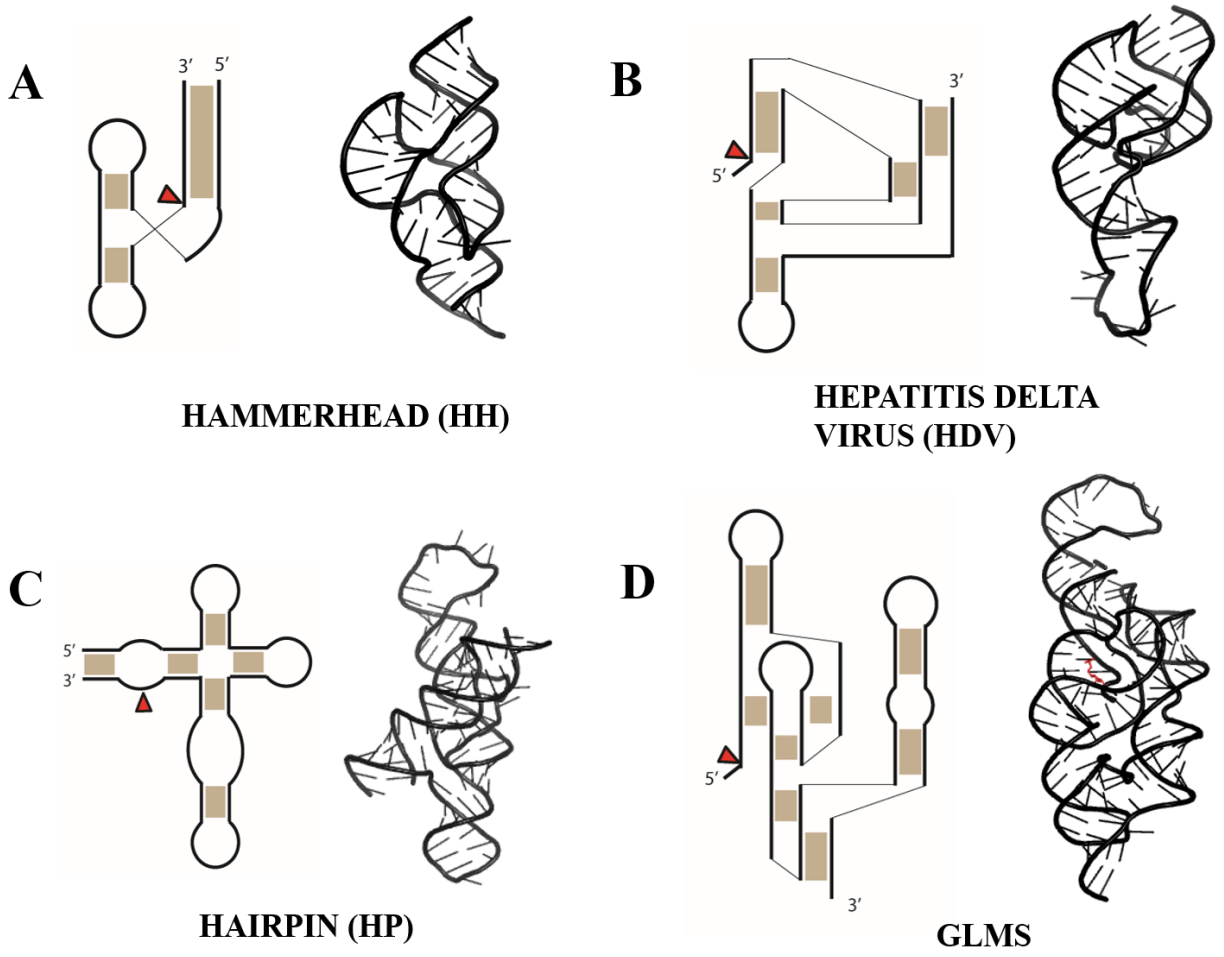

Figure 1.9 Natural endonucleolytic ribozymes: secondary and tertiary structures. Shaded regions denote helices that are mostly base-paired. Cleavage sites are denoted by red triangles. Secondary structures of ribozymes in **A**-**G** and **I** reflect the three-dimensional folds as obtained from crystal structures, while that in **H** does not in the absence of a high-resolution structure of the ribozyme.

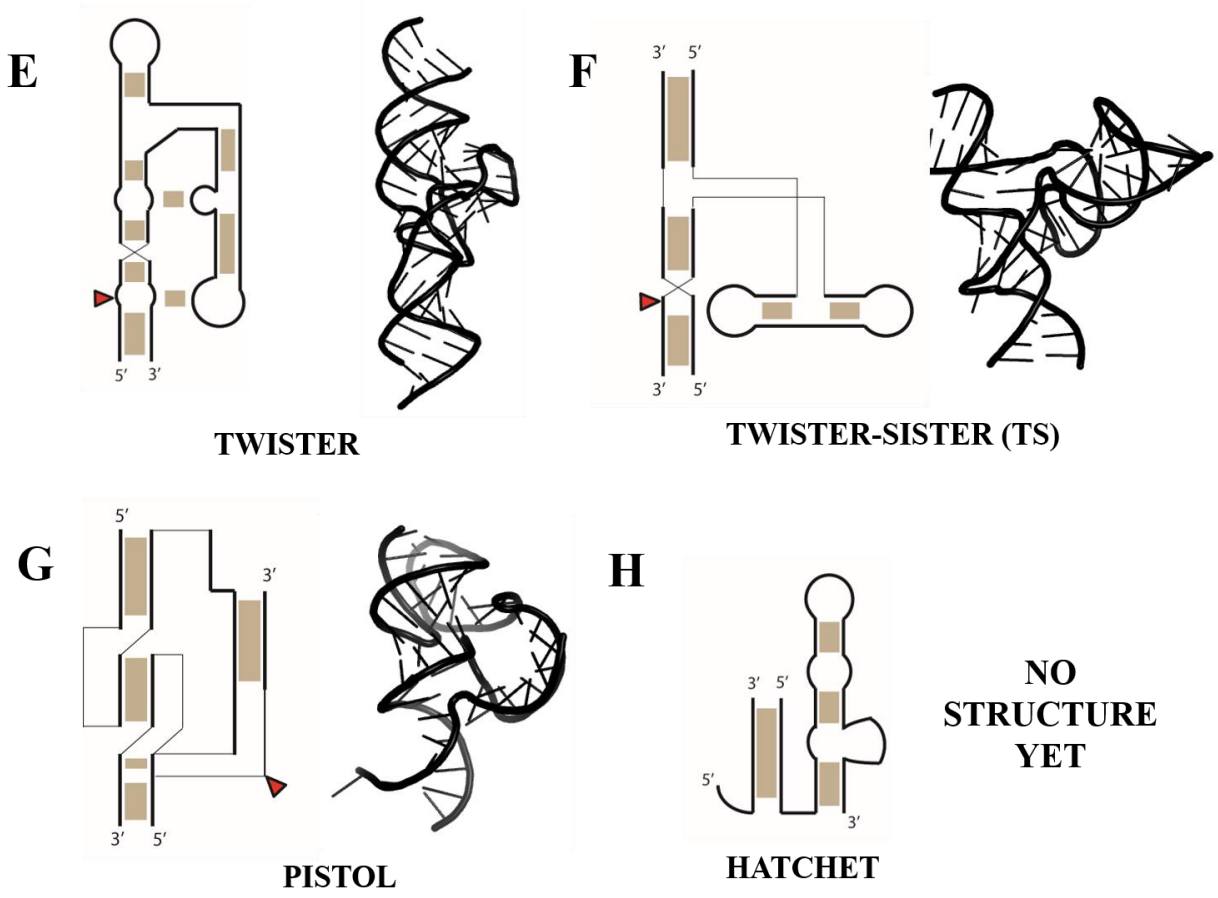

Figure 1.9 Natural endonucleolytic ribozymes. Continued from previous page.

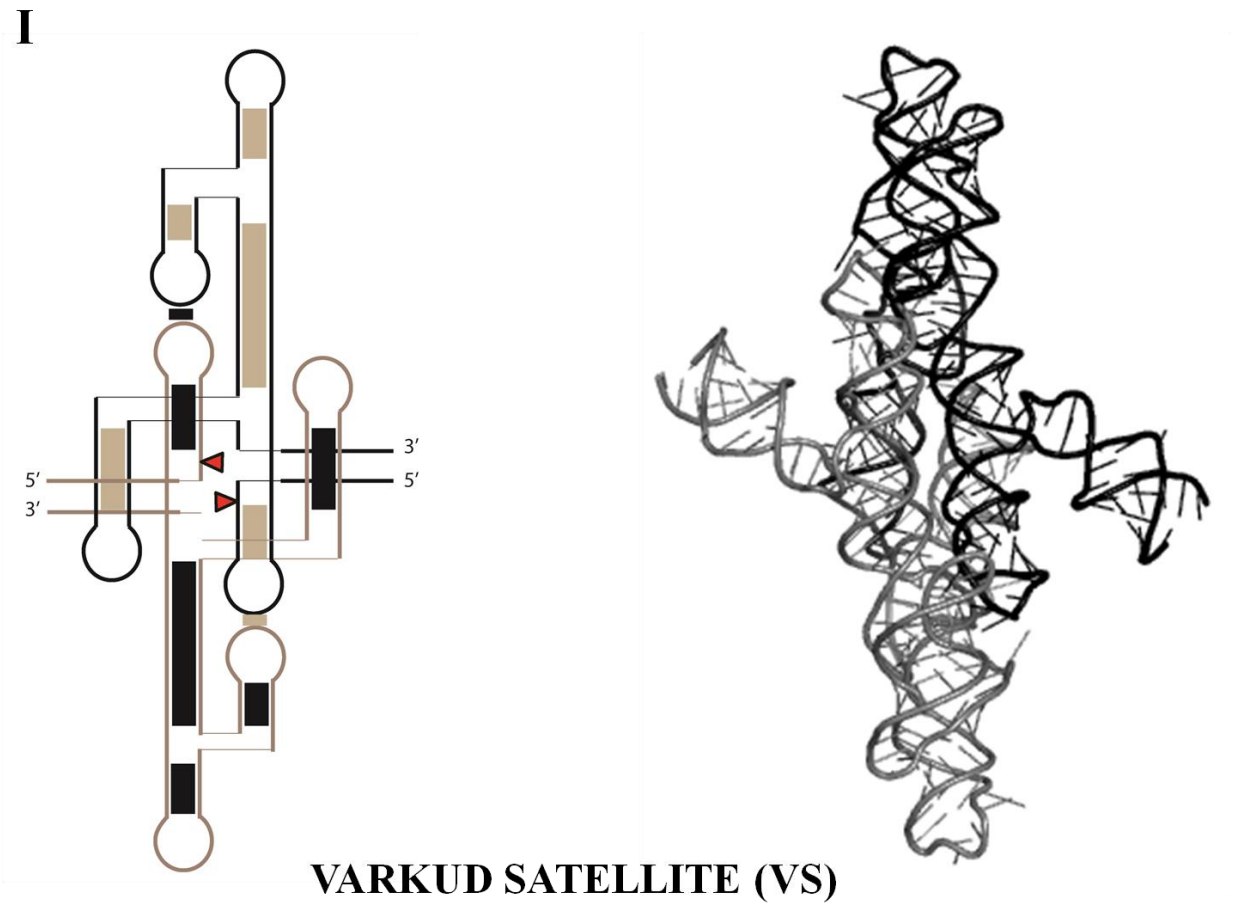

Figure 1.9 Natural endonucleolytic ribozymes. Continued from previous page. This structure is a result from the work discussed in chapters 2 and 3.

Ribozyme class	Cleavage site Alignment (α catalysis)	Transition state stabilization (β catalysis)	Nucleophile Activation (γ catalysis)	Leaving group stabilization (δ catalysis)	Structural organization	Cleavage Site (wild-type)
Hammerhead	148° (4RGE)	Mg^{2+}	Guanine (N1)	Guanine (2'OH), Mg^{2+}	3-way junction	C-C
Hairpin	158° (1M5O)	Guanine (N2)	Guanine (N1)	Adenine (N1)	3 or 4-way junction	A-G
HDV	140° (5DI2)	Mg^{2+}	$Mg^{2+}(H_2O)_n$	Cytosine (N3)	Pseudoknot	U-G
Varkud satellite (VS)	**128° (4R4P)**	**Guanine (N2)**	**Guanine (N1)**	**Adenine (N1)**	**3-way junction**	**G-A**
glmS	140° (3G8S)	Glc6NP cofactor, Guanine (N2)	Guanine (N1)	Glc6NP cofactor	Pseudoknot	A-G
Twister	148° (4RGE)	Guanine (N2), Mg^{2+*}	Guanine (N1)	Adenine (N3)	Pseudoknot	U-A
Twister-sister	71° (5Y85)	Guanine (N1), Mg^{2+*}	$Mg^{2+}(H_2O)_n$	Not assigned	3 or 4-way junction	C-A
Pistol	167° (4RGE)	Mg^{2+*}	Guanine (N1) *	Adenine or guanine (N3, 2'OH) *	Pseudoknot	G-U
Hatchet	No structure. Exhibit essential requirement for Mg^{2+}. In addition to implicating a function group with a functional pKas of ~7, interactions with the non-bridging oxygen by RNA moieties is suggested.					C-U

Table 1.1 Summary of catalytic strategies proposed for endonucleolytic ribozymes. Assignments are subject to alterations in the light of new data; however, this is more applicable to the ribozymes discovered in the last three years due to the inadequacy of functional data. The moieties without particularly strong evidence for their catalytic roles are marked with asterisks. (for latest reviews see Lilley, 2017; Ren et al., 2017; Lee and Lee, 2017). See figure 1.9 for structures.

1.2.6 Perspectives

Each field of scientific research has its own evolutionary history, which when rewound in its chronological order provides, a glimpse of the contemporary scientific zeitgeist. The discipline of RNA catalysis emerged as a result of serendipity. None of the discoveries following this were results of proactively looking for catalytic RNA, but rather sprung from unusual observations while studying other biological phenomena. The burgeoning field of RNA biochemistry and enzymology was quick to delve into studying new ribozyme molecules generally resulting in incremental additions to the subject. High-resolution crystal structures arrived later in the scene as RNA structural biology came to its own, arguably spurned by the discovery of ribozymes and the

unsolved problems in RNA catalysis. The relevance of these structures was judged on their ability to explain functional data accumulated over years of biochemical research. The 'saga of the hammerhead' is a great example of the struggles between functional and structural data that raged for more than two decades (De la Peña et al., 2017). As the well-spring of serendipitously discovered catalytic RNA molecules ran dry in the first decade of the twenty-first century, a new pipeline of ribozyme discovery has been opened though bioinformatic approaches that can scan the non-coding transcriptome for potentially functional sequences. This paradigm in ribozyme discovery has delivered four new ribozymes in three years and promises many more in the years to come (Ren et al., 2017). About twenty years ago, an *in vitro* selection experiment in the Breaker lab looking for self-cleaving RNA motifs in randomized libraries of RNA yielded several unique sequences out of which two classes possessed sufficient activities to function in biology (Tang and Breaker, 2000). One of them was the ubiquitous hammerhead motif, and the other named the 'X-motif' has still not surfaced in life. It is not unreasonable (although optimistic) to expect that deep searches in the genome might uncover the existence of sequences similar or related to the X-motif. In addition to the discovery pipeline, the process of studying these new ribozymes has changed. Multiple crystal structures of these new ribozymes have emerged within months of their first reports. These structures have informed more directed biochemical interrogations into their catalytic mechanisms but the narrow focus on catalysis due to the ready availability of their global structures, has discouraged extensive biochemistry to study entire sequences of these ribozymes, an effort that had often yielded new information not directly sought but ultimately useful in broadening the horizons of ribozyme research. High-throughput sequencing (HTS) techniques can provide functional maps of the residues and their variants in the ribozyme and the probable roles

and dispensability of each nucleotide, in the broader context of the ribozyme (Kobori et al., 2015; Kobori and Yokobayashi, 2016).

High-resolution structures are still of paramount importance as HTS experiments do not provide any three-dimensional information. A host of crystal structures often published within a few months of each other have provided complementary information regarding different aspects of catalysis. In some instances, however, multiple structures of the same ribozyme captured states that suggested contradictory mechanisms (Breaker, 2017). Even similar structures have elicited divergent interpretations. One can hope that the friction between competing hypotheses will eventually forge paths leading to functional models that are consistent with most observations. While the recently discovered ribozymes have promptly yielded to structure determination, the Varkud satellite (VS) ribozyme had eluded crystallization for over two decades. With 25 years of functional data derived from elegant biochemical and biophysical experiments available, the crystal structures appeared to be a grand culmination; however, in addition to providing a structural framework to corroborate functional data, the structures of the VS ribozyme have revealed unique features in RNA folding – both local and global, with implications in VS biology. These structures have provided the first glimpses of the enzyme active site, unveiling key players in catalysis, and their contributions. These high-resolution structures promise to provide reference points for future mechanistic investigations through atomic perturbations to the catalytic residues in the form of experiments and simulations. The structure and catalytic function of the VS ribozyme is the focus of the next two chapters.

1.3 Ribozymes and evolution of catalytic function

The four billion year-long journey of life could very well be a journey of the ribozyme out of the effervescent RNA World into placid modern biology. The success of this journey depended upon the capacities of ribozymes to adapt to new conditions and biochemical requirements and adopt novel catalytic functions. If life passed through an 'RNA-only' phase before the arrival of LUCA, ribozymes had to carry out a plethora of diverse functions ranging from replication, ligation and cleavage to biosynthesis and processing of metabolites (Chen et al., 2007; Pressman et al., 2015) (section 1.2.1). In the absence of a sophisticated replication machinery made of proteins, error rates for copying or synthesizing RNA from various RNA templates were probably high (Horning and Joyce, 2016) and the accumulation of lethal mutations posed a serious threat to the continuity of primitive life. However, the chemical nature of the RNA alphabet and the forces that create RNA structures make RNA robust to mutational perturbations (see chapter 4) leading to the preservation of its global functional fold and catalytic apparatus even in the face of most instances of copying errors. Such a robust catalyst would likely be made of only a few essential residues, usually located close to the active site, and the rest of the molecule would provide the scaffold with identities of its building blocks not as important as the structures they create. This design is exemplified by catalytic RNA, especially endonucleolytic ribozymes that consist of a few conserved nucleotides scaffolded by structures primarily created with base-paired helices and structural motifs like loops and bulges (Cochrane and Strobel, 2008) (section 1.2.2). This robustness of RNA probably enabled functional preservation across generations underlying the conservative aspect of evolution. The progressive aspect of evolution, however, demands change. Mutations are vehicles of evolution, which suggests that exploration of sequence (genotype) space eventually leads to the discovery of new structure/function (phenotype). While robustness allows

for exploration of sequence space, one of its consequences is the existence of isolated (and relatively small) regions of activity, populated by RNA sequences having distinct functions (Wagner, 2005). So, how can these sequences 'escape' their functional islands to diversify their catalytic portfolio? What are the sources of increasing complexity in ribozymes? The large gap in evolutionary time and the absence of 'RNA fossils' have made it impossible to take the 'dig and discover' approach to seek out answers to these fundamental questions, and we may never know the details of how RNA transformed itself and with it the course of our biological history. But the crime scene may be re-constructed using the little information that is available and by assessing the plausibility of different feasible paths we hope to discover mechanisms that are logically consistent with the current framework of RNA evolution. The discovery of novel classes of endonucleolytic ribozymes have garnered renewed interest in ribozyme mechanisms. An equally (if not more) interesting question is that of their origins. The distinct structures and mechanisms of these ribozymes might suggest independent evolutionary origins yet the structural complexities that some of these RNAs exhibit drastically reduce the probabilities of independent emergence through random mutations from non-functional sequences. As observed with proteins, novel functional sequences arise from existing functional sequences that are often simpler than their evolutionary offspring (Meier and Özbek, 2007). Incremental increase in molecular complexity has been observed in RNA as well; rRNAs have expanded in size by adding more nucleotides across phyla enabling evolutionary geneticists to map the organisms to their positions in the tree of life using the composition of their rRNA (Bokov and Steinberg, 2009; Fox et al., 1977). This principle is reflected more prominently at the organismic level with the transformation of the 3.5 billion-year-old, microscopic and single-celled archaea to the sentient and creative *Homo sapiens*, capable of pondering its own origins. Therefore, it is not unlikely that different classes of

endonucleolytic ribozymes evolved from simpler self-cleaving motifs that through mutual association with other RNA modules led to their diversification.

1.4 Questions addressed in the thesis

Ribozymes present the most primitive catalytic systems in biology and are thought to have played a central role in jumpstarting primitive life. The purpose of this thesis is to understand three main aspects of RNA catalysis: (1) structural principles that are exploited to enable catalysis, (2) catalytic strategies to facilitate the chemical step at the active site and (3) plausible evolutionary origins of distinct catalytic functions. What does it take to make an enzyme purely out of RNA? How does RNA accomplish the canonical steps of enzyme catalysis with its limited resources? Are the chemical principles of catalysis conserved across polymers i.e. does RNA use the same catalytic strategies a protein or DNA uses when catalyzing a similar reaction? Chapter 2 and 3 uses crystal structures of the VS ribozyme to answer these questions. Although ribozymes catalyze a limited set of chemical reactions in extant biology, in the context of the evolution of biocatalysis and the RNA World, RNA most likely a supported a diverse chemistry. How did a polymer made of just four kinds of chemically similar building blocks acquire such wide-ranging capabilities when distinct functional sequences are generally isolated from each other in evolutionary space? How do complex RNA catalysts emerge from simpler catalytic motifs? How did the diversity in catalytic function emerge in RNA? These are the questions I have explored in chapters 4 and 5.

RNA motifs can selectively bind small molecules as demonstrated by riboswitches, which has been exploited to create nano-devices called aptamers (Ellington and Szostak, 1990). Aptamers provide excellent recognition modules that have been used to design genetic circuits (Berens et al., 2015), sensors (Cho et al., 2009) and therapeutics (Keefe et al., 2010), bringing functional RNA

from the annals of Origins research and domains of biology into the rapidly growing marketplace of biomolecular devices. In appendix 1, I have discussed efforts to create a sensitive and selective sensor for Pb^{2+} using a fluorescent RNA aptamer called 'Spinach' (Paige et al., 2011).

Less than 2% of all RNA transcribed from DNA code for proteins. While a major fraction of the remaining could be 'junk' remnants of blind tinkering by evolutionary forces, this 'dark matter' of the genome is slowly but surely revealing some of its secrets to the scientific enterprise. This thesis is a minuscule part of it.

REFERENCES

1. Acharya, S., Foldesi, A. and Chattopadhyaya, J., 2003. The pK(a) of the internucleotidic 2'-hydroxyl group in diribonucleoside (3'→5') monophosphates. *J. Org. Chem.*, 68, 1906-1910.

2. Adamala, K. and Szostak, J. W., 2013. Nonenzymatic template-directed RNA synthesis inside model protocells. *Science*, 342, 1098-1100.

3. Attwater, J., Wochner, A. and Holliger, P., 2013. In-ice evolution of RNA polymerase ribozyme activity. *Nat. Chem.*, 5, 1011-1008.

4. Avery, O. T., MacLeod, C. M. and McCarty, M., 1944. Studies on the chemical nature of the substance inducing transformation of Pneumococcal types: Induction of transformation by a deoxyribonucleic acid fraction isolated from Pneumococcus type III. *J. Exp. Med.* 79, 137–158.

5. Babbitt, P. C. and Gerlt, J. A., 1997. Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities. *J. Biol. Chem.*, 272, 30591–30594.

6. Ban, N., Nissen, P., Hansen, J., Moore, P. B. and Steitz, T. A., 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, 289, 905-920.

7. Benner, S. A., Ellington, A. D. and Tauer, A., 1989. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci. U. S. A.*, 86, 7054–7058.

8. Berens, C., Groher, F. and Suess, B., 2015. RNA aptamers as genetic control devices: the potential of riboswitches as synthetic elements for regulating gene expression. *Biotechnol. J.* 10, 246-257.

9. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E., 2000. The Protein Data Bank. *Nucleic Acids Res.*, 28, 235-242.

10. Bernhardt, H. S., 2012. The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). *Biol. Direct*, 7, 23-33.

11. Bingaman, J. L., Zhang, S., Stevens, D. R., Yennawar, N. H., Hammes-Schiffer, S. and Bevilacqua, P. C., 2017. The GlcN6P cofactor plays multiple catalytic roles in the *glmS* ribozyme. *Nat. Chem. Biol.*, 13, 439-445.

12. Bokov, K. and Steinberg, S. V., 2009. A hierarchical model for evolution of 23S ribosomal RNA. *Nature*, 457, 977-980.

13. Bork, P. and Koonin, E. V., 1998. Predicting functions from protein sequences—where are the bottlenecks? *Nat. Genet.*, 18, 313–318.

14. Breaker, R. R., 2012. Riboswitches and the RNA World. *Cold Spring Harb. Perspect. Biol*., 4: a003566.

15. Breaker, R. R., 2017. Mechanistic debris generated by twister ribozymes. *ACS Chem. Biol.*, 12, 886-891.

16. Brenner, S., Jacob, F., and Meselson, M., 1961. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature*, 190, 576–581.

17. Buzayan, J. M., Hampel, A. and Bruening, G., 1986. Nucleotide sequence and newly formed phosphodiester bond of spontaneously ligated satellite tobacco ringspot virus RNA. *Nucleic Acids Res.*, 14, 9729-9743.

18. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K. H., Golden, B. L., Kudrot, C. E., Cech, T. R. and Doudna, J. A., 1996. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, 273, 1678-1685.

19. Chen H, Giese, T. J., Golden, B. L. and York, D. M., 2017. Divalent metal ion activation of a guanine general base in the hammerhead ribozyme: Insights from molecular simulations. *Biochemistry*, 56, 2985-2994.

20. Chen, X., Li, N. and Ellington, A. D., 2007. Ribozyme catalysis of metabolism in the RNA World. *Chem. Biol.*, 4, 633- 655.

21. Cho, E. J., Lee, J. W. and Ellington, A. D., 2009. Applications of aptamers as sensors. *Annu. Rev. Anal. Chem.*, 2, 241-264.

22. Cochrane, J. C. and Strobel, S. A., 2008. Catalytic Strategies of self-cleaving ribozymes. *Acc. Chem. Res*., 41, 1027-1035.

23. Collins, C. A. and Guthrie, C., 2000. The question remains: Is the spliceosome a ribozyme? *Nature*, 7, 850-854.

24. Cornish, P. V. and Giedroc, D. P., 2006. Pairwise coupling analysis of helical junction hydrogen bonding interactions in luteoviral RNA pseudoknots. *Biochemistry*, 45, 11162-11171.

25. Crick, F. H., 1958. On protein synthesis. *Symp. Soc. Exp. Biol.*, 12, 138-163.

26. Crick, F. H., 1966. The Genetic Code-Yesterday, Today? and Tomorrow. *Cold Spring Harb. Symp. Quant. Biol.*, 31, 1-9.

27. Crick, F. H., 1968. The origin of the genetic code. *J. Mol. Biol.*, 38, 367-379.

28. Das, S. R. and Piccirilli, J. A., 2005. General acid catalysis by the hepatitis delta virus ribozyme. *Nat. Chem. Biol.* 1, 45-52.

29. DasGupta, S., Suslov, N. B. and Piccirilli, J A., 2017. Structural basis for substrate helix remodeling and cleavage loop activation in the Varkud satellite ribozyme. *J. Am. Chem. Soc.*, 139, 9591-9597.

30. De la Peña, M., García-Robles, I. and Cervera, A., 2017. The Hammerhead ribozyme: A long history for a short RNA. *Molecules*, 22, 78-90.

31. Drum, C. L., Yan, S. Z., Bard, J., Shen, Y. Q., Lu D, Soelaiman ,S., Grabarek, Z., Bohm, A. and Tang, W. J., 2002. Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin. *Nature*, 415, 396-402.

32. Eguchi, Y. and Tomizawa, J., 1991. Complexes formed by complimentary RNA stem loops their formations, structures and interaction with ColE Rom protein. *J. Mol. Biol.*, 220, 831-842.

33. Elledge, S. J., Zhou, Z. and Allen, J. B., 1992. Ribonucleotide reductase: regulation, regulation, regulation. *Trends Biochem. Sci.* 17, 119–123.

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

35. Emilsson, G. M., Nakamura, S., Roth, A. and Breaker, R. R., 2003. Ribozyme speed limits. *RNA*, 9, 907-918.

36. Engelhart, A. E., Powner, M. E. and Szostak, J. W., 2013. Functional RNAs exhibit tolerance for non-heritable 2′–5′ vs. 3′–5′ backbone heterogeneity. *Nat. Chem.*, 5, 390–394.

37. Fedor, M. J., 2009. Comparative Enzymology and Structural Biology of RNA Self-Cleavage. *Annu. Rev. Biophys.*, 38, 271–299.

38. Ferré-D'Amaré, A. R., Zhou, K. and Doudna, J. A., 1998. Crystal structure of a hepatitis delta virus ribozyme. *Nature*, 395, 567-574.

39. Fica, S. M., Tuttle N, Novak T., Li N.-S., Lu, J., Koodathingal, P., Dai Q., Staley, J. P. and Piccirilli, J. A., 2013. RNA catalyses nuclear pre-mRNA splicing. *Nature*, 503, 229-234.

40. Fox, G. E., Magrum, L. J., Balch, W. E., Wolfe, R. S. and Woese, C. R., 1977. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc. Natl. Acad. Sci. U. S. A.*, 74, 4537-4541.

41. Galej, W.P., Wilkinson, M.E., Fica, S.M., Oubridge, C., Newman, A.J. and Nagai, K., 2016. Cryo-EM structure of the spliceosome immediately after branching. *Nature*, 537, 197-201.

42. Gesteland, R. F., Cech, T. R. and Atkins, J. F., 2006. *The RNA World: the nature of modern RNA suggests a prebiotic RNA world (3rd)* Cold spring harbour, NY. Cold Spring Harbour Laboratory Press.

43. Gilbert, W., 1986. The RNA World. *Nature*, 319, 618.

44. Gordon, P. M., Sontheimer, E. J. and Piccirilli, J. A., 2000. Kinetic characterization of the second step of group II intron splicing: role of metal ions and the cleavage site 2'-OH in catalysis. *Biochemistry*, 39, 12939-12952.

45. Griffith, F., 1928. The significance of pneumococcal types. *J. Hyg.* 27, 113-159.

46. Gros, F., Hiatt, H., Gilbert, W., Kurland, C.G., Risebrough, R.W., and Watson, J.D., 1961. Unstable ribonucleic acid revealed by pulse labelling of Escherichia coli. *Nature*, 190, 581–585.

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

48. Harris, M. E., Piccirilli, J. A. and York, D. M., 2015. Enzyme transition states from theory and experiment. *Biochim. Biophys. Acta*, 1854, 1727-1728.

49. Haselhoff, J. and Gerlach, W. L., 1988. Simple RNA enzymes with new and highly specific endoribonuclease activities. *Nature*, 334, 585-591.

50. Heldenbrand, H., Janowski , P. A., Giambaşu, G., Giese, T. J., Wedekind, J. E., and York, D. M., 2014. Evidence for the role of active site residues in the hairpin ribozyme from molecular simulations along the reaction path. *J. Am. Chem. Soc.*, 136, 7789-7792.

51. Herschlag, D., Eckstein, F. and Cech, T. R., 1993. The importance of being ribose at the cleavage site in the tetrahymena ribozyme reaction. *Biochemistry*, 32, 8312-8321.

52. Hirao, I. And Ellington, A. D., 1995. Re-creating the RNA world. *Curr. Biol.*, 5, 1017-1022.

53. Horning, D. P., Joyce, G. F., 2016. Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci. U. S. A.*, 113, 9786-9791.

54. Hougland, J. L., Sengupta, R. N., Dai, Q., Deb, S. K. and Piccirilli, J. A., 2008. The 2' hydroxyl group of the guanosine nucleophile donates a functionally important hydrogen bond in the tetrahymena ribozyme reaction. *Biochemistry*, 47, 7684-7694.

55. Izatt, R. M., Christensen, J. J., and Rytting, J. H., 1971. Sites and thermodynamic quantities associated with proton and metal ion interaction with ribonucleic acid, deoxyribonucleic acid, and their constituent bases, nucleosides, and nucleotides. *Chem. Rev.*, 71, 439-481.

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

57. Jimenez, R. M., Polanco, J. A. and Lupták, A., 2015. Chemistry and Biology of Self-cleaving ribozymes. *Trends Biochem. Sci.*, 40, 648-661.

58. Jones, C. P. and Ferré-D'Amaré, A. R., 2015. RNA quaternary structure and global symmetry. *Trends Biochem. Sci.*, 40, 211–220.

59. Joyce, G. F., 2002. The antiquity of RNA-based evolution. *Nature*, 418, 214-221.

60. Kawamura, K., 2012. Drawbacks of the ancient RNA-based life-like system under primitive earth conditions. *Biochimie*, 94, 1441-1450.

61. Keefe, A. D., Pai, S. S. and Ellington A. D., 2010. Aptamers as therapeutics. *Nat. Rev. Drug Discov.*, 9, 537-550.

62. Kim, S. H., Quigley, G. J., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Windier, J. and Rich, A., 1973. Three-dimensional structure of yeast phenylalanine transfer RNA: Folding of the polynucleotide chain. *Science*, 179, 285-288.

63. Klein, D. J. and Ferré-D'Amaré, A. R., 2006. Structural basis of *glmS* ribozyme activation by glucosamine-6-phosphate. *Science*, 313, 1752-1756.

64. Kobori, S. and Yokobayashi, Y., 2016. High-throughput mutational analysis of a twister ribozyme. *Angew. Chem Int. Ed. Engl.*, 55, 10354-10357.

65. Kobori, S., Nomura, Y., Miu, A. and Yokobayashi, Y., 2015. High-throughput assay and engineering of self-cleaving ribozymes by sequencing. *Nucleic Acids Res.*, 43, e85-93.

66. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. and Cech, T.R., 1982. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, 31, 147–157.

67. Kuo, M. Y., Sabourin, J. R., Dinter-Gottleib, G .and Taylor, J., 1988. Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *J. Virol.*, 62, 4439-4444.

68. Lau, M. W. and Ferré-D'Amaré, A. R., 2013. An *in vitro* evolved *glmS* ribozyme has the wild-type fold but loses coenzyme dependence. *Nat. Chem. Biol.*, 9, 805-810.

69. Lee, K.-Y. and Lee, B.-J., 2017. Structural and Biochemical Properties of Novel Self-Cleaving Ribozymes. *Molecules*, 22, 678-696.

70. Li, Y., and Breaker, R.R., 1999. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2 '-hydroxyl group. *J. Am. Chem. Soc.*, 121, 5364-5372.

71. Liberman J. A., Guo, M., Jenkins, J. L., Krucinska, J., Chen, Y., Carey, P. R., Wedekind, J. E., 2012. A transition-state interaction shifts nucleobase ionization toward neutrality to facilitate small ribozyme catalysis. *J. Am. Chem. Soc.*, 134, 16933-16936.

72. Lilley, D. M. J. and Eckstein F. 2008. *Ribozymes and RNA Catalysis*. RSC publishing.

73. Lilley, D. M. J., 2011. Mechanisms of RNA catalysis. *Phil. Trans. R. Soc. B*, 366, 2910-2917.

74. Lilley, D. M. J., 2017. How RNA acts as a nuclease: some mechanistic comparisons in the nucleolytic ribozymes. *Biochem. Soc. Trans.*, 45, 683–691.

75. Lindskog, S., 1997. Structure and mechanism of carbonic anhydrase. *Pharmacol. Ther.*, 74, 1-20.

76. Lodish, H., Berk, A., Matsudaira, P., Kaiser, C. A., Scott, M. P., Zipursky, L. and Darnell, J., 2003. *Molecular Cell Biology (5th Ed)*. Macmillan.

77. Martick, M. and Scott, W. G., 2006. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell*, 126, 309–320.

78. Meier, S. and Özbek, S., 2007. A biological cosmos of parallel universes: does protein structural plasticity facilitate evolution? *BioEssays*, 29, 1095–1104.

79. Milewski, S., 2002. Glucosamine-6-phosphate synthase-the multi-facets enzyme. *Biochim. Biophys. Acta*, 1597, 173-192.

80. Mir, A. and Golden, B. L., 2016. Two Active Site Divalent Ions in the Crystal Structure of the Hammerhead Ribozyme Bound to a Transition State Analogue. *Biochemistry*, 55, 633-636.

81. Mir, A., Chen, J., Robinson, K., Lendy, E., Goodman, J., Neau, D. and Golden, B. L., 2015. Two Divalent Metal Ions and Conformational Changes Play Roles in the Hammerhead Ribozyme Cleavage Reaction. *Biochemistry*, 54, 6369-6381.

82. Moore, P. B., 1999. Structural motifs in RNA. *Annu. Rev. Biochem.* 68, 287–300.

83. Murray J. B., Seyhan, A. A., Walter, N. G., Burke, J. M. and Scott, W. G., 1998. The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone. *Chem. Biol.*, 5, 587-595.

84. Nakano, S., Chadalavada, D. M., Bevilacqua, P. C., 2000. General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science*, 287, 1493-1497.

85. Narlikar, G. J., and Herschlag, D., 1997. Mechanistic aspects of enzymatic catalysis: lessons from comparison of RNA and protein enzymes. *Annu. Rev. Biochem.*, 66, 19-59.

86. Nelson, J. W. and Breaker, R. R., 2017. The lost language of the RNA World. *Sci. Signal*, 10, eaam8812.

87. Noller, H. F., Hoffarth, V. and Zimniak, L., 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science*, 256, 1416-1419.

88. Paige, J. S., Nguyen-Duc, T., Song, W. and Jaffrey, S. R., 2012. Fluorescence imaging of cellular metabolites with RNA. *Science*, 33, 1194-1197.

89. Paige, J. S., Wu, K. Y. and Jaffrey, S. R., 2011. RNA mimics of green fluorescent protein. *Science*, 333, 642-646.

90. Palliart. J. C., Marquet, R., Skripkin, E., Ehresmann, C. and Ehresmann, B., 1996. Dimerization of retroviral genomic RNAs: structural and functional implications. *Biochimie*, 78, 639-653.

91. Parsons, S. M. and Raftery, M. A., 1972. Ionization behaviour of the cleft carboxyls in lysozyme-substrate complexes. *Biochemistry*, 11, 1633-1638.

92. Piccirilli, J. A., Vyle, J. S., Caruthers, M. H. and Cech, T. R., 1993. Metal ion catalysis in the Tetrahymena ribozyme reaction. *Nature*, 361, 85-88.

93. Pley, H. W., Flaherty, K. M. and KcKay, D. B., 1994. Three dimensional structure of a hammerhead ribozyme. *Nature*, 372, 68-74.

94. Powner, M. W. and Sutherland, J. D., 2011. Prebiotic chemistry: a new modus operandi. *Phil. Trans. R. Soc. B.*, 366, 2870–2877.

95. Pressman, A., Blanco, C. and Chen, I. A., 2015. The RNA World as a model system to study the origin of life. *Curr. Biol.*, 25, R953–R963.

96. Pyle, A. M., 1993. Ribozymes: a distinct class of metalloenzymes. *Science*, 261, 709-714.

97. Pyle, A. M., 2010. The tertiary structure of group II introns: implications for biological function and evolution. *Crit. Rev. Biochem. Mol. Biol.*, 45, 215-232.

98. Radzicka, A. and Wolfenden, R., 1995. A proficient enzyme. *Science*, 267, 90-92.

99. Radzicka, A. and Wolfenden, R., 1996. Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases. *J. Am. Chem. Soc.*, 118, 6105–6109.

100. Raines, R. T., 1998. Ribonuclease A. *Chem. Rev.*, 98, 1045-1066.

101. Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., Kumar, V., Lee, C.-T., Urlaub, J., Kastner, B., Stark, H. and Lührmann, R., 2016. Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science*, 353, 1399-1405.

102. Ren, A., Micura, R. and Patel, D. J., 2017. Structure-based mechanistic insights into catalysis by small self-cleaving ribozymes. *Curr. Opin. Chem. Biol.*, 41, 71–83.

103. Ren, A., Vušurović, N., Gebetsberger, J., Gao, P., Juen, M., Kreutz, C., Micura, R. and Patel, D. J., 2016. Pistol ribozyme adopts a pseudoknot fold facilitating site-specific in-line cleavage. *Nat. Chem. Biol.*, 12, 702-708.

104. Robertson, M. P. and Joyce, G. F., 2012. The origins of the RNA World. Cold *Spring Harb. Perspect. Biol.*, 4: a003608.

105. Roth, A., Weinberg, Z., Chen, A. G., Kim, P. B., Ames, T. D. and Breaker, R. R. 2014. A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat. Chem. Biol.*, 10, 56-60.

106. Rupert, P. B. and Ferré-D'Amaré, A. R., 2001. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, 410, 780-786.

107. Salehi-Ashtiani, K., Lupták, A., Litovchick, A. and Szostak, J. W., 2006. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science*, 313, 17880-17889.

108. Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartles, H., Agmon, I., Franceschi, F.and Yonath, A., 2000. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, 102, 615-623.

109. Schopf, J. W., Kitajima, K., Spicuzza, M. J., Kudryavtsev, A. B. and Valley, J. W., 2017. SIMS analyses of the oldest known assemblage of microfossils document their taxon-correlated carbon isotope compositions. *Proc. Natl. Acad. Sci. U. S. A.*, 115, 53–58.

110. Shan, S., Kravchu, A. V., Piccirilli, J. A. and Herschlag, D., 2001, Defining the catalytic metal ion interactions in the tetrahymena ribozyme reaction. *Biochemistry*, 40, 5161-5171.

111. Shih, I. H. and Been, M. D., 2001. Involvement of a cytosine side chain in proton transfer in the rate-determining step of ribozyme self-cleavage. *Proc. Natl. Acad. Sci. U. S. A.*, 98, 1489-1494.

112. Sontheimer, E. J., Gordon, P.M. and Piccirilli, J. A., 1999. Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes Dev.*, 13, 1729-1741.

113. Stahley, M. R. and Strobel, S. A., 2005. Structural evidence for a two-metal ion mechanism of group I intron splicing. *Science*, 309, 1587-1590.

114. Stryer, L., Berg, J. M., Tymoczko, J. L., 2002. *Biochemistry (5th Ed.).* San W.H. Freeman.

115. Suslov, N. B., DasGupta, S., Huang, H., Fuller, J. R., Lilley, D. M. J., Rice, P. A. and Piccirilli, J. A., 2015. Crystal structure of the Varkud satellite ribozyme. *Nat. Chem. Biol.*, 11, 840-846.

116. Tang, J. and Breaker, R. R., 2000. Structural diversity of self-cleaving ribozymes. *Proc. Natl. Acad. Sci. U. S. A.*, 97, 5784-5789.

117. Theimer, C. A. and Feigon, J., 2006. Structure and function of telomerase RNA. *Curr. Opin. Struct. Biol.*, 16, 307-318.

118. Thompson, J. E., Kutateladze, T. G., Schuster, M. C., Venegas, F. D. Messmore, J. M. and Raines, R. T., 1995. Limits to catalysis by Ribonuclease A. *Bioorg. Chem.* 23, 471–481.

119. Vasiliev, V. D., Selivanova, O. M. and Koteliansky, V. E., 1978. Specific self-packing of the ribosomal 16 S RNA. *FEBS Lett.* 95, 273-276.

120. Vasiliev, V. D., Serdyuk, I. N., Gudkov, A. T. and Spirin, A. S., 1986. *Structure, Function, and Genetics of Ribosomes*. Springer-Verlag, New York.

121. Wagner, A., 2005. Robustness, evolvability, and neutrality. *FEBS Lett.* 579, 1772–1778.

122. Watson, J., Hopkins, N., Roberts, J., Steitz, J. A. and Weiner, A., 1987. *Molecular Biology of the Gene (4th Ed.).* The Benjamin/Cummings Publishing Company, Inc.

123. Webb, C. H. and Lupták, A., 2011. HDV-like self-cleaving ribozymes. *RNA Biol.*, 8, 719-727.

124. Weinberg, Z., Kim, P. B., Chen, T. H., Li, S., Harris, K. A., Lünse, C. E. and Breaker, R. R., 2015. New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nat. Chem. Biol.*, 11, 606-610.

125. White III, H. B., 1976. Coenzymes as fossils of an earlier metabolic state. *J.Mol. Evol.*, 7, 101-104.

126. Wilson, T. J. and Lilley, D. M. J., 2011. Do the hairpin and VS ribozymes share a common catalytic mechanism based on general acid-base catalysis? A critical assessment of available experimental data. *RNA*, 17, 213-221.

127. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrherin, C., Hartsch, T. and Ramakrishnan, V., 2000. Structure of the 30S ribosomal subunit. *Nature*, 407, 327-339.

128. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. and Breaker, R. R., 2004. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, 428, 281-286.

129. Woese C., 1967. *The Genetic Code: The Molecular Basis for Genetic Expression.* Harper & Row; New York, 179–195.

130. Yan, C., Hang, J., Wan, R., Huang, M., Wong, C.C. and Shi, Y., 2015. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science*, 349, 1182-1191.

131. Zhang, J. and Ferré-D'Amaré, A. R., 2015. Structure and mechanism of the T box riboswitches. *WIREs RNA*, 6, 419-433.

132. Zhao, C. and Pyle, A. M. 2017. Structural insights into the mechanism of group II Intron Splicing. *Trends Biochem. Sci.*, 6, 470-482.

Chapter - 2

CRYSTAL STRUCTURE OF THE VS RIBOZYME: ARCHITECTURE OF AN RNA ENZYME

CONTENTS OF THIS CHAPTER HAVE BEEN PUBLISHED AS AN ARTICLE IN NATURE CHEMICAL BIOLOGY (***Nat. Chem. Biol.*** **2015, 11, 840–846**). *ALL MATERIALS OF THE ARTICLE HAVE BEEN ADAPTED WITH THE COPYRIGHT PERMISSION FROM THE NATURE PUBLISHING GROUP.*

2.1. INTRODUCTION

2.1.1 Biological context

The Varkud satellite (VS) ribozyme is part of a ~0.9 kb transcript generated from a plasmid (VS plasmid) of the same length that is found in several isolates of *Neurospora* (Saville and Collins, 1990; Kennell et al., 1995). The VS plasmid is possibly a satellite of a more abundant ~3.7 kb Varkud (V) plasmid (named after a small village in Kerala, India where the *Neurospora* isolates were discovered) that also encodes a reverse transcriptase (RT) instrumental in synthesizing cDNA from the transcripts of this plasmid (Kennell et al., 1995; Wilson and Lilley, 2015). The VS plasmid produces a ~900 nucleotide VS RNA that includes ~200 nucleotide VS ribozyme motifs separated by ~700 nucleotide stretches of transcript. The existence of the ribozyme motif was a serendipitous discovery by Saville and Collins (Saville and Collins, 1990), when they found that this motif was capable of site-specific autocatalytic cleavage and ligation. This activity was thought to be useful in resolving circular concatemeric intermediates in the rolling-circle replication of the VS plasmid (Figure 2.1). VS plasmid is transcribed by mitochondrial RNA polymerase to general multimeric transcripts harboring multiple ribozyme motifs that generate cuts along the transcript to produce linear RNA. This is followed by autocatalytic ligation at the same site to generate circular RNA, the predominant form in which VS RNA is found in biology.

Circular monomers are reverse transcribed by mitochondrial RT to generate VS plasmids completing the replication cycle. The process is similar to rolling circle replication of other viral satellite plasmids, the difference is that VS RNA is transcribed from templates that are DNA and not RNA (Kennell et al., 1995).

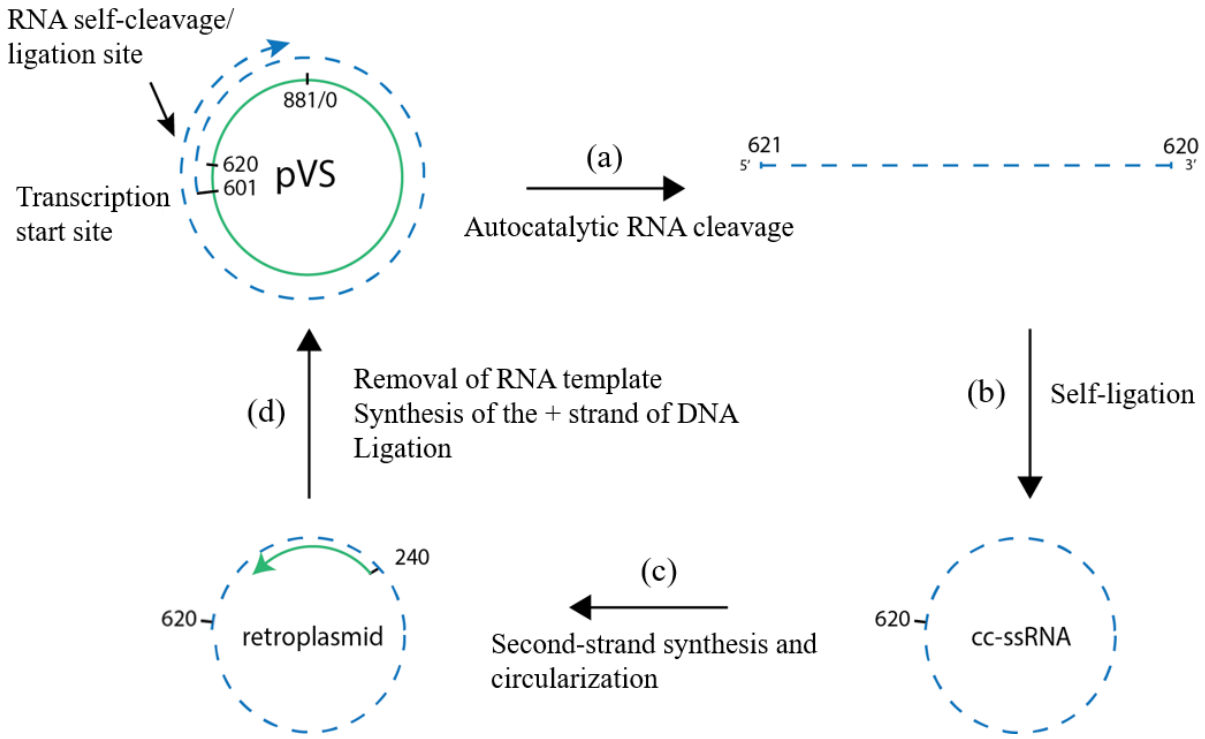

Figure 2.1 Replication cycle of the VS plasmid. The Varkud satellite plasmid (pVS) is transcribed by the mitochondrial RNA polymerase from a *Neurospora* mtDNA promoter, with the major start site corresponding to position 601 (top-left). In the absence of a transcription termination sequence, RNA polymerase synthesizes a single continuous transcript containing multiple repeats of the 881 nucleotide pVS sequence. Ribozyme-catalyzed self-cleavage (a) and ligation (b) at position 620 produce circular single stranded RNA equivalent in length to pVS. VS RNA is then reverse transcribed by a RT from the V plasmid (c) starting from position 240 to yield full-length (-) stand cDNA. RNA template is then displaced or degraded, allowing for the synthesis of (+) strand DNA (d). Adapted from Suslov, 2012.

2.1.2 Varkud satellite ribozyme: pieces of a puzzle

2.1.2.1 Structural features

The study of VS literature provides another fascinating glimpse into the progress of science and the scientific method that is filled with unexpected discoveries, unusually accurate predictions driven by sharp intuition, and occasional false starts and dead ends. The last catalytic RNA discovered by 'accident', the VS ribozyme was part of an 881-nucleotide transcript that could catalyze site-specific RNA strand scission *in vitro* when it was discovered in 1990 (Saville and Collins, 1990). The ability of the VS ribozyme to ligate at the same location supported its potential biological role in rolling circle replication of the VS plasmid (Saville and Collins, 1991). Its catalytic function was triangulated to a 154-nucleotide motif (within the 881 nucleotide VS RNA transcript), with 153 nucleotides downstream of the cleavage site (making the VS ribozyme the largest endonucleolytic RNA motif in biology) and only a single upstream nucleotide whose identity is irrelevant to cleavage activity (Guo et al., 1993). This observation made the VS ribozyme more tractable and laid the groundwork for future functional investigations.

Preliminary kinetic analyses established a requirement of Mg^{2+} in VS cleavage, in addition to demonstrating that a single VS ribozyme molecule is capable of self-cleavage (i.e. cis cleavage) (Collins and Olive, 1993). Through mutagenesis and structure probing by chemical modification provided the next major step in understanding the VS ribozyme structure. Extensive site-directed mutations to the ribozyme sequence in isolation and in combination to other complimentary mutations provided the first functional map of the ribozyme sequence. Mutagenesis, coupled with secondary structure probing by a battery of chemicals that modify nucleobases depending on their extent of solvent exposure, it was possible to develop an initial secondary structure of the VS ribozyme (Figure 2.2A). This consisted of six primarily base-paired helices interrupted by internal

bulges and closed by terminal loops, including the helix that contains the ribozyme cleavage site (Beattie at al., 1995). Several Mg^{2+}-dependent interactions occur during VS folding to create the active site suggesting structural complexity in its three-dimensional fold. A rate-determining conformational change in the RNA was proposed to be important for VS catalysis (Beattie at al., 1995). The substrate-helix when physically disconnected from the rest of the ribozyme (referred to as the catalytic domain) can be cleaved *in trans* with cleavage rates that are higher than that for *cis* cleavage (0.7-1/min versus 0.08-0.12/min) (Figure 2.2B) (Beattie at al., 1995; Guo and Collins, 1995). Interestingly the substrate sequence was identified to be a stable stem-loop structure with the loop nucleotides being important for cleavage. In the absence of single-stranded regions capable of base-pairing in the stems of either the substrate or the catalytic domain, substrate recognition and binding was expected to occur via multiple tertiary interactions, a suggestion supported by low K_m values for catalytic domain-substrate association. This proposed interaction was found to be mediated by a pseudoknot between stem-loops of helix 5 and the helix 1 (substrate-helix), consisting of three WC base-pairs (Rastogi et al., 1996). Damage selection experiments localized important residues to a subset of the VS sequence and identified sequences upstream (part of helix 1a) of the cleavage site that attenuated cleavage (Beattie and Collins, 1997).

VS cleavage proceeds efficiently in the presence of molar concentrations of monovalent cations like Li^+, NH_4^+ and trivalent $Co(NH_3)_6^{3+}$, a mimic of a hydrated Mg^{2+} (Murray et al. 1998), suggesting structural and not catalytic roles for Mg^{2+} in those conditions. Cleavage rates were drastically enhanced (from 0.08/min to ~5/min) in the presence of spermine, a polyamine that potentially assists in folding by stabilizing the polyanionic backbone of RNA (Olive and Collins, 1998). This fast cleavage was in fact a *trans* reaction between two complete ribozymes (substrate plus catalytic domain) as established by mutation complementation studies. A symmetric

association (Figure 2.2C) was proposed where the catalytic domain from one ribozyme binds and cleaves the substrate-helix from the other via the formation of a dimer with two active sites created *in trans* (Olive and Collins, 1998). A kissing-loop interaction between substrate and stem-loop 5 was thought to be responsible for this dimerization (Figure 2.2C). Although this suggestion was not supported by experimental evidence at that time, crystal structures described in this work confirm this dimeric model for trans reactivity (section 2.2.3, Figure 2.12).

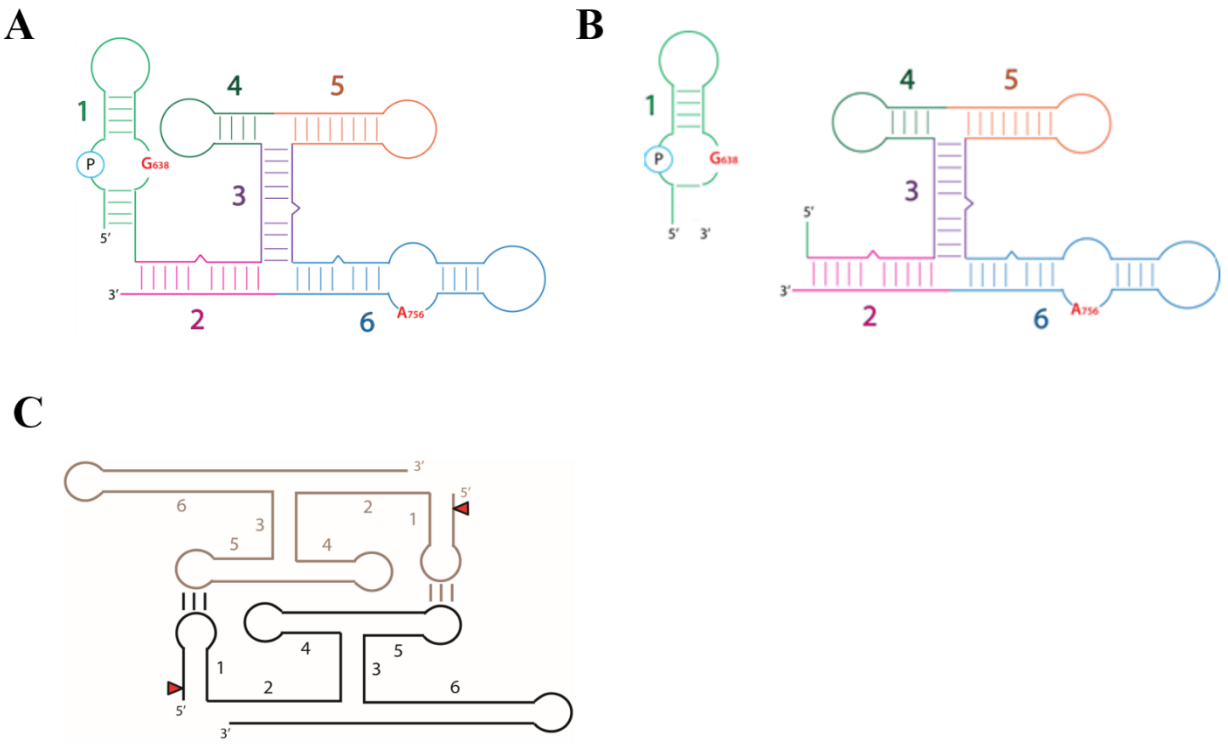

Figure 2.2 VS ribozyme: cis, trans and trans dimer forms. A. The canonical cis-form of the ribozyme containing helices 1-6. **B.** The canonical trans-form of the ribozyme containing helices 2-6 and physically disconnected helix 1. **C.** Symmetric dimer model for trans activity as proposed by Olive and Collins, 1998. Red triangles indicate cleavage site and kissing loop interactions are shown by parallel lines.

Preference for trans catalysis and the mode of substrate binding is unique to the VS ribozyme. Other self-cleaving ribozymes like the hammerhead and hairpin can be converted to trans cleaving systems by breaking the secondary structure (base-paired helix) that is created by the substrate and the rest of the ribozyme. However, the VS ribozyme appears to be active *in trans* without any such engineering. This places it in an elite group of 'true RNA enzymes' (capable of multi-turnover) with RNase P, spliceosome and the ribosome. The catalytic sequence of the VS ribozyme was further abridged by trimming parts of helices 1, 2, 4 and 6, leading to increase in reactivity. This was primarily due to the deletion of cleavage attenuation elements upstream the cleavage site and helix 2 nucleotides at the junction of helices 1 and 2 (Rastogi and Collins, 1998). These alterations presumably make the connection between substrate and catalytic domain more flexible leading to more efficient substrate docking. Nucleotides upstream of the cleavage site constitute a base-paired helix that is part of the wild-type VS sequence (Jones et al., 2001). This helix, named helix 7 was eliminated as a consequence of truncation studies performed to identify the minimal ribozyme sequence capable of catalysis. The sequences in helix 7 can likely adopt two distinct secondary structures and the toggling between the two involves nucleotides from junction 1-2 (Figure 2.3). This conformational switch between helices 2 and 7 possibly plays a vital role in regulating the cleavage-ligation equilibrium.

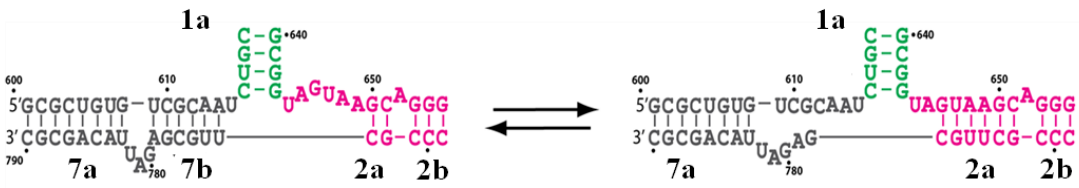

Figure 2.3 Secondary structure rearrangement in helices 7b and 2a. Conformational rearrangement in junction 1-2-7 via structural rearrangements of helices 7b and 2a as proposed by Jones et al., 2001. The structure on the left that consists of a tighter junction with a fully formed helix 7b (including cleavage attenuation elements upstream of the cleavage site) favors ligation, whereas the structure on the right favors cleavage with a looser junction. Adapted from Suslov, 2012.

Paralleling observations with truncated constructs of the hammerhead and hairpin ribozymes, the minimal VS construct favors cleavage over ligation, but full-length VS (that includes helix 7) favors ligation similar to the full-length versions of both hammerhead and hairpin ribozymes (Lilley and Eckstein, 2008). The propensity of wild-type sequences to favor ligation is perhaps reflected in the preponderance of the ligated circular form of these RNAs in biology (Kennell et al., 1995; Saville and Collins, 1990). In addition to regulating the cleavage-ligation equilibrium in VS catalysis, the conformational flexibility of the three-way junction formed by helices 1, 2 and 7 is likely important for controlling cis versus trans docking of the substrate that has profound implications in VS biology (see chapter 3 for a model outlining the role of VS catalysis in the resolution of circular RNA intermediates in the rolling-circle replication of the VS plasmid).

2.1.2.2 Chasing the active site

With the structural features of the functional fold of the VS ribozyme largely mapped out, efforts were directed toward identifying the active site of the ribozyme. This identified hot spots contained one or more nucleotides that when mutated resulted in considerable reduction in VS cleavage. These include a bulged adenosine in helix 2 (A652), nucleotides at a three-way junction formed by helices 2, 3 and 6 that includes the two unpaired adenosines, an unpaired guanosine, an unpaired adenosine and three nucleotides (identities not important) just upstream of it (Lafontaine et al., 2001b) (Figure 2.4). Despite the dispensability of a major part of helix 6, two internal bulges in this helix were identified to harbor nucleotides that appeared to be important for activity. The first, referred to as the A_2 bulge contains two tandem adenosines. Deleting both nucleotides, resulting in the removal of this bulge leads to substantial loss of activity (Lafontaine et al., 2001b), although the identities of these nucleotides are mutable. This observation discounts the possibility

of catalytic contributions of these residues. The second bulge referred to as the A730 loop, appeared to be essential for VS cleavage (Figure 2.4).

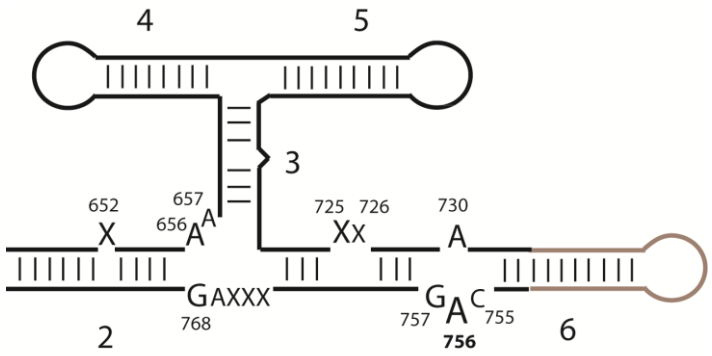

Figure 2.4 Secondary structure of the VS ribozyme circa 2001. Secondary structure proposed by Lafontaine et al., 2001b derived from biochemical and biophysical data. The overall structure primarily consists of base-paired helices, interspersed with conserved nucleotides in internal bulges. 'X' indicates that the presence of a nucleotide is necessary at that location; however, identity is not important. The size of the letters representing the nucleotides are proportional to their importance for VS activity.

Mutating nucleotides belonging to the A730 loop resulted in cleavage defects to different degrees; however, mutations to one residue, A756, produced the largest reduction in cleavage. A756 mutants did not have any effect on substrate binding or global folding of the VS ribozyme which hinted at the possibility of a catalytic role for A756 (Lafontaine et al., 2001b). Results supporting this claim came from cross-linking experiments, where the nucleotide just downstream of the scissile phosphodiester (A621) was replaced with a 4-thiouridine nucleotide. Under cleavage conditions this 4-thio-U nucleotide cross-linked to A756 in both cleaved and ligated forms of the VS ribozyme (Hiley et al., 2002). This observation provides strong evidence for the involvement of A756 in forming the active site. Base analog substitutions to A756 revealed the importance of the Watson-Crick (WC) edge of the adenine, especially the imino N1 atom, in VS catalysis

(Lafontaine et al., 2002b). Direct involvement in catalysis requires the adenine to take part in proton transfer. A756 displayed a pH-dependent interference pattern when replaced by nucleotide analogs in nucleotide analog interference mapping (NAIM) experiments. Ligation defect in mutants containing 8-azaadenosine (p*K*a 2.2) and purine riboside (p*K*a 2.1) instead of A756 were rescued at lower pH (Jones and Strobel, 2003). These observations suggest that ionization of the A756 nucleobase probably plays an important role in catalysis. Although the ligation defect of an abasic A756 mutant could not be rescued by adding exogenous adenine or imidazole (a moiety often used by protein enzymes for general acid-base catalysis), substituting A756 with an imidazole analog resulted in partial rescue of trans cleavage (Zhao et al., 2005).

A756 was found to be important for trans ligation as well (McLeod and Lilley, 2004). Studying the pH dependence of ligation rates revealed the importance of a protonation/deprotonation event involving a moiety with a pKa of 5.6. Although the pKa of free adenine is ~3.5, a shift of 2 pKa units within folded RNA is not unprecedented therefore, these results present A756 as a strong candidate for participating in nucleobase catalysis. pH dependence on cis cleavage revealed another ionization event involving a group with a pKa of 8.3, in addition to one with pKa = 5.8 (Smith and Collins, 2007). The unperturbed pKa of guanine (9.2) is closest to 8.3 and its pKa can conceivably be lowered by its local environment. A bell-shaped dependence of cleavage rate on pH is indicative of two nucleobase ionizations hence this result suggested a general acid-base mechanism involving two nucleobases in VS catalysis. One candidate, G638 is a conserved nucleotide in an internal bulge within the substrate (called the internal cleavage loop as it also contains the cleavage site). Mutating this guanine to any other natural nucleotide lowers cleavage rates by four orders of magnitude underscoring its importance. Base analog substitutions revealed the importance of the imino N1 atom and keto group which provides further

circumstantial evidence for direct roles of G638 in VS catalysis (Wilson and Lilley, 2007; Jaikaran et al., 2008). The identification of nucleotides A756 and G638 as the ionisable groups corresponding to pKa values of 5.6-5.8 and 8.3 respectively, suggested probable roles in general acid/base catalysis; however, the exact nature of proton transfer was inconclusive (Jaikaran et al., 2008). Since the usual role of a general acid is to stabilize the 5'- oxygen leaving group thereby assisting in cleavage of the covalent bond between it and phosphorous atom of the scissile phosphate, artificial labilization of this bond is expected to rescue the cleavage defect of a VS ribozyme construct that has a mutant general acid nucleotide. Replacing the 5'-oxygen atom with sulfur increases the P-O (S) bond length and weakens it considerably. This leads to an artificially labile bond, thereby lowering its dependence on proton transfer by a general acid. An expected rescue of cleavage defect was observed in a VS construct with an A756G mutation (Wilson et al., 2010), providing strong evidence that A756 acts as a general acid in VS catalysis. This leaves G638 to act as a general base, although in the absence of a similarly decisive experiment, its role is still debatable.

2.1.3 Global fold

Structural features responsible for assembling the catalytic apparatus in the VS ribozyme are organized by three three-way junctions. Lilley and co-workers developed a low-resolution model of the global fold of the complete VS ribozyme (helices 1-7), using a 'divide and conquer' approach. With the secondary structures of the three-way junctions at hand, they investigated ion-dependent folding of these junctions using electrophoretic mobility and FRET (Forster resonance energy transfer)-based assays. They introduced asymmetry to these junctions by artificially elongating two of the three helices allowing distinction between different configurations of the

junction (Lafontaine et al., 2001a). For example, junction configurations in which the two long helices subtend an obtuse angle or are stacked (subtending an angle of 180° between them) travel faster than configurations in which they subtend an acute angle. Comparative gel mobility assays revealed a Mg^{2+}-induced conformation change in junction 2-3-6 that results in a coaxial stacking between helices 3 and 6, with helix 2 protruding out from the junction titled toward helix 6. This observation was supported by FRET studies. A similar approach was taken to elucidate the structure and folding of junction 3-4-5 (Lafontaine et al., 2002a). Helices 3 and 4 appeared to be coaxially stacked in the presence of Mg^{2+} with helix 5 protruding out of the stack toward helix 3. Helix 3 clearly acts as the 'axle' around which both junctions are 'fitted'; therefore, this helix was used as a reference to bring the global structures of junctions 2-3-6 and 3-4-5 together to define the global fold of the catalytic domain (helices 2-6) of the VS ribozyme (Figure 2.5A) (Lafontaine et al., 2002a). The cleft created by helices 2 and 6 that are appropriately oriented by junction 2-3-6, was proposed to be the site of substrate docking. This would bring the cleavage site in close proximity to the catalytically important A730 loop that contains the putative general acid A756 (Lafontaine et al., 2001b). Junction 3-4-5 orients helix 5 to interact with the docked substrate, stabilizing it in the catalytic cleft (Lafontaine et al., 2002a). A model of the substrate-helix obtained from NMR (Flinders and Dieckmann, 2001) was successfully docked into a model of the catalytic domain discussed above, underscoring its viability (Figure 2.5A) (Lafontaine et al., 2002a). Point mutations to the residues at these junctions result in cleavage defects and disrupt native folding of domains spatially separated from these junctions (Sood and Collins, 2001). Lilley and co-workers went further in investigating the VS global fold by including helix 7 and the substrate as part of their cis ribozyme system. This sequence representing the VS catalytic motif in it's entirely, was studied by small-angle x-ray scattering (SAXS) (Lipfert et al., 2008). The global structure of the

VS ribozyme as revealed by SAXS consists of a central coaxial stack created by helices 3 and 4 and a small segment of helix 6 with the rest of the helix bent at an angle of 130°. Helices 2 and 5 emanate from two opposite sides of the central stack, with the latter making an angle of 40° with helix 3 (Figure 2.5B). This architecture likely creates a suitable environment for the substrate to dock into and assembles the VS active site by bringing the putative catalytic nucleobases G638 (as a part of the substrate) and A756 (as a part of helix 6), and the cleavage site in close proximity to each other (Figure 2.5B). The model generated from SAXS data is the first model to define the spatial orientation of the substrate relative to the catalytic domain and describe the global fold of junction 1-2-7, although the absence of higher resolution data rendered these assignments provisional.

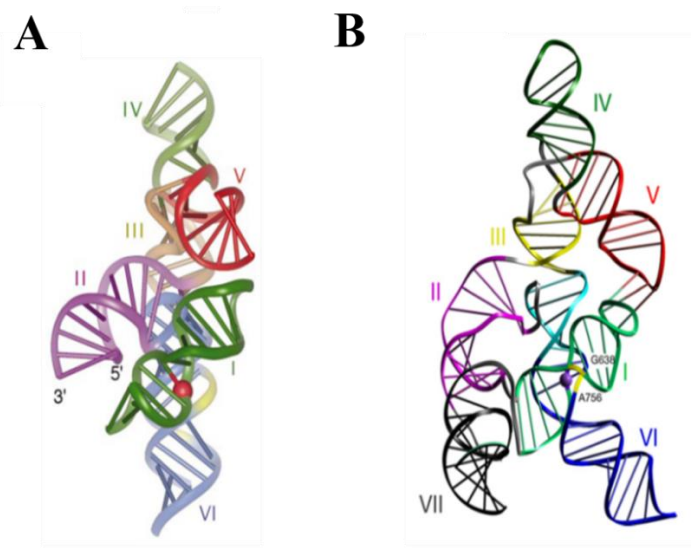

Figure 2.5 Low resolution models of the global structure of the VS ribozyme. A. Global structure of the catalytic domain (helices 2-6) derived from electrophoretic mobility and FRET-based assays. Substrate-helix obtained from NMR has been docked into this catalytic domain. **B.** Global structure obtained from SAXS. Reprinted with permission from Macmillan Publishers Ltd., EMBO J (Lafontaine et al., 2002a), copyright 2002 and Elsevier Ltd., Structure (Lipfert et al., 2008), copyright 2008.

The modular nature of these junctions highlighted by their ability to fold into their native conformations independent of context (Lafontaine et al., 2001a; Lafontaine et al., 2002a), is illustrated by the ability of the catalytic cleft created by junction 2-3-6 to bind the substrate independent of junction 3-4-5. A minimal construct where helices 3, 4 and 5 are replaced by a single stem-loop structure that preserves junction 2-3-6, retains cleavage activity, when supplemented with higher Mg^{2+} concentrations (Sood and Collins, 2002). However, deleting this stem-loop thereby removing junction 2-3-6 abolishes activity, highlighting the role of junction 2-3-6 in substrate binding (Sood and Collins, 2002). A simple structure similar to a stem-loop preserving the structural features of the junction could have given rise to a more sophisticated apparatus (helices 4 and 5 held together by junction 3-4-5) that utilized tertiary kissing-loop interactions specific to the VS substrate sequence for substrate binding (see chapter 4).

2.2. RESULTS

2.2.1 Isolating a homogenous dimer population

One of the major breakthroughs in molecular biology came in the form of *in vitro* transcription, a technique that enables production of abundant quantities of RNA from corresponding DNA templates by the action of viral RNA polymerases in a relatively short time. Transcribed RNA is generally purified by denaturing polyacrylamide gel electrophoresis (dPAGE). Since this process disrupts the three-dimensional structure of RNA, a suitable refolding step precedes crystallization trials in order to capture the functionally relevant conformation of the RNA. However, RNA is known to adopt multiple non-functional conformations in addition to its functional fold and denaturation-renaturation cycles have been known to introduce conformational heterogeneity to the folding landscape of RNA (Edwards et al., 2009; Golden and Kundrot, 2003; Ke and Doudna, 2004; Wong et al., 2007). Conformational heterogeneity is detrimental to efficient crystal packing therefore confirming the folding landscape of the RNA sample prior to setting up crystallization trials is essential.

VS ribozyme constructs (Figure 2.6A) to be used for crystallization exhibited conformational heterogeneity after going through denaturation-renaturation steps. Native size exclusion chromatography (nSEC) traces indicated that the VS sequences populated multiple conformations on refolding after dPAGE purification (Figure 2.7A). These included VS dimers, oligomers and other undesirable aggregates. Dimerization and oligomerization during transcription (Saville and Collins, 1990) and in the presence of positively charged species like viomycin (Olive et al., 1995) and spermine (Olive and Collins, 1998) has been observed for wild-type VS ribozymes. To bypass the renaturation steps, we adopted an FPLC-based method for purifying transcribed RNA that avoided denaturation entirely by using high resolution nSEC. This

reduced the conformational heterogeneity to three distinct species that were identified as monomer, dimer and oligomer of dimers, respectively (Figure 2.7B). Light scattering experiments confirmed the identities of the monomer and dimer peaks (Table 2.1),

Technique	Peak 1 (M. W. in kDa)	Peak 2 (M. W. in kDa)
Static light scattering	153±3.2	71±1.9
Dynamic light scattering	142±4.0	68±2.0
Calculated (IDT Oligoanalyzer)	~114	~57

Table 2.1 Analysis of peaks by static and dynamic light scattering (Suslov, 2012).

Interestingly, substituting the non-essential wild-type GUAC tetraloop of stem-loop 4 with an AAACA pentaloop, an epitope for fragment of antigen binding (Fab), BL3-6 (Koldobskaya et al., 2011) to facilitate crystallization (Figure 2.6B), eliminated the aggregation problem and allowed us to isolate the monomer and dimer peaks while retaining cleavage activity (Figure 2.7C, E). dPAGE analysis showed the presence of both precursor and product forms of the ribozyme in the monomer fraction; however, the dimer fraction predominantly consisted of precursors (Figure 2.7D). The preponderance of monomers in product RNA suggests reduced affinity of the cleavage product toward dimerization. Cleaved RNA was found to dimerize >10-fold slower than its uncleaved counterpart, explaining the difference in precursor/product concentrations in monomers versus dimers (Suslov, 2012). The presence of both precursor and cleavage products in crystallization drops as a result of self-cleavage *in situ* is undesirable as it could inhibit formation of crystal lattice, thus preventing crystallization.

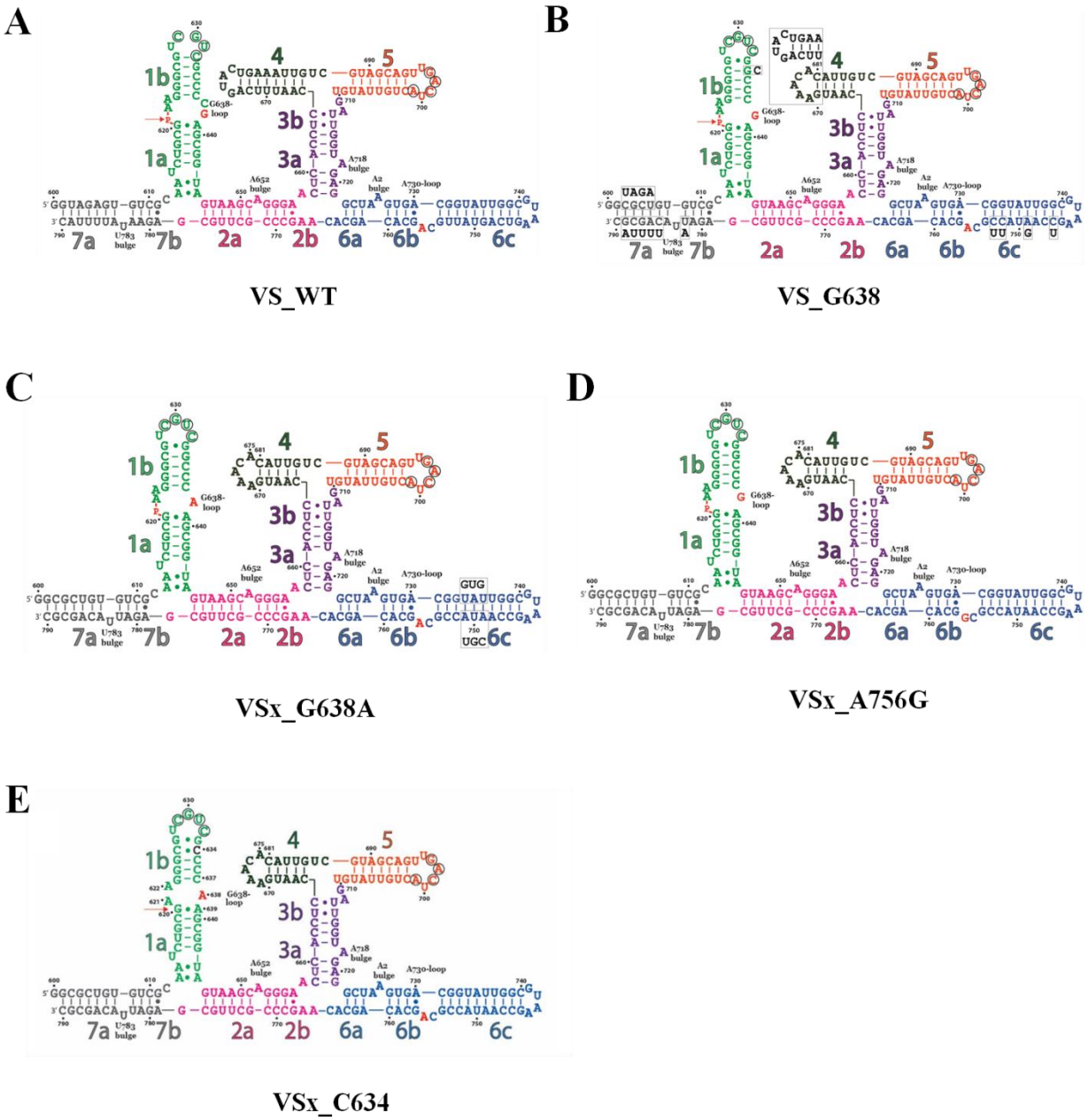

Figure 2.6 Crystallization constructs of the VS ribozyme. A. Secondary structure of the *Neurospora intermedia* wild-type VS ribozyme in the canonical representation. Numbers indicate helices according to the traditional VS nomenclature; letters indicate sub-helices. Internal loops and bulges are labeled for clarity. The scissile phosphate P is shown in red. The ribozyme is depicted in the colors used for subsequent figures. Nucleotides participating in the loop 1 – loop 5 kissing interaction are circled. **B.** Secondary structure of the VS ribozyme construct that is identical to the crystallization constructs VSx_G638A and VSx_A756G but contains wild-type catalytic nucleobases G638 and A756 (shown in red); the wild-type sequences, shown in black and boxed, are mapped onto the secondary structure. Nucleotide numbering as in A. Mutations made in the VSx_G638A_tGU construct are shown in black, boxed and mapped onto the secondary structure.

Figure 2.6 Crystallization constructs of the VS ribozyme (continued). C. Secondary structure of the crystallization construct VSx_G638A containing a wild-type A756 and a G638A mutation. **D.** Secondary structure of the crystallization construct VSx_A756G containing a wild-type G638 and an A756G mutation. **E.** Secondary structure of the crystallization construct VSx_C634. C634 is shown in black.

General strategies to prevent spontaneous cleavage in catalytic RNA have included mutations to the catalytic nucleobases (Ferré-D'Amaré et al., 1998; Salter et al., 2006), replacing the 2'-oxygen nucleophile with 2'-O-methoxy or 2'-deoxy groups (Rupert and Ferré-D'Amaré, 2001; Martick and Scott, 2006; Liu et al., 2014; Ren et al., 2016,) or designing an unreactive RNA backbone (Torelli et al., 2007; Rupert et al., 2002; Mir and Golden, 2016). The latter strategies involve introduction of non-natural RNA sequences into crystallization constructs that require the ligation of RNA oligonucleotides containing these modifications to *in vitro* transcribed/chemically synthesized RNA. The products of these ligation reactions have to be purified by dPAGE for efficient separation of the desired product from side-products of ligation, consequently introducing denaturation-renaturation steps. In light of these potential pitfalls, we decided to make complimentary mutations to the catalytic nucleobases to inactivate self-cleavage. Coupling the information derived from these two constructs, one with a G638A (putative general base) mutation, but a wild-type A756 (putative general acid) referred to as VSx_G638A, and the other with a wild-type G638 but an A756G mutation VSx_A756G would provide a near-complete picture of the active site (Figures 2.6C, D, 2.8C, D). Catalytic nucleobase mutants primarily produced dimers (Figures 2.7F, 2.8B) that were kinetically stable at room temperature when purified by nSEC, unlike wild-type dimers that gradually converted to monomers when incubated for the same duration (Figure 2.9A). Similar results were obtained for another crystallization construct VSx_C634 (Figures 2.6E, 2.8C) (section 2.2.2 and chapter 3).

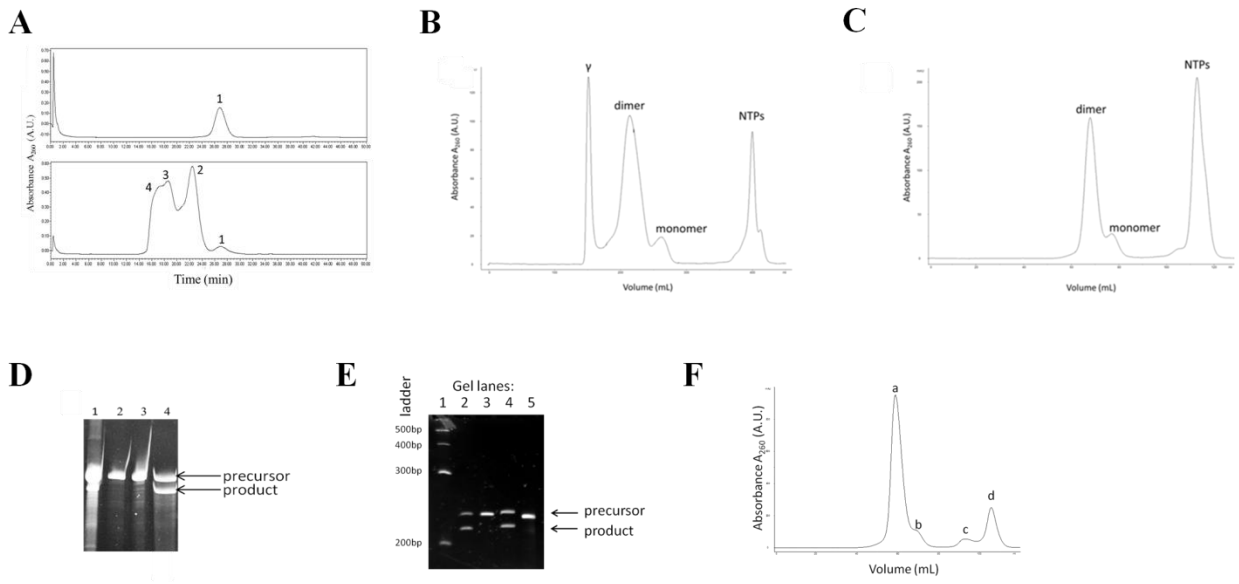

Figure 2.7 Exploring the conformational landscape of the VS ribozyme. A. nSEC trace of the VS ribozyme after denaturation (top) and denaturation followed by renaturation (bottom). RNA conformation corresponding to peak 1 elutes at the same time in both cases, indicating that peaks 2-4 are due to conformational heterogeneity introduced by renaturation. **B.** Native purification of wild-type VS ribozyme captures three distinct species – monomer, dimer and oligomer formed by the aggregation of dimers (γ). **C.** Native purification of a VS ribozyme construct with stem-loop GUAC replaced by AAACA yields two species– monomer and dimer. The abolition of aggregation with this stem-loop substitution suggests possible roles of the wild-type stem-loop 4 nucleotides in oligomerization. **D.** dPAGE (8%) analysis of monomer, dimer and oligomer γ from B. The lanes were loaded as follows: 1 – input, 2 – oligomer γ, 3 – dimer, 4 – monomer. The monomer contains both precursor and product states, but the dimer is exclusively the precursor. **E.** Shortening of helix 4 and AAACA hairpin graft do not interfere with VS cleavage/ligation equilibrium. VS transcription reactions were loaded on an 8% dPAGE and stained with EtBr. The lanes were loaded as follows: 1- RNA ladder (NEB), 2 - VS_WT, 3 - VS_G638A, 4 - VS_G638, 5 - crystallization construct VSx_G638A. The slower migrating species corresponds to the precursor form of the ribozyme (scissile bond intact), whereas the faster migrating species corresponds to the cleaved form of the ribozyme (scissile bond broken). G638A mutants (lanes 3 and 5) contain only the precursor form of the ribozyme. **F.** VS dimer was purified directly from transcription reactions by size exclusion chromatography under native conditions at 4°C. The chromatogram shows four distinct peaks in order of size (left-to-right and largest-to-smallest): (a) VS dimer, (b) VS monomer, (c) DNase I digestion products of the transcription template and (d) NTPs.

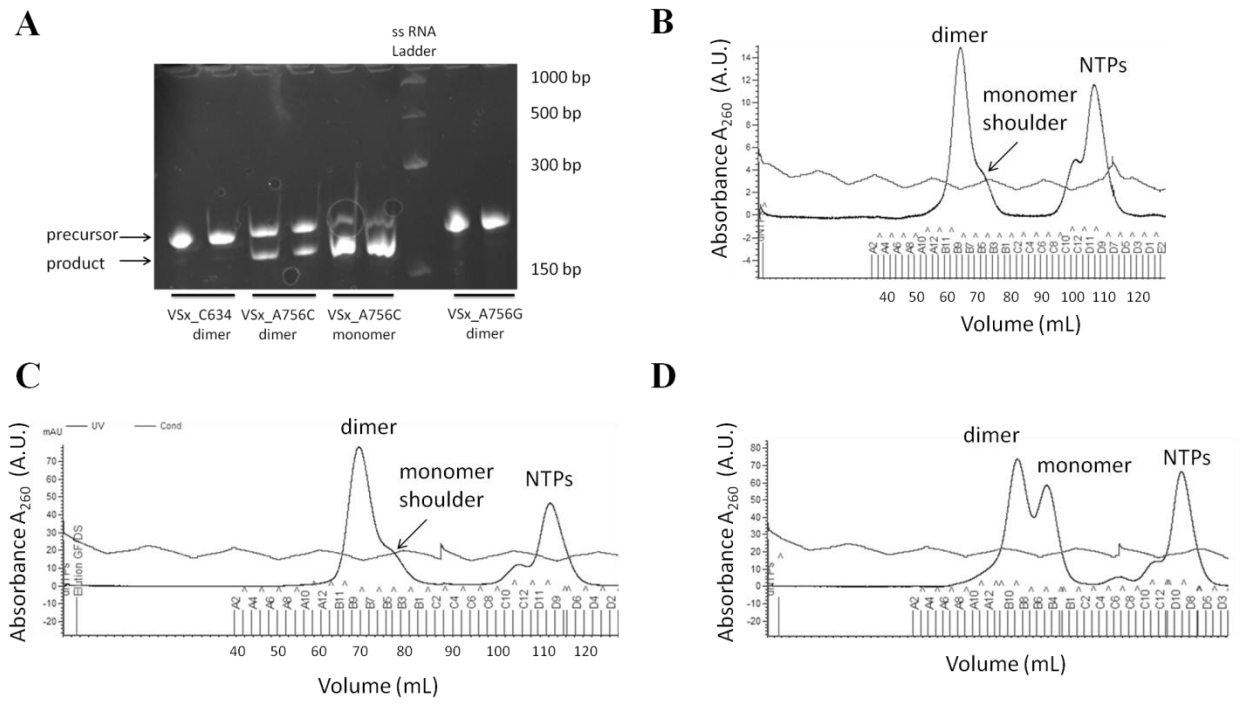

Figure 2.8 Monomer-dimer equilibrium for VSx_A756G, VSx_A756C and VSx_C634. A. dPAGE (10%) analysis of crystallization constructs from *in vitro* transcription. VSx_C634 and VSx_A756G show negligible cleavage and populate dimer states. VSx_A756C populates both monomeric and dimeric states. VSx_A756C consists of both precursor and product RNAs with dimers comprising ~1:1 precursor: product while monomers primarily comprised of products (section 2.2.2). **B.** Native purification of VSx_A756G primarily captures the dimer with a monomer shoulder that was not collected during pooling fractions from FPLC. **C.** Native purification of VSx_C634 primarily captures the dimer with a monomer shoulder that was not collected during pooling fractions from FPLC. **D.** Native purification of VSx_A756C captures both dimer and monomer species. Only elution fractions corresponding to the dimer peak were collected.

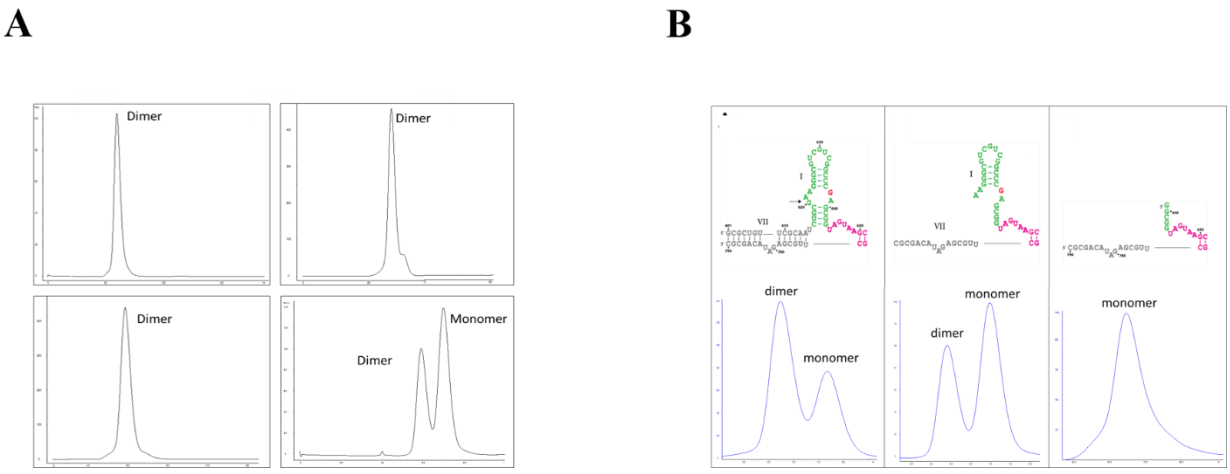

Figure 2.9 Stability of the dimer and effect of helices 1 and 7 on dimerization. A. VS_G638A (left) and VS_G638 (right) diluted to 0.01 μM before (bottom) and after (top) incubation at room temperature for 96 h. Ribozyme with wild-type active sites self-cleave during the incubation. Products have cleaved 5' ends hence preferably populate the monomeric species. **B.** Deleting the 5' end of the substrate-helix 1a and helix 7 (cleavage attenuation elements) shifts the equilibrium toward monomer (VS_G638_Δ1aΔ7). Complete deletion of the substrate stem-loop inhibits dimerization to yield monomers exclusively (VS_G638_Δ1aΔ1bΔ7). Helices 2b - 6 are omitted for simplicity (Suslov, 2012).

In the light of intermolecular interactions between two complete *cis*-ribozymes via interactions between stem loop 5 and substrate stem-loops from both monomers (Olive et al., 1995; Rastogi et al., 1996, Oullet et al., 2009), we investigated the role of helix 1b in dimerization. Shortening the 5' end of the substrate gradually convert the dimeric species to monomers suggesting its role in dimerization (Figure 2.9B). We proceeded to set up crystallization trials with the kinetically stable dimeric VS ribozyme species containing mutations to either catalytic nucleobases thereby capturing the precursor forms of the ribozyme (Figure 2.9A).

2.2.2 Crystallization trials (and tribulations)

nSEC purification by FPLC provided RNA samples that were stable homogenous precursors. We screened a series of constructs with varied helix lengths with various buffer conditions to obtain a single crystal 'hit' from a condition (100 mM bis-tris pH 6.5, 2.0 M ammonium sulfate) from the INDEX screen (Hampton). RNA-Fab complexes crystallized in the same morphology but the diffraction data could not be phased using the coordinates of the Fab. The inability of Fab coordinates to provide any phasing information for the diffraction data coupled with the observation that RNA-only crystals adopted the same space group as the supposed RNA-Fab crystals led us to believe that the 'RNA-Fab' crystals were in fact just RNA crystals. We abandoned the use of Fabs and screened the 'RNA-only' VSx_G638A construct against crystallization screens to improve our resolution to 3.1/3.3/3.8 angstroms (see Materials and methods, section 7.5.1 for additional crystallization conditions) (Figure 2.10A). The VSx_A756G and VSx_A756C constructs (designed to capture an active site with wild-type general base G638) were less amenable to yielding single crystals. nSEC of VSx_A756C showed a substantial monomer peak (Figure 2.8D) that contained both precursor and product forms (Figure 2.8A), consistent with functional data (Sood and Collins, 2002; Eckstein and Lilley, 2008). This sequence heterogeneity might have contributed to the difficulties in obtaining well-ordered single crystals. We actively pursued the VSx_A756G construct as it almost exclusively populated the dimeric state and explored crystallization condition space to obtain needle-shaped or twinning crystals (Figure 2.10B). Even after improving crystal morphology the best resolution achieved from these conditions was of ~5Å (Figure 2.10C). Trials with additional conditions from additive screens (Hampton) resulted in large single crystals with well-defined edges that diffracted to ~3Å resolution (Figure 2.10D). In the absence of a Fab module of known structure to facilitate

molecular phasing, we used iridium soaking experiments to introduce iridium into rationally-designed binding sites in the RNA constructs that provide strong anomalous signal enabling solution by single-wavelength anomalous dispersion (SAD) (Figure 2.11) (Keel et al., 2007). For this purpose, we created a special VS construct with tandem GU base-pairs known to bind iridium-hexamine in helix 6 (Figure 2.6C) (Keel et al., 2007).

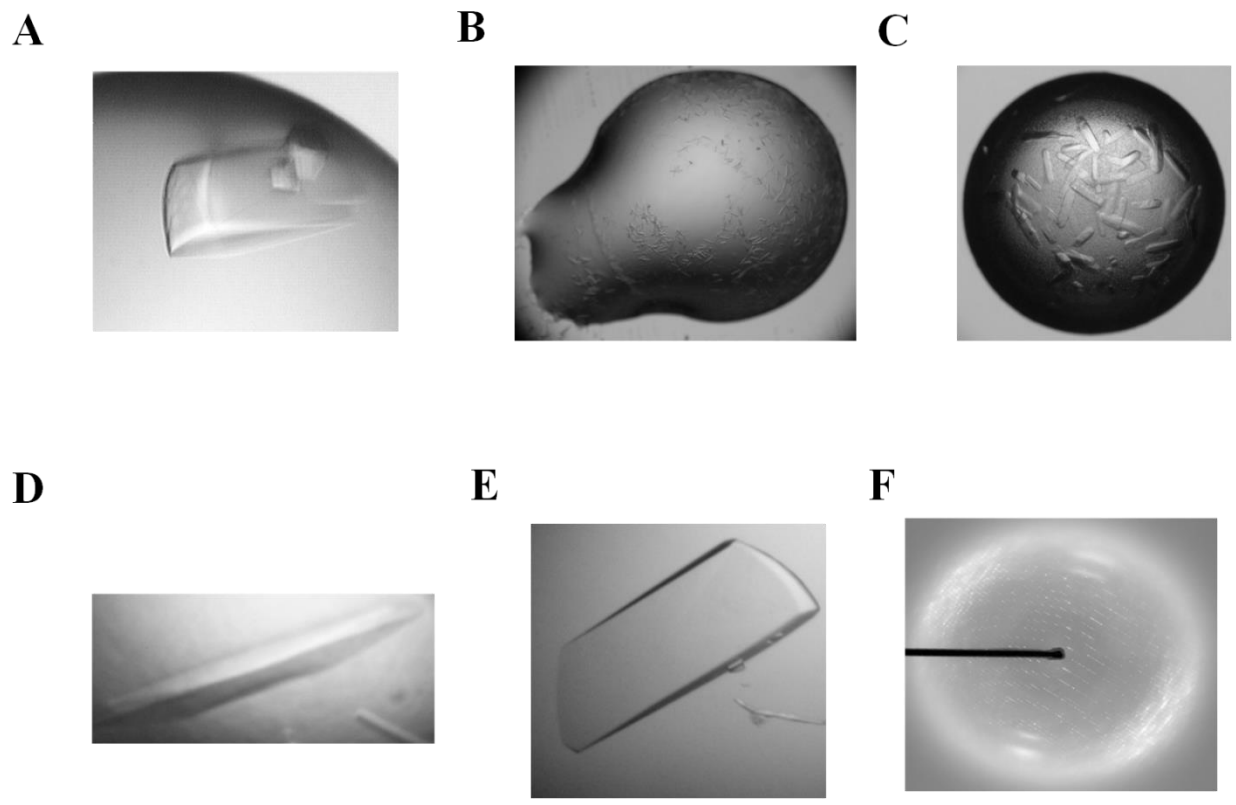

Figure 2.10 Crystallization of VS ribozyme constructs. A. Optimized single crystal of VSx_G638A (diffracted to 3.1Å). **B.** Needle-shaped crystals of VSx_A756G. **C.** Plate crystals of VSx_A756G (diffracted to 4.7Å). **D.** Optimized single crystal of VSx_A756G (diffracted to 3.1Å). **E.** Optimized single crystal of VSx_C634 (diffracted to 3.3 Å). **F.** Typical diffraction pattern of optimized crystals.

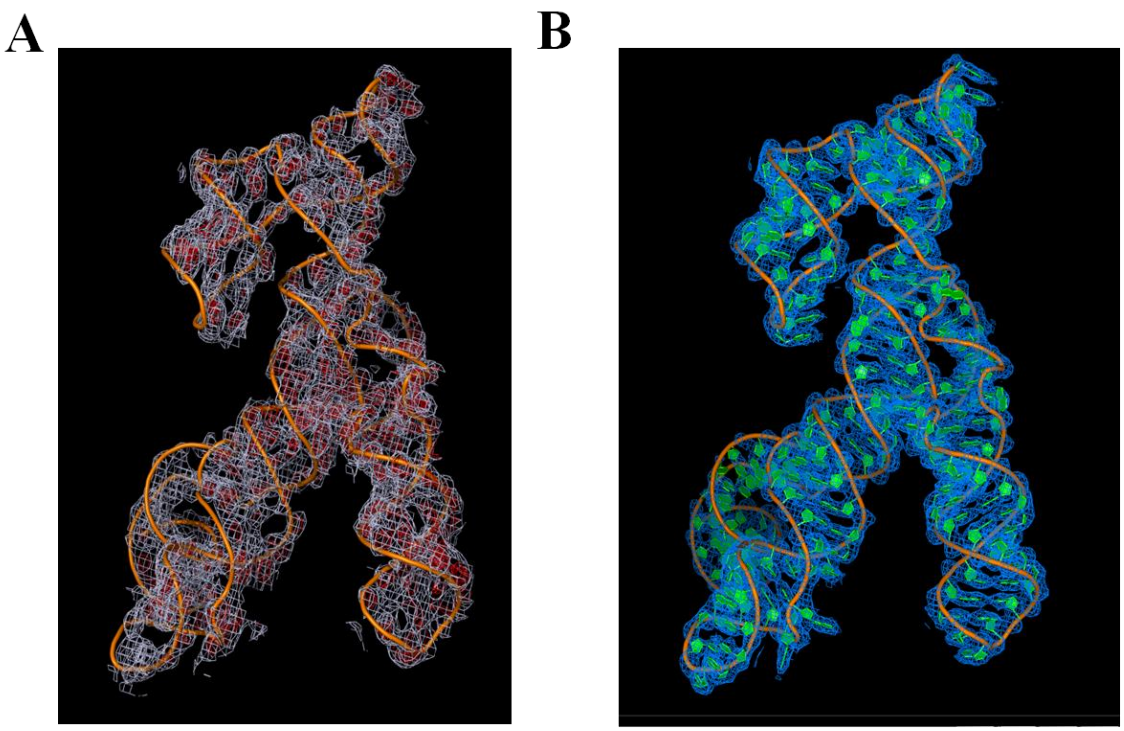

Figure 2.11 The model and accompanying electron density maps. A. Experimental, unbiased electron density map contoured at 1.3σ (white) or 3.0σ (red) and carved within 1.6 Å of selected atoms. The trace of the phosphodiester backbone is shown in orange. **B.** VS ribozyme juxtaposed with the final $2|F_{obs}|-|F_{calc}|$ electron density map contoured at 1.0σ and carved within 1.6 Å of selected atoms. The RNA is rendered in cartoon mode.

The final refined model for the VSx_G638A (PDB ID code: 4R4V) construct had an R_{work} of 0.17 and R_{free} of 0.21 and coordinate and phase errors were 0.36 Å and 25.2°, respectively. The structure of VSx_A756G (PDB ID code: 4R4P) was solved by molecular replacement using VSx_G638A as the search model. This construct was refined to R_{work} of 0.23 and R_{free} of 0.27. Coordinate and phase errors were 0.55 Å and 40.3°, respectively. These crystallization constructs contain a mutation that pre-organizes the substrate-helix into the 'shifted' secondary structure (Figure 2.6B-D) and thereby eliminates the thermodynamic penalty for shifting incurred upon substrate binding. Hence, we solved a third structure of the VS ribozyme dimer (helices 1–7), containing a substrate-helix that lacks this pre-organizing mutation (with a

G638A mutation to prevent self-cleavage). We refer to this construct as VSx_C634 indicating the presence of a wild-type C634 instead of the C634G mutation that constitutively shifts the substrate-helix (Figure 2.6E). This structure would capture the shifted conformation in its natural context, and comparison with NMR structures of isolated substrate helices (Micheal et al., 2000; Flinders and Dieckmann, 2001; Hoffman et al., 2003) would allow us to map the changes in substrate conformation before and after docking to the catalytic domain (see chapter 3 for a detailed description of substrate remodeling on docking). The conditions that produced robust single crystals were similar to that in VSx_G638A crystallization, but we obtained a best resolution of 3.3Å for this construct (Figure 2.10E) (PDB ID code: 5V3I). Diffraction data was phased using VSx_G638A as a search model and the structure was refined to R_{work} of 0.21 and R_{free} of 0.24. Coordinate and phase errors were 0.38 Å and 32.4°, respectively. Simulated–annealing omit maps were created for VSx_A756G and VSx_C634, the constructs solved by molecular replacement, to remove initial model bias. The diffraction, phasing and refining statistics for VSx_G638A, VSx_A756G and VSx_C634 are included in the crystallographic data table 2.2 (see Materials and methods, section 7.5.2 for detailed procedures).

	VSx_G638A_tGU	VSx_G638A	VSx_A756G	VSx_C634
Data Collection				
Space Group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Cell dimensions				
a, *b*, *c* (Å)	104.1, 104.1, 213.0	102.2, 102.2, 215.0	102.7, 102.7, 211.0	104.3, 104.3, 211.9
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	1.10548	1.07812	0.97921	0.97921
Resolution (Å)	200-3.40(3.52-3.40)	50-3.07(3.12-3.07)	41.47-3.07(3.12-3.07)	93.58-3.29 (3.41-3.29)
R_{merge} (%)	8.8(46.5)	13.1(66.8)	0.218(0.032)	0.115 (0.049)
$I/\sigma I$	31.4(2.5)	25.2(2.0)	9.2(1.0)	14.08(3.14)
Completeness (%)	98.4(87)	99.7(96.8)	99.4(94)	98.89(92.03)
Redundancy	9.8(6.8)	10.2(7.4)	9.5(7.6)	9.8 (9.0)
Molecules in ASU	1	1	1	1
SAD Phasing				
No. of heavy atom-(Ir) sites	16			
Mean FOM, after initial	0.47			
Mean FOM, after DM	0.76			
BAYES-CC	47.7±24.6			
Refinement				
No. of reflections		31653	22499	18250
R_{work}/R_{free} (%)		17.7/21.5	23.5/26.9	21.6/24.6
Coordinate error		0.36	0.55	0.38
Phase error		25.2	40.32	32.4
No. of atoms				
RNA		5991	5983	3976
Mg^{2+}		9	26	
Water		35	1	1
K^+		5		
Mean B-factors (Å^2)				
RNA		60.37	132.00	73.20
Mg^{2+}		50.62	158.20	21.00
K^+		85.84		
Water		46.52	66.50	
RMS deviation				
Bond lengths (Å)		0.009	0.011	0.003
Bond Angles (°)		1.752	1.90	0.70

Table 2.2 Data collection and refinement statistics of VSx_G638A_tGU, VSx_G638A, VSx_A756G and VSx_C634. High resolution shell is shown in parenthesis.

2.2.3 Global structure

2.2.3.1 Overall structure from crystallography

All three ribozyme constructs (VSx_G638A, VSx_A756G, and VSx_C634) fold into identical global folds with the only differences localized in the substrate helix (Figure 2.12C) (see chapter 3). These constructs have shortened stem 4 (by 3 bp) closed by the aforementioned AAACA pentaloop and mutated stems 7a and 6c to enhance stability (Figure 2.6). The three altered stems are remote from the active site. The AAACA pentaloop, in addition to preventing aggregation, mediates vital crystal contacts in the lattice (Figure. 2.20). The crystal structures revealed that the VS ribozyme forms a symmetric dimer (Figures 2.12, 2.13), with an intricate interdigitation of helical segments from the two subunits that is unprecedented among known ribozymes. Dimerization creates two hybrid active sites, in which each protomer donates its substrate-helix to the catalytic domain of the other (Figures 2.12, 2.13). This structural exchange (Figure 2.13) resembles the process of domain swapping observed in proteins where protein segments exchange part of their structures to form an intertwined dimer or higher-order oligomer (Rousseau et al., 2012). The importance of substrate nucleotides in the formation of dimers is consistent with the absence of dimerization in VS constructs that lack substrate helices (section 2.2.1).

The secondary structure of the VS ribozyme in the crystal conforms to the model derived from biochemical footprinting and mutagenesis (Beattie et al., 1995), with the exception of junction 1-2-7, whose secondary structure had not been defined (Figure 2.12B). Electrophoretic mobility shift, fluorescence resonance energy transfer (FRET) and small-angle x-ray scattering (SAXS) data on the monomeric species generally agreed with our structure (section 2.3.1) (Lafontaine et al., 2002a; Lipfert et al., 2008).

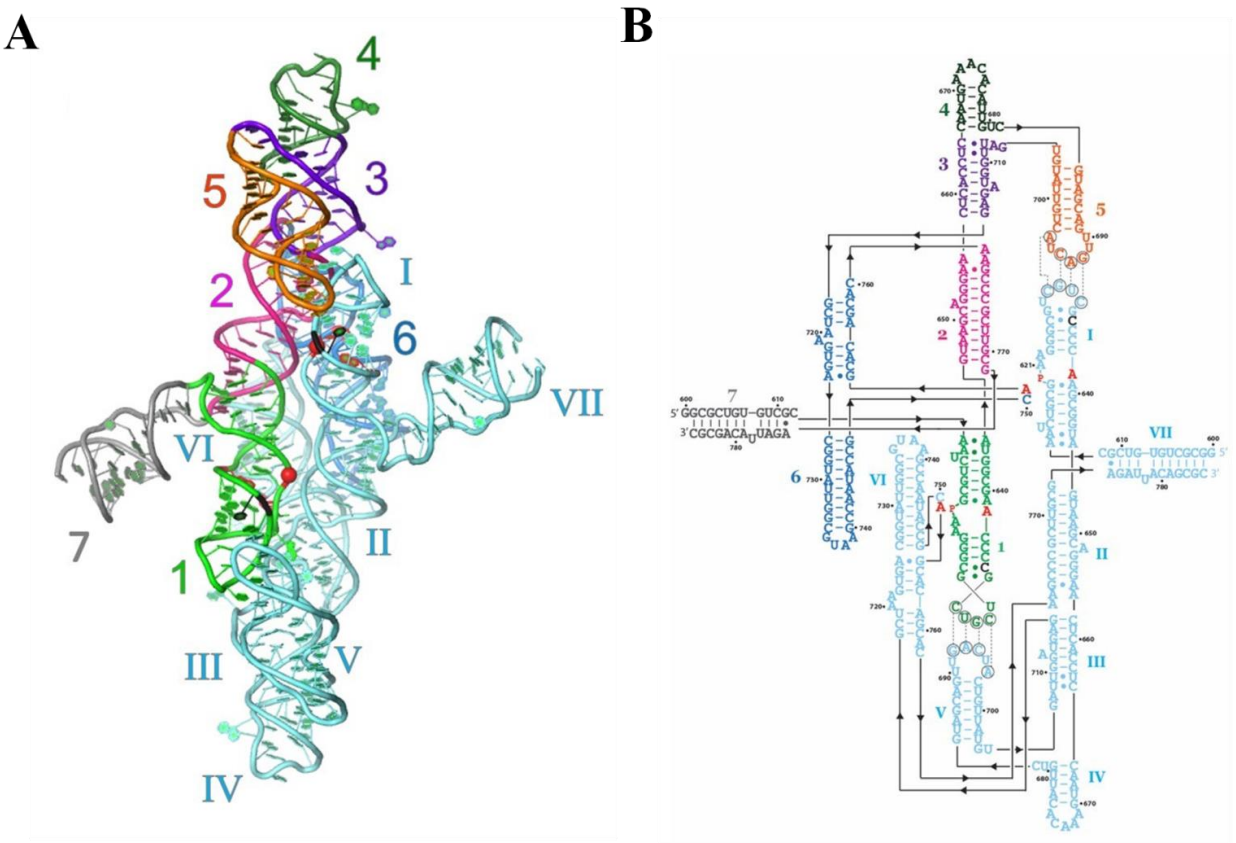

Figure 2.12 Crystal structure of the Varkud satellite (VS) ribozyme: dimer. A. Three-dimensional structure of the VS ribozyme dimer in cartoon representation. Color scheme of the top RNA as in figure 2.6, bottom RNA is shown in light blue. Scissile phosphates and catalytic nucleobases are represented as red spheres and red sticks, respectively, and C634 nucleotides are shown in black. **B.** Revised secondary structure of the VS ribozyme as obtained from the crystal structure in A. Substrate-helix from one protomer interacts with stem-loop 5 of the other via kissing-loop interactions. Tertiary interactions between protomers are shown as dotted lines. Helices 5, 6 and 7 emanate from a central coaxial stack involving helices 1, 2, 3 and 4. Color scheme matches that in the crystal structure A.

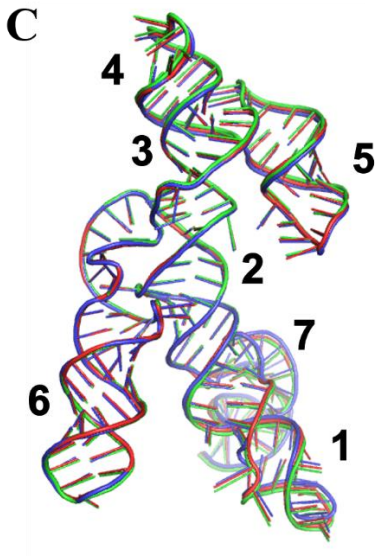

Figure 2.12 Crystal structure of the Varkud satellite (VS) ribozyme: dimer (continued). **C.** Monomers corresponding to three VS constructs used in crystallographic studies: VSx_G638A (red), VSx_A756G (blue) and VSx_C634 (green) are globally superimposable.

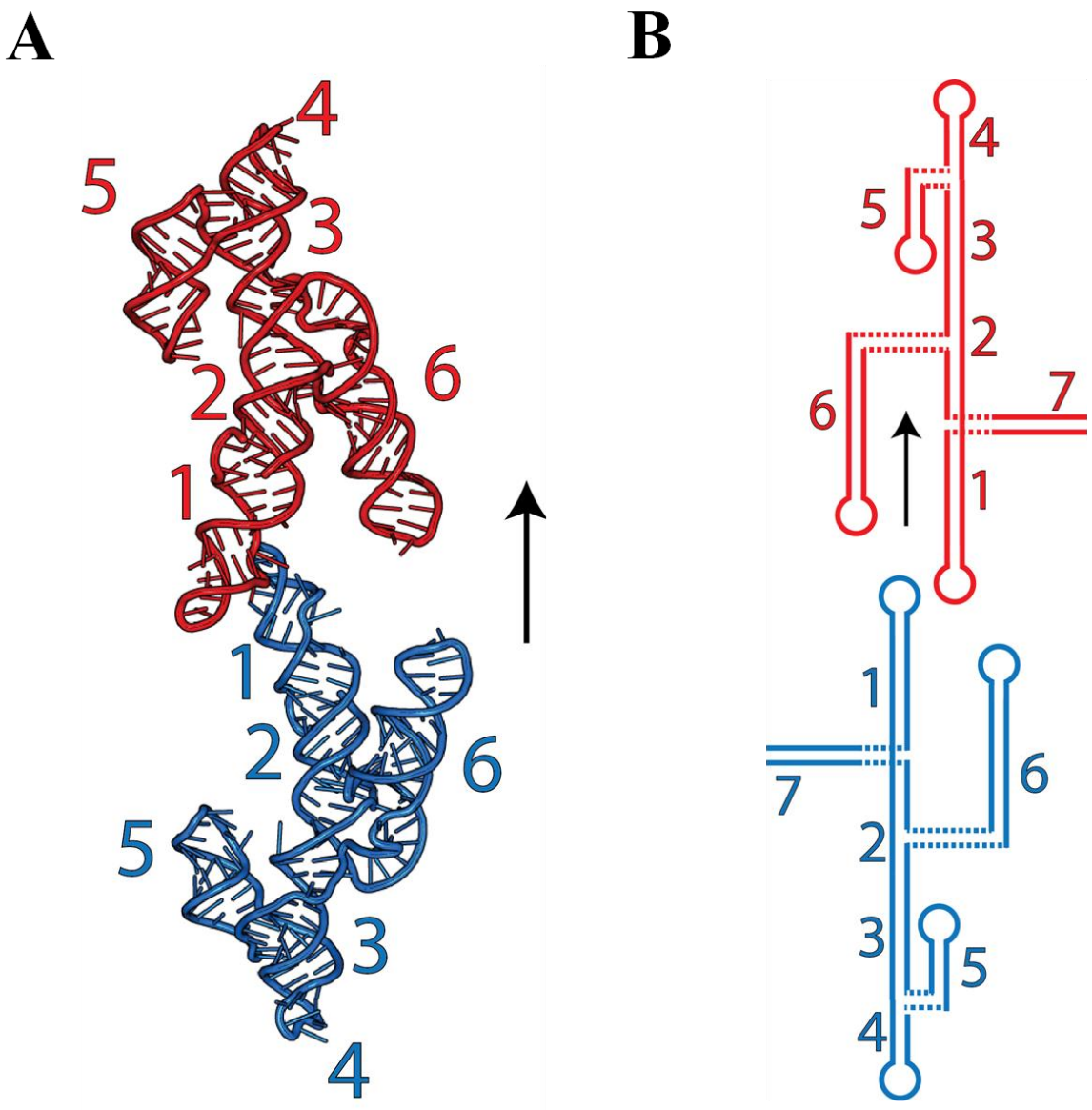

Figure 2.13 Formation of VS ribozyme dimer by association of two protomers *in trans*.
A. Protomers in cartoon representation. Helices 7 are removed for clarity. The top protomer is translated away from the bottom protomer along the y-axis (black arrow). **B.** Secondary structure of the VS ribozyme showing the relative orientation of the helices in each protomer. Length of helices is not to scale. Arrow indicates insertion of the substrate-helix 1 of the blue subunit into the cleft made by helices 2 and 6 of the red subunit.

Two adjacent multihelix stacks run along the central axis of the dimer. Each stack consists of helices 1−4 of each protomer, extended further to include helix 5 of the other protomer via an intermolecular 'kissing' interaction with stem-loop 1. Helices 5, 6 and 7 emerge from each stack at their respective three-way junctions, organizing the global architecture of the RNA (Figure 2.12) (Lafontaine et al., 2002a; Lipfert et al., 2008). As helix 5 runs roughly parallel to the coaxial stack from which it protrudes (helices 3 and 4 subtends an angle of 170° between them and helix 5 bends toward helix 3 at an angle of 11°), the kissing interactions lead to a close lateral association of the two multihelix stacks. Helix 7, which pairs the 3′ and 5′ ends of the RNA chain, protrudes from the otherwise globular structure. Helices 2 and 3 form an approximate coaxial stack (that connects the stacks formed by helices 1 and 2, and 3 and 4) creating an angle of 151° between them (Figure 2.13A). An angle of 55° between helix 2 and helix 6 (Figure 2.13A) creates a cleft for docking of the substrate-helix (helix 1) of the other protomer (Figure 2.13B). This is mediated by an intramolecular tertiary contact between the A652 bulge in helix 2 and the A_2 bulge (A725-A726) in helix 6 (Figures 2.16B, 2.18). As inferred from low resolution models created by the Lilley group (Lipfert et al., 2008), helix 3 is probably the most important domain in the structure that governs the relative spatial orientation of each domain. The A718 bulge in helix 3 introduces a kink in the stem separating the helix into helix 3a and 3b that are held at an angle of 168° between them (Figures 2.6, 2.13). Helices 1 and 2 subtend an angle of 161° with helix 7 bending away from the coaxial stack toward helix 1 creating an angle of 65° between them (Figures 2.12, 2.19A).

2.2.3.2 Solution-phase small-angle x-ray scattering (SAXS)

Preliminary analysis using nSEC had indicated that VS ribozyme constructs dimerize in solution, an observation that was reflected in the crystal. However, domain-swapped dimers have been found to be artifacts of crystal packing in protein and RNA structures (Liu et al., 1998; Chen et al., 2015). Small-angle x-ray scattering (SAXS) can be used to obtain information regarding the global fold of a biomolecule in solution. To confirm the nature of association in solution, we obtained SAXS data for the VSx_G638A construct and created a low-resolution SAXS envelope that agrees well with the global conformation of the symmetric VS ribozyme dimer obtained from the crystal structure (Svergun et al., 1995; Franke and Svergun, 2009; Volkov and Svergun, 2003; Kozin and Svergun, 2001) (Figures 2.14, 2.15). The reverse approach in which theoretical scattering was calculated from the crystal structure also resulted in agreement between the dimer and SAXS data (Figures 2.14, 2.15) (Schneidman-Duhovny et al., 2013). A dimer model of the ribozyme fits the data significantly better than a monomer in terms of parameters such as radius of gyration (R_g) and maximum dimension (D_{max}) (Figure 2.14C) (Svergun, 1992) of the molecule. The results from SAXS measurements and the software suites used are summarized in table 2.3. SAXS data have been deposited in the small angle scattering biological data bank (SASBDB) under accession code SASDAC9.

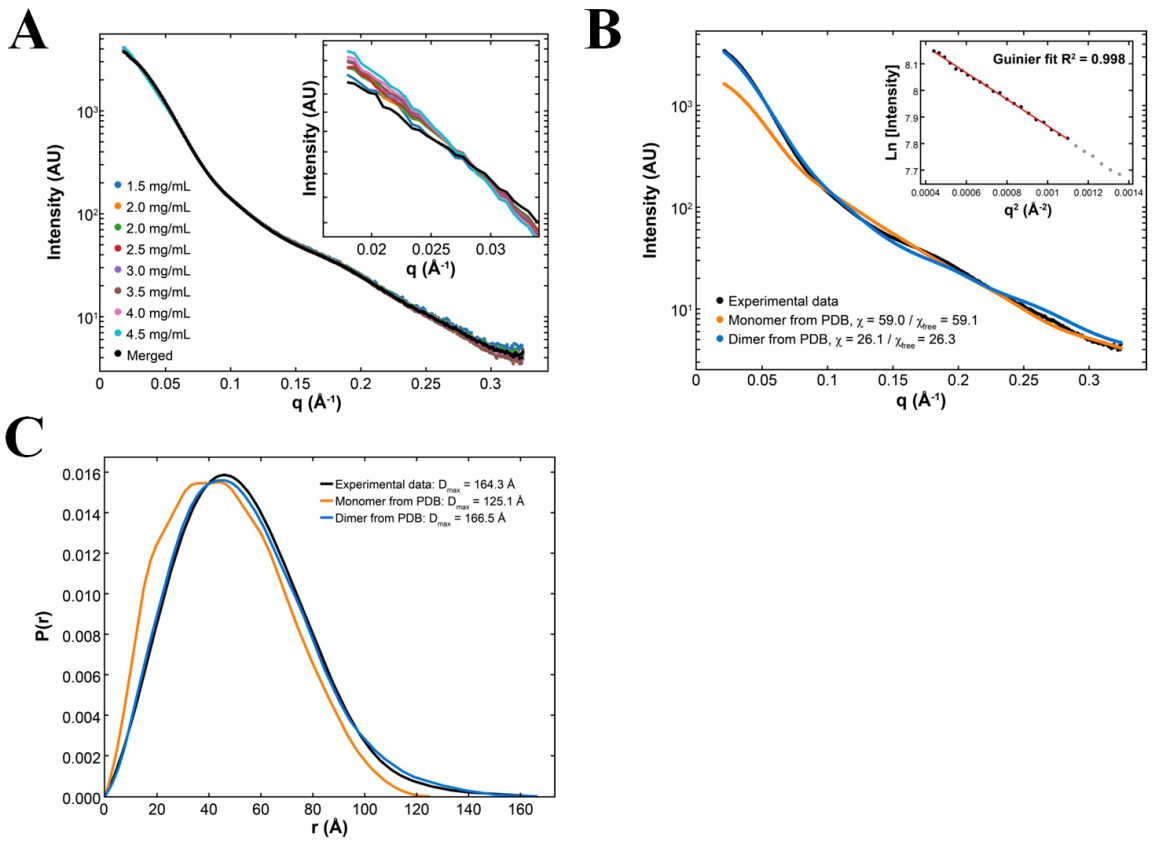

Figure 2.14 Solution phase studies with SAXS. A. SAXS datasets from 1.5 mg/mL to 4.5 mg/mL sample concentrations, scaled to match the 4.5 mg/mL sample intensities. All eight datasets show a high level of agreement except for minor systematic variation at very low q (shown in the zoomed top right inset). This was ameliorated in the final merged dataset by extrapolation to infinite dilution. The resulting merged dataset was used for all subsequent structural comparisons. Error bars representing the uncertainties (1 standard deviation) in the measured intensities are shown, but most are too small to be visible. The horizontal axis q is the scattering vector = $[4\pi \sin \theta / \lambda]$. **B.** The experimental SAXS curve (black) was fitted to theoretical curves (orange: monomer; blue: dimer) calculated for one or both subunits of our dimeric crystal structure (PDB ID code 4R4V) using FOXS (Schneidman-Duhovny et al., 2013). A dimer of the ribozyme fits the data significantly better than a monomer (χ / χ_{free} = 26.1 / 26.3 for the dimer, in comparison to 59.0 / 59.1 for the monomer), particularly at low q (Rambo and Tainer, 2013). Discrepancies between the experimental data and the dimer model at higher q may indicate conformational differences or flexibility in solution. We suspect this is largely due to movement helix 7, which makes relatively few interactions with the rest of the structure, and is thus likely to be flexible in solution. Although χ and χ_{free} should approach 1 for an ideal model, the extremely high Intensity/error ratio of the experimental data makes the χ and χ_{free} metrics very sensitive to any differences between the model and experimental data. The top right inset shows a Guinier plot of the experimental data, including the Guinier fit (red line) and Guinier region (black circles) as calculated by AUTORG (Petoukhov et al., 2007). The Guinier region shows no apparent curvature and the fit indicates an R_g of 38.7 ± 1.3 Å.

Figure 2.14 Solution phase studies with SAXS (continued). C. Pair-distance distribution curves calculated from experimental (black) and model-derived (orange and blue) scattering by DATGNOM (Svergun, 1992). The horizontal axis r is the distance measured within the molecule. The maximum r (D_{max}) of the distributions is shown. Here it is also apparent that a dimer of the ribozyme structure is a good match to the experimental data. Summary values from the real space transforms shown here can be found in table 2.3.

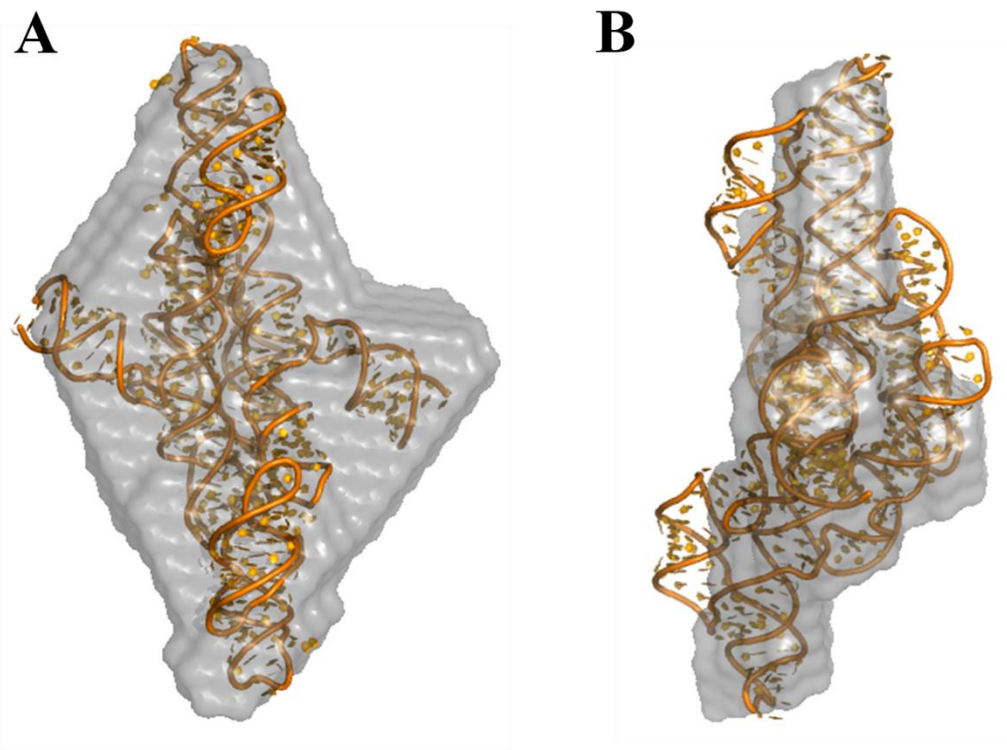

Figure 2.15 SAXS envelope (SASBDB ID code SASDAC9) super-imposition with crystal structure (PDB ID code 4R4V). The envelope generated by averaging the bead models calculated from SAXS data is shown as a gray surface and the RNA dimer is shown as an orange cartoon, in two different perspectives. The envelope shown is the result of the averaging and filtering of 20 DAMMIF (Franke and Svergun, 2009) calculations (Normal Spatial Discrepancy, NSD = 0.587 ± 0.156) by DAMAVER (Volkov and Svergun, 2003) and was aligned to the crystal structure by SUPCOMB (Kozin and Svergun, 2001). A and B panels show different perspectives.

	VSx_G638A	Monomer from PDB	Dimer from PDB
Data collection			
Instrument	SIBYLS beamline Advanced Light Source Berkeley CA		
Beam Geometry	Superbend synchrotron light 100x100 micron		
Wavelength (Å)	1.03		
Q range ($Å^{-1}$)	0.012 – 0.324		
Exposure times (s)	0.5, 1, and 6		
Sample concentrations (mg/mL)	1.5, 2.0, 2.0, 2.5, 3.0, 3.5, 4.0, and 4.5		
Structural parameters			
R_g (Å) from Guinier fit	38.7 ± 1.3	36.2	41.1
I(0) (AU) from Guinier fit	4298 ± 18		
D_{max} (Å) from P(r)	164.4	125.1	166.5
Reciprocal space R_g (Å) from P(r) fit	41.3	36.5	41.6
I(0) (AU) from P(r) fit	4491		
Real space R_g (Å) from P(r)	41.3	37.0	41.6
Ab initio molecular envelopes NSD	0.587 ± 0.155		
Software employed			
Data merging	ALMERGE		
Model intensity curve fitting	FOXS		
Guinier analysis	AUTORG		
Pair-distance distribution	DATGNOM		
Ab initio molecular envelopes	DAMMIF and DAMAVER		
Structure superposition	SUPCOMB		

Table 2.3 SAXS data collection, scattering derived parameters and software used.

2.2.3.3 Trans docking of the substrate-helix

The dimer structure can be thought of as two enzyme-substrate complexes, with each ribozyme core consisting of helices 2−6 plus the substrate-helix (helix 1) of the opposite subunit (Figures 2.13, 2.16B). Helix 1 contains the putative general base (G638) as well as the scissile phosphate, whereas helix 6 contains the putative general acid (A756). The active site is formed *in trans* by close melding of helix 2 with the internal loops of helices 1 and 6 (the G638 and A730 loops, respectively) (Figure 2.12). The loop-loop kissing interactions with helix 5 further stabilize the docking of helix 1. A U-turn motif in each stem-loop organizes the nucleotides into an A-form conformation to promote formation of the inter-loop base-pairs (Figures 2.12, 2.16), consistent with mutational analysis, 2′-deoxyribose and phosphorothioate interference and nuclear magnetic resonance (NMR) data (Beattie et al., 1995; Rastogi et al., 1996; Bouchard et al., 2008; Bouchard and Legault, 2014) (see chapter 3 for a detailed discussion on tertiary interactions between substrate-helix and the catalytic domain).

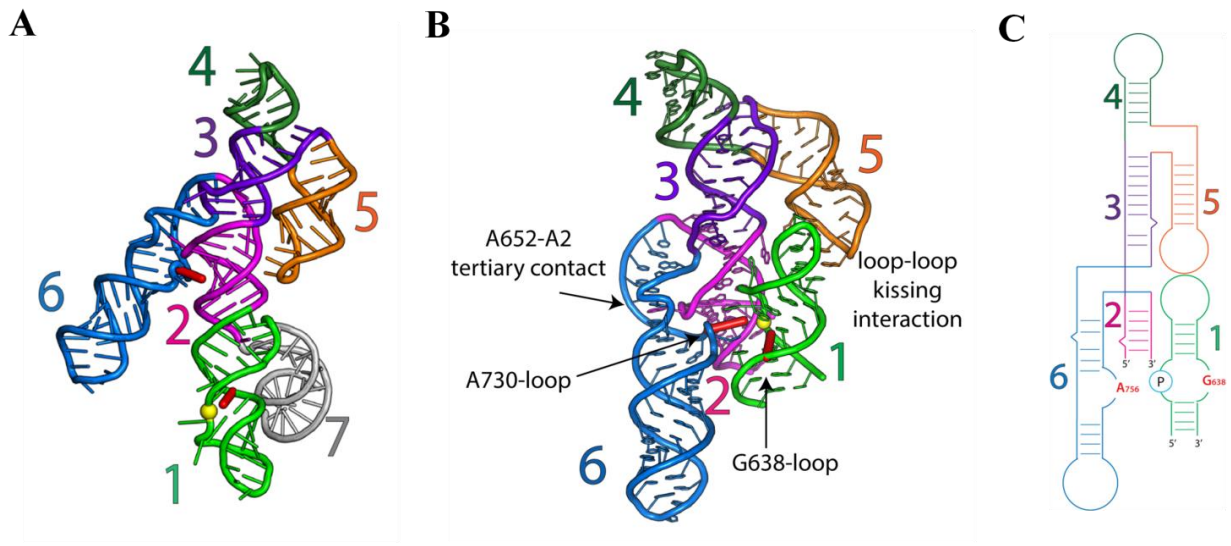

Figure 2.16 Crystal structure of the Varkud satellite ribozyme: protomer with substrate. A. VS ribozyme protomer in cartoon form. Substrate-helix (helix 1) projects away from the catalytic domain of its own protomer. The scissile phosphate is shown as a yellow sphere. Catalytic nucleobases are shown as red sticks. **B.** Catalytic domain (helices 2−6) of one protomer and the substrate-helix (helix 1) donated by the other protomer docked *in trans* (shown in green for consistency in coloring scheme; see figure 2.6). The three-way helical junctions 2-3-6 and 3-4-5 organize the overall fold of the catalytic domain. The yellow sphere depicts the scissile phosphate. Red sticks correspond to the catalytic nucleobases. Junction 1-2-7 and accompanying helices 1 and 7 have been omitted for clarity. **C.** Trans-form of the ribozyme revealed in our crystal structure: helices 2-6 of one protomer and helix 1 of the second protomer (see figure 2.12).

2.2.3.4 Structures of three-way junctions

Three C-family junctions (Lescoute and Westhoff, 2006) assemble the moving parts of the VS ribozyme into its catalytically-competent structure.

Junctions 3-4-5 and 2-3-6 organize the catalytic domain. The global architecture of the core is dictated by two three-way helical junctions (3-4-5 and 2-3-6) connected through the common helix 3 (Figure 2.16A). Junction 3-4-5 is stabilized by the formation of three base-triples: U713-U665-U686, U684-A667-G711 and C666-G685-A712 created by nucleotides from helices 3, 4 and 5 (Figure 2.17). The formation of these base-triples and the noncanonical U-U pair is consistent with the NMR structure of the isolated junction (Bonneau and Legault, 2014). This junction orients helix 5 for formation of the kissing-loop interaction (Figure 2.16B).

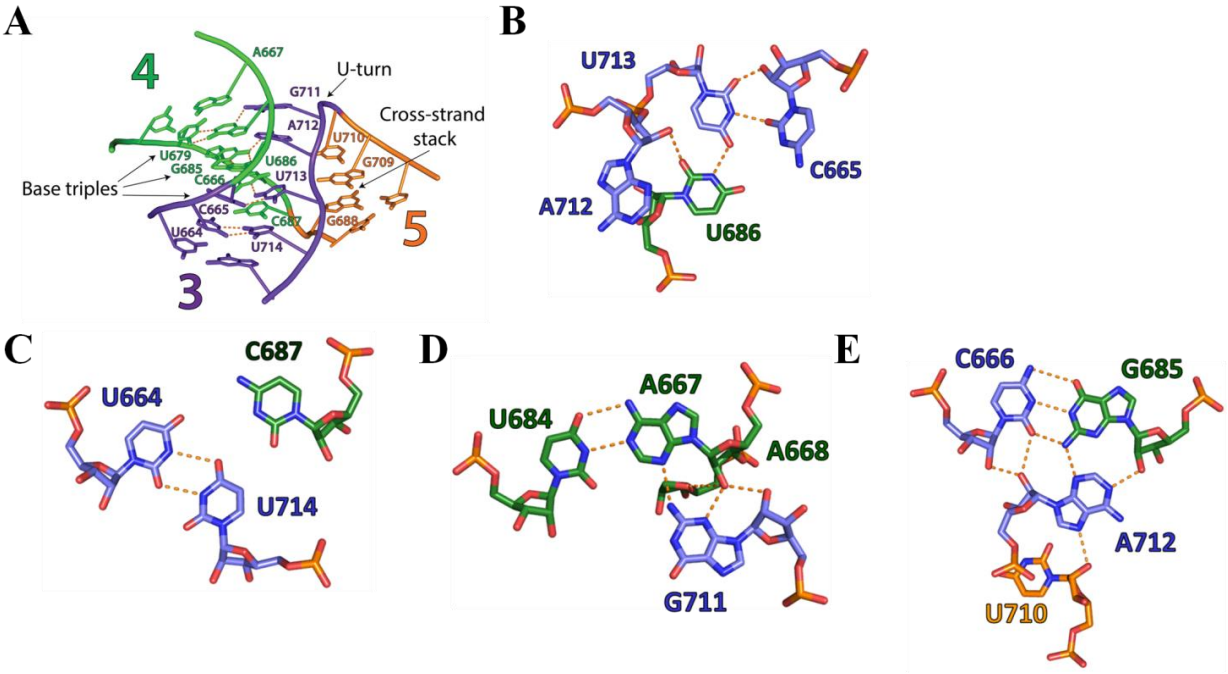

Figure 2.17 Structural features stabilizing junction 3-4-5. A. Local interactions in junction 3-4-5. **B.** C665-U686-U713 base-triple involves a *cis* WC-sugar edge base-pair between U713 and C665. **C.** U664:U714 base-pair and the nearby co-planar C687. **D.** A667-U684-G711 base-triple. **E.** C666-G685-A712 base-triple.

Junction 2-3-6 (Figure 2.16) orients helices 2 and 6 in a conformation that facilitates docking of the substrate-helix and formation of the active site. It juxtaposes the minor groove of helix 2 with the 3′ strand of the A730 loop (containing A756) within helix 6. Docking of the substrate-helix via the minor groove of its G638 loop positions the scissile phosphate for catalysis (Figure 2.16). Junction 2-3-6 includes several noncanonical base-pair interactions and a divalent metal ion (Figure 2.18). Helices 2 and 3 are coaxial ($\theta_{2\text{-}3}$ = 151°) (Figure 2.16C), although one nucleotide (A657) at their junction is extruded from the stack (it stacks against A622 extruded from helix 1, as described in chapter 3). The last base-pair of helix 2 is an imino G-A pair as predicted by biochemical data. (Figure 2.18D) (Lafontaine et al., 2001b). Base stacking is continued by two purine nucleobases in junction J_{62} that form base-triples, A720-

A766-U659 and G721-A767-C658 with the minor groove face of helix 3 (Figure 2.18B, C). Junction J_{36} contains no unpaired nucleotides and its phosphate backbone makes a sharp turn, likely stabilized by the metal ion coordinated to its own phosphodiester backbone and that of helix 2 (Figure 2.18A). The orientation of helices 2 and 6 established by junction 2-3-6 is reinforced by additional, tertiary interactions between bulges within each of these helices (Figures 2.16, 2.18E). A652 flips out of helix 2 and inserts into helix 6, stacking between paired bases and forming a *cis* WC-sugar edge base-pair with A725 of the 'A_2 bulge' motif of helix 6. The backbone of the A_2 bulge forms an S-turn motif, in which the directions of ribose sugars invert twice, extruding A726 into the solvent and positioning A725 for pairing with A652 (Figure 2.18E). The A652−A_2 bulge tertiary interaction is consistent with a large body of solution data (Beattie et al., 1995; Sood and Collins, 2002) and the NMR structure of the isolated junction (Bonneau et al., 2014).

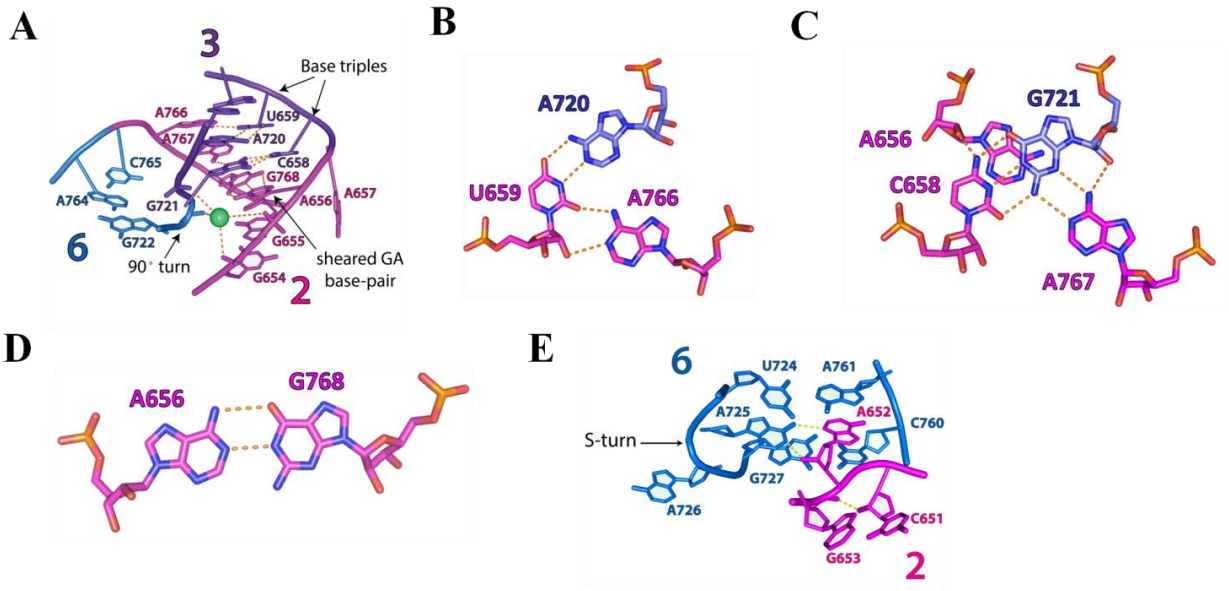

Figure 2.18 Structural features stabilizing junction 2-3-6. A. Local interactions in junction 2-3-6. **B.** U659-A720-A766 base-triple consists of a *cis* WC-sugar edge base-pair between A766 and U659. **C.** C658-G721-A767 base-triple. **D.** A656-G768 WC-WC base-pair. **E.** Tertiary interaction between the A652-bulge of helix 2 and the A_2-bulge of helix 6.

Junction 1-2-7 projects the substrate-helix outward. Nucleotides in junction 1-2-7 form an intricate network of non-canonical interactions (Figure 2.19A, B). These include two purine-purine base-pairs (G611-A779 and A645-A613) (Figure 2.19C, E) and a base-triple (U644-A614-G778) (Figure 2.19D), as well as cross-strand purine stacks that provide base-stacking continuity between helices 1 and 2 (Figure 2.19B, F). J_{127} orients helix 7 outward from the dimerization interface and directs the substrate-helix (helix 1) away from its own catalytic domain and toward that of the partner protomer (Figure 2.12, 2.19A). The biological relevance of this junction remains unconfirmed; however, circumstantial evidence suggests possible important roles in regulating cis versus trans activity in addition to modulating the cleavage-ligation equilibrium (see chapter 3 for a detailed discussion on the importance of this junction in the rolling circle replication of the VS plasmid).

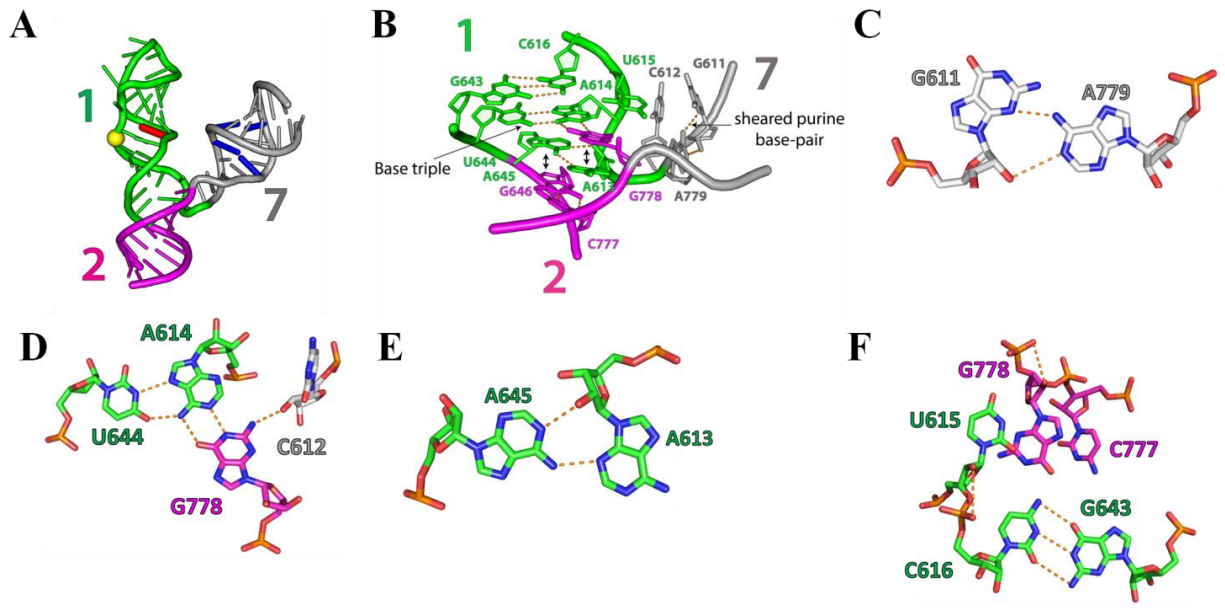

Figure 2.19 Structural features stabilizing junction 1-2-7. A. Junction 1-2-7 formed by the coaxial stack consisting of helices 1 and 2 and helix 7 that subtends an angle of 65° with helix 1. Cartoon expanded for clarity. The U782-G608 and C785-G606 base-pairs, known to favor ligation, are depicted in blue (Jones et al., 2001). The yellow sphere and the red stick correspond to the scissile phosphate and the catalytic nucleobase, G638, respectively. **B.** Local interactions in junction 1-2-7. Orientation corresponds to that shown in A. Double arrowheads indicate stacking interactions. **C.** A779-G611 *cis* WC-sugar edge base-pair. **D.** U644-A614-G778 base-triple involves a *cis* WC-Hoogsteen pair between U644 and A614. **E.** *Cis* WC-sugar edge A645-A613 base-pair. **F.** U615 bulge. U615 is extruded from helix 1 where it stacks against G778.

2.2.3.5 Structure of helix 7: roles in ligation

Nucleotides upstream of the cleavage site in the substrate-helix reduce cleavage rates and increase efficiencies of ligation, whereas VS ribozyme constructs with a single nucleotide upstream of the cleavage site exhibit almost negligible ligation capabilities. Extending the 5' end of helix 1a into a sequence capable of base-pairing with a similar extension at the 3' of helix 2 results in the formation of a distinct domain referred to as helix 7. The role of helix 7 in ligation was first identified by Strobel and co-workers (Jones et al., 2001). The presence of a base-paired helix 7, most likely holds the cleaved strands 5' of the cleavage site close to the rest of the

ribozyme, thereby facilitating the reverse reaction (ligation). This is in sharp contrast to the HDV ribozyme that contains a single nucleotide upstream of the cleavage site and is thus unable to catalyze ligation. This is most likely due to the facile dissociation of the 5' piece of the RNA required for the ligation reaction. Helix 7 also creates junction 1-2-7 with helices 1 and 2. This junction although a part of the wild-type VS ribozyme sequence was largely neglected as a result of truncation studies. In its biological context J_{127} probably regulates the equilibrium governing cis and trans activities (see chapter 3). This reductionist approach of eliminating domains that do not affect cleavage, in efforts to create minimal forms of ribozymes, have proved problematic with regards to understanding the broader physiological contexts of catalytic function in the past. The hammerhead ribozyme is an excellent case study where all functional data accumulated for about a decade was on a minimal form of the ribozyme, but its crystal structures were incompatible with its function (Pley et al., 1994; De la Pena et al., 2017). It was observed much later that nucleotides removed during developing the minimal ribozyme, mediated loop-loop tertiary interactions that not only reorganized the active site thus enhancing cleavage, but also played vital roles in ligation (De la Pena et al., 2003; Khvorova et al., 2003). Similarly, eliminating domains C and D of the wild-type hairpin ribozyme results in a hinged construct that is capable of cleavage but not ligation (Fedor, 2000). Since, these ribozymes are predominantly found as ligated circles in their natural context, it can be assumed that the wild-type ribozymes favor ligation over cleavage which underscores the importance of 'ligation auxiliary elements' in these endonucleolytic ribozymes (Suslov, 2012).

The secondary structure of J_{127} obtained from crystallography resembles both conformations proposed by Strobel and co-workers (Figure 2.3) (Jones et al., 2001). In their model, both conformations in equilibrium contained one extensively base-paired helix and one helix with

unpaired nucleotides near the junction it shared with the substrate-helix. This model suggested the existence of a switch that could modulate the rigidity of junction 1-2-7, organizing the ligation auxiliary elements in helices 1a and 7. A more rigid junction would favor ligation and a 'looser' junction would favor cleavage. Thus, the local restructuring of the junction could provide a mechanism for regulating the cleavage-ligation equilibrium (Jones et al., 2001). The crystal structure reveals a tighter junction structure with extensive base-pairing in both helices 7 and 2b supporting the greater propensity for ligation in our VS constructs. The local structure of junction 1-2-7 in cis ribozymes, where the substrate folds back and docks into the catalytic domain of its own RNA molecule, is still unknown; however, since greater flexibility at this junction is required for cis docking, it is reasonable to speculate that the cis junction would resemble the conformation with a weak J_{17}.

2.2.3.6 AAACA pentaloop mediates essential crystal contacts

A common approach in RNA construct design entails replacing flexible, nonessential loops with structured GNRA tetraloops. These loops impart stability and local rigidity and provide an opportunity to create lattice contacts through interactions with the minor groove of RNA helices. In this work, we found that a nonessential tetraloop (in stem-loop 4) was contributing to RNA aggregation and that replacing the tetraloop with a more flexible AAACA pentaloop eliminated the aggregation problem and allowed isolation of the ribozyme as a dimer (Figure 2.7B, C). Moreover, the pentaloop, which was intended to provide an epitope for Fab binding (Koldobskaya et al., 2011), adjusted its conformation (relative to the one observed previously in complex with the Fab) to mediate lattice contacts with two other VS ribozymes (Figure 2.20). In light of these observations, it seems that in some cases of RNA crystallization

it may prove advantageous to screen constructs in which nonessential loops, including tetraloops, are replaced with flexible pentaloops (Koldobskaya et al., 2011).

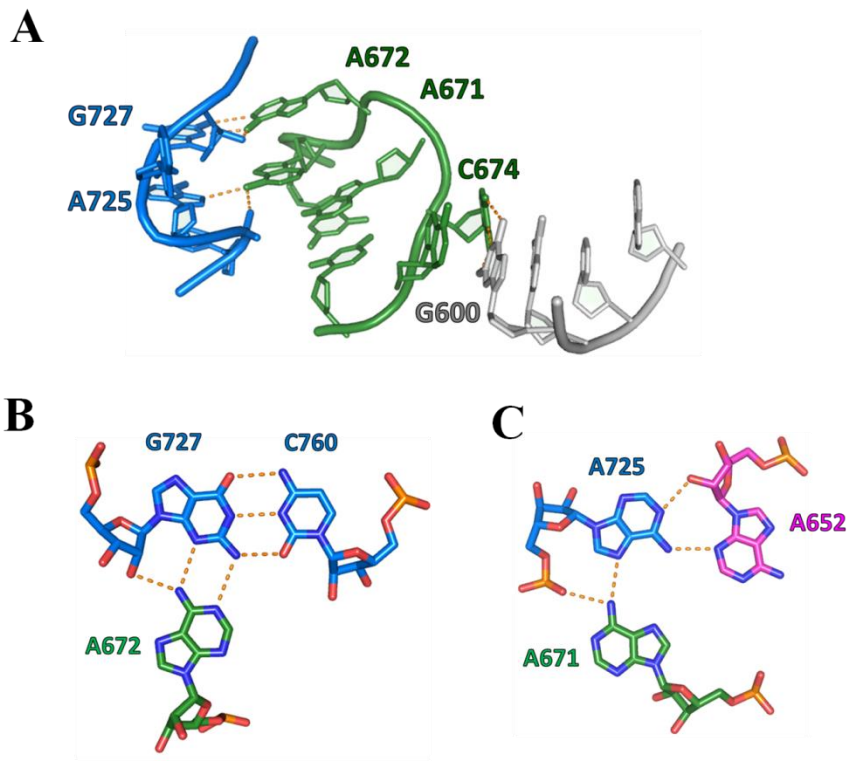

Figure 2.20 Crystal contacts mediated by the AAACA pentaloop. A. Crystal contacts between the AAACA hairpin and helices 6 (blue) and 7 (grey). C674 makes a WC base-pair with G600. Stacked A671 and A672 interact with the minor groove of helix 6. **B.** Interaction between A672 and G727-C760 base-pair. A672 makes a *trans* WC-sugar edge pair with G727. **C.** Interaction of A671 with a *cis* WC-sugar edge pair between A725 and A652.

2.3 DISCUSSION

2.3.1 Global fold: comparison with electrophoretic mobility shift/FRET/SAXS data

The global structure of the VS ribozyme obtained from crystallographic data largely agrees with the model generated by electrophoretic mobility shift and FRET-based assays, and solution-phase small-angle x-ray scattering (SAXS) studies of the VS ribozyme monomer (Lafontaine et al., 2001a; Lafontaine et al., 2002a; Lipfert et al., 2008). Relative orientation and stacking patterns observed in the crystal structure (Figure 2.16) are also consistent with bioinformatic predictions (Liang et al., 2012; Lescoute and Westhoff, 2006; Tyagi and Mathews, 2007). Despite the agreements, the crystallographic model and previous low-resolution models differ in two fundamental aspects. Helices 2 and 3 are coaxially stacked in the crystal structure ($\theta_{2b/3a}$ = 151^o and $\theta_{3a/6b}$ = 146^o); however, calculations from electrophoretic mobility experiments suggest stacking between helices 3 and 6 ($\theta_{2/3}$ = 136^o and $\theta_{3/6}$ = 158^o). Both conformations are consistent with computational data (Liang et al., 2012; Lescoute and Westhoff, 2006; Tyagi and Mathews, 2007).

2.3.2 Three-way junctions: comparison with NMR structures

Junction 3-4-5 adopts similar topologies in both crystal and solution (from NMR) (Figure 2.21). The canonical uridine turn (Ashraf et al., 1009; Quigley and Rich, 1976) formed by U710, G706 and A712 facilitates a backbone reversal between helices 3 and 5. The sharp turn occurs after U710 and includes a G711 and A712 base stack in both structures (Bonneau and Legault, 2014). A sudden turn in the backbone between G687 and G688 facilitates the orientation of J_{45} (Bonneau and Legault, 2014). J_{34} is wider in the crystal than in the NMR structure which could be a reflection of sequence differences between the constructs used for NMR and crystallography

(Figure 2.21A) (Bonneau and Legault, 2014; Dagenais et al., 2017). The NMR structure revealed the presence of six Mg^{2+} ions (determined by Mn^{2+}-induced PRE), with two Mg^{2+} ions bound at the junction that have possible roles in stabilizing the active conformation of this junction (Bonneau and Legault, 2014). The presence of Mg^{2+} ions is consistent with Mn^{2+} rescue of phosphorothioate interference (Sood et al., 1998). No Mg^{2+} ion was observed to be bound to this junction in the crystal possibly because of the high ionic strength of crystallization conditions (~2M $(NH_4)_2SO_4$) (Figure 2.21A). Identical base-triples stabilize the junction in both structures; however, the crystal and NMR structures differ in the details of these interactions. For example, in the U713-C665-U686 triple, a hydrogen bond between the ribose 2'-hydroxyl group of C665 and the carbonyl oxygen of U713 is captured only in the crystal structure. Similarly, a hydrogen bond between the carbonyl oxygen of U686 and the exocyclic amino group of C665 is present exclusively in the NMR structure (Figure 2.21B). The only difference in the A712-C666-G685 triple between the NMR and crystal structures lie in presence of a viable A-minor interaction that involves a hydrogen bond between the N1 atom of A712 and the 2'-hydroxyl group of G685. These groups are just beyond hydrogen bonding distance in the NMR model (Figure 2.21C).

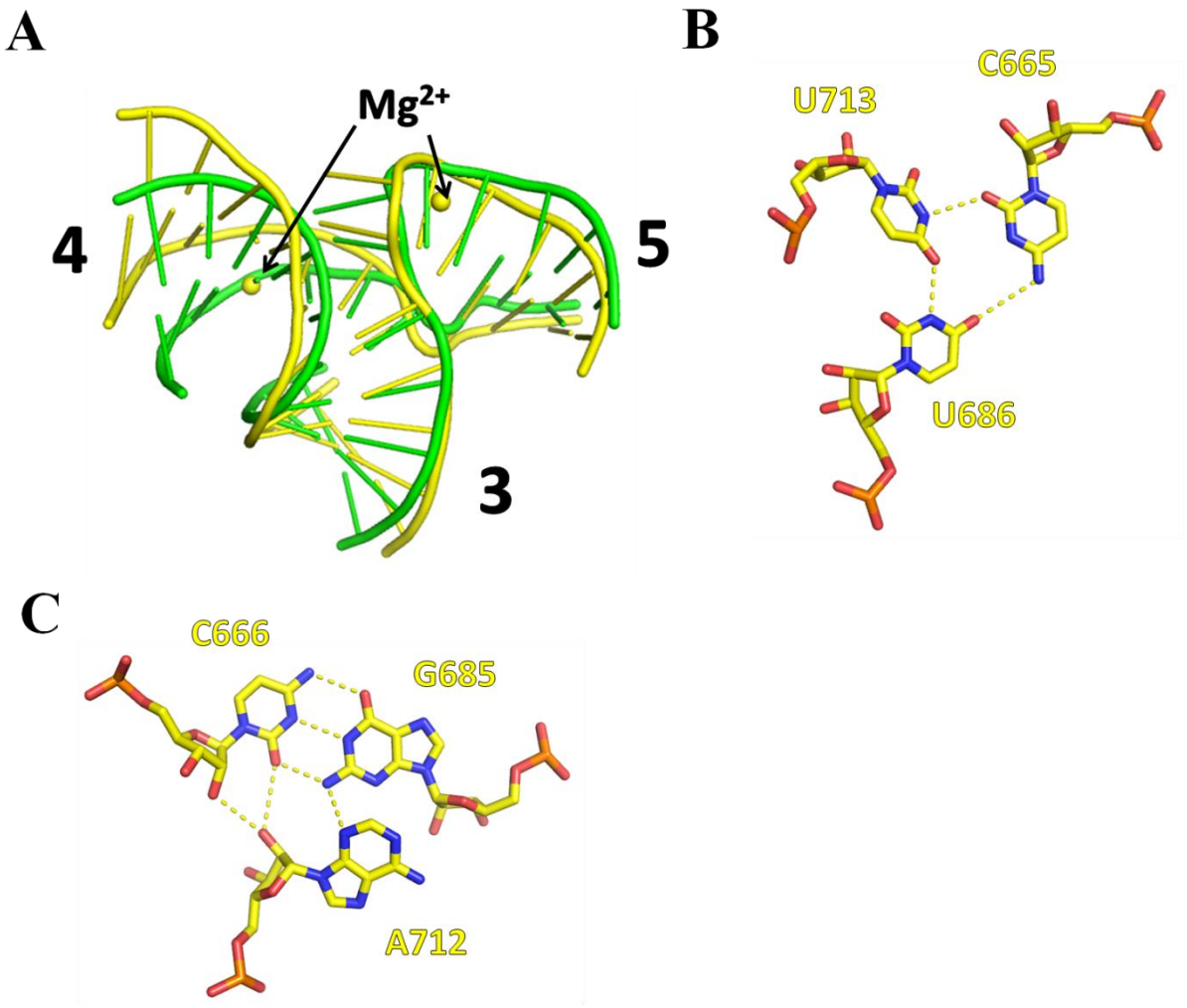

Figure 2.21 Comparison of junction 3-4-5 as revealed by NMR and crystal structures.
A. Alignment of junction 3-4-5 from crystal structure (green, PDB ID code: 4R4P, this work) with the NMR structure (yellow, PDB ID code: 2MTK, Bonneau and Legault, 2014) of an isolated junction shows overall agreement of global conformations. Mg^{2+} ions observed in the NMR structure are depicted as yellow spheres. **B.** U713-C665-U686 base-triple from NMR. **C.** U712-C666-U685 base-triple from NMR (For corresponding base-triples in the crystal structure see figure 2.17 B, E).

Junction 2-3-6 adopts a type C topology (similar to the other two junctions), in both NMR and crystal structures (Figure 2.22A). Both NMR and crystallography reveal an intramolecular tertiary interaction between the bulged A652 and nucleotides from the A_2 loop (Figure 2.18E) that is consistent with functional data (Beattie et al., 1995; Sood and Collins, 2002; Bonneau et al.,

2015). This vital interaction brings helices 2 and 6 together to stabilize the cleft created by junction 2-3-6 into which the substrate-helix docks before catalysis. A sharp turn between G721 and G722 alters the strand direction between helices 3 and 6 (Dagenais et al., 2017) and the junction core is stabilized by two base-triples and a WC-WC G-A non-canonical base-pair. The A767-C658-G721 triple exhibits subtle differences in the NMR and crystal structures (Figure 2.22B). The hydrogen bond between the exocyclic amino group of A767 and the N3 atom of G721 is present in both structures; however, the crystal captures a possible hydrogen bond between the same exocyclic amino group and the 2'-hydroxyl group of G721, an interaction that is absent in the NMR structure of the junction. Similarly, the existence of a hydrogen bond between the N1 atom of A767 and the exocyclic amino group of G721 is revealed only in the NMR structure (Bonneau et al., 2015). The bases involved in the triple obtained from the NMR structure are shifted slightly from perfect coplanarity unlike the coplanar structure of the base-triple in the crystal structure. Outside the base-triples another hydrogen bond, between the 2'-hydroxyl group of C765 and the N7 atom of A766 is exclusively present in the NMR structure (Bonneau et al., 2015). Both structures reveal the presence of a Mg^{2+} ion bound to the core of the junction, although their exact positions relative to the junction nucleotides are not identical. Another difference between the structures involves A718, a nucleotide that introduces a bend in helix 3 in the crystal structure. A718 is stacked inside the helix in the NMR structure of the isolated junction but assumes an extrahelical position in the crystal structure (Figure 2.22A) (Bonneau et al., 2015; Dagenais et al., 2017). This local change in backbone conformation is instrumental in creating a more closed junction, in addition to introducing the bend in helix 3 which likely plays a role in rearranging the relative orientations of helices 2 and 6, thereby preparing the docking site for the incoming substrate (Dagenais et al., 2017). The narrower conformation of J_{23} in the crystal structure further illustrates the ‘open to

closed' conformation change of junction 2-3-6 in an active fold of the substrate-bound catalytic domain. Additional roles of the A718 bulge in helix 3 could be to adjust the orientation of junction 3-4-5 and junction 2-3-6 relative to each other resulting in the catalytically competent global structure of the VS ribozyme.

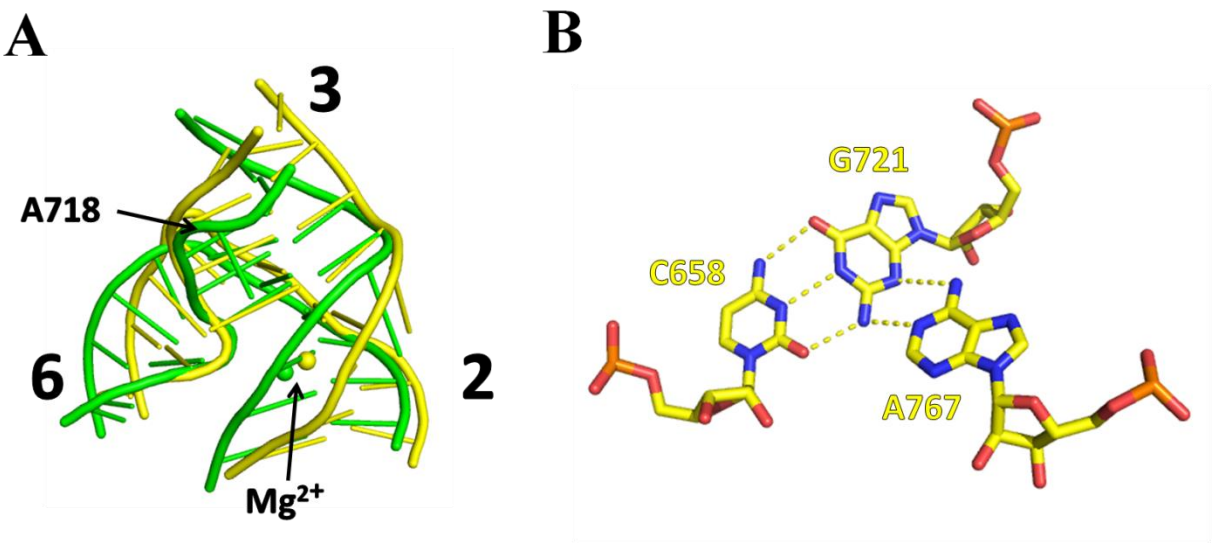

Figure 2.22 Comparison of junction 2-3-6 as revealed by NMR and crystal structures. A. Alignment of junction 2-3-6 from crystal structure (green, PDB ID code: 4R4P, this work) with the NMR structure (yellow, PDB ID code: 2N3R, Bonneau et al., 2015) of an isolated junction shows overall agreement. A718 is extruded from the helical stack in the crystal structure suggesting a structural rearrangement during folding. Mg^{2+} ions bound to the junction in the crystal and NMR structures are depicted by green and yellow spheres, respectively. **B.** A767-C658-G721 base-triple from NMR (For the corresponding base-triple in the crystal structure see figure 2.18C).

2.3.3 Dimerization and domain-swapping could lead to higher-order structures in RNA

(See chapter 3 for the biological significance of dimerization)

Our structure also reveals a mode of RNA association, in which one domain from a monomeric RNA replaces the same domain from an identical RNA chain, resulting in an intertwined dimer. This process, referred to as domain swapping, occurs commonly among proteins. Domain-swapping provides a mechanism to form higher-order oligomers and regulate folding and function (Rousseau et al., 2012). Together with gene-duplication events, domain swapping may have provided a mechanism for proteins to evolve larger, complex folds from smaller, simpler ones (Rousseau et al., 2012). Examples of functional RNA that utilize multiple units of the same molecule include the φ29 bacteriophage pRNA that consists of a homopentamer and finds use in designing RNA nanostructures (Simpson et al., 2000). The glycine riboswitch consists of a pair of tandem riboswitch molecules connected by a kink-turn motif and binds its ligand cooperatively (Butler et al., 2011). The cyclic-diAMP riboswitch binds two molecules of its ligand by using two RNA modules that are related by pseudosymmetry (Jones and Ferré-D'Amaré, 2014).

Perhaps this mode of association served analogous roles in regulation of RNA function and the emergence of complex folds in the RNA World. The peptidyl transferase center of the ribosome contains A and P sites that form symmetrical pocket-like structures with RNA backbone folds related by internal twofold symmetry, prompting speculation that the peptidyl transferase center evolved from a dimeric RNA (Belousoff et al., 2010). Our findings reveal domain swapping as a dimerization mechanism through which a symmetrical structure could emerge in RNA.

2.3.4 Building an RNA enzyme: lessons from the VS ribozyme

The VS ribozyme, the largest endonucleolytic ribozyme in nature, presents an excellent model to study the details of RNA structure and folding in relation to function. Once thought to be structurally too simple for enzyme catalysis, RNA has now been established as an efficient catalytic system, bringing out essential biochemical transformations in the cell. Crystallographic analyses of the VS ribozyme and more specifically its catalytic domain, described in this chapter, provide the structural framework to revisit the general features of enzyme catalysis in the context of RNA. The four primary steps of enzyme action (section 1.1) namely folding, substrate recognition and binding, catalysis and product release are carried out by the VS ribozyme as described below.

RNA, an intrinsically flexible molecule, must be organized into a definite three-dimensional shape in order to accomplish catalysis. The VS ribozyme is a fairly large molecule (186 nucleotides, M.W. of ~114kDa) made up of multiple domains that are capable of folding independent of the others. The modular nature of these domains likely provides evolutionary advantage as local misfolding at regions distant to the active site does not greatly affect the global fold of the RNA or the local folds of essential domains. These 'moving parts' of the 'RNA machine' are assembled by structural motifs such as three-way junctions in case of the VS. Three-way junctions organize the helices into spatial orientations that constitute the catalytically-active global fold. A network of hydrogen bonds, base stacks and RNA-bound metal ions stabilize the cores of these junctions. Helices 2 and 6, harboring nucleotides that constitute or stabilize the active site are brought together in this active fold, despite being separated in the secondary structure of the ribozyme. Helix 5 projects its stem-loop out of the central coaxial stack made of helices 1, 2, 3 and 4, toward the docking cleft created by the relative positioning of helices 2 and

6. The former structural feature is supported by junction 3-4-5 and the latter, by junction 2-3-6. The two junctions and the domains they organize are connected by helix 3, a base-paired stem that acts as the axle around which the catalytic domain is oriented in three-dimensional space.

The incoming substrate docks into the cleft and is positioned to interact with stem-loop 5 via a kissing-loop interaction that primarily involves WC base-pairing. This interaction, in addition to recognizing the correct substrate (via complimentary base-pairing between hairpin nucleotides of stem-loop 5 and the substrate), keeps it bound to the enzyme and remodels its structure that includes local restructuring of the active site, a priming step that precedes catalysis. The active site is created by cleavage site nucleotides and catalytic nucleobases, one of which resides in the substrate strand opposite to the cleavage site (G638) and the other in an internal bulge in helix 6 (A756). The entire process of activating the substrate for catalysis and the chemical step of catalysis itself is the subject of chapter 3.

Post-catalysis, the product needs to be released by the catalytic domain to catalyze the next round of RNA cleavage. Once cleaved the affinity of the VS substrate-helix toward the catalytic domain (and toward dimerization), decreases by 10-fold, enabling the catalytic domain to preferentially bind uncleaved substrates, thereby completing the fourth step of enzyme catalysis (product release). The possible reasons for this reduced affinity and its biological implications are discussed in chapter 3.

RNA SEQUENCES USED IN THE EXPERIMENTS DISCUSSED IN THIS CHAPTER ARE LISTED IN MATERIALS AND METHODS (7.9.1, CHAPTER 7).

REFERENCES

1. Andersen, A. A. and Collins, R. A., 2000. Rearrangement of a stable RNA secondary structure during VS ribozyme catalysis. *Mol. Cell*, 5, 469-478.

2. Andersen, A. A. and Collins, R. A., 2001. Intramolecular secondary structure rearrangement by the kissing interaction of the Neurospora VS ribozyme. *Proc. Natl. Acad. Sci. U. S. A.*, 98, 7730-7735.

3. Ashraf, S. S., Ansari, G., Guenther, R., Sochacka, E., Malkiewicz, A. and Agris, P. F., 1999. The uridine in "U-turn": contributions to tRNA-ribosomal binding. *RNA*, 5, 503-511.

4. Beattie, T. L. and Collins, R. A., 1997. Identification of functional domains in the self-cleaving Neurospora VS ribozyme using damage selection. *J. Mol. Biol.*, 267, 830-840.

5. Beattie, T. L., Olive, J. E. and Collins, R. A., 1995. A secondary-structure model for the self-cleaving region of Neurospora VS RNA. *Proc. Natl. Acad. Sci. U. S. A.*, 92, 4686-4690.

6. Belousoff, M. J., Davidovich, C., Zimmerman, E., Caspi, Y., Wekselman, I., Rozenszajn, L., Shapira, T., Sade-Falk, O., Taha, L., Bashan, A., Weiss, M. S. and Yonath, A., 2010. Ancient machinery embedded in the contemporary ribosome. *Biochem. Soc. Trans.*, 38, 422–427.

7. Bonneau, E. and Legault, P., 2014. Nuclear Magnetic Resonance Structure of the III−IV−V three-way junction from the Varkud satellite ribozyme and identification of magnesium-binding sites using paramagnetic relaxation enhancement. *Biochemistry*, 53, 6264−6275.

8. Bonneau, E., Girard, N., Lemieux, S. and Legault, P., 2015. The NMR structure of the II–III–VI three-way junction from the Neurospora VS ribozyme reveals a critical tertiary interaction and provides new insights into the global ribozyme structure. *RNA*, 21, 1621–1632.

9. Bouchard, P. and Legault, P., 2014. Structural insights into substrate recognition by the *Neurospora* Varkud satellite ribozyme: Importance of U-turns at the kissing-loop junction. *Biochemistry*, 53, 258–269.

10. Bouchard, P., Lacroix-Labonte, J., Desjardins, G., Lampron, P., Lisi, V., Lemieux, S., Major, F. and Legault, P., 2008. Role of SLV in SLI substrate recognition by the Neurospora VS ribozyme. *RNA*, 14, 736-748.

11. Butler, E. B., Xiong, Y., Wang, J. and Strobel, S. A., 2011. Structural basis of cooperative ligand binding by the glycine riboswitch. *Chem. Biol.*, 18, 293–298.

12. Collins, R. A. and Olive, J. E., 1998. Spermine switches a Neurospora VS ribozyme from slow cis cleavage to fast trans cleavage. *Biochemistry*, 37, 6476-6484.

13. Dagenais, P., Girard, N., Bonneau, E. and Legault, P., 2017. Insights into RNA structure and dynamics from recent NMR and X-ray studies of the Neurospora Varkud satellite ribozyme. *WIREs RNA*, 8: e1421. doi: 10.1002/wrna.1421.

14. De la Pena, S., Gago, S. and Flores, R., 2003. Peripheral regions of natural hammerhead ribzoymes greatly increase their self-cleavage activity. *EMBO J.*, 22, 5561-5570.

15. Khvorova, A., Lescoute, A., Westhot, E. and Jayasena, S. D., 2003. Sequence elements outside the hammerhead ribozyme catalytic core enabe intracellular activity. *Nat. Struct. Biol.*, 10, 708-712.

16. Edwards, A. L., Garst, A. D. and Batey, R. T., 2009. Determining structures of RNA aptamers and riboswitches by X-ray crystallography. *Methods Mol. Biol.*, 535, 135-163.

17. Fedor, M. J., 2000. The structure and function of the hairpin ribozyme. *J. Mol Biol.*, 297, 268-291.

18. Ferré-D'Amaré, A. R., Zhou, K. and Doudna, J. A., 1998. Crystal structure of a hepatitis delta virus ribozyme. *Nature*, 395, 567-574.

19. Flinders, J. and Dieckmann, T., 2001. A pH controlled conformational switch in the cleavage site of the VS ribozyme substrate RNA. *J. Mol. Biol.*, 308, 665-679.

20. Franke, D. and Svergun, D. I., 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, 42, 342-346.

21. Golden, B. L. and Kundrot, C. E., 2003. RNA crystallization. *J. Struct. Biol.*, 142, 98-107.

22. Guo, H. C. and Collins, R. A., 1995. Efficient trans-cleavage of a stem-loop RNA substrate by a ribozyme derived from neurospora VS RNA. *EMBO J.*, 14, 368-376.

23. Guo, H. C., De Abreu, D. M., Tillier, E. R., Saville, B. J., Olive, J. E. and Collins, R. A., 1993. Nucleotide sequence requirements for self-cleavage of Neurospora VS RNA. *J. Mol. Biol.*, 232, 351-361.

24. Hiley, S. L., Sood, V. D., Fan, J. and Collins, R. A., 2002. 4-thio-U cross-linking identifies the active site of the VS ribozyme. *EMBO J.*, 21, 4691-4698.

25. Hoffmann, B., Mitchell, G. T., Gendron, P., Major, F., Andersen, A. A., Collins, R. A., and Legault, P., 2003. NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site. *Proc. Natl. Acad. Sci. U. S. A.*, 100, 7003-7008.

26. Jones C. P. and Ferré-D'Amaré, A. R. 2014. Crystal structure of a c-di-AMP riboswitch reveals an internally pseudo-dimeric RNA. *EMBO J.*, 33, 2692-2703.

27. Jones, F. D. and Strobel, S. A., 2003. Ionization of a critical adenosine residue in the Neurospora Varkud satellite ribozyme active site. *Biochemistry*, 42, 4265-4276.

28. Jones, F. D., Ryder, S. P. and Strobel, S. A., 2001. An efficient ligation reaction promoted by a Varkud Satellite ribozyme with extended 5'- and 3'-termini. *Nucleic Acids Res.* 29, 5115-5120.

29. Ke, A. and Doudna, J. A., 2004. Crystallization of RNA and RNA-protein complexes. *Methods*, 34, 408-414.

30. Kennell, J. C., Saville, B. J., Mohr, S., Kuiper, M. T., Sabourin, J. R., Collins, R. A. and Lambowitz, A.M., 1995. The VS catalytic RNA replicates by reverse transcription as a satellite of a retroplasmid. *Genes Dev.*, 9, 294-303.

31. Koldobskaya, Y., Duguid, E. M., Shechner, D. M., Suslov, N. B., Ye, J., Sidhu, S. S., Bartel, D. P., Koide, S., Kossiakoff, A. A. and Piccirilli, J. A., 2011. A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination. *Nat. Struct. Mol. Biol.*, 18, 100-106.

32. Kozin, M. B. and Svergun, D. I., 2001. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.*, 34, 33-41.

33. Lafontaine, D. A., Norman, D. G. and Lilley, D. M. J., 2001a. Structure, folding and activity of the VS ribozyme: importance of the 2-3-6 helical junction. *EMBO J.*, 20, 1415-1424.

34. Lafontaine, D. A., Norman, D. G. and Lilley, D. M. J., 2002a. The global structure of the VS ribozyme. *EMBO J.*, 21, 2461-2471.

35. Lafontaine, D. A., Wilson, T. J., Norman, D. G. and Lilley, D. M. J., 2001b. The A730 loop is an important component of the active site of the VS ribozyme. *J. Mol. Biol.*, 312, 663-674.

36. Lafontaine, D. A., Wilson, T. J., Zhao, Z. Y., and Lilley, D. M. J., 2002b. Functional group requirements in the probable active site of the VS ribozyme. *J. Mol. Biol.*, 323, 23-34.

37. Lai, M. M., 1995. The molecular biology of hepatitis delta virus. *Annu. Rev. Biochem.*, 64, 259-286.

38. Lescoute, A., and Westhof, E., 2006. Topology of three-way junctions in folded RNAs. *RNA*, 12, 83-93.

39. Lipfert, J., Ouellet, J., Norman, D. G., Doniach, S. and Lilley, D. M. J., 2008. The complete VS ribozyme in solution studied by small-angle X-ray scattering. *Structure*, 16, 1357-1367.

40. Liu, Y., Wilson, T. J., McPhee, S. A. and Lilley, D. M. J., 2014. Crystal structure and mechanistic investigation of the twister ribozyme. *Nat. Chem Biol*., 10, 739-744.

41. Liu, Y., Hart, P. J., Schlunegger, M. P. and Eisenberg, D., 1998. The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-Å resolution. *Proc. Natl. Acad. Sci. U. S. A.*, 95. 3437–3442.

42. Martick, M. and Scott, W. G., 2006. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell*, 126, 309-320.

43. McLeod, A. C., and Lilley, D. M. J., 2004. Efficient, pH-dependent RNA ligation by the VS ribozyme in trans. *Biochemistry*, 43, 1118-1125.

44. Michiels, P. J., Schouten, C H., Hilbers, C. W. and Heus, H. A., 2000. Structure of the ribozyme substrate hairpin of Neurospora VS RNA: a close look at the cleavage site. *RNA*, 6, 1821-1832.

45. Mir, A. and Golden, B. L., 2016. Two active site divalent ions in the crystal structure of the hammerhead ribozyme bound to a transition state analogue. *Biochemistry*, 55, 633-636.

46. Mir, A., Chen, J., Robinson, K., Lendy, E., Goodman, J., Neau, D. and Golden, B. L., 2015. Two divalent metal ions and conformational changes play roles in the hammerhead ribozyme cleavage reaction. *Biochemistry*, 54, 6369-6381.

47. Murray, J. B., Seyhan, A .A., Walter, N. G., Burke, J. M. and Scott, W. G., 1998. The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone. *Chem. Biol.*, 5, 587-595.

48. Olive, J. E. and Collins, R. A., 1998. Spermine switches a Neurospora VS ribozyme from slow cis cleavage to fast trans cleavage. *Biochemistry*, 37, 6476-6484.

49. Olive, J. E., De Abreu, D. M., Rastogi, T., Andersen, A. A., Mittermaier, A. K., Beattie, T. L. and Collins, R. A., 1995. Enhancement of Neurospora VS ribozyme cleavage by tuberactinomycin antibiotics. *EMBO J.*, 14, 3247-3251.

50. Ouellet, J., Byrne, M., and Lilley, D. M. J., 2009. Formation of an active site in trans by interaction of two complete Varkud Satellite ribozymes. *RNA*, 15, 1822-1826.

51. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. and Svergun, D. I., 2007. ATSAS 2.1 – towards automated and web-supported small-angle scattering data analysis. *J. Appl. Crystallogr.*, 40, s223-s228.

52. Quigley, G.J. and Rich, A., 1976. Structural domains of transfer RNA molecules. *Science*, 194, 796-806.

53. Rambo, R. P. and Tainer, J. A., 2013. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, 496, 477-481.

54. Rastogi, T. and Collins, R. A., 1998. Smaller, faster ribozymes reveal the catalytic core of Neurospora VS RNA. *J. Mol. Biol.*, 277, 215-224.

55. Rastogi, T., Beattie, T. L., Olive, J. E. and Collins, R. A., 1996. A long-range pseudoknot is required for activity of the Neurospora VS ribozyme. *EMBO J.*, 15, 2820-2825.

56. Ren, A., Vušurović, N., Gebetsberger, J., Gao, P., Juen, M., Kreutz, C., Micura, R. and Patel, D. J., 2016. Pistol ribozyme adopts a pseudoknot fold facilitating site-specific in-line cleavage. *Nat. Chem. Biol.*, 12, 702-708.

57. Rousseau, F., Schymkowitz, J. and Itzhaki, L. S., 2012. Implications of 3D domain swapping for protein folding, misfolding and function. *Adv. Exp. Med. Biol.*, 747, 137–152.

58. Rupert, P. B. and Ferré-D'Amaré, A. R., 2001. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, 410, 780-786.

59. Rupert, P. B., Massey, A. P., Sigurdsson, S. T. and Ferré-D'Amaré, A. R., 2002. Transition state stabilization by a catalytic RNA. *Science*, 298, 1421-1424.

60. Salter, J., Krucinskam J., Alam, S., Grum-Tokars, V. and Wedekind, J. E., 2006. Water in the active site of an all-RNA hairpin ribozyme and effects of Gua8 base variants on the geometry of phosphoryl transfer. *Biochemistry*, 45, 686-700.

61. Saville, B. J. and Collins, R. A., 1990. A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell*, 61, 685-696.

62. Saville, B. J. and Collins, R. A., 1991. RNA-mediated ligation of self-cleavage products of a Neurospora mitochondrial plasmid transcript. *Proc. Natl. Acad. Sci. U. S. A.* 88, 8826-8830.

63. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. and Sali, A., 2013. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* 105, 962- 974.

64. Simpson, A. A., Tao, Y., Leiman, P. G., Badasso, M. O., He, Y., Jardine, P. J., Olson, N. H., Morais, M. C., Grimes, S., Anderson, D. L., Baker, T. S. and Rossmann, M. G., 2000. Structure of the bacteriophage phi29 DNA packaging motor. *Nature*, 408, 745–750.

65. Smith, M. D. and Collins, R. A., 2007. Evidence for proton transfer in the rate-limiting step of a fast-cleaving Varkud satellite ribozyme. *Proc. Natl. Acad. Sci. U. S. A.*, 104, 5818-5823.

66. Sood, V. D. and Collins, R. A., 2001. Functional equivalence of the uridine turn and the hairpin as building blocks of tertiary structure in the Neurospora VS ribozyme. *J. Mol. Biol.*, 313, 1013-1019.

67. Sood, V. D. and Collins, R. A., 2002. Identification of the catalytic subdomain of the VS ribozyme and evidence for remarkable sequence tolerance in the active site loop. *J. Mol. Biol.*, 320, 443-454.

68. Sood, V. D., Beattie, T. L., and Collins, R. A., 1998. Identification of phosphate groups involved in metal binding and tertiary interactions in the core of the Neurospora VS ribozyme. *J. Mol. Biol.*, 282, 741-750.

69. Suslov, N. B. The crystal structure of the Varkud satellite ribozyme: Implications for RNA catalysis, Biology and evolution (PhD. Thesis), The University of Chicago.

70. Svergun D. I., Barberato C. and Koch M. H. J., 1995. CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Cryst.*, 28, 768-773.

71. Svergun, D. I., 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.*, 25, 495-503.

72. Torelli, A. T., Krucinska, J. and Wedekind, J. E. 2007. A comparison of vanadate to a 2'-5' linkage at the active site of a small ribozyme suggests a role for water in transition-state stabilization. *RNA*, 13, 1052-1070.

73. Tyagi, R. and Mathews, D. H., 2007. Predicting helical coaxial stacking in RNA multibranch loops. *RNA*, 13, 939-951.

74. Volkov, V. V. and Svergun, D. I., 2003. Uniqueness of ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.*, 36, 860-864.

75. Wilson T. J. and Lilley D. M. J., 2015. RNA catalysis-is that it? *RNA*, 21, 534-537.

76. Wilson, T. J., Li, N-S., Lu, J., Frederiksen, J. K., Piccirilli, J. A. and Lilley, D. M. J. Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme. *Proc. Natl. Acad. Sci. U. S. A.*, 107, 11751-11756.

77. Wilson, T. J., McLeod, A. C. and Lilley, D. M. J., 2007. A guanine nucleobase important for catalysis by the VS ribozyme. *EMBO J.*, 26, 2489-2500.

78. Wong, T. N., Sosnick, T. R. and Pan, T., 2007. Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proc. Natl. Acad. Sci. U. S. A.*, 104, 17995-18000.

79. Zhao, Z. Y., McLeod, A., Harusawa, S., Araki, L., Yamaguchi, M., Kurihara, T. and Lilley, D. M. J., 2005. Nucleobase participation in ribozyme catalysis. *J. Am. Chem. Soc.*, 127, 5026-5027.

Chapter - 3

CRYSTAL STRUCTURE OF THE VS RIBOZYME: SUBSTRATE BINDING AND CATALYSIS

CONTENTS OF THIS CHAPTER HAVE BEEN PUBLISHED AS ARTICLES IN THE JOURNAL OF AMERICAN CHEMICAL SOCIETY (***J. Am. Chem. Soc.* 2017, 139, 9591–9597**) *AND NATURE CHEMICAL BIOLOGY* (***Nat. Chem. Biol.* 2015, 11, 840–846**). *ALL MATERIALS OF THE ARTICLE HAVE BEEN ADAPTED WITH THE COPYRIGHT PERMISSION FROM THE AMERICAN CHEMICAL SOCIETY AND THE NATURE PUBLISHING GROUP.*

3.1 INTRODUCTION

3.1.1 Secondary structure rearrangements in RNA

RNA folding frequently occurs in a hierarchical manner, in which secondary structure forms before tertiary structure. The process of folding occurs through two distinct steps. Stable secondary structures like stem-loops are formed rapidly on a microsecond scale (Micura and Höbartner, 2003). After the formation of a stable secondary structure, formation of tertiary structure occurs by assembling stable secondary structure motifs through long-range base-pairing interactions. Tertiary folding takes place in the millisecond to second scale (Micura and Höbartner, 2003). However, local secondary structures can be altered through the formation of tertiary interactions facilitated by RNA folding or upon complexation with ligands or proteins (Wu and Tinoco Jr., 1998; Silverman et al., 1999; Knitt et al., 1994; Zhao et al., 2015; Savva et al., 1995; Gilbert et al., 2006). The ability of a single RNA sequence to adopt multiple secondary structures have resulted in the creation of molecular switches to regulate biological processes. Alternate structures of RNA hairpins modulate gene expression in *Escherichia coli* and *Bacillus subtilis* (Fayat et al., 1983; Putzer et al., 1992; Babitzke and Yanofsky, 1993). The elongation cycle of protein synthesis involves elongation factors EF-Tu and EF-G that bind to alternative conformations of the 28S rRNA thus potentially presenting a layer of regulation (Wool et al. 1992).

Spliceosomal snRNAs undergo dramatic changes in secondary structure during snRNP maturation, spliceosome assembly, and splicing brought about by protein binding and the action of ATP-dependent helicases (Nguyen et al., 2016). An RNA switch is thought to regulate tRNA recognition at the ribosomal A-site by modulating the codon-anticodon arrangement between mRNA and tRNA (Lodmell and Dahlberg,1997; von Ahsen, 1998). Riboswitches are examples of natural RNA switches that regulate gene expression at the transcriptional or translational levels by switching between its apo and holo-conformations (Tucker and Breaker, 2005). Artificial RNA switches that can adopt two distinct secondary structures by virtue of their sequence design have been constructed and studied by NMR spectroscopy (Höbartner and Micura, 2003). Although secondary structure rearrangements are not uncommon in RNA folding and function, only in a few cases have the pre- and post-rearrangement structures been defined. Herein, we focus on the substrate helix of the Varkud satellite (VS) ribozyme, which rearranges upon binding to the catalytic domain of the ribozyme to create the active site for catalysis.

3.1.2 VS substrate has a unique sequence

Helix 1 of the VS ribozyme harbors the cleavage site and is therefore referred to as the substrate-helix. The overall substrate sequence can be divided into four distinct segments (Figures 2.6A, 3.3). The hairpin loop consists of unpaired nucleotides that participate in a tertiary kissing-loop interaction with the hairpin nucleotides of stem-loop 5. This kissing-loop interaction is responsible for substrate recognition and binding (Rastogi et al., 1996). Hairpin nucleotides are supported by a helix (helix 1b) that consists of four G-C base-pairs. The stem following the C-G pair located immediately under the substrate hairpin comprises an array of three G-C pairs created by complementary base-pairing between the contiguous guanine and cytosines in the 5' and 3'

strands, respectively. Helix 1b ends in an internal bulge containing the cleavage site (scissile phosphate flanked by G620 and A621) and a catalytic guanine (G638). This bulge, referred to as the internal cleavage loop is followed by helix 1a that connects helix 1 to helices 2 and 7 via junction 1-2-7 (Figure 2.6A). The substrate-helix undergoes extensive structural remodeling on binding to the catalytic domain of the ribozyme which spans the substrate hairpin, helix 1b and the internal cleavage loop. Dimethyl sulfate (DMS) probing unveiled a Mg^{2+}-induced rearrangement in the secondary structure of substrate-helix 1b that involve a single-nucleotide register shift in the 3' strand of the substrate (Andersen and Collins, 2000). This helix shift is induced by the kissing-loop interaction between the substrate loop and stem-loop 5 (Andersen and Collins, 2001). In addition to being an essential precondition for cleavage, 'shifted' substrates were found to bind to the catalytic domain with greater affinity (Zamel and Collins, 2002). While the requirement for substrate rearrangement for cleavage might derive partly from tighter binding to the catalytic domain, no mechanistic framework was suggested to explain the relationship between the kissing-loop interaction, register shift in helix 1b and catalytic cleavage.

Substrate remodeling is accompanied by restructuring of other regions in the catalytic domain that interact with the active substrate. These include the hairpin of stem-loop 5 that participates in the kissing-loop interaction hence directly plays a role in inducing substrate rearrangement and the A730 loop in helix 6 that stabilizes the catalytically-active conformation of the internal cleavage loop in the substrate-helix.

3.2 RESULTS

3.2.1 Remodeling of the A730 loop

In the course of substrate docking, the substrate-helix is juxtaposed against helix 2 and 6, which brings the catalytic adenine, A756 close to the cleavage site as a part of the catalytically primed active site. This involves local restructuring of the A730 loop in helix 6 (Figure 3.1A, B) as revealed by comparing the NMR structure of an isolated stem-loop 6 with the crystal structure of the full-length ribozyme. The overall features of the A730 loop within helix 6 are similar in both structures. An A730-G757 *cis* WC-WC base-pair widens helix 6 (Figure 3.1C, D), allowing formation of an S-turn motif that extrudes C755 and the catalytic nucleobase, A756 from the helical stack (Figure 3.1 A, B). The crystallographic model exhibits a more pronounced widening of the minor groove and a difference in overall curvature of the backbone near the S-turn, with the extruded nucleobases projecting further outward from the helical stack to mediate interactions with nucleotides in the active site region of helix 1 (Figure 3.1 A, B). S-turn is a common structural motif in RNA that was first identified in eukaryotic 5S rRNA (Wimberly et al., 1993) and the sarcin-ricin loop of 28S rRNA (Szewczak et al., 1993). This backbone feature enables extrusions of nucleotides from helical stacks allowing them to interact with residues otherwise distant in the secondary structure of the RNA.

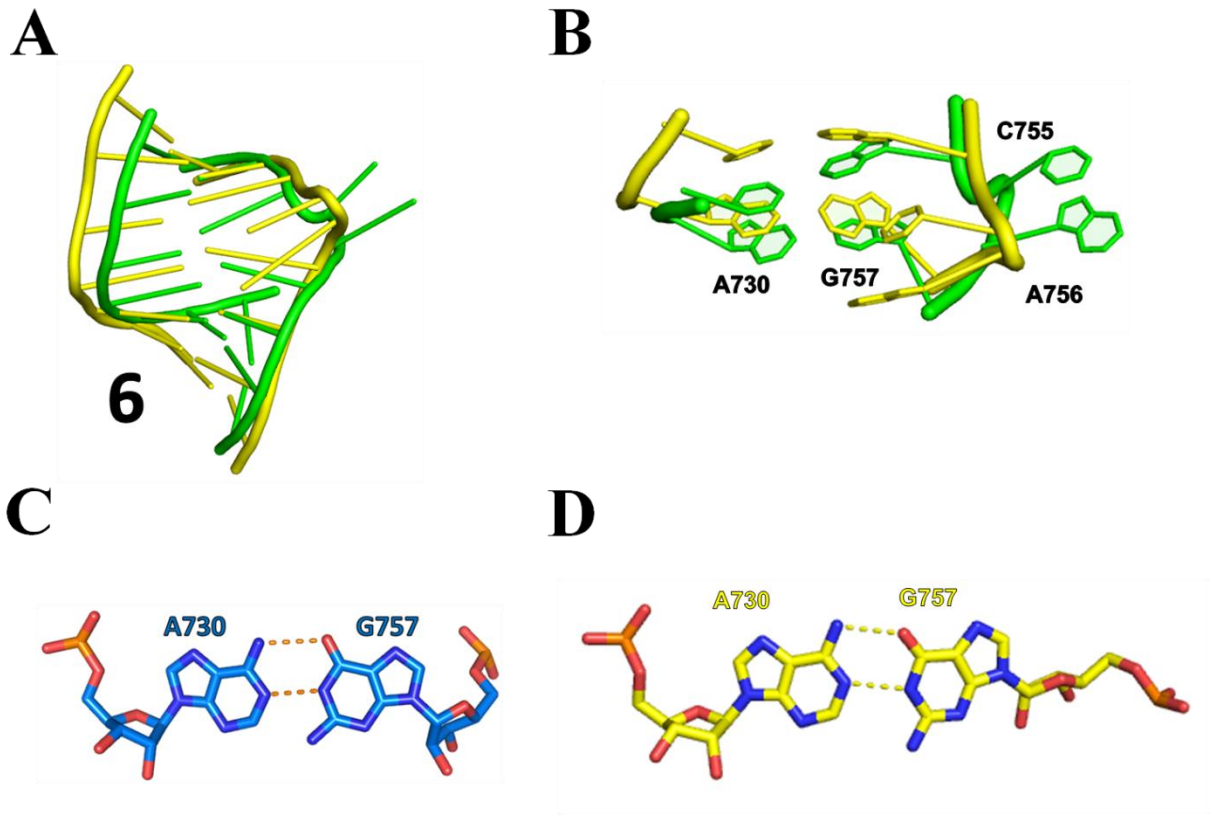

Figure 3.1 Comparison of helix 6 as revealed by NMR and crystal structures. A. Alignment of helix 6 obtained from the crystal structure (green, PDB ID code: 4R4P, this work) with the NMR structure (yellow, PDB ID code: 2L5Z, Desjardins et al., 2011) of an isolated stem-loop show overall agreement. C755 and A756 are extruded from the helical stack in the crystal structure suggesting a structural rearrangement during folding and probably substrate binding. **B.** Alignment of the A730 loop from crystal and NMR structures of helix 6. C755 and A756 assume extrahelical positions, supported by an S-turn motif. **C.** *Cis*-WC A730-G757 base-pair in the crystal structure. **D.** *Cis*-WC A730-G757 base-pair in the NMR structure.

3.2.2 Remodeling of stem-loop 5

Nucleotides involved in the kissing-loop interaction that are located in the hairpins of stem-loop 1 and 5, are unpaired and generally flexible allowing them to orient themselves into a configuration favorable for the formation of WC pairs. A U-turn motif that organizes the stem-loop 5 hairpin uses a Mg^{2+} ion to stabilize its conformation (Campbell et al., 2006). There is subtle reorganization in the hairpin on binding Mg^{2+} that is most pronounced at nucleotides C699 and U700 leading to the formation of a more compact structure, an observation that has been supported by computational studies (Figure 3.2) (Bergonzo et al., 2015). In the presence of Mg^{2+}, nucleotides U696, G697 and A698 constitute the U-turn fold while residues U695 and A701 form a canonical base-pair that closes the hairpin (Figure 3.2A-C, E). This U-A base-pair persists even in the absence of Mg^{2+} as a part of an isolated stem-loop (Figure 3.2D). However, in the crystal structure that captures the complete ribozyme-substrate complex, U695 and A701 are not close enough to form a stable WC base-pair. Their lack of coplanarity further prevents the formation of this pair. This frees A701 to participate in the kissing-loop interaction with C629 from the substrate hairpin (see below).

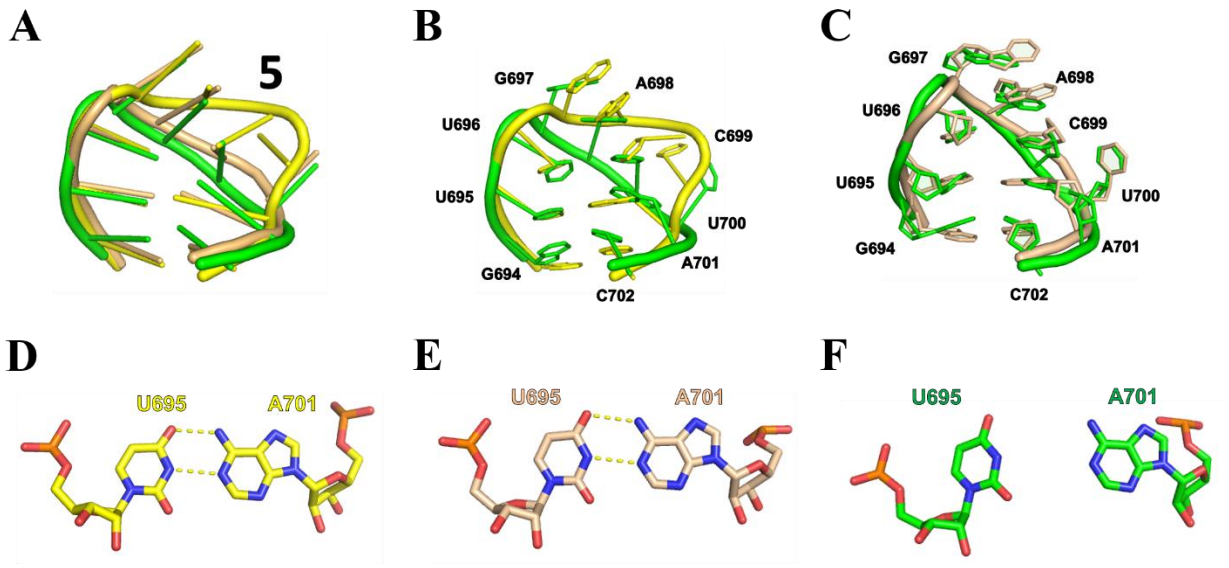

Figure 3.2 Comparison of the hairpins of stem-loop 5 as revealed by NMR and crystal structures. A. Alignment of stem-loop 5 from crystal structure (green, PDB ID code: 4R4V, this work) with the NMR structures of isolated stem-loops in the presence (wheat, PDB ID code: 1YN1, Campbell et al., 2006) and absence of Mg^{2+} (yellow, PDB ID code: 1TBK, Campbell et al., 2005). Structures in the presence of Mg^{2+} are in agreement with each other but differ from the structure of the hairpin in the absence of Mg^{2+}. **B.** Alignment of stem-loop 5 hairpins from crystal structure in the presence of Mg^{2+} and NMR structure in the absence of Mg^{2+}. There is a pronounced narrowing of the hairpin structure in the presence of Mg^{2+} with extrusion of U700 from the stack. **C.** Alignment of stem-loop 5 hairpins from crystal and NMR structures in the presence of Mg^{2+}. Both structures are in good agreement with each other. **D.** WC base-pair: U695-A701 in the absence of Mg^{2+} as revealed by NMR. **E.** WC base-pair: U695-A701 in the presence of Mg^{2+} as revealed by NMR. **F.** U695 and A701 do not form a base-pair in the crystal structure.

3.2.3 Remodeling of the substrate

3.2.3.1 Structure of the isolated substrate-helix

Dieckmann and co-workers have described the NMR structure of an RNA oligonucleotide corresponding to the natural sequence of the substrate-helix (nucleotides U617 to G642), which encompasses a portion of helix 1a, the internal cleavage loop, helix 1b and the terminal loop (PDB ID code: 1HWQ, Flinders and Dieckmann, 2001). Their analysis showed that helix 1b adopts the so-called "unshifted" secondary structure, consisting of four WC base-pairs (G623-C636, G624-C635, G625-C634, and C626-G633) capped by a dynamic hexanucleotide loop (Figure 3.3A); the predicted base-pair between G627 and C632 nucleotides is not stable.

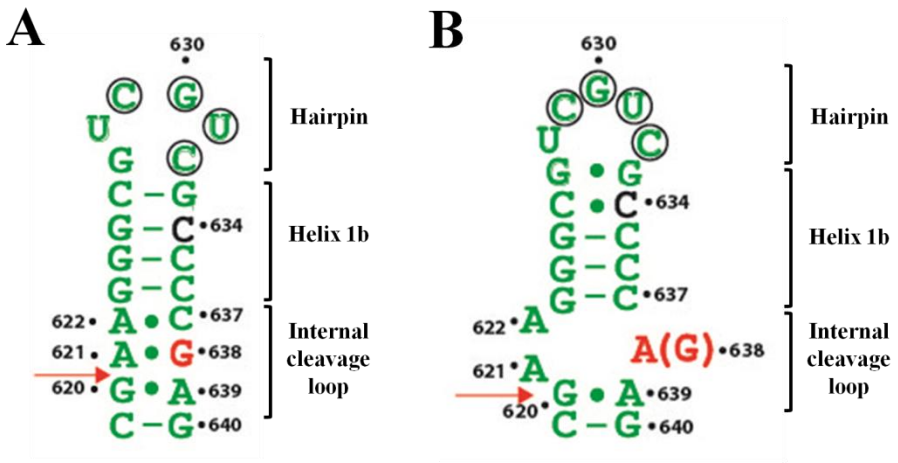

Figure 3.3 Secondary structures of the substrate-helix derived from NMR and crystallography. A. Secondary structure of the undocked substrate derived from the NMR structure of an isolated stem-loop (PDB ID code: 1HWQ, Flinders and Dieckmann, 2001). **B.** Secondary structure of the substrate derived from the crystal structure of the full-length ribozyme with docked substrate (PDB ID code: 5V3I, this work). Nucleotides involved in the loop−loop kissing interaction with stem-loop 5 are circled, catalytic nucleobases are shown in red, C634 is shown in black, and the cleavage sites are indicated by red arrows.

The internal cleavage loop samples a compact 'closed' conformation and a more dynamic 'open' conformation in a pH-dependent manner (Flinders and Dieckmann, 2001). The closed conformation characterized by three non-canonical base-pairs, two tandem sheared G-A pairs (G620-A639 and A621-G638) and a sheared A^{+}-C pair (A622-C637), in which A622 is protonated at N1, predominates at lower pH (Figure 3.3A) (Michealis et al., 2000; Flinders and Dieckmann, 2001). In this closed conformation, A639 forms a cross- strand stack with nucleotide A621, which likely stabilizes the non-canonical pairs and disfavors the splayed conformation at the cleavage site (Figure 3.4A). At higher pH values, A622 deprotonates and the open conformation predominates. In this conformation, A622 disengages from its pairing with C637 and adopts an extra-helical position. The internal loop becomes more dynamic, but the sheared G-A pairs and base stacking of nucleotides opposite A622 remain intact. Disruption of the A622-C637 pair by deletion of C637 (ΔC637) also shifts the internal cleavage loop to a more open conformation (Hoffmann et al., 2003), suggestive of the role of the A622-C637 base-pair in the conformational dynamics of the cleavage loop (Figure 3.4B). The ΔC637 substrate-helix (PDB ID code: 1OW9, Hoffmann et al., 2003) preserves the G620-A639 sheared pair, but G638 forms a shared-sheared base-pair with both A621 and A622 (Figure 3.4B). This conformation likely represents an intermediate state *en route* to the catalytically relevant conformation of the cleavage loop captured in the crystal (Figure 3.4C).

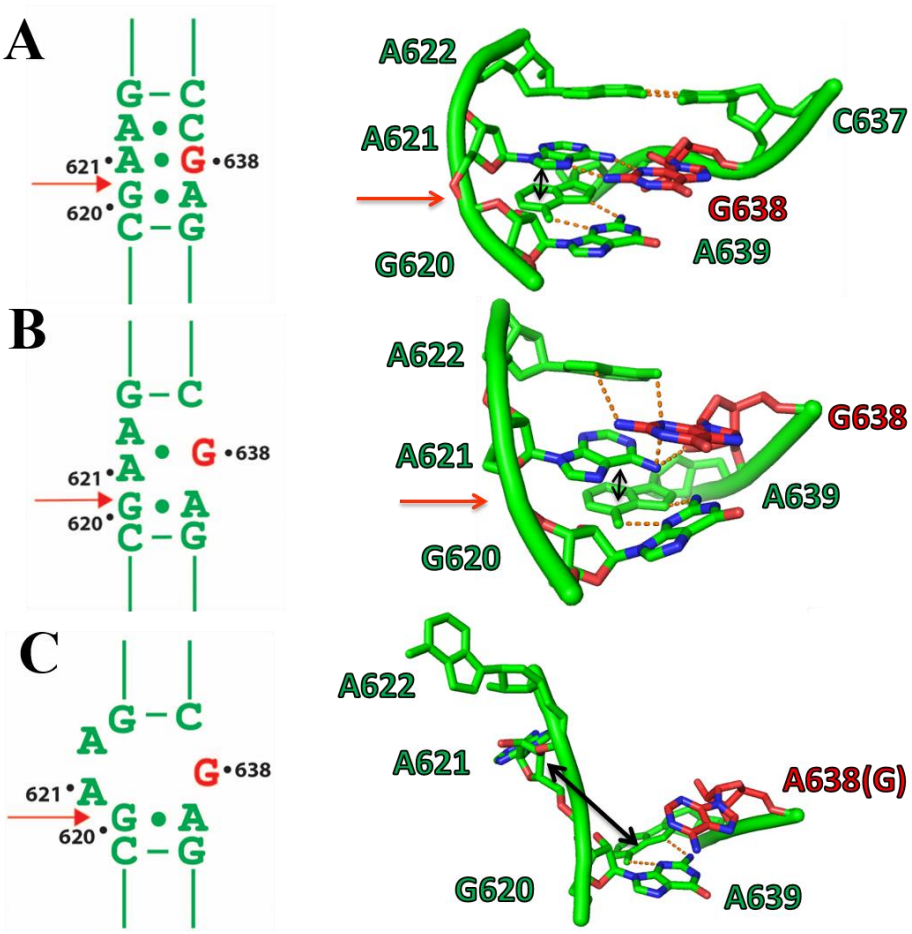

Figure 3.4 Internal cleavage loops in the VS substrate. A. NMR structure (right) and secondary structure derived from the NMR structure (left), of the internal cleavage loop from an isolated wild-type substrate (PDB ID code: 1HWQ). Cleavage site and catalytic guanine are shown in red. Cross-strand stack is shown by a double-headed arrow. The cleavage loop consists of three non-canonical base-pairs, one of them involving the catalytic G (G638) and the N+1 nucleotide (A621). **B.** NMR structure (right) and secondary structure derived from the NMR structure (left), of the internal cleavage loop from an isolated mutant substrate with an 'open' cleavage loop as a result of deleting the C637 nucleotide (PDB ID code: 1OW9). G638 interacts with both A621 and A622 and is positioned in a plane between the two nucleotides. Coloring and labeling scheme as in A. **C.** Crystal structure (right) and secondary structure derived from the crystal structure (left), of the internal cleavage loop from a docked substrate in the context of the full-length ribozyme, activated for cleavage (PDB ID code: 5V3I). This structure contains a G638A mutation. The cleavage loop is flexible as a result of disruption of the base-pairs in A. A621 and A622 are extruded from the helix and G638 is unpaired. Coloring and labeling scheme as in A.

3.2.3.2 Structure of the docked substrate-helix

The VS substrate docks into a catalytic cleft created by helices 2 and 6, oriented accordingly by junction 2-3-6. The location of the substrate relative to these helices allows formation of the active site and is locked into position by stem-loop 5 via the kissing-loop interaction. The kissing-loop interaction consists of three WC base-pairs, G630-C699, U631-A698, and C632-G697, and a non-canonical base-pair C629-A701 (Figure 3.5) arranged in an A-form helical conformation that renders stacking between helix 5 and the substrate-helix contiguous (Figure 3.5A). Two unpaired nucleotides in the docked substrate-helix, G633 and C634, contribute to this stacking (Figure 3.5A). The base-paired region under the hairpin loop, referred to as helix 1b, consists of three WC base-pairs: G623-C637, G624-C636, and G625-C625 (Figure 3.3B), and one non-canonical C626-C634 pair (Figure 3.3B) (Nagaswamy et al., 2000). The internal cleavage loop contains the nucleotide A638 (catalytic G638 in wild-type), positioned in close proximity to the scissile phosphodiester that links G620 and A621 (Figures 3.3B, 3.4C). The G623-C637 base-pair (Figure 3.4C) and a sheared G620-A639 base-pair (Figure 3.3B, 3.4C) close the internal loop from above and below, respectively. Within the internal loop immediately following G620, two nucleotides, A621 and A622, flip out of the helix and insert into individual binding pockets involving nucleotides from helices 2 and 6 (Figure. 3.4C) (see section 3.2.4.2 for a detailed description of the active site and tertiary interactions between substrate and catalytic domains).

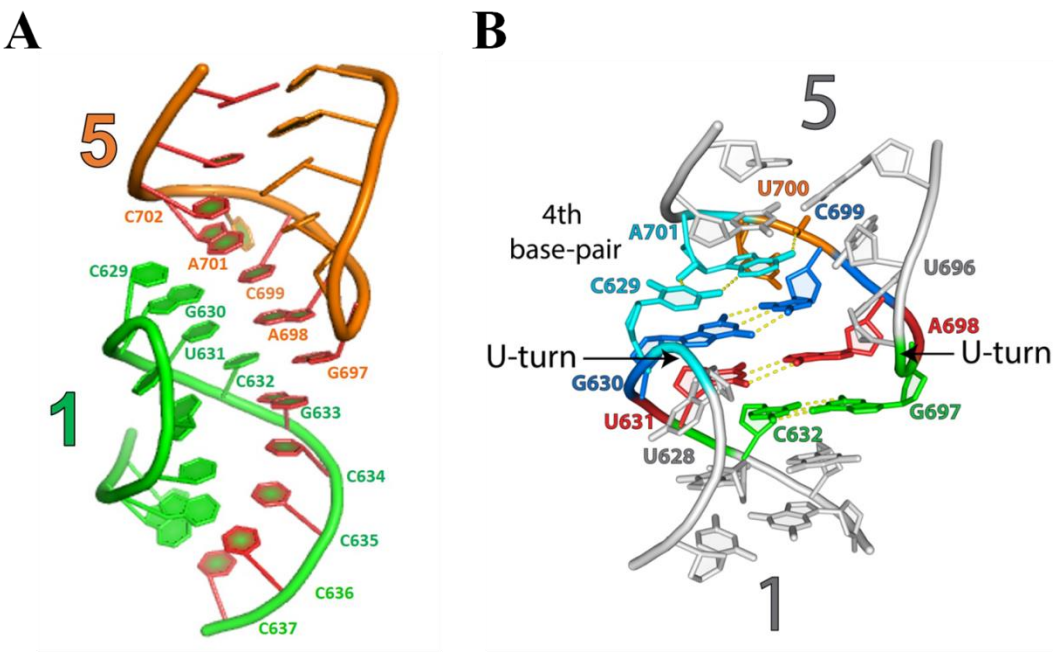

Figure 3.5 Kissing-loop interaction between substrate and stem-loop 5. A. Extended stacking between nucleotides of the substrate (green) and stem-loop 5 (orange), shown in red. This adds stability to the kissing-loop interaction between the substrate and catalytic domain. **B.** Kissing-loop interaction creates an A-form helix, comprising three WC base-pairs and a non-canonical pair.

3.2.3.3 Rearrangements in the substrate hairpin

Comparison of the NMR structure of the isolated substrate-helix (PDB ID code: 1HWQ) with the crystal structure of the full-length ribozyme, which captures the conformation of the substrate engaged in the kissing interaction, reveals the conformational rearrangements that occur upon docking of the substrate-helix into the ribozyme (Figure 3.6). Hairpin nucleotides undergo substantial reorganization upon substrate binding resulting in a more structured loop. Nucleobases that make up the mini-helix formed as a result of the kissing-loop interaction between the substrate and stem-loop 5, go from being randomly oriented to being stacked with their WC edges pointing outward, suitably poised for the tertiary interaction (Figure 3.5B). Beginning with the closing base-pair of the hairpin loop in the docked structure, the WC edge of C632 faces away from G627,

its potential pairing partner in the undocked helix, and instead participates in the kissing interaction by forming a base-pair with G697 in stem-loop 5 (Figures 3.5B, 3.6B). In this respect, a dynamic or absent G627-C632 base-pair in the undocked helix as implied by the NMR data (Figure 3.3A) likely facilitates substrate binding relative to a stable base-pair (Zamel and Collins, 2002); nevertheless, the exocyclic amines of G627 and C632 reside within 3.1 Å of one another (Figure 3.6C). The nucleotide immediately downstream of C632, G633, forms a base-pair with C626 in the unbound substrate-helix, closing the hexaloop (Figure 3.3A) (Flinders and Dieckmann, 2001). This base-pair was predicted to persist in the bound form of the substrate based on the lack of DMS reactivity at C626 (Andersen and Collins, 2000; Andersen and Collins, 2001). C634, a nucleotide that forms a base-pair with G625 in the unbound substrate, was predicted to be unpaired and extruded out of the helical stack in the bound substrate, on the basis of its increased reactivity to DMS (Andersen and Collins, 2000; Andersen and Collins, 2001). Contrary to these expectations, C626 and C634 disengage from the WC base-pairs with G633 and G625 respectively and form a non-canonical base-pair (Nagaswamy et al., 2000), in which the N3 atom of C626 accepts a hydrogen bond from the exocyclic amino group of C634 (Figure 3.6D). This rearrangement directs the WC edge of C626 toward the interior of helix 1b (Figure 3.6B), consistent with its observed protection from DMS modification in the presence of saturating ribozyme (Andersen and Collins, 2000; Andersen and Collins, 2001). In contrast, the WC edge of C634 becomes solvent exposed in the docked state, consistent with the observed increase in DMS reactivity caused by the presence of saturating ribozyme or an isolated stem-loop 5 (Andersen and Collins, 2000; Andersen and Collins, 2001).

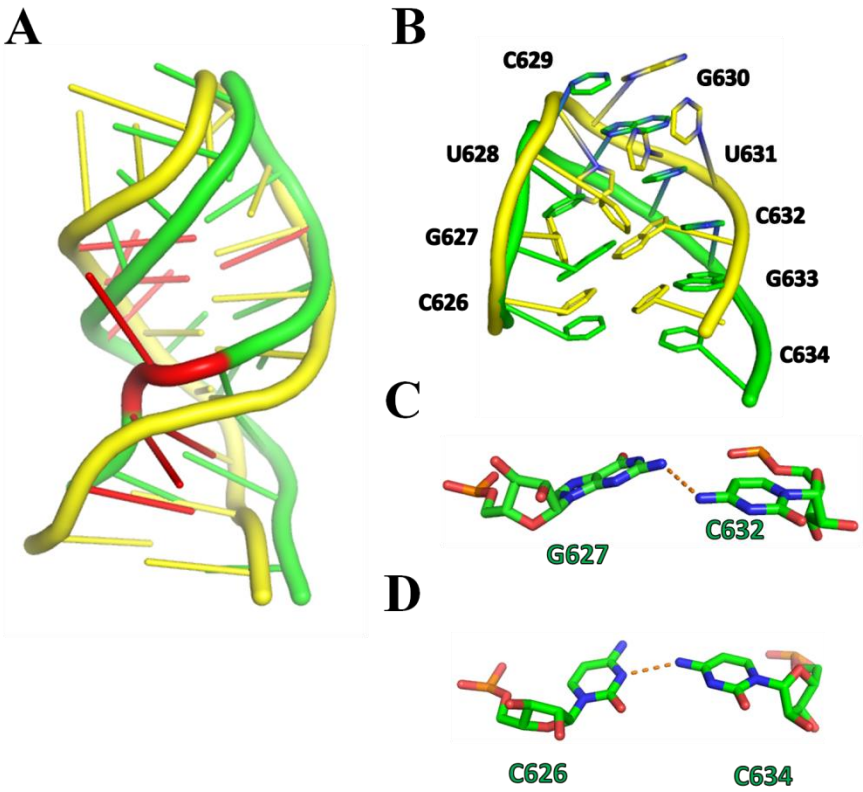

Figure 3.6 Overlay of undocked (yellow) and docked (green) substrates: substrate hairpin. A. Complete substrate helices obtained from crystallography (PDB ID code: 5V3I, this work) and NMR (PDB ID code: 1HWQ, Flinders and Dieckmann, 2001) superimposed on each other shown in cartoon mode. Helices have been anchored on the U631 nucleotide for best superposition results. Nucleotides are depicted in ladder mode. Nucleotides that undergo the most drastic changes on docking are colored in red. The backbone region that undergoes the largest conformational change is shaded red. **B.** Alignment of stem-loop 1 from crystal structure (green, PDB ID code: 5V3I, this work) with the NMR structure of an isolated stem-loop (PDB ID code: 1HWQ, Flinders and Dieckmann, 2001). **C.** Previously uncharacterized H-bond between C632 and G627. **D.** Previously uncharacterized non-canonical base-pair between C626 and C634.

Disengaged from C626, G633 sits between C634 and G697, rendering the stacking between helix 1b and the kissing duplex contiguous (Figure 3.5A) (Bouchard and Legault, 2014). This extended stacking stabilizes the kissing-loop interaction, thereby facilitating substrate binding to the catalytic domain of the ribozyme. The kissing interaction expands the hairpin loop (Figure 3.3B) and causes significant changes in backbone pseudotorsional angles $\Delta(\eta,\theta)$ as compared to the undocked state (section 3.2.3.5). Upon docking, the backbone undergoes a sharp turn between

C632 and C635 on the 3′ side of the substrate-helix (Figure 3.6A) with nucleotides G633 and C634 remaining inside the helical stack (Figures 3.3B, 3.7A). Concomitant with the disruption of the C626-G633 base-pair in this loop expansion, helix 1b undergoes a reorganization that involves three GC pairs, consistent with the biochemically inferred secondary structure (Andersen and Collins, 2000; Andersen and Collins, 2001). In this reorganization, the three GC base-pairs in the unbound substrate (G625-C634, G624-C635, and G623-C636) are replaced by three new GC pairs (G625-C635, G624-C636, and G623-C637), which represent a change in nucleotide register in helix 1b creating the so-called 'shifted' substrate-helix (Figures 3.3, 3.7A, 3.16A for electron density omit map).

3.2.3.4 Rearrangements in the internal cleavage loop

Conformational changes associated with substrate docking extend beyond the terminal loop and helix 1b and into the internal cleavage loop, disrupting two of the sheared pairs (A622-C637 and A621-G638), and adjusting the conformation of the third (G620-A639) (Figure 3.4C). A622 disengages from C637; the latter forms the G623-C637 base-pair in helix 1b of the docked substrate as part of the register change. *En route* to the docked state, A622 rotates out of the helical stack from the major groove face into the minor groove of helix 2 near junction 2-3-6 (Figure 3.7B). With its nucleobase now oriented nearly parallel to the substrate-helix, A622 stacks with A657 and forms hydrogen bonds with G654 and the G655-C770 base-pair (section 3.2.4.2). A621 disengages from G638, its pairing partner in the undocked state and rotates out of the helical stack from the major groove face of the substrate-helix into the minor groove of helix 2 (Figure 3.7B), where it forms a base-triple with the G653-C771 base-pair in the docked state (section 3.2.4.2). In a similar manner, A639 rotates out of the helical stack maintaining both its interaction with its

pairing partner G620 and its coaxial orientation with A621 (Figure 3.7B). However, strand shifting upon docking displaces G620 from the plane of A639 and creates a 10.6 Å space between the A639 and A621 nucleobases (Figures 3.12, 3.14). Filling this gap are two nucleobases from the A730 loop, C755 and A756, that project from an S-turn (Figure 3.1B) to form a contiguous (A621-A756-C755-A639) stack (Figures 3.12, 3.14). Restructuring of the internal loop creates a sharp turn in the phosphodiester backbone thereby splaying the nucleotides flanking the scissile phosphate (Figures 3.4C, 3.6A, 3.14, 3.15). As a result, the docked state brings the reactive groups closer to the ideal in-line geometry. G638 (A638 in this structure), released by the extrusion of A621, stacks against G620, and A756 inserts beneath the extruded A621, poised to mediate general acid/base catalysis (section 3.2.5).

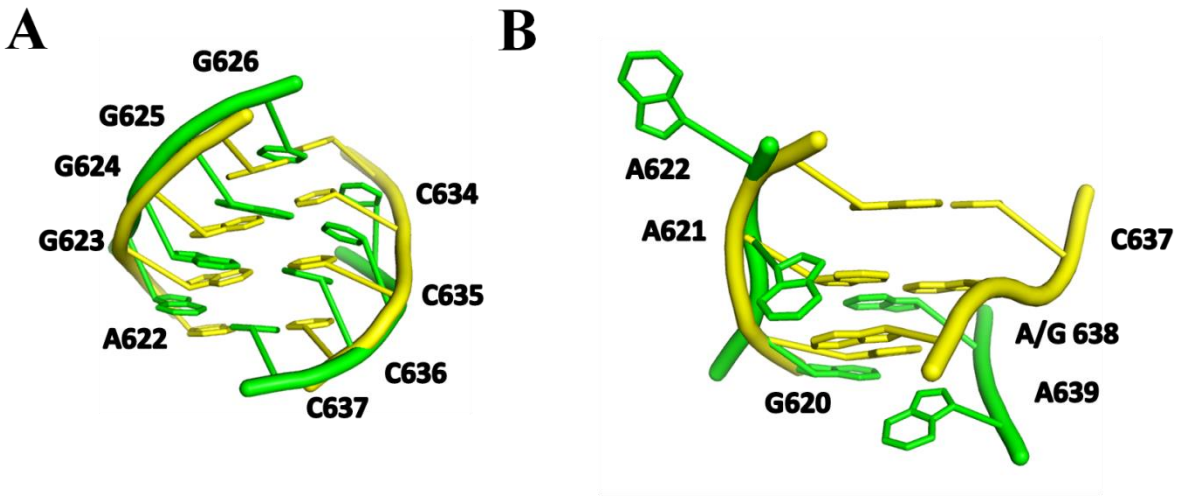

Figure 3.7 Overlay of undocked (yellow) and docked (green) substrate: helix 1b and internal cleavage loop. A. Helix 1b superimposed on each other. The docked helix reveals a shift in register, replacing three WC base-pairs in the isolated substrate with three new WC base-pairs. **B.** Internal cleavage loops superimposed on each other. The cleavage loop becomes more flexible on substrate docking primarily as a result of the disruption of non-canonical base-pairs in the cleavage loop of the isolated substrate and the extrusion of bases, A621 and A622 from the loop. PDB ID codes of undocked and docked structures are 1HWQ and 5V3I, respectively.

3.2.3.5 Structural rearrangements in the substrate-helix: a semi-quantitative analysis

Local conformational changes in the RNA backbone can be quantitatively analyzed using a computational method developed by Pyle and co-workers (Keating et al., 2011 and Zhao et al., 2015). This method is based on the descriptions of RNA backbone conformation defined by a set of angular coordinates (η and θ) analogous to dihedral angles (φ and ψ) used to describe peptide bond conformation in proteins (Keating et al., 2011). η and θ angles define the angular position of vectors connecting P and C4′ atoms of the RNA backbone (Duarte and Pyle, 1998; Duarte et al., 2003; Keating et al., 2011). To compare the extents of remodeling at different segments of the substrate-helix, we parameterized the structures of the substrate-helix as obtained by NMR (unshifted) and crystallography (shifted) with a set of η and θ values for each nucleotide (see Keating et al., 2011 and Zhao et al., 2015 for the mathematical treatment for calculating η and θ). Difference in pseudo-torsion angles, Δ(η,θ) between the unshifted and shifted structures for each nucleotide position provides a measure of backbone restructuring.

Values of Δ(η,θ) are the largest at the cleavage loop and hairpin (Figure 3.8). This is consistent with both functional and structural data (Beattie et al., 1995). High values at G620 and A621 support the reorientation of G620 and the splayed conformation created by G620 and A621 that facilitates in-line attack by the 2'-oxygen of G620. The largest difference in pseudo-torsion angles is observed at A622, (Figure 3.8) which is due to the sharp turn observed in the crystal structure of a docked substrate (Figures 3.4A, C, 3.7B). Reorganization of the cleavage loop into a catalytically active structure is reflected in the high Δ(η,θ) values at nucleotides G638 and A639. Backbone restructuring at these nucleotides widens the cleavage loop, thereby creating space between A639 and the extruded A621, which allows cross-strand stacks (C755 and A756) from helix 6 (Figure 3.4A, C).

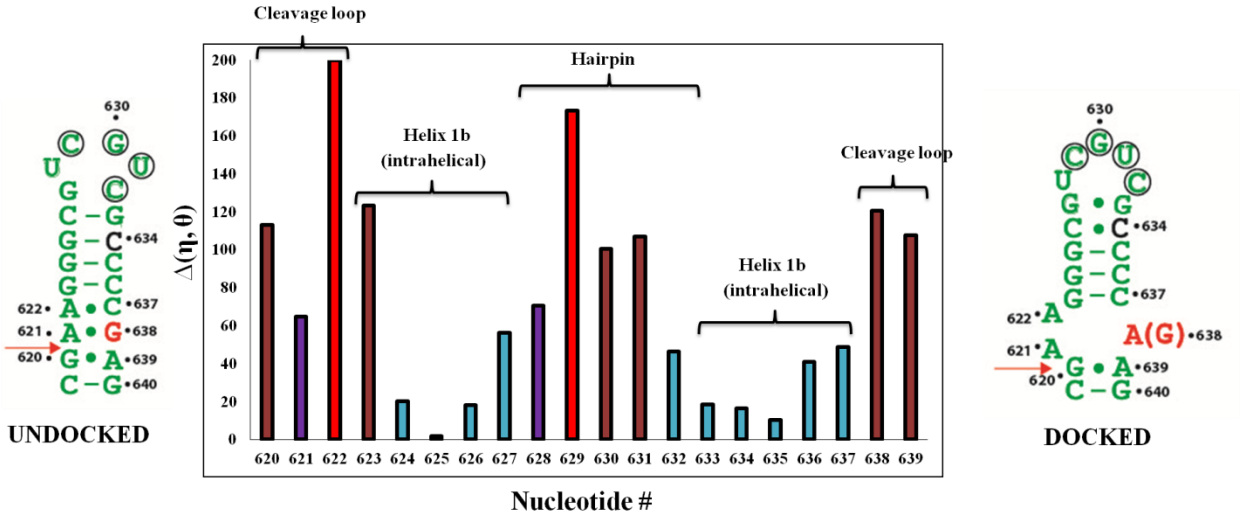

Figure 3.8 Analysis of changes in backbone geometry of the substrate on docking. Major changes in backbone conformations are observed in the internal cleavage loop and stem loop of the substrate, after it docks into the catalytic domain of the ribozyme. Nucleotides involved in the kissing-loop interaction with stem-loop 5 are circled, catalytic nucleobase G is shown in red, C634 is shown in black, and the cleavage site is indicated by a red arrow. Δ(η,θ) (Duarte and Pyle, 1998; Duarte et al., 2003; Keating et al., 2011) indicates the differences in pseudo-torsion angles between docked and undocked substrate helices; the greater its value, the greater is the conformational change. Color-coding as follows: red bars, Δ(η,θ) > 150; brown bars, 150 > Δ(η,θ) < 80; purple bars, 80 > Δ(η,θ) < 50; blue bars, Δ(η,θ) < 50.

Nucleotides in the hairpin, especially those involved in mediating the kissing-loop interaction, undergo structuring during substrate docking. This is consistent with high values of Δ(η,θ) at C629, G630 and U631. A low Δ(η,θ) value at C632 probably reflects the effect of pre-organization of this nucleotide due to potential base-pairing within the substrate-helix (G627-C632). Re-organization of nucleotides at the hairpin could potentially use C632 as a guide to assemble the helical stack seen in the docked substrate hairpin (Figure 3.6B). A sharp change in backbone conformation at C629 suggested by a high Δ(η,θ) value positions the nucleotide to interact with A701 of stem-loop 5. The importance of this nucleotide for the formation of the kissing interaction and consequently cleavage activity is highlighted by 10-100-fold cleavage defects when mutations were made to this nucleotide (Rastogi et al., 1996). Nucleotides in helix

1b that maintain base-pairing and intrahelical stacking throughout docking and rearrangement of the substrate-helix exhibit lower values of $\Delta(\eta,\theta)$.

3.2.4 Tertiary interactions stabilize substrate binding and activation

In the dimer structure, substrate-helix of each protomer forms binding interactions with the other protomer in two regions: the terminal stem loop engages the loop of helix 5 to form a kissing-loop interaction and the internal loop, which harbors the cleavage site and the putative general base, G638, interacts with a cleft formed by helix 6 (containing the general acid, A756) and helix 2 (Figures 2.12, 3.9). These interactions rearrange the substrate-helix and prime the active site for catalysis. The kissing-loop interaction plays vital roles in substrate binding and extensively remodels the substrate-helix. As discussed above, substrate remodeling spans the hairpin (section 3.2.3.3), helix 1b (section 3.2.3.3) and the cleavage site (section 3.2.3.4). The remodeled structure is thermodynamically less stable due to a decrease in base-stacking and base-pairing within the substrate-helix (Figure 3.3). This energy penalty is offset by additional stabilization provided by the nucleotides from helices 2 and 6, by forming binding pockets to accommodate the pair of adenine bases extruded from the helical stack during substrate docking, in addition to the kissing-loop interaction with stem-loop 5 as described below.

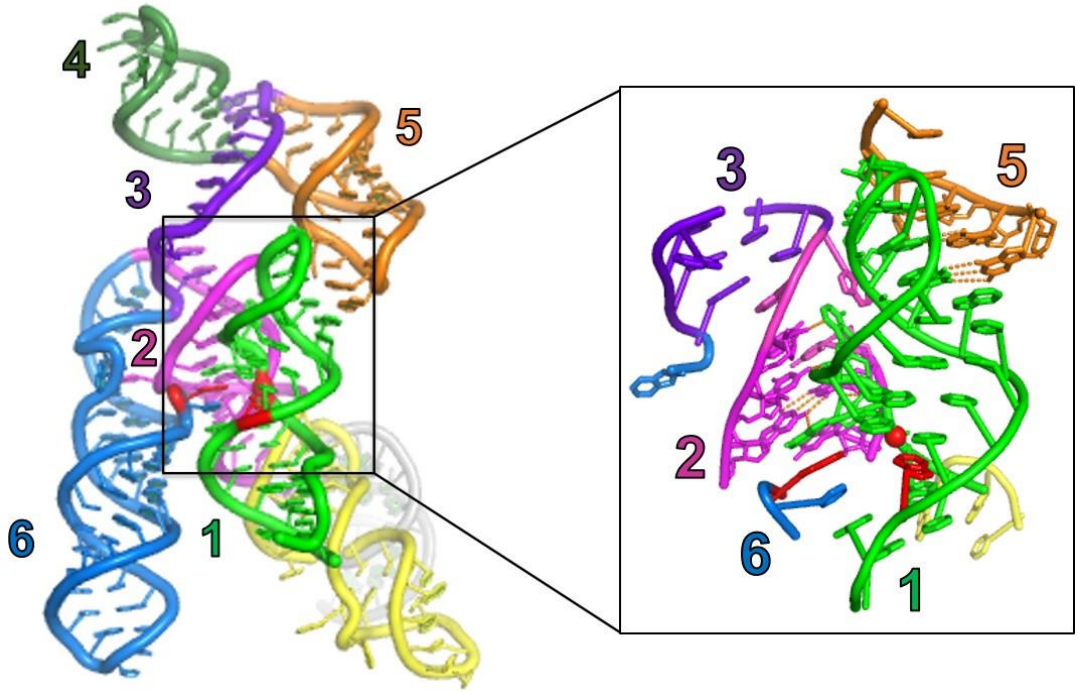

Figure 3.9 Substrate binding in VS catalysis. The VS substrate (green) docks into the catalytic core of the ribozyme via tertiary interactions between the substrate and nucleotides from helix 2 (magenta), helix 5 (orange), and helix 6 (blue). Catalytic nucleobases are shown in red. Cleavage site is shown in red and in the zoomed structure as a red sphere. H-bond interactions between substrate and catalytic domain are shown as orange dashed lines.

3.2.4.1 Interactions with stem-loop 5: binding the substrate

The first set of tertiary interactions between the substrate and the catalytic domain is included in the kissing-loop interaction between the stem loops of the substrate and helix 5 that consists of three WC base-pairs and a non-canonical pair as described above. Mutations to the bases that are involved in these interactions reduce cleavage activity by one to three orders of magnitude (Beattie et al., 1995; Rastogi et al., 1996; Rastogi and Collins, 1998). Nucleotides G630, U631, G697, A698 and C699 all involved in WC base-pairing show protection from modification by KE/CMCT and DMS and DEPC (Beattie et al., 1995; Anderson and Collins, 2001). Hydrazine-induced removal of the pyrimidine ring of C632 from the substrate does not affect cleavage, which

is consistent with the dynamic nature of the G627-C632 pair (Beattie et al., 1995). This pair has to be broken in order for the C632 to pair with G697 (from stem-loop 5) as the first WC pair of mini-helix created by the kissing-loop interaction (Figures 3.9, 3.10A, 3.16A for electron density omit map). The disruption of this terminal base-pair of the substrate hairpin increases the flexibility of the loop-stem junction thereby accommodating the 'shifted' state of the substrate-helix. Removal of the cytosine base of nucleotide C632, likely introduces similar flexibility that ultimately favors the active fold of the ribozyme. The existence of a set of three WC pairs: C632-G697, U631-A698 and G630-C699 were inferred from biochemical data; however, the C629-A701 non-canonical pair was not identified, although its presence is consistent with functional data (Figure 3.10) (Rastogi et al., 1996; Beattie and Collins, 1997; Jones and Collins, 2003; Sood et al., 2002).

The kissing-loop structure from crystallography is consistent with the NMR structure of kissing-loop complex comprising isolated stem loops 1 and 5 (Figure 3.11A). The sequence corresponding to stem-loop 1 in the NMR study lacked a cleavage loop, contained an artificially enlarged hairpin by disrupting a potential G627-C632 WC pair with a G627A mutation and introducing a C634G mutation in helix 1b (Bouchard and Legault, 2014). While both structures captured the 'shifted' substrate, only the crystal structure captured the register shift in its native context. The C634G mutation in the substrate stem-loop used in the NMR study is known to constitutively shift the substrate-helix (Andersen and Collins, 2001; Zamel and Collins, 2002), therefore the register shift captured in the NMR structure of the substrate-stem-loop 5 'kissing' complex is artificially induced unlike the substrate sequence in the crystallization construct (VSx_C634) that does not contain any pre-organizing mutation. The nucleobase at position 634 (G634) is flipped out of the helix in the NMR structure, in contrast to the intrahelical position of

C634 in the crystal structure. This inconsistency most likely stems from the difference in sequences corresponding to the substrate-helix used in NMR and crystallographic studies.

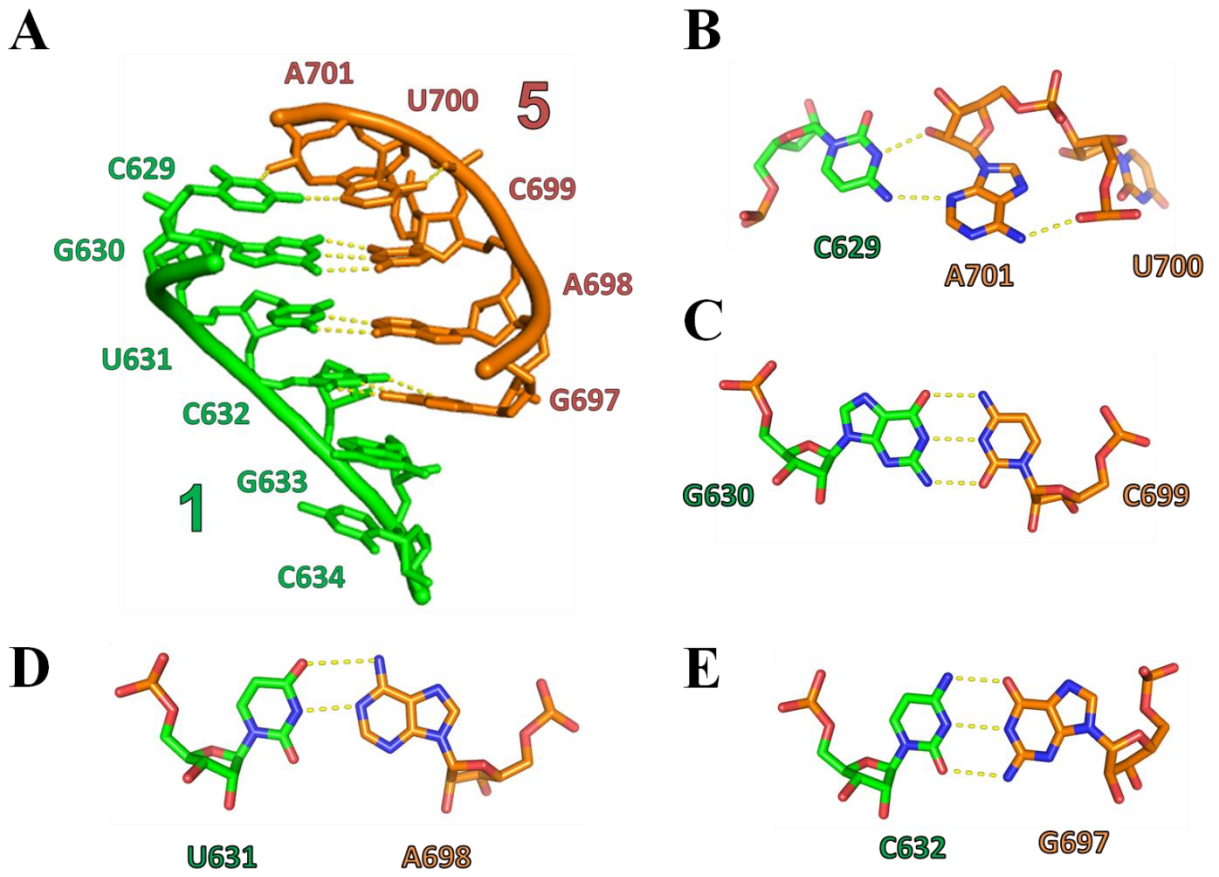

Figure 3.10 A-form helix formed in the kissing-loop interaction between substrate and stem-loop 5. A. Hairpin nucleotides of the substrate (green) form three WC base-pairs and one non-canonical base-pair with hairpin nucleotides of stem-loop 5 (orange) to constitute an A-form helix. Substrate nucleotides G633 and C634 stack underneath G697 extending the helix. **B.** C629-A701 non-canonical base-pair. The exocyclic amino group of A701 is in hydrogen bonding distance to the pro-*Rp* oxygen atom of the flipped out U700. **C.** G630-C699 WC base-pair. **D.** U631-A698 WC base-pair. **E.** C632-G697 WC base-pair.

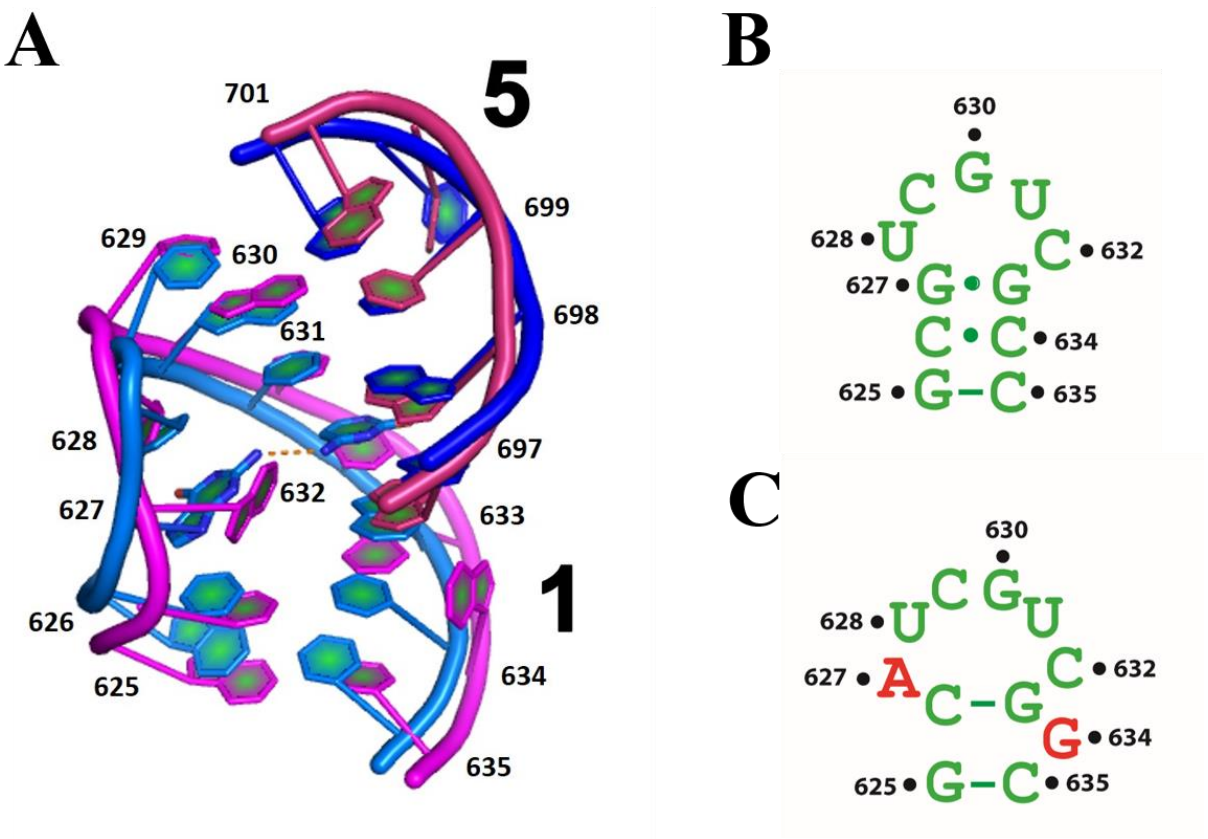

Figure 3.11 Comparison of the kissing-loop complexes as revealed by NMR and crystal structures. A. Overlay of NMR (PDB ID code: 2MI0, Bouchard and Legault, 2014) and crystal (PDB ID code: 5V3I, this work) structures of the kissing-loop complexes. NMR: substrate stem-loop 1 (light pink), catalytic domain stem-loop 5 (deep pink). Crystal: substrate stem-loop 1 (light blue), catalytic domain stem-loop 5 (deep blue). **B.** Secondary structure of the hairpin in the crystallization construct. This sequence is identical to the wild-type sequence. **C.** Secondary structure of the hairpin construct used for NMR in A. Two mutations in this construct, G627A and C634G are indicated in red.

3.2.4.2 Interactions with helices 2 and 6: stabilizing the active site

Substrate docking and remodeling involves the extrusion of nucleobases immediately downstream of the cleavage site. Upon docking, two adenines, A621 and A622 disengage from their respective base-pairs and assume extrahelical positions (Figure 3.12, 3.16C for electron density omit map). This creates a more 'open' active site, in addition to creating a splayed cleavage site involving an intrahelical G620 and an extrahelical A621. A621 is accommodated in a binding pocket created with nucleotides from helices 2 and 6 that form H-bonding and base-stacking interactions with A621. The N3 atom of A621 is in H-bonding distance to the exocyclic amino group of G653 and the A621 ribose 2'-hydroxyl forms a H-bond with ribose 2'-hydroxyl group of C771. These interactions are further stabilized by fixing the relative orientation of G653 and C771 with respect to A621 by the formation a WC base-pair between these nucleotides (Figures 3.12, 3.13A). Mutations to these nucleotides (G653C, C771G-independent mutations) that disrupt this base-pair in addition to compromising interactions with A621, introduce cleavage defects of over a 100-fold (Beattie et al., 1995) and an A621G mutation reduces cleavage by 40-fold (Wilson et al., 2007). These observations support the importance of these tertiary interactions. The open space created between A621 and A639 is occupied by the putative general acid, A756 that projects out of an S-turn motif in helix 6 (Figure 3.1B) and stacks against A621. C755 stacks between A756 and A639 creating a four-nucleobase stack in the active site. The exocyclic amino group of C755 is in H-bonding distance with the 2'-hydroxyl group of G620, an interaction that probably orients the nucleophile for in-line attack (Figure 3.13E).

A622, the other adenine nucleobase that is extruded from the helical stack is similarly accommodated in a binding pocket created exclusively by nucleotides from helix 2 (Figure 3.13B). Interactions that stabilize the extrahelical position of A622 in this pocket include H-bonding

interactions between the N3 atom of A622 and the ribose 2'-hydroxyl group of G655, and ribose 2'-hydroxyl group of A622 and the exocyclic amino group of G654. The exocyclic amino group and the N7 atom of A622 are both in H-bonding distance with ribose 2'-hydroxyl group of C770. WC pairing between G655 and C770 positions these nucleotides to mediate interactions with A622. A657, a bulged nucleotide at J_{23} stacks against A622 providing further stabilization. An A657U mutation that substitutes a purine with a pyrimidine reduces cleavage by ~15-fold (Lafontaine et al., 2001b), most likely reflecting the reduced stacking efficiency of pyrimidine rings. A strong cleavage defect of over three orders of magnitude introduced by an A622U mutation (Wilson et al., 2007) underscores the importance of this nucleotide in organizing a catalytically-competent active site.

Structural organization of the binding pockets for accommodating A621 and A622 resemble that in protein enzymes. Enzymes like uracil-DNA glycosylase recognize the uracil nucleobases in a DNA strand, flip them out of the double helix and remove them as a part of base-excision repair. Amino acid residues donate hydrogen bonds and provide stacking against the aromatic ring to hold the extruded uracil in place (Figure 3.13C) (Savva et al., 1995). Structural contraptions to facilitate base-flipping utilized by RNA enzymes might be ancient precursors to more modern nucleotide binding pockets designed with amino acids, although the strategies for binding the extruded nucleotide remain the same.

In addition to stabilizing the A621 and A622 nucleotides, nucleotides from helix 2 at junction 2-3-6 donate H-bonds to the phosphate backbone connecting these nucleotides to the rest of the substrate. This region exhibits the largest reorganization in backbone geometry (Figures 3.6A, 3.8) first, to create a splayed conformation for endonucleolytic cleavage between G620 and A621, and second, to revert the backbone to its natural conformation corresponding to a standard

A-form helix between A622 and G623 (Figure 3.12). This enables the formation of helix 1b that plays important roles in substrate activation (section 3.3.1.3). These sharp turns in the backbone geometry are facilitated by interactions between the backbone nbOs and nucleotides, G654 and A771 from helix 2 (Figure 3.13D). The backbone conformation between A621 and A622 is stabilized by a hydrogen bond between the pro-*Rp* nbO atom and the ribose 2'-hydroxyl group of C771. The second turn in the RNA backbone between A622 and G623 is stabilized by putative hydrogen bonds between both nbO atoms and the ribose 2'-hydroxyl group of G654. These interactions enable the RNA backbone to undergo such drastic changes at the internal cleavage loop of the substrate. Functional data from mapping 2'-deoxy and phosphorothioate interference at helix 2 and the substrate-helix respectively, support the ribose-backbone interactions between helix 2 and the substrate-helix (Sood et al., 2002). 2'-deoxy interference at nucleotides G654, G655, C770 and C771 highlight the importance of their ribose moieties. Corresponding replacements of nbO atoms at G620pA621, A621pA622, and A622pG623 with sulfur, result in varying degrees of cleavage defects. A G620psA621 mutation largely abolishes cleavage, an A621psA622 mutation lowers activity by 5-fold and an A622psG623 mutation retains wild-type activity (Kovacheva et al., 2004). However, all phosphorothioate mutations lower the extent of cleavage by ~3-5-fold (Kovacheva et al., 2004). This observation might reflect the reduced fraction of substrate-helices that are in their catalytically active conformation in the absence of backbone stabilization.

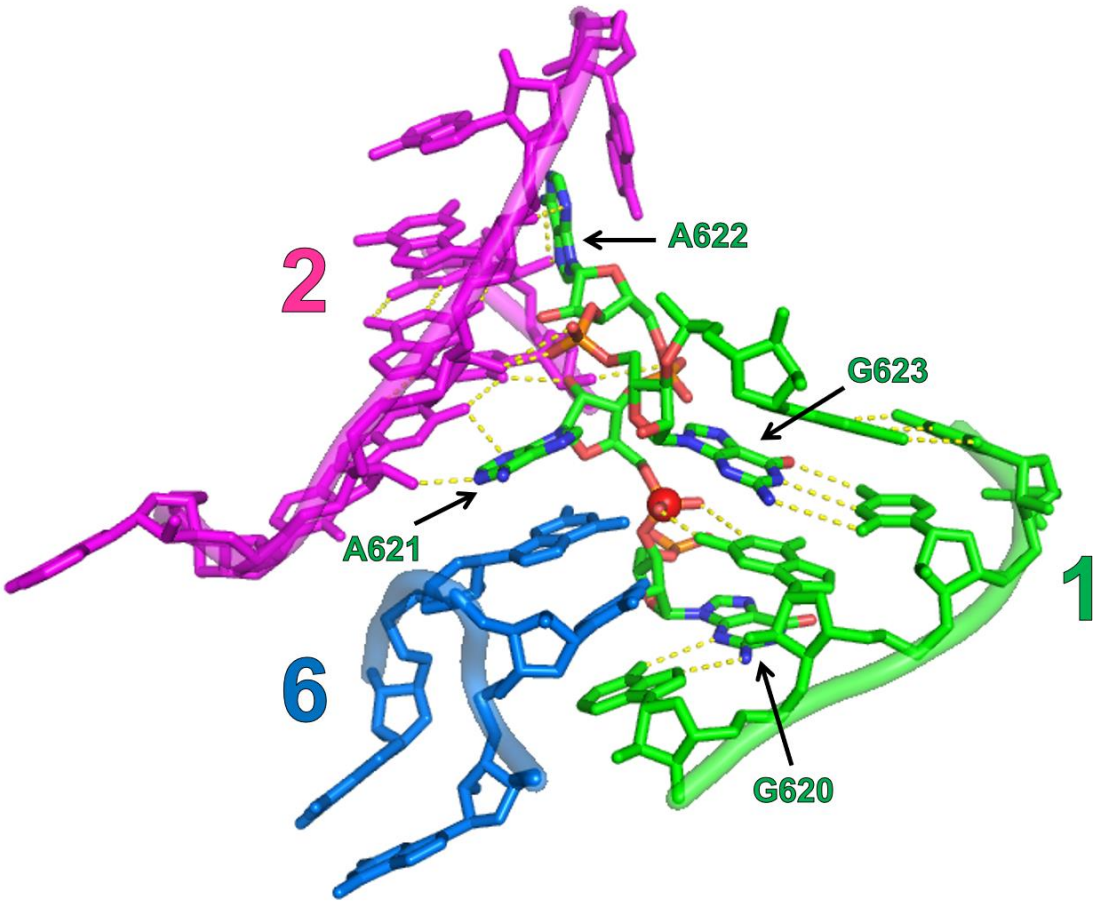

Figure 3.12 Tertiary interactions between substrate and helices 2 and 6: global perspective. Nucleotides A621 and A622 extruded from the helix interact with nucleotides from helices 2 and 6. The splayed conformation between G620 and A621 that defines the ribozyme cleavage site is stabilized by hydrogen bonds donated by G638 to the non-bridging oxygen (nbO) atoms of the scissile phosphate. The backbone-turn between A621 and A622 is stabilized by a hydrogen bond between the pro-*Rp* nbO atom and the ribose 2'-hydroxyl group of C771. The second turn in the RNA backbone between A622 and G623 is stabilized by hydrogen bonds between both nbO atoms and the ribose 2'-hydroxyl group of G654.

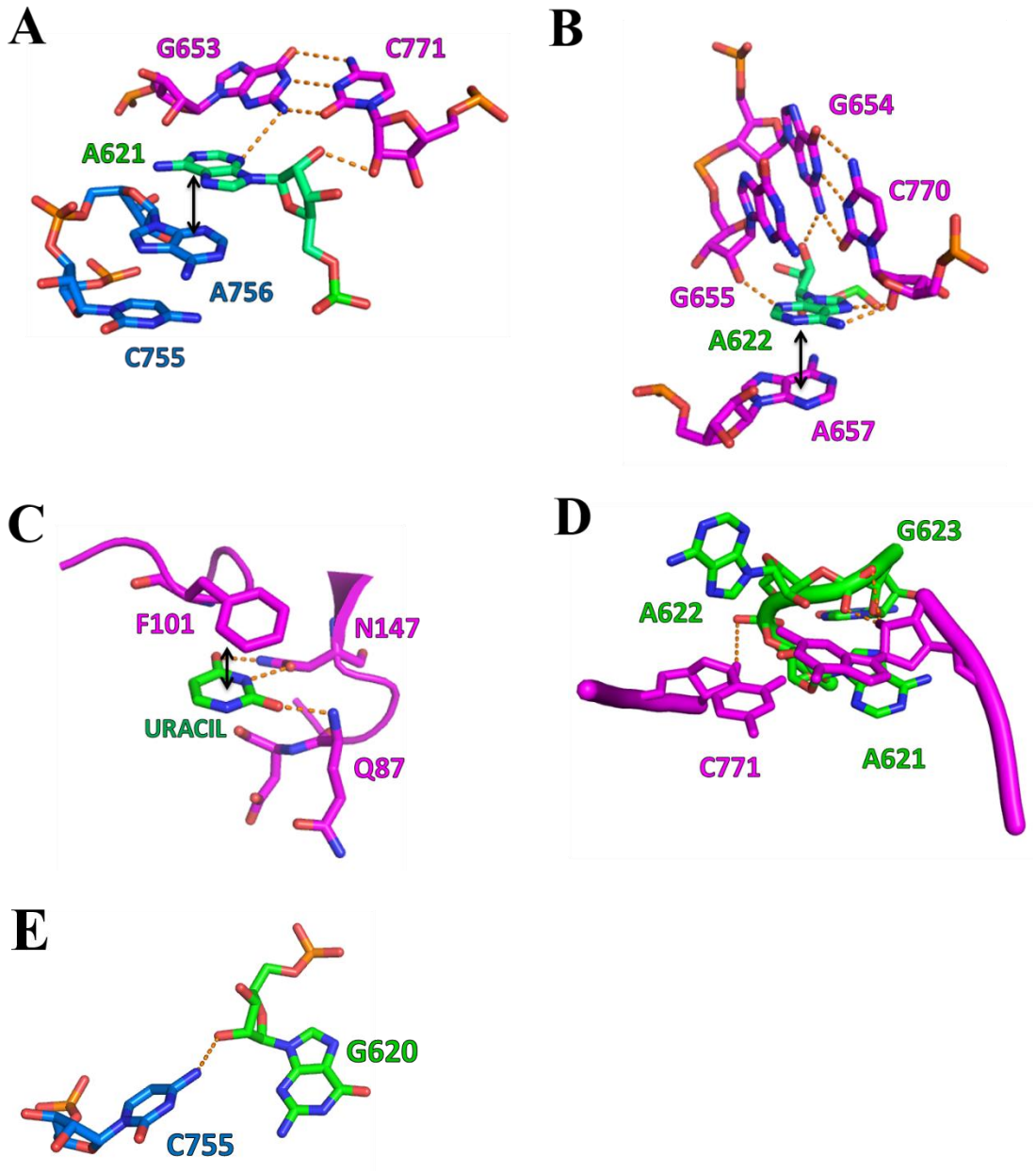

Figure 3.13 Tertiary interactions between substrate and helices 2 and 6: individual interactions. A. Interactions of A621 of the substrate with nucleotides from helices 2 and 6 of the catalytic domain consist of H-bonds (shown by orange dashed lines) and a base stack (shown by a double-headed arrow). **B.** Interactions of A622 of the substrate with nucleotides from helix 2 of the catalytic domain consist of H-bonds and a base stack. **C.** Binding pocket for uracil in protein enzyme uracil DNA glycosylase utilizes similar strategies of H-bonding and base-stacking to stabilize the flipped uracil nucleobase (PDB ID code: IUDH, Savva et al., 1995). **D.** Interactions between helix 2 and the backbone corresponding to nucleotides G620 to G623. Non-bridging oxygen atoms form hydrogen bonds with ribose moieties of G654 and C771 to stabilize the pair of sharp turns in the backbone between nucleotides A621 and A622, and A622 and G633. **E.** H-bond between the exocyclic amino group of C755 and 2'-hydroxyl group of G620.

3.2.5 Structure of the active site and catalysis

A network of tertiary interactions between the substrate and the catalytic domain of the ribozyme activates the former for catalysis. The active fold of the VS ribozyme brings the internal loops of helices 1 and 6 (G638 loop and A730 loop, respectively) in close proximity, positioning the catalytic nucleobases from these two loops near the cleavage site. Association of these loops and the minor groove of helix 2b buries 1,247 $Å^2$ of solvent-accessible surface area, creating the catalytic core (Figures 2.16, 3.14). Tertiary interactions facilitating this association are detailed in section 3.2.4. At the active site, two separate stacks of nucleobases create a V-shaped arrangement that organizes the catalytic residues around the scissile phosphate and splays the cleavage site nucleotides G620 and A621 (Figures 3.14). One edge of this 'V' contains the proposed general base, G638 (A638 in VSx_G638A and VSx_C634) and consists exclusively of nucleotides from helix 1. Stacking on the 5′ side of G638 is intrastrand, whereas on the 3′ side, G638 forms a cross-strand stack with G620, which harbors the 2′-O nucleophile. The other edge of the V contains the proposed general acid, A756 (G756 in VSx_A756G), and consists of nucleotides from both loops, including the nucleotide immediately downstream of the scissile phosphate, A621. The direct stacking between A756 and A621 is consistent with uv-crosslinking observed between nucleotides 756 and 621 when A621 was replaced by a 4-thio uridine (Hiley et al., 2002) Hydrogen-bonding interactions with the cleavage-site nucleotides further facilitate the splayed conformation and include the sheared base-pair between G620 and A639 near the apex of the 'V' and an A-minor interaction between A621 and helix 2b. (Figures 3.14, 3.4C, 3.13A). Catalytic nucleobases G638 and A756 are both unpaired and extend toward the scissile phosphate that lies between them.

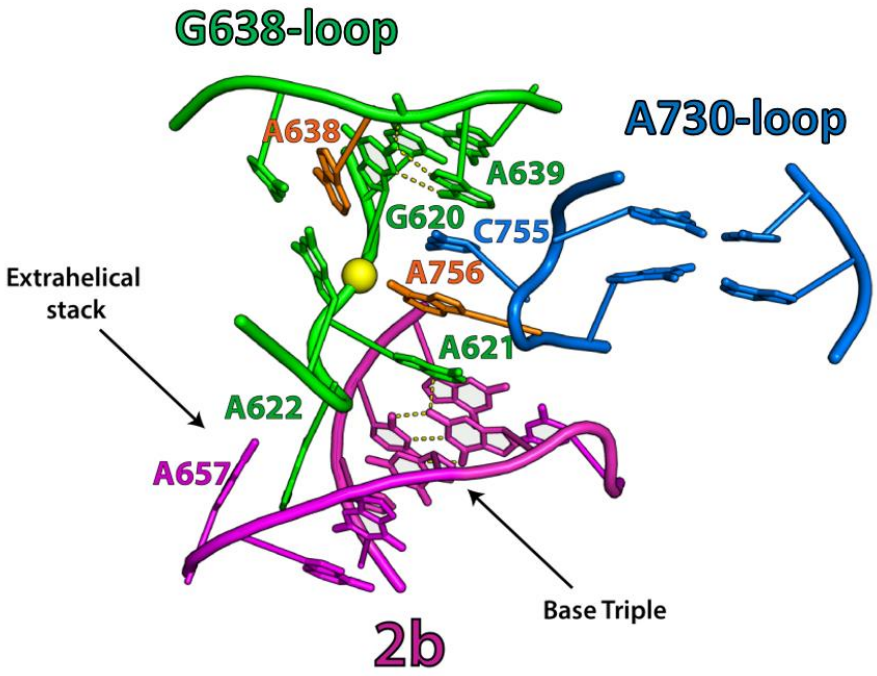

Figure 3.14 Local environment of the VS ribozyme active site. Global architecture of the VS ribozyme active site reveals a 'V' configuration of nucleotides that splays the cleavage site. The proposed general acid, A756 and general base, G638 (A638 in this structure) in the cleavage reaction are shown in orange. The scissile phosphate is shown as a yellow sphere. One edge of the 'V' is made up of C637, A638 (G638 in WT), G620 and C619 and the other edge consists of nucleotides G640, G639, C755 and A756.

The splayed conformation, a feature observed in the active sites of the other nucleolytic ribozymes (Cochrane et al., 2008; Ren et al., 2017), allows acquisition of the in-line geometry required for the reaction: ideally a 180° angle (τ) defined by the nucleophile, the phosphorus reaction center and the leaving group (this angle is on average 70° for a standard A-form RNA helix). Our structures are consistent with a wealth of biochemical data implicating nucleobase catalysis by G638 and A756 in the cleavage mechanism of the VS ribozyme (section 2.1.2.2) (Wilson et al., 2007; Sood and Collins, 2002; Lafontaine et al., 2002b; Wilson and Lilley, 2011, Wilson et al., 2010), (Figure 3.15). Nevertheless, in our precatalytic structures containing a G638A mutation (VSx_G638A, VSx_C634), the reacting groups deviate substantially from in-line

geometry (τ = ~100°), and the catalytic nucleobases reside slightly beyond hydrogen-bonding distance of the scissile phosphate (Figure 3.15A, B). Clearly adjustments must occur to bring the ribozyme active site into a fully activated state. Inspection of the A756G structure suggests that such adjustments could be restricted to local atomic movements. The A756G ribozyme, which adopts the same overall fold and active-site architecture as the G638A ribozymes (Figure. 2.12C), configures the reacting groups closer to the in-line geometry (τ = 129°). In addition, the imino group of G638 resides within hydrogen-bonding distance of the scissile phosphate (Figure 3.15C). This interaction, first revealed by the crystal structure is consistent with recent functional data (Weissman et al., unpublished results). These results suggest that G638 may stabilize the in-line conformation and facilitate catalysis directly via hydrogen bonding or proton transfer, analogous to that proposed for the hairpin ribozyme (Kath-Schorr et al., 2012; Pinard et al., 2001; Heldenbrand et al., 2014). Nevertheless, the molecular strategy by which the active site G mediates catalysis in VS and other ribozymes requires further functional analysis. Regarding residue A756, in both structures N1 of the purine nucleobase sits close to the 5′-oxygen leaving group, poised to mediate proton transfer (Figure 3.15A-C, 3.16D, E for electron density omit map).

Our structures provide a powerful framework for evaluating mechanistic proposals. Metal rescue (Cd^{2+}) of cleavage defects introduced by phosphorothioate substitutions in the scissile phosphate, suggest a possible role of inner-sphere Mg^{2+} coordination to the pro-*Sp* nbO, which might reveal an additional player in VS catalysis (Weissman et al., unpublished results). RNA-bound metal ions can potentially modulate pKas of the catalytic nucleobases, making them better general acid/base catalysts. In addition to possible roles of metal ions in pKa shifting, a plausible interaction between the exocyclic amino group of A756 and the pro-*Rp* nbO atom (Figure 3.15E) might raise the pKa of the adenine nucleobase. Similar interactions involving catalytic G and A

and non-bridging oxygen atoms of the scissile phosphate have been observed in the hairpin ribozyme (Liberman et al., 2012; Heldenbrand et al., 2014). Modifications that remove the exocyclic amino group from the catalytic guanine (G8) in the hairpin ribozyme, result in departures from the ideal 'in-line' conformation in the cleavage site (Fedor, 2000). Similarly, replacing A638 by wild-type guanine in the VS ribozyme brings the cleavage site geometry closer to the ideal 'in-line' arrangement (from 97° in VSx_G638A to 129° in VSx_A756G). In addition to the H-bond between the exocyclic amino group of G638 and the pro-*Rp* nbO, the N1 atom is also in H-bonding distance to the pro-*Sp* nbO. Similar interactions have been observed in crystal structures of the twister and twister-sister ribozymes (Liu et al., 2014; Ren et al., 2014; Liu et al., 2017; Zheng et al., 2017). These interactions stabilize the transition state of the ribozyme with possible roles in maintaining in-line geometry.

Considering the use of inactivating mutations to trap the precursor, the deviations from in-line geometry observed in VSx_G638A and VSx_A756G could reflect local misfolding of these mutant RNAs, as substrate helices bearing mutations at G638 have a severe cleavage defects, but retain in-line probing profiles and binding affinities similar to those of the corresponding helix lacking the mutation (Wilson et al., 2007). In addition, the A756G ribozyme retains full activity against a substrate bearing a 5′-sulfur leaving group (Kath-Schorr et al., 2012). These data and the strong agreement between the structural and biochemical data support the functional relevance of our structures. Modeling the wild-type catalytic nucleobases into each structure by isosteric replacement suggests that the active-site configuration of VSx_A756G forms potential hydrogen bonds to the scissile phosphate from the amino and guanidino groups of A756 and G638, respectively (Figure 3.15D, E). More speculatively, if our structures reflect progressive stages along the reaction coordinate, then formation of new hydrogen bonds might accompany acquisition

of the in-line geometry. This is supported by computation (Ganguly et al., unpublished results). Structural analysis of the hairpin ribozyme has also implicated a catalytic strategy involving increased hydrogen bonding to the scissile phosphate as the reaction progresses (Rupert et al., 2002).

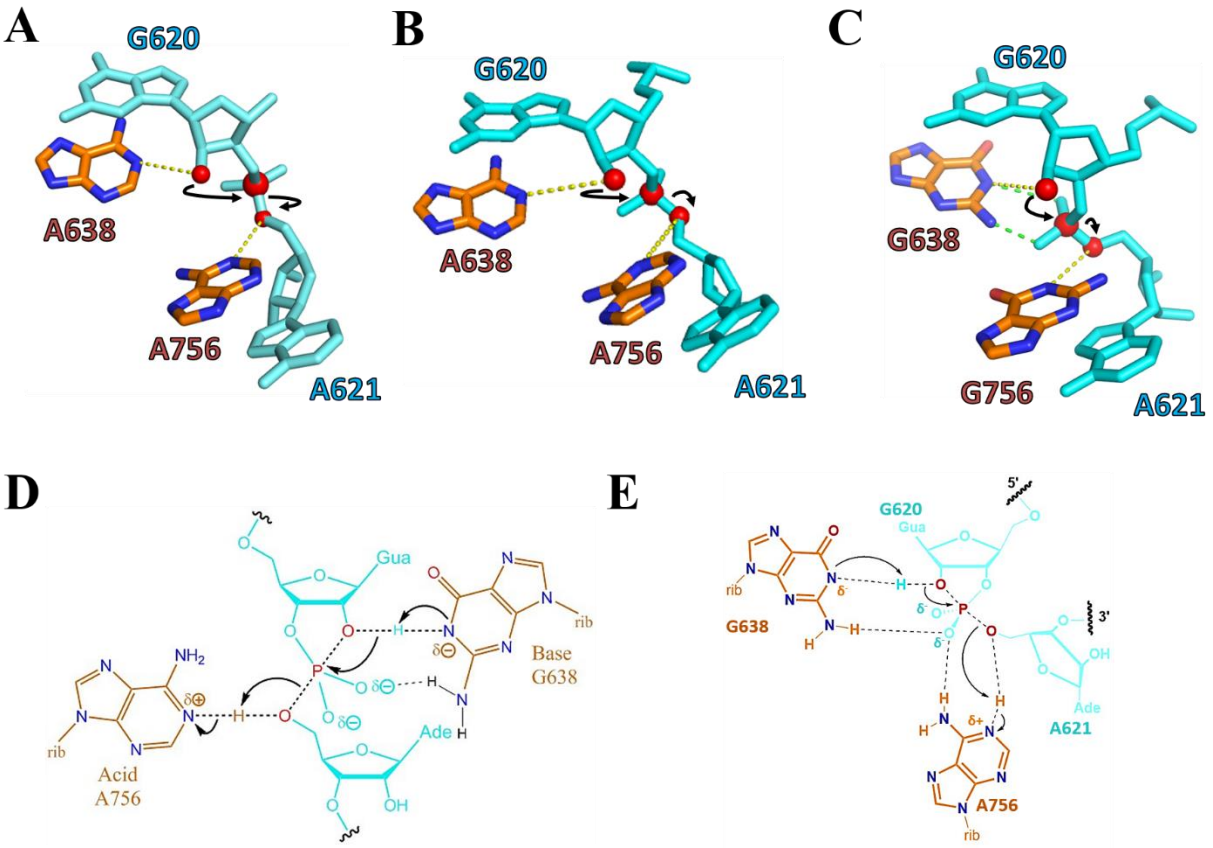

Figure 3.15 Cleavage site showing splayed nucleotides and catalytic nucleobases. A. Cleavage site in the VSx_G638A structure (PDB ID code: 4R4V). 2′-OH nucleophile, scissile phosphate and leaving group are depicted by red spheres. Arrows depict movement of electrons in the cleavage reaction. Yellow dashes indicate interactions involved in general acid/base catalysis. **B.** Cleavage site in the VSx_C634 structure (PDB ID code: 5V3I). Color scheme as in A. **C.** Cleavage site in the VSx_A756G structure (PDB ID code: 4R4V). Color scheme as in A. Green dashes indicate interactions between the G638 and non-bridging oxygen atoms of the scissile phosphate. **D.** Schematic of the biochemically inferred mechanism of general acid-base catalysis by the VS ribozyme involving nucleobases of G638 and A756. Arrows indicate bond formation/cleavage during the transition state.

Figure 3.15 Cleavage site showing splayed nucleotides and catalytic nucleobases (continued). E. Interactions in the active site obtained after modeling the wild-type adenine in place of G756 in the VSx_A756G structure. This reveals additional interactions between A756 and nbO atoms of the scissile phosphate. Dashes indicate potential H-bonds formed between catalytic nucleobases and cleavage site nucleotides.

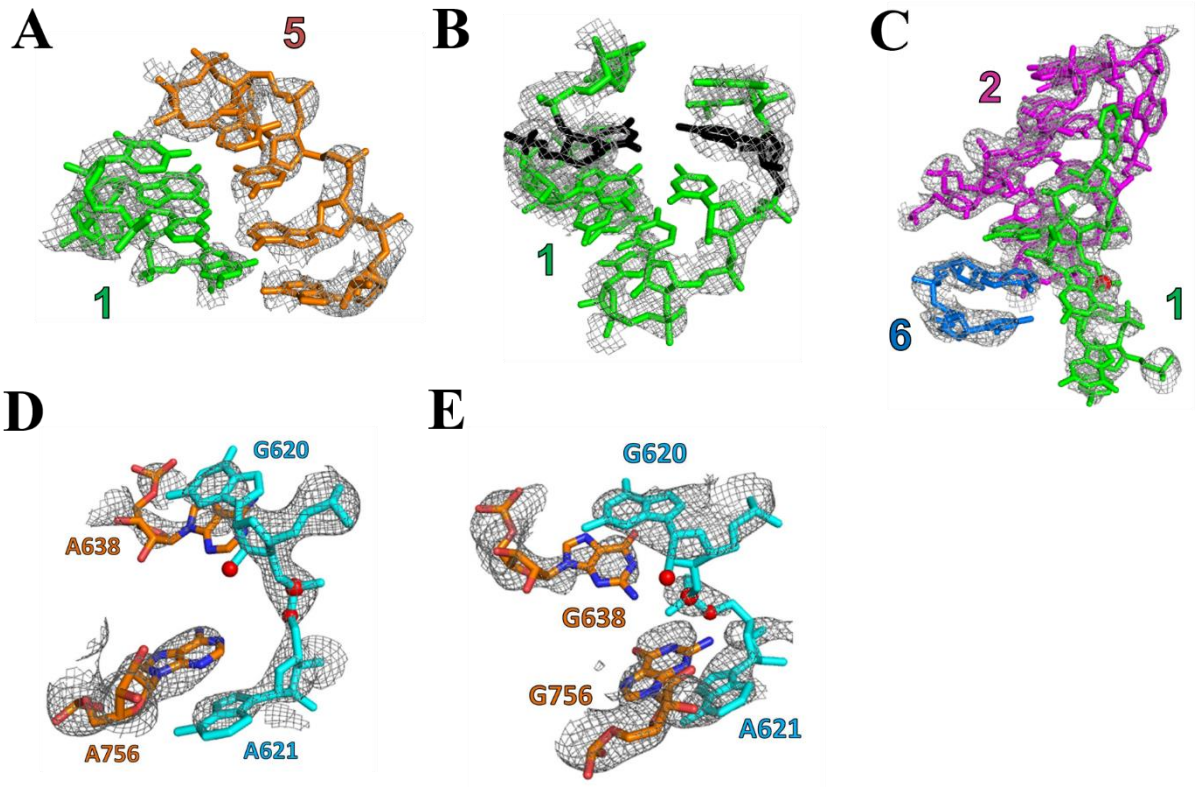

Figure 3.16 Electron density omit maps of important regions. 2|Fo|-|Fc| composite omit maps from simulated annealing, contoured at 2.0σ and carved within 1.6 Å of selected atoms (A-C) and within 2 Å of selected atoms (D-E). **A.** Nucleotides from the substrate and stem-loop 5 of the catalytic domain involved in the kissing interaction. **B.** Docked substrate helix 1b revealing the shifted register. New base-pairs formed as a result of this rearrangement: G623-C637, G624-C636, G624-C636 and C626-C634 (shown in black). **C.** Active site nucleotides in a docked substrate: interactions between the substrate and the catalytic domain. Scissile phosphate is shown as a red sphere. **D.** Cleavage site and catalytic nucleobases in VSx_G638A. Coloring scheme as in figure 3.15. **E.** Cleavage site and catalytic nucleobases in VSx_A756G. Coloring scheme as in figure 3.15.

3.3 DISCUSSION

3.3.1 Substrate rearrangement plays a central role in VS function

Functional data suggest that a shift in nucleotide register of helix 1b is a prerequisite for cleavage and shifted substrate-helices bind to the catalytic domain of the ribozyme with greater affinity. While the requirement of substrate 'shifting' for cleavage may derive partially from the latter observation: substrates need to be bound tightly to the ribozyme to be cleaved efficiently, no model was available to explain these observations. As substrate rearrangement provides a regulatory checkpoint for VS catalysis, it is essential to develop a structural framework to understand its role in the process. The problem can be broken down into the following questions:

1. How does the kissing-loop interaction at the hairpin induce a register shift in helix 1b when the two are physically separated by the C626-G633 base-pair that is expected to persist in the rearranged substrate according to biochemical data?
2. Why do substrates with shifted helix 1b bind to the ribozyme catalytic domain with greater affinities than their unshifted counterparts?
3. What is the mechanism by which the kissing-loop interaction activates a distant internal cleavage loop that is separated from the hairpin by helix 1b? How does rearrangements in helix 1b lead to activation of the internal cleavage loop?

In short, how are the kissing-loop interaction at the hairpin, substrate shifting in helix 1b and activation of the cleavage loop connected?

We have defined the structural differences between the docked and undocked states of the VS ribozyme substrate helix. We used a complete VS ribozyme crystallization construct containing a substrate helix that adopts an unshifted secondary structure in isolation and a shifted secondary structure upon docking to form the active site. This structure enabled us to map the

structural changes and tertiary interactions that lead to the creation of an active substrate from an inactive one.

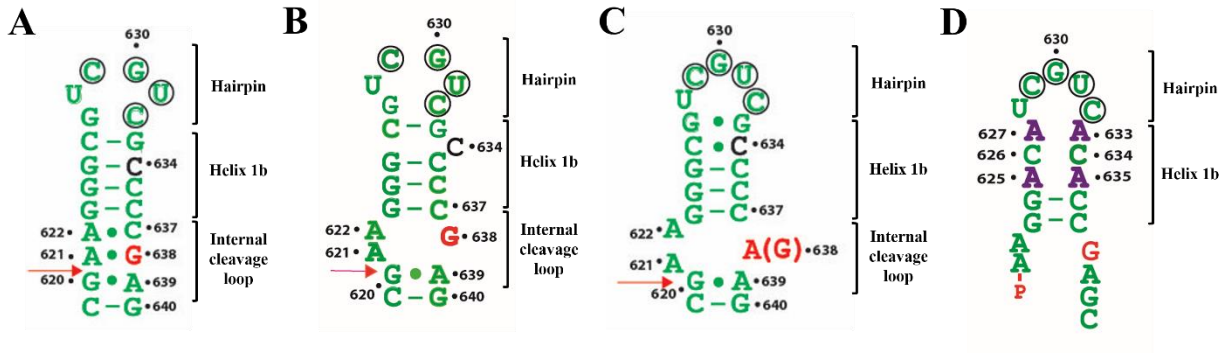

Figure 3.17 Secondary structures of the substrate-helices in their unshifted and shifted states. **A.** Secondary structure of the unshifted substrate derived from NMR (Flinders and Dieckmann, 2001). **B.** Secondary structure of the shifted substrate derived from DMS probing (Andersen and Collins, 2000; Andersen and Collins, 2001). The C626-G633 pair underneath the hairpin persists in this structure, while C634 is bulged out of the helix. **C.** Secondary structure of the shifted substrate derived from crystallography (this work). The C626-G633 pair is replaced by a C626-C634 non-canonical pair that consists of an intrahelical C634. **D.** Secondary structure of ZA10, a substrate variant with high binding affinity toward the catalytic domain (Zamel and Collins, 2002). Nucleotides that differ from wild-type are shown in purple. The secondary structure of ZA10 resembles that of the substrate obtained from crystallography in C.

3.3.1.1 Kissing-loop interaction expands the junction between substrate hairpin and helix 1b, thereby accommodating a shifted helix

The G627-C632 base-pair that closes the substrate hairpin is dynamic by virtue of its terminal position in the stem-loop, as revealed by NMR (Flinders and Dieckmann, 2001). On substrate docking, this base-pair is disrupted (or the unpaired state is favored) as C632 points away from the interior of the helix and base-pairs with G697 of stem-loop 5. This stabilizes a more expanded conformation of the hairpin, making the junction between the hairpin and the 3' strand of helix 1b more flexible. The unique arrangement of tandem guanines and cytidines in the 5' and 3' strands of helix 1b allows for an alternate secondary structure where the G623-C636, G624-C635 and G625-C634 WC base-pairs in the unshifted stem (Figure.3.17A) are replaced by a new set of WC base pairs comprised of G623-C637, G624-C636, G625-C635 (Figure 3.17C), representing the 'shifted' conformation. Therefore, the expansion of the hairpin directly caused by the kissing-loop interaction, is responsible for accommodating the alternate shifted conformation.

3.3.1.2 A shifted helix stabilizes the kissing-loop interaction resulting in stronger binding

The shifted structure of helix 1b consists of two nucleotides, G633 and C634 that do not participate in WC base-pairing, in addition to a fully paired stem. These nucleotides though not paired in the WC sense, are located inside the helix, most likely due to a favorable stacking environment and in case of C634, the presence of a single H-bond it shares with C626 (Figure 3.6 C, D). The shifted conformation of helix 1b requires an expanded hairpin, a conformation that is inhibitory to the re-formation of a G627-C632 base-pair. In addition, the register shift disengages G633 and C634 from their respective base-pairs, which allows them to mediate contiguous stacking of the kissing helix and helix 1b (Figures 3.5, 3.10A). This stabilizes the kissing-loop

interaction by extending the minihelix stack associated with it. This structural feature, in addition to other factors like Mg^{2+}-binding to the U-turn motifs in hairpins of helices 1 and 5, accounts for the additional stabilization of 7-8 kcal/mol compared to a regular RNA duplex with three WC base-pairs (Bouchard and Legault, 2014). Our structure is consistent with the observation that substrate variants that contain preshifted helix 1b bind to stem-loop 5 with a higher affinity (~1.8-3 kcal), reflecting the energetic cost for the helix shift (Bouchard and Legault, 2014). Interestingly, the stability of substrate-helix/ribozyme complexes are similar to that of substrate-helix/stem-loop 5 complexes. This suggests that the energetic contribution from the kissing-loop interaction accounts for most of binding energy provided by substrate docking (Bouchard and Legault, 2014).

Binding studies involving substrate-helix variants with mutations in their hairpins and helix 1b, and the complete catalytic domain of the ribozyme provide functional support for our structure (Zamel and Collins, 2002). Substrate variants with mutations that constitutively shift helix 1b, bind to the catalytic domain with 10-100-fold higher affinities than variants that lack this pre-organizing mutation, and with ~300-1000-fold higher affinities than substrate helices locked in a conformation that prohibits any register shift (Zamel and Collins, 2002). A G627A mutation that prevents the formation of a G627-C632 base-pair results in a conformation characterized by an expanded loop with a flexible hairpin-stem junction. Pre-shifted substrate variants do not show increased binding on expanding their hairpins by a G627A mutation since an irreversibly shifted helix already supports the formation of the kissing-loop interaction and hairpin expansion does not provide additional stabilization. Substrate variants with wild-type helix 1b (capable of assuming a shifted conformation), show increased binding when their hairpins are expanded, supporting the link between hairpin expansion and substrate rearrangement. Substrate variants with helix 1b locked in an unshiftable conformation are unresponsive to mutations that expand their hairpins

which underscores the importance of the shifted conformation in stabilizing the kissing-loop interaction. In these binding studies, the substrate variant that exhibit the strongest binding consists of a shifted stem and a hairpin loop that was expanded beyond its terminal nucleotides (627 and 632), to include unpaired nucleotides G633, C634 and C626 (Zamel and Collins, 2002). This construct bears striking resemblance to the secondary structure of the shifted substrate captured in the crystal structure (Figure 3.17 C, D).

3.3.1.3 Register shift in helix 1b organizes the active cleavage loop

In the undocked substrate, helix 1b consists of base-pairs formed by three guanines (623-625) in the 5' strand with three cytosines (C636-C634) in the 3' strand. The fourth cytosine (C637) forms a sheared base-pair with A622 in the internal cleavage loop (Figure 3.17A). The nucleotide immediately downstream of the scissile phosphate, A621, forms a sheared pair with the putative general base G638. The unique arrangement of guanines and cytosines allows the formation of an alternate secondary structure of helix 1b when accommodated by the expanded hairpin. This brings C635, previously paired with G624, in a pairing position with G625. This register shift ultimately ratchets C637 out of the internal cleavage loop into helix 1b, as part of a newly formed G-C base-pair. This frees A622 to assume an extrahelical position (section 3.2.3.4). Loss of stacking as a result of base flipping allows A621 to assume a similar extrahelical position (section 3.2.3.4). In addition to freeing G638 to mediate catalytic interactions, the extrusion of A621 splays the cleavage site nucleotides to bring the reactant atoms closer to the ideal in-line geometry (Figure 3.14). The extrusion of A621 coupled with local rearrangement of the sheared G620-A639 pair, creates a space between A639 and A621 that is large enough to accommodate C755 and A756 from helix 6 to create an extended cross-strand stack. The distance between the planes of A639

and A621 is 10.6 Å which is just large enough to accommodate two stacked nucleobases (the distance between two base-rings in a base-paired helix is ~3.4 Å; four bases would occupy 10.2 Å in vertical space). Therefore, extrusion of A621 has three vital consequences: (1) attainment of near in-line geometry, (2) freeing the putative general base, G638 (Wilson et al., 2007) and (3) accommodating the putative general acid, A756 (Kath-Schorr et al., 2012) near the cleavage site. Collectively, these structural changes create an 'open' configuration of the internal cleavage loop that splays the cleavage-site nucleotides, allowing juxtaposition of the catalytic nucleobases and the scissile phosphate (Figures 3.14, 3.15). The creation of this open configuration could occur stochastically in an isolated substrate as suggested by NMR (Flinders and Dieckmann, 2001), but the available data are not sufficient to suggest stochasticity in the register shift in helix 1b. Notwithstanding the physical nature of rearrangement, the energetic cost of reorganizing helix 1b and the active site that involves a net loss in the number of H-bonds and the extrusion of two bases out of the helical stack, is supplanted by favorable interactions in the kissing-loop motif and binding pockets for the flipped bases. Docking of the substrate into the catalytic domain accompanies the formation of six additional stacking interactions (9 as compared to 3), two additional WC base- pairs (6 as compared to 4), and five additional non-canonical interactions (7 as compared to 2).

Although base complementarity between substrate hairpin and stem-loop 5 does not extend beyond the hairpin nucleotides, local interactions within the kissing-loop complex are sufficient to induce structural rearrangements in helix 1b as well as activate the cleavage loop for catalysis. Interaction between isolated substrate stem-loop and stem-loop 5 can trigger shifting in the substrate reflecting the modular nature of VS ribozyme folding. Structural modularity has been discussed in the context of the global fold of the RNA, which is organized by independently folding

three-way junctions that assemble each domain into a catalytically-competent structure. This modularity that exists at the global level and at the level of tertiary interactions between domains, extends to intramolecular restructuring within the substrate-helix. Stabilization energy from tertiary interactions at the substrate hairpin is transmitted to activate the cleavage loop via reorganization of the base-paired helix 1b that connects the two segments of the substrate. Identification of the correct secondary structure from crystallography enabled us to develop a structural framework to study the various aspects of substrate activation. It reveals how the shifted register of helix 1b physically couples the kissing loop interaction and the activation of the internal cleavage loop (Figure 3.18). Additionally, comparison between the undocked and docked substrates explains the specificity of endonucleolytic cleavage at the phosphodiester linkage between G620 and A621 in the VS substrate, an observation that has come to define VS catalysis.

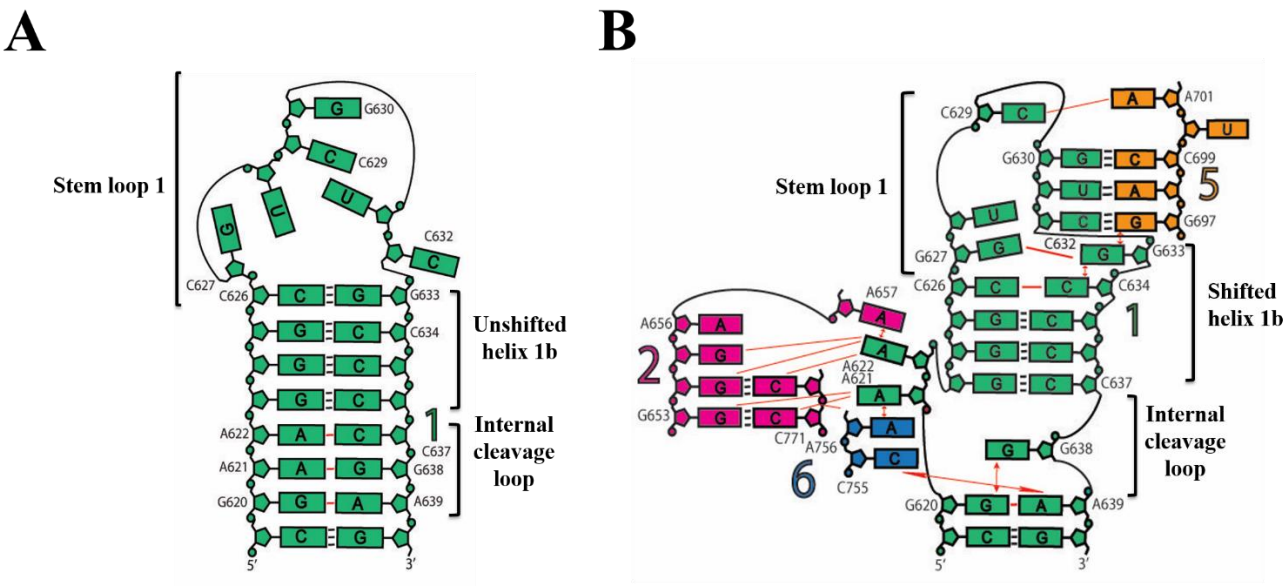

Figure 3.18 Changes in the secondary structure of the VS substrate on binding. A. Isolated substrate (as obtained from NMR structure PDB ID code: 1HWQ). The hairpin nucleotides are disorganized, helix 1b is unshifted, and the internal cleavage loop is 'closed' and consists of three noncanonical base-pairs that keep cleavage site nucleotides, G620 and A621, inside the helix. **B.** Substrate docked to the catalytic domain of the ribozyme. The docked substrate contains a reorganized stem loop that interacts with the stem loop nucleotides of helix 5, a shifted helix 1b with a new set of WC base-pairs, and a more flexible internal cleavage loop. WC H-bonds are shown by black lines, noncanonical interactions are shown by red lines, and red double-headed arrows indicate additional stacking interactions. On docking the substrate internal cleavage loop becomes more flexible as A621 and A622 get extruded and the distance between the planes of A621 and A639 increases to 10.6 Å.

3.3.2 Rearrangements in secondary structure induced by the formation of tertiary structure in other catalytic RNA

Similarities with the hairpin ribozyme. Remodeling of the internal cleavage loop observed in our structure bears similarity to that observed in the hairpin ribozyme. Both VS and hairpin ribozymes feature a sheared pair between the nucleotide upstream of the scissile phosphate (N−1) and the nucleotide directly downstream of the general base (A639 in VS, A9 in hairpin) (Figures 3.4C, 3.19A) (Cai and Tinoco Jr., 1996). In the unbound substrate of both ribozymes, the nucleotide immediately downstream of the scissile phosphate (N+1) forms a noncanonical pair (N+1A-G638 in VS, N+1G-A9 in hairpin) (Figures 3.4A, 3.19B). This pair gets disrupted upon docking, forcing the N+1 nucleotide to flip out of the helix thereby creating the splayed conformation necessary for nucleolytic cleavage (Figures 3.4C, 3.19) (Cai and Tinoco Jr., 1996; Rupert and Ferré-D'Amaré, 2001; Rupert et al., 2002). Both ribozymes stabilize the extruded N+1 nucleotide through stacking interaction with the general acid (A756 in VS, A38 in hairpin) (Rupert and Ferré-D'Amaré, 2001; Rupert et al., 2002). Thus, analogous reorganization strategies configure the active sites of the hairpin and VS ribozymes for catalysis (Figures 3.4C, 3.19). Molecular dynamics studies of the hairpin ribozyme support the inferred structural reorganization at the active site (Ocheing et al., 2016; Heldenbrand et al., 2014). Similar cleavage site reorganization accompanied by structuring of the substrate helix upon Mg binding in the twister ribozyme has been demonstrated using 2-aminopurine fluorescence (Ren et al., 2014). Internal loop rearrangements likely occur frequently in RNA folding and offer a mechanism to enhance folding fidelity (Xue et al., 2016).

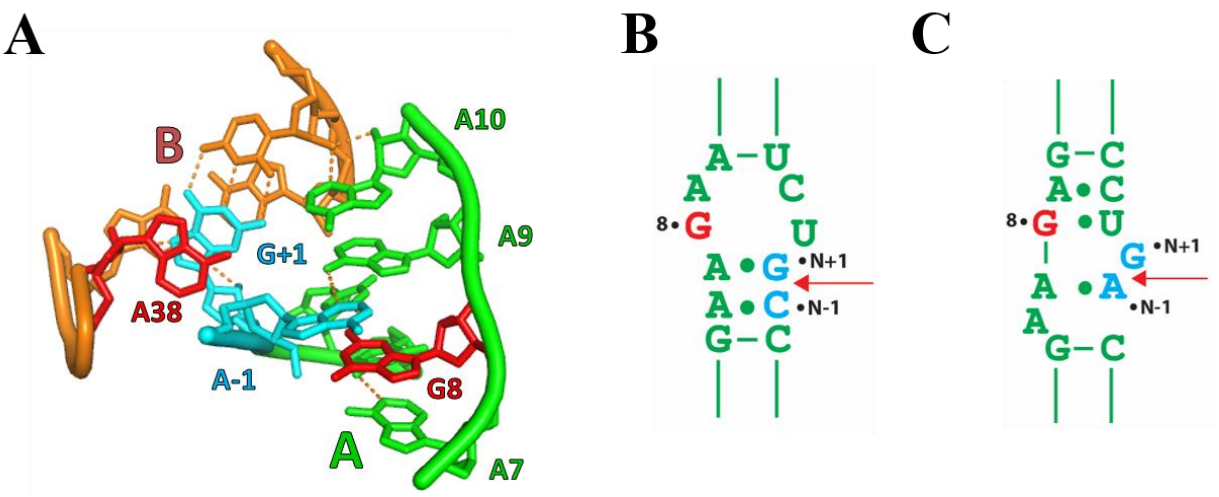

Figure 3.19 Active site of the hairpin ribozyme. A. Crystal structure of the pre-catalytic active site of the hairpin ribozyme (PDB ID code: 1M5O, Rupert and Ferré-D'Amaré, 2001) from docked substrate. Cleavage site nucleotides are shown in cyan, catalytic nucleotides in red, and loop A and loop B nucleotides in green and orange, respectively. **B.** Secondary structure of the internal cleavage loop of an isolated substrate as revealed by NMR (Cai and Tinoco, 1996). Coloring scheme as in A, cleavage site indicated by a red arrow. The loop consists of two non-canonical base-pairs, one of them involves the N+1 guanine. N-1 adenine is mutated to a cytidine in this construct. **C.** Secondary structure of the internal cleavage loop of a docked substrate in the context of the full-length ribozyme, as revealed by the crystal structure (Rupert and Ferré-D'Amaré, 2001). Coloring and labeling scheme as in B. The loop consists of three non-canonical base-pairs, with an unpaired and flipped out N+1 guanine.

Similarities with the group I intron ribozyme. Remodeling through a register shift in RNA secondary structure also occurs during the folding of the independently-folding P5abc subdomain derived from the Tetrahymena group I intron ribozyme (Wu and Tinoco Jr, 1998; Silverman et al., 1999). This subdomain adopts a non-native secondary structure in the absence of Mg^{2+}. Upon addition of Mg^{2+}, the bulge in P5a and the three-way junction between P5a and P5c undergo extensive secondary structure rearrangement involving the formation of two tandem GA pairs from P5b and a single residue shift in the base pairing register of P5c as a result of a reorganization of tandem guanine−pyrimidine base-pair reminiscent of the register shift in the VS substrate helix 1b (Figure 3.20A) (Wu and Tinoco Jr, 1998; Silverman et al., 1999). A recent study combining ^{15}N relaxation dispersion NMR with chemical probing revealed a transient, "excited-state" intermediate in the folding pathway for the P5abc subdomain (Xue et al., 2016). Specifically, P5c rapidly samples a native-like, register shifted state at a low level (∼3%) that facilitates acquisition of the native state.

Similarities with the group II intron ribozyme. Secondary structure reorganization activates the branch-point in domain 6 of the group II intron ribozyme (Costa et al., 2016). This domain toggles between a 1-nucleotide bulge (adenosine branch-point) and a 2-nucleotide bulge that includes the branch-point, utilizing a reshuffling of tandem guanine−pyrimidine base-pairs (Figure 3.20B). The conformation harboring a 1-nucleotide bulge is specific for branching, whereas the conformation with a 2-nucleotide bulge is specific to exon ligation. This restructuring of the branch-point therefore regulates the first and second steps of splicing. However, the role of tertiary interactions in this structural reorganization remains unclear.

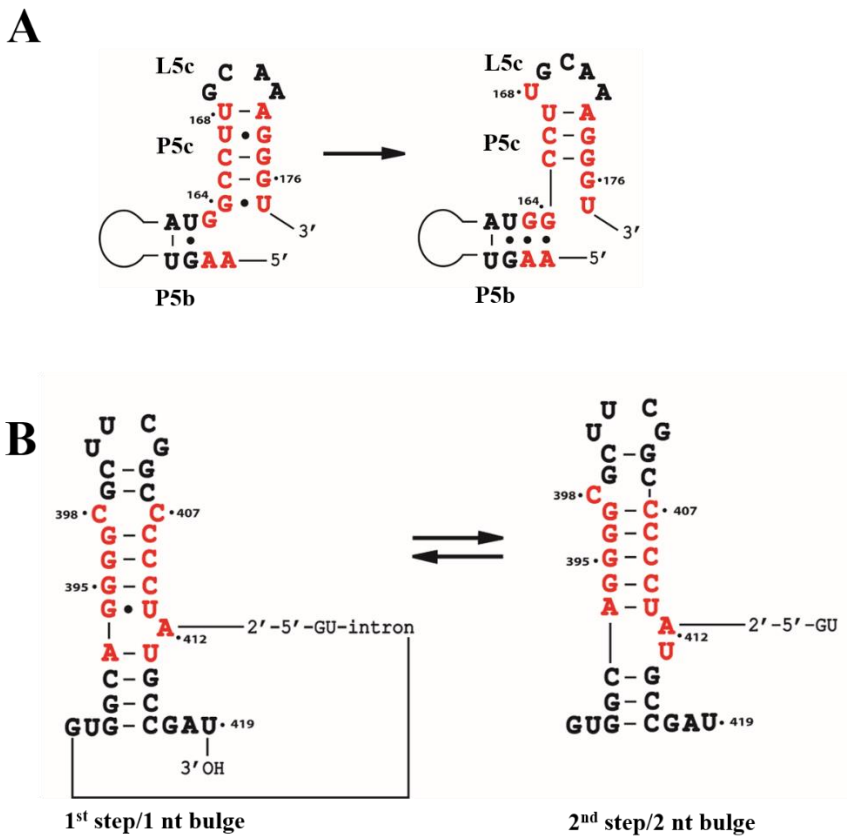

Figure 3.20 Secondary structure rearrangements in the group I intron and group II intron ribozymes. A. Local remodeling during Mg^{2+}-induced folding of the group I intron ribozyme includes changes in secondary structure of P5b and P5c (Figure adapted from Wu and Tinoco Jr., 1998). **B.** Rearrangements in the secondary structure of domain 6 of the group II intron ribozyme during the second step of splicing (Figure adapted from Costa et al., 2016).

3.3.3 Strategies for RNA cleavage are conserved across different biopolymers

The primary catalytic strategies for facilitating an S_N2-type attack leading to RNA cleavage are: attainment of an in-line geometry, stabilization of the transition state, activation of the nucleophile, stabilization of the leaving group (chapter 1). The VS ribozyme, a representative of RNA-based ribonucleases, uses two nucleobases positioned close to the scissile phosphate to mediate general acid/base catalysis. A guanine can potentially abstract a proton from the 2'-hydroxyl group unleashing a much stronger nucleophile, while an adenine can donate a proton to the 5'-oxygen with a building negative charge as a result of the cleavage reaction. In addition to mediating proton transfer between the nucleophile and leaving groups, these nucleobases provide stacking interactions to the cleavage site nucleotides and donate H-bonds to non-bridging oxygens of the scissile phosphate potentially stabilizing the negatively-charged transition state. These interactions likely stabilize the near in-line conformation at the cleavage site (Figure 3.21A).

Ribonuclease A, a protein enzyme that represents the prototypic ribonuclease, has been studied extensively for more than half a century. RNase A uses imidazole moieties of two histidine residues to mediate general acid/base catalysis (Raines, 1998; Vitagliano et al., 2002) and a lysine side chain for stabilization of the negative charge on one of the nbOs (Figure 3.21C). Recent crystal structures and biochemical investigations of an *in vitro* selected RNA-cleaving DNAzyme (DNAzyme 8-17) revealed similar catalytic strategies (Liu et al., 2017; Cepeda-Plaza et al., 2018). A guanine is positioned close to the scissile phosphate, with its N1 atom in H-bonding distance with the 2'-hydroxyl nucleophile and its exocyclic amino group in H-bonding distance with the scissile phosphate (Figure 3.21B). The catalytic guanine stacks against the nucleotide containing the ribose 2'-hydroxyl nucleophile, presumably providing further stabilization to the near in-line

geometry. This arrangement is reminiscent of the active site of the VS ribozyme and is similar to the catalytic architecture of most endonucleolytic ribozymes.

Active sites of natural ribonucleases (ribozymes and protein enzymes) and synthetic ribonucleases (DNAzymes) reveal interesting similarities in catalytic strategies. The use of a guanine residue for catalysis in both DNA and RNA ribonucleases possibly underscores the intrinsic lack of variability in the catalytic repertoire of nucleic acids and consequently their limitations as enzymes. Overall, the principles of biocatalysis are conserved across different biopolymers, highlighting the simplicity of the chemical processes that make life possible, often hidden under layers of sophistication added by billions of years of evolution.

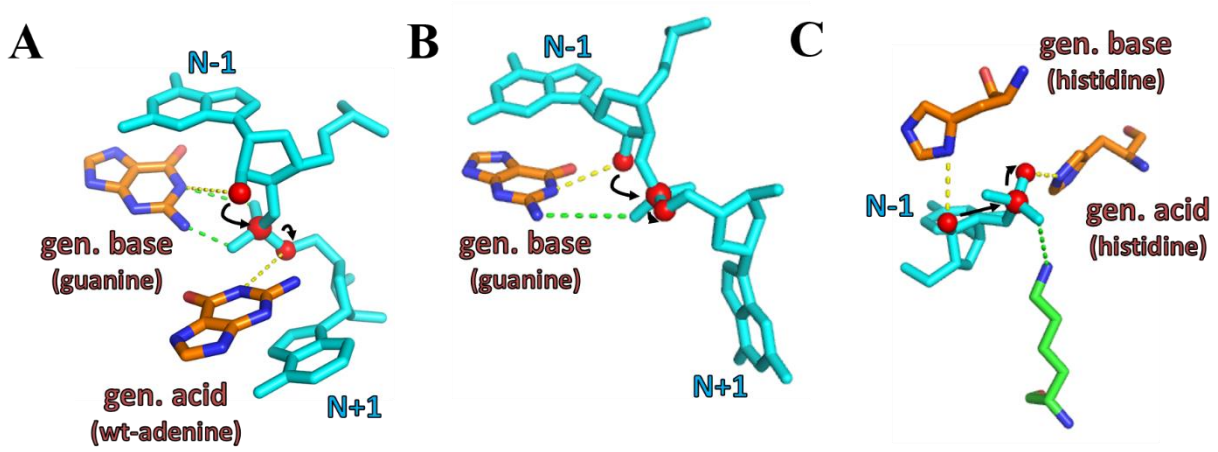

Figure 3.21 Catalytic architecture of the active sites of ribonucleases. A. Active site of the VS ribozyme consists of a splayed cleavage site with catalytic residues positioned close to the scissile phosphate (PDB ID code: 4R4P, this work). **B.** Active site of the 8-17 RNA-cleaving DNAzyme, with a splayed cleavage site and a catalytic guanine suitably positioned to participate in proton transfer with the cleavage site (PDB ID code: 5XMA, Liu et al., 2017). **C.** Active site of RNase A, bound to cytidine 2'-monophosphate. Two histidine residues are positioned for proton transfer, and a lysine forms a H-bond with the γ-phosphate (PDB ID code: IJVU, Vitagliano et al., 2002).

3.3.4 Similarities in active site architecture across different classes of endonucleolytic ribozymes

Based on biochemical data noted above and now structural data, the VS and hairpin ribozymes both appear to facilitate self-cleavage via direct participation of adenine and guanine nucleobases, serving analogous roles in catalysis (Figures 3.19, 3.22A). It is possible that the mechanistic similarity simply reflects the limited collection of side chains that RNA has available for catalysis. However, the similarities between the VS and hairpin ribozymes extend beyond identity of catalytic nucleobases and remodeling of the active site (section 3.3.2) and into architectural features of the active sites. First, in addition to the splayed conformation of the nucleotides flanking the scissile phosphate, in both ribozymes the catalytic G stacks below the N-1 nucleotide and the catalytic A stacks above the N+1 nucleotide (Figures 3.22B, 3.23A, B, D). Second, in both ribozymes an S-turn motif (Rupert and Ferré-D'Amaré, 2001; Desjardins et al., 2011) directs the catalytic A toward the leaving group and positions an upstream nucleotide coaxially to complete a stacking sandwich (Figures 3.22, 3.23A, B, D). Third, the two ribozymes use analogous motifs, referred to here as the L-platform, to position the catalytic guanine nucleobase toward the nucleophilic 2′-OH and complete a stacking sandwich (Figures 3.22, 3.23A, B, D) with the base of the L formed by a noncanonical pair between the N−1 nucleotide and the nucleotide immediately downstream of the general base. The conservation of these features in both ribozymes highlights their functional importance. The similar active-site motifs employed by the VS and hairpin ribozymes, despite having low sequence and structural homology (Figure 3.22A**)**, uphold the possibility that these ribozymes represent an example of convergent evolution (Wilson and Lilley, 2011). In the classic case of convergent evolution among protein enzymes that catalyze analogous reactions, the proteases chymotrypsin and subtilisin form a common Asp-His-Ser

catalytic triad organized by distinct tertiary scaffolds (Kraut et al., 1977). In the context of nucleases, a similar His(Tyr)-His-Lys catalytic triad at the active sites of RNase A and the unrelated tRNA endoRNase EndA may reflect convergent evolution (Bujnicki and Rychlewski, 2001). In contrast to the case in protein, RNA probably has fewer catalytic motifs available for building a nuclease active site. The VS and hairpin ribozymes could have accessed the same catalytic motif from distinct sequences and structures; however, the possibility of a shared evolutionary history between the two distinct catalytic motifs is a topic of investigation (see chapters 4 and 5). Evolutionary history notwithstanding, two dissimilar ribozymes sharing mechanistic and active-site features implicate RNA as a robust scaffold that can access functional architectures from distant regions of sequence space.

Of the other ribozymes that have been extensively studied in terms of catalysis and structure, the hammerhead ribozyme active site bears similarities with that of the VS and hairpin ribozymes. The hammerhead ribozyme forms its active site through interactions within a three-way helical junction rather than through interactions between internal loops and facilitates self-cleavage in a distinct manner (Martick and Scott, 2006). Like the VS and hairpin ribozymes, nucleophile activation involves a guanine nucleobase, but a 2′-hydroxyl group from another guanosine residue (rather than an adenine nucleobase) appears poised to stabilize the leaving group (Martick and Scott, 2006). In addition, a divalent metal ion facilitates catalysis via interactions with the scissile phosphate (Mir et al., 2015; Mir and Golden, 2016). Despite these mechanistic differences, the hammerhead ribozyme organizes its active site using motifs similar to the hairpin and VS ribozymes. The catalytic G stacks between the N−1 nucleotide and the nucleotide 5′ of it as a part of an L-platform. The guanine nucleobase harboring the catalytic 2′-OH stacks with the

N+1 nucleotide and emerges from an approximate S-turn motif that positions the upstream nucleobase coaxially (Martick and Scott, 2006) (Figure 3.23C, D).

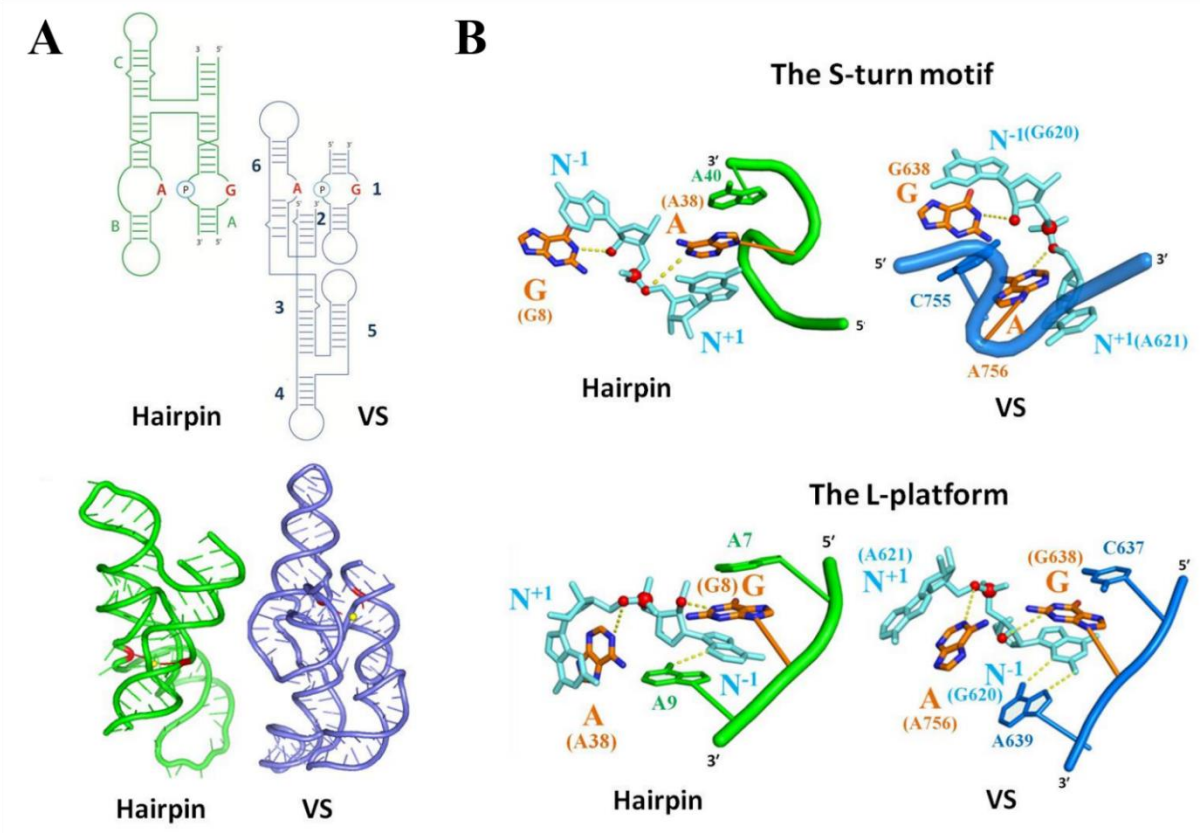

Figure 3.22 Similarities in the active sites of the VS and hairpin ribozymes. A. Secondary and tertiary structures of the hairpin (green, PDB ID code: 1M5O, Rupert and Ferré-D'Amaré, 2001) and VS (blue, PDB ID code: 4R4P, this work) ribozymes. The yellow sphere and red sticks represent the scissile phosphate and catalytic nucleobases, respectively. **B.** Similarity in the local active site structure of the hairpin and VS ribozymes. The proposed catalytic nucleobases form stacking interactions with nucleotides flanking the scissile phosphate (cyan). Red spheres correspond to oxygen atoms of the 2′-OH nucleophile, scissile phosphate and leaving group. The S-turn motif projects the adenine general acid toward the leaving group and provides an auxiliary nucleobase that stacks against the opposite side of the general acid. The L-platform positions the catalytic G nucleobase. The base of the L corresponds to a purine-purine base pair between the N−1 nucleotide and the nucleotide 3′ of the catalytic G that rotates the N−1 nucleobase toward the major groove. This facilitates formation of the stem of the L: a three-base stack between the N−1 nucleobase, the catalytic G and the nucleotide 5′ of the catalytic G (see figure 3.23D for illustration of the S-turn and L-platform). VS structures were derived from the VSx_G638A ribozyme. A guanosine nucleobase is modeled in place of A638, preserving the position of the purine base itself. VSx_C634 and VSx_A756G variants have the same motifs with the only differences indicated in Figure 3.15 A-C.

Thus, the L-platform and S-turn motifs, associated coaxial alignments, and mutualistic positioning of the cleavage site and catalytic groups through stacking appear to be a general architectural feature that can facilitate nucleolytic cleavage using different catalytic strategies (Figure 3.23D**)**. The presence of these architectural features in the context of what appears to be a distinct mechanism of catalysis underscores their functional importance and bolsters the case for convergent evolution.

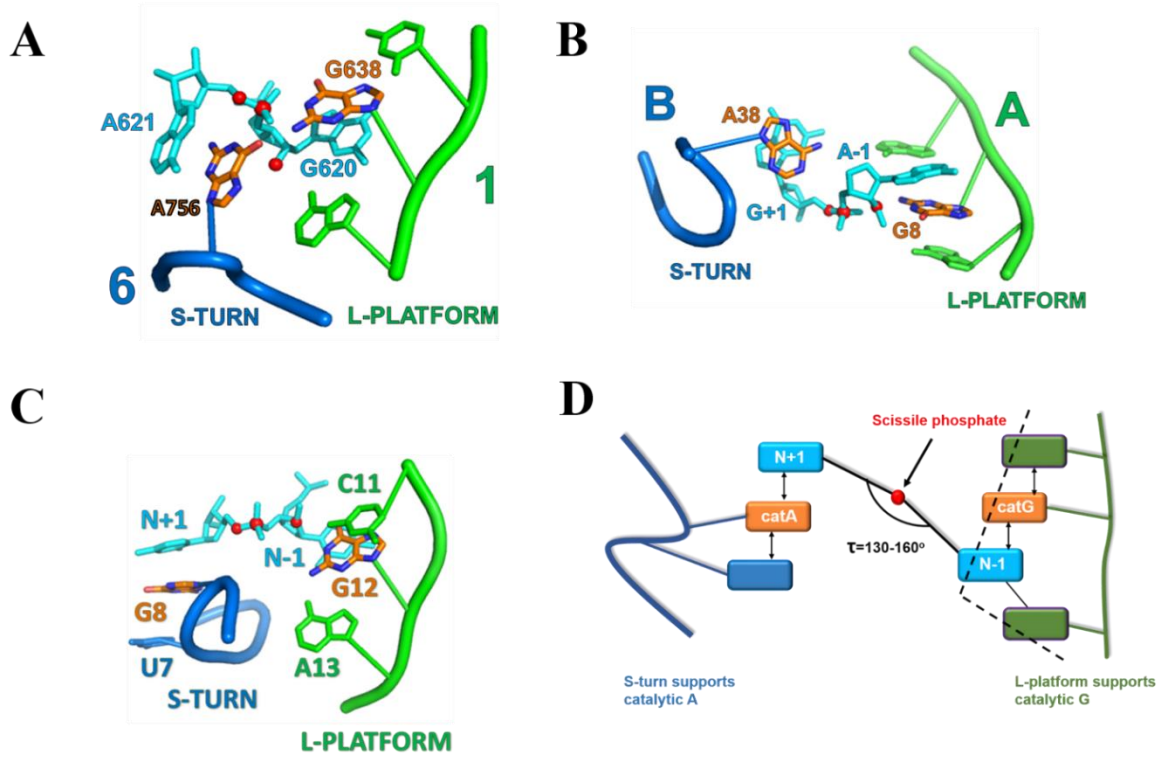

Figure 3.23 Common structural motifs in the VS, hairpin and hammerhead ribozyme active sites. A. Active site of the VS ribozyme reveals the use of an S-turn to project the general acid, and an 'L-platform' to support the general base. Nucleotides representing the general acid and base stack against the N+1 and N-1 nucleotides, respectively. **B.** Active site of the hairpin ribozyme reveals the use of an S-turn to project the general acid, and an 'L-platform' to support the general base. Nucleotides representing the general acid and base stack against the N+1 and N-1 nucleotides, respectively. **C.** The hammerhead ribozyme contains an approximate S-turn and L-platform for positioning the catalytic groups and splaying the nucleotides that flank the cleavage site (PDB ID code: 3ZD5, Martick and Scott, 2006). **D.** Schematic illustration of the S-turn and 'L-platform' motifs in the context of the active sites of the VS, hairpin and hammerhead ribozymes. Color scheme as in figure 3.22.

3.3.5 Role of VS catalysis in biology: rolling circle replication

The VS ribozyme motif exists as part of a multimeric transcript of the VS plasmid. VS RNA is interspersed with ribozyme motifs that extend to ~200 nt and are separated by non-catalytic transcripts that are ~700 nt long. The autocatalytic motifs resolve multimeric transcripts generated from the continuous transcription of the VS plasmid by catalyzing site-specific cleavage, followed by self-ligation at the same site producing RNA circles. Although considered a self-cleaving motif, the VS ribozyme can also catalyze trans cleavage/ligation reactions efficiently. Cis cleavage of the wild-type ribozyme proceeds with a moderate rate (~0.1/min) (Guo et al., 1993); however, the corresponding trans reaction occurs approximately ten times faster (Guo et al., 1995). Helix 1a that includes nucleotides upstream of the cleavage site favors ligation over cleavage and extending upstream nucleotides to an entire base-paired helix (helix 7), further reduces cleavage rates (~0.025/min), but increase ligation efficiency (Oullet et al., 2009; Jones et al., 2001). The strand 5' of the cleavage site formed as a result of endonucleolytic cleavage has a higher tendency to remain base-paired with complimentary nucleotides in the 3’ strand of helix 1a, in the presence of a full-length helix 7, which allows the terminal 5'-OH to be in close proximity to the 2'-3' cyclic phosphate, which is the other product of cleavage. This allows for the reverse reaction, ligation. The preponderance of the circular form of VS transcripts support the propensity of full-length sequences toward ligation (Saville and Collins, 1990, Kennell et al., 1995).

The crystal structure of the VS ribozyme dimer revealed how enzyme-substrate complexes are created by trans docking of the substrate of one protomer into the catalytic domain of the other. This was enabled by the hitherto unknown conformation of junction 1-2-7, that projects the substrate (helix 1) away from the catalytic domain of its own protomer. This model of trans interaction between two full-length ribozymes is supported by functional data that demonstrated

that two cis-active ribozymes can cleave each other's substrate *in trans* (Oullet et el. 2009). Reconstitution of VS cleavage *in trans*, however, showed similar cleavage rates as cis cleavage (~0.025/min), underscoring the significance of helix 1a and more importantly, helix 7 in cleavage attenuation (Oullet et el. 2009). This dimer arrangement captured in the crystal explains why trans cleavage is favored over cis cleavage by the VS ribozyme. Local restrictions on folding imposed by junction 1-2-7, makes trans docking easier than cis docking, as the former involves a complete physical separation of the substrate from the catalytic domain introducing greater flexibility for docking. Artificial VS constructs with circular permutations that connect the substrate to the 3' end of the catalytic domain via a single-stranded linker are more efficient in cis cleavage, further supporting the roles of junction 1-2-7 in regulating cis versus trans activity (Zamel et al., 2004). The preference for trans cleavage is emphasized when the catalytic domain has the option of cleaving either an upstream or a downstream substrate. The catalytic domain prefers to cleave a substrate-helix that is ~700 nucleotide downstream, over the substrate-helix immediately upstream of it in a model construct containing a single catalytic domain flanked by two substrates designed to mimic the biological context of the VS ribozyme (Figure 3.24) (Poon et al., 2006). This suggests that the trans docking captured in the dimer crystal structure is probably biologically relevant for VS cleavage.

We observed that uncleaved precursors had higher affinity toward dimerization than cleaved products (Figures 2.7D, 2.9). Helix 1a is weakened upon substrate cleavage, which can transiently unstructure junction 1-2-7. Unstructuring this junction allows the substrate-helix more flexibility rather than lock it in a conformation suitable for trans docking, and consequently dimerization. Therefore, monomers contain both precursor and product species, while dimers contain only precursors (Figures 2.7D, 2.9). Since the dimer can be used as a proxy for trans

docking, the reduced affinity of products for dimerization could explain product release from the enzyme-substrate complex after cleavage.

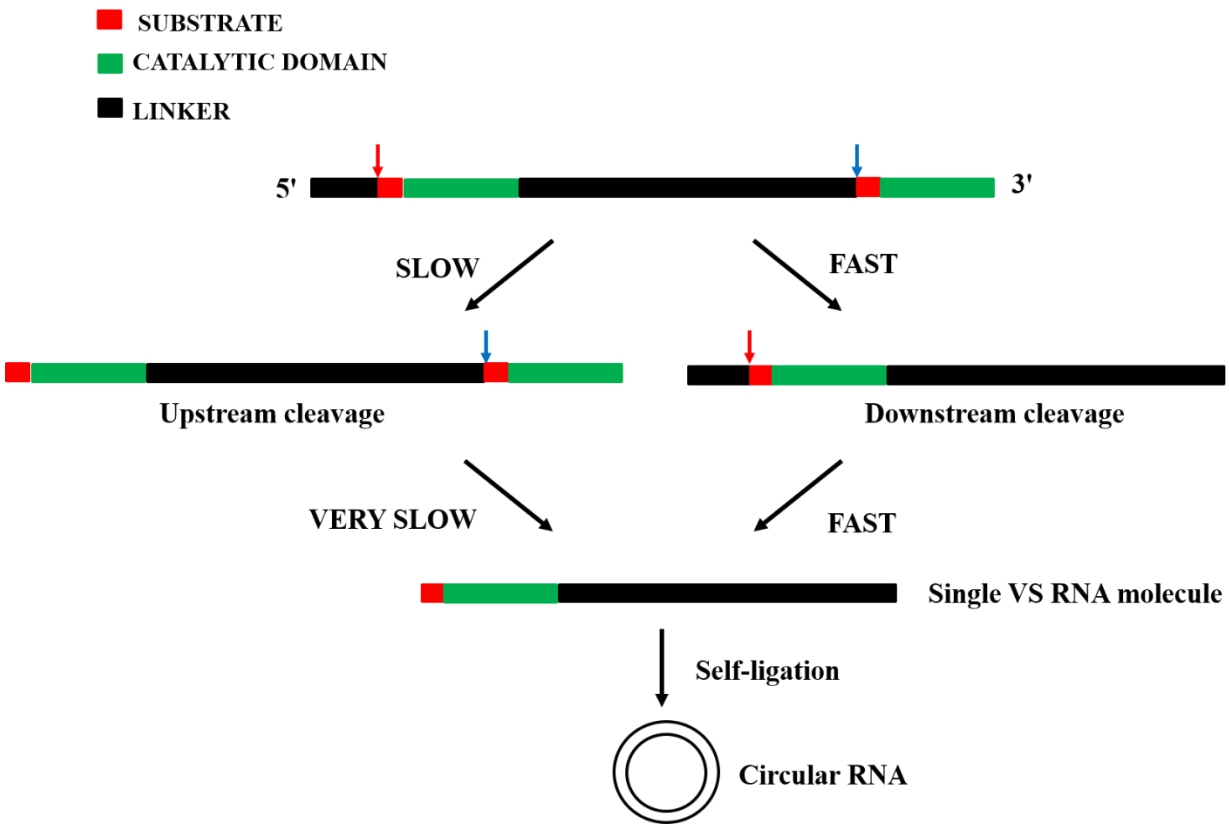

Figure 3.24 Experiment designed to mimic the native multimeric context for VS cleavage. The construct consists of a catalytic domain sequence flanked by a substrate immediately upstream of it and a downstream substrate separated by a ~700 nt linker. Upstream, and downstream cleavage sites are indicated by red and blue arrows, respectively. Appearance of products resulting from the cleavage of an upstream substrate is slower compared to that from downstream cleavage. The difference in rates of further cleavage of these products probably results from the continued occupancy of the upstream substrate in the docking site in the initial product of upstream cleavage (left) (Pereira et al., 2008). The second cleavage reaction produces a single transcript of VS RNA which self-ligates to form circular RNA.

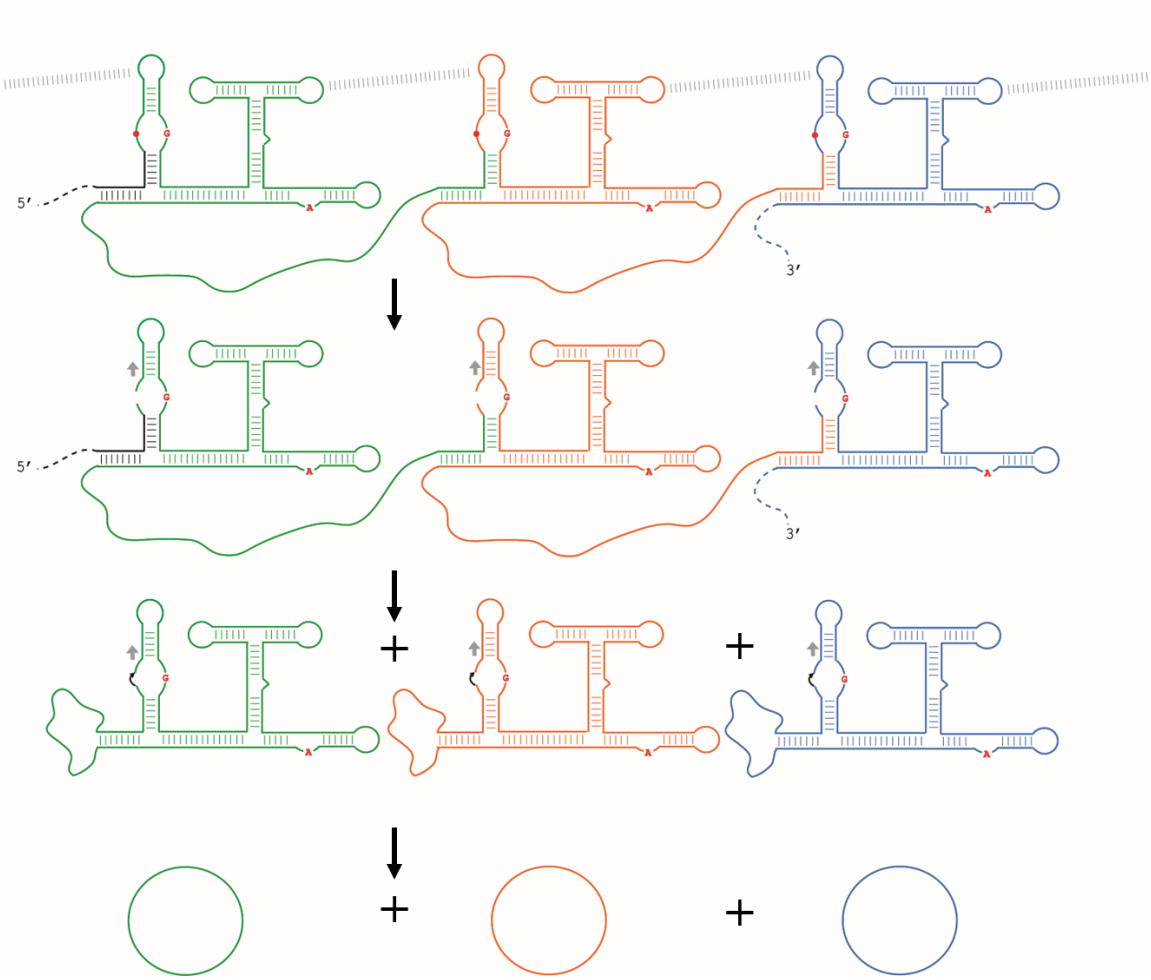

Figure 3.25 Role of the VS ribozyme in rolling-circle replication. Catalytic domains bind downstream substrates through kissing-loop interactions (shown as dashed lines), shift the substrates (gray arrow) and catalyze trans cleavage. After cleavage, junctions connecting substrates to its catalytic domains are relaxed allowing strand exchange. Following strand exchange the VS catalyzes cis-ligation in a substrate already activated from the initial catalytic step, producing RNA circles. This is a hypothetical model awaiting validation (see figure 3.26).

The role of cleavage in 'relaxing' junction 1-2-7 has implications in regulating the cleavage-ligation equilibrium that is central to the formation of circular RNA intermediates during rolling-circle replication of VS plasmids. Since a structured junction favors trans docking, a VS substrate in a multimeric transcript probably docks into the catalytic domain of the VS ribozyme motif upstream of it resulting in substrate rearrangement and trans cleavage (Figure 3.25). The severed connection in the cleaved substrate-helix frees helix 1b from the constraints imposed by the structure of junction 1-2-7 allowing strand exchange between two adjacent ribozyme motifs necessary for the formation of closed circular transcripts. 5' strands from helices 1a and 7, disconnected from the rest of the RNA, participates in the strand-swap (Figure 3.25), which involves extensive unstructuring and restructuring of junction 1-2-7. The disruption of this junction in the product-helix lowers its affinity toward the catalytic domain resulting in product release. This frees up the catalytic domain for cis docking. A transiently relaxed conformation of junction 1-2-7 could allow docking *in cis*, thereby positioning the 5'-hydroxyl and 2'-3' cyclic phosphate groups close to one another at the active site for cis ligation. An already shifted helix 1b (shifting occurs due to trans docking) favors ligation (Figure 3.25). Ligation could occur after formation of the new helix 1a and helix 7, before junction 1-2-7 becomes fully structured. After ligation, restructuring of the junction would draw the substrate-helix away from its own active site, preventing self-cleavage and therefore producing stable RNA circles (Figure 3.25).

The majority of endonucleolytic ribozymes resolve multimeric intermediates in rolling-circle replication to produce RNA circles by the ligation of intermediate cleavage products. The generation of a 2'-3' cyclic phosphate product from ribozyme cleavage probably has an evolutionary advantage over the products generated from complete cleavage of a phosphodiester (involving hydrolysis of the cyclic product) in the context of re-ligation at the

same site. Ligation that involves a 2'-3' cyclic phosphate does not require ATP, unlike a ligation reaction that involves a phosphate group making the cleavage-ligation paradigm energetically more facile in ribozyme biology.

3.4 FUTURE DIRECTIONS

High-resolution crystal structures of the VS ribozyme have provided a structural framework for >25 years of functional data. In addition, these structures have revealed aspects of VS catalysis that were previously unknown. These insights provide impetus for new investigations into VS structure and function. Some of the follow up experiments that emerge from our structural studies are discussed below.

Conformation and role of junction 1-2-7. Our structures capture a domain-swapped dimer where the substrate from one protomer points away from its catalytic domain. This orientation is presumably stabilized by the local conformation of junction 1-2-7. However, the substrate must dock into the catalytic domain of the same molecule for cis cleavage which requires changes in junction 1-2-7. Single point mutations to nucleotide G778, which mediates the most interactions in this junction do not affect either trans or cis cleavage (DasGupta et al., unpublished results). This might suggest that further disruptions are needed to 'relax' the conformation of this junction for enabling cis cleavage. This problem could be approached by meticulous mutagenesis to the nucleotides in junction 1-2-7, or through a randomization-selection experiment, where junction 1-2-7 would be randomized, and complete ribozymes with this randomized library at J_{127} would be challenged with cis and trans activities. Comparing junction sequences of molecules active for cis and trans cleavage/ligation would enable us to map the differences in the junction that enable one form of activity over the other.

Randomization-selection experiments should also be complimented with computation. Simulations of substrate docking in the background of different J_{127} sequences could elucidate the difference in docking dynamics between cis and trans forms of the ribozyme, in addition to thermodynamic requirements for docking for each. This has implication for domain swapping in

multidomain RNAs and might reveal principles governing the formation of quaternary structures in RNA.

Sequences with cis activity could be further investigated structurally. This might yield structures of the monomeric form of the VS ribozyme. A monomer structure is important because it would likely provide structural insights into cis docking that are currently unavailable. Electrophoretic mobility shift assays (EMSA) and FRET-based studies used to determine the global folds of junctions 2-3-6 and 3-4-5 (Lafontaine et al., 2001a; Lafontaine et al., 2002a) could be used to determine the global fold of an isolated J_{127}. Comparing global folds of junction 1-2-7 in isolation and as a part of the full-length ribozyme could speak to the modular nature of VS domains and provide information regarding the relative orientation of helix 1 (substrate-helix) in this junction.

Rearrangement of substrate secondary structure: structural investigations. Shifting of helix 1b nucleotides triggers a rearrangement of the cleavage loop into its catalytically-competent conformation, which includes splayed cleavage site nucleotides, an unpaired catalytic base, G638 and the intrusion of the general acid, A756 that is otherwise distant from the cleavage site. This conformation is stabilized by tertiary interactions between extruded nucleotides from the substrate cleavage loop and nucleotides from helices 2 and 6. The abrogation of cleavage in the absence of a register shift in helix 1b is likely due to the inability of the cleavage loop to rearrange. Structure determination of the full-length VS ribozyme with a substrate sequence that forces helix 1b into its unshifted conformation (Zamel and Collins, 2002), would provide a glimpse into the active site of an unshifted substrate docked into the catalytic domain. However, the kissing-loop interaction between the catalytic domain and an unshifted substrate is weak therefore, formation of the VS dimer would be difficult. Since, the

crystallization conditions were optimized for the dimer, crystallizing the VS monomer with an unshifted substrate would require additional exploration of crystallization condition space.

Rearrangement of substrate secondary structure: computational and biophysical investigations. Crystal structure of VSx_C634 captured the conformation of the VS ribozyme with a docked substrate containing wild-type helix 1b. This crystal structure and the NMR structure of an isolated substrate stem-loop (Flinders and Dieckmann, 2001) present static coordinates of the initial and final states of the substrate during binding but do not provide information regarding the dynamics of this process. This involves understanding the sequence of molecular motion corresponding to binding, shifting, and remodeling of the cleavage loop. Calculating the path of least energy using computational approaches would provide an approximate thermodynamically feasible trajectory for substrate rearrangement.

Complimentary experiments using time-resolved fluorescence spectroscopy would provide experimental insights into the changes in local environments of nucleotides that are involved in substrate rearrangement. C632 could play an important role in the transition from unshifted to shifted conformation by switching between the G627-C632 and C632-G697 base-pairs. The former base-pair favors a unshifted conformation, while the latter stabilizes the kissing-loop interaction with stem-loop 5, triggering shifting of helix 1b. Nucleotides, C626 and C634 unpair from their respective WC base-pairs during the register shift that introduces minor changes in the stacking environment; however, C634 is solvent exposed as inferred from DMS probing data (Anderson and Collins, 2001). Since these cytosines undergo prominent changes with respect to local environment, substituting them with a fluorescent base analog, pyrrolo-cytosine (pyrroloC) (Tinsley and Walter, 2006; Zhao and Xia, 2009) would enable us to follow the structural rearrangements in real-time. The fluorescence of pyrroloC is quenched when located inside a base-

paired helix, due to base-stacking and base-pairing. Unpairing and unstacking have additive effects on its fluorescence signal, therefore pyrroloC could be a marker for measuring degrees of unpairing/unstacking during the course of substrate rearrangement. A similar approach could be used to probe the transition of C637 from being involved in a non-canonical base-pair in the cleavage loop to being part of the G623-C637 WC pair in the shifted helix 1b. Other nucleotides that undergo substantial molecular motion are A621 and A622. Since these bases are flipped out of the helix, introducing fluorescent probes at these positions would presumably provide clear indication of the structural rearrangements that define an active cleavage loop. 2-aminopurine (2AP), a fluorescent analog of adenosine has fluorescent properties that are similar to pyrroloC and is probably the most studied fluorescent base analog (Walter, et al., 2002; Zhao and Xia, 2009). 2AP substitutions at positions 621 and 622 would enable us to probe the local conformational dynamics of the internal cleavage loop. This three-pronged approach involving structural, biophysical and computational investigations should be able to test the model suggested in section 3.3.1 regarding the roles of the kissing-loop interaction and register shift in assembling the VS active site. Some of the specific questions to be addressed are as follows:

a. How is the kissing loop interaction coupled to shifting and cleavage loop remodeling?
b. Is the active conformation of the cleavage loop due to base flipping or is base-flipping a stochastic process that stabilizes the active conformation of the cleavage loop?
c. What are the energetics of substrate rearrangement?
d. What structural features of an unshifted helix 1b inhibit the formation of a catalytically-competent active site?

Investigations into secondary structure rearrangements in the VS ribozyme could set up a framework for other structural rearrangements induced by RNA-RNA tertiary interactions.

Catalysis in the active site. Structure of the active site as revealed by the two complementary crystal structures (VSx_G638A and VSx_A756G) provides a glimpse into the catalytic architecture of the ribozyme. In addition to the interactions corresponding to general acid-base catalysis, new catalytic interactions between the putative general base, G638 and the pro-*Rp* and pro-*Sp* oxygen atoms, and between a modeled general acid, A756 (in the VSx_A756G structure) with the pro-*Sp* oxygen atom of the scissile phosphate, were revealed. These interactions should be tested by incorporation of nucleobase analogs of G638 and A756 that exhibit a variety of functional groups in the background of a phosphorothioate linkage at the scissile phosphate. Preliminary results confirm the interaction between the exocyclic amino group of G638 and the scissile phosphate, observed in the crystal structure (Weissman et al., unpublished results). Quantum Mechanics/Molecular Mechanics (QM/MM) studies investigating the roles of catalytic nucleobases in the active site support these structural and functional data (Ganguly et al., unpublished results). Our crystal structures promise to be a great starting point for further computational efforts.

Testing the trans-cleavage/cis ligation model in vitro. Formation of RNA circles require strand exchange between nucleotides upstream and downstream of the cleavage site. Each circle is a result of two cleavage reactions and one ligation. Trans cleavage by the catalytic domain of one RNA releases the downstream strand (which is the upstream strand of the next RNA), while trans cleavage of its substrate by the catalytic domain of an upstream RNA releases the upstream strand (Figure 3.26). Strand exchange is presumably followed by ligation *in cis.* This model can be tested using a concatemeric RNA construct consisting of three complete VS ribozyme molecules each separated by a ~700 nt linker (Figure 3.26). Substrate docking involves WC base-pairing between the hairpins of the substrate and stem-loop 5 of the catalytic domain (kissing-loop

interaction). It is possible to selectively inhibit cis/trans cleavage by the disruption of this intermolecular interaction and rescue activity by complimentary mutations. The active site is assembled using a catalytic guanine from the substrate and a catalytic adenine from the catalytic domain. By similar mutation-and-rescue experiments on the catalytic residues it would be possible to distinguish cis and trans cleavage/ligation thereby defining the nature of interactions between substrates and catalytic domains (upstream or downstream) with respect to cleavage and ligation in the multimeric construct. Experimental readout would be provided by the formation of circles, which indicates two successful cleavage reactions and one ligation reaction.

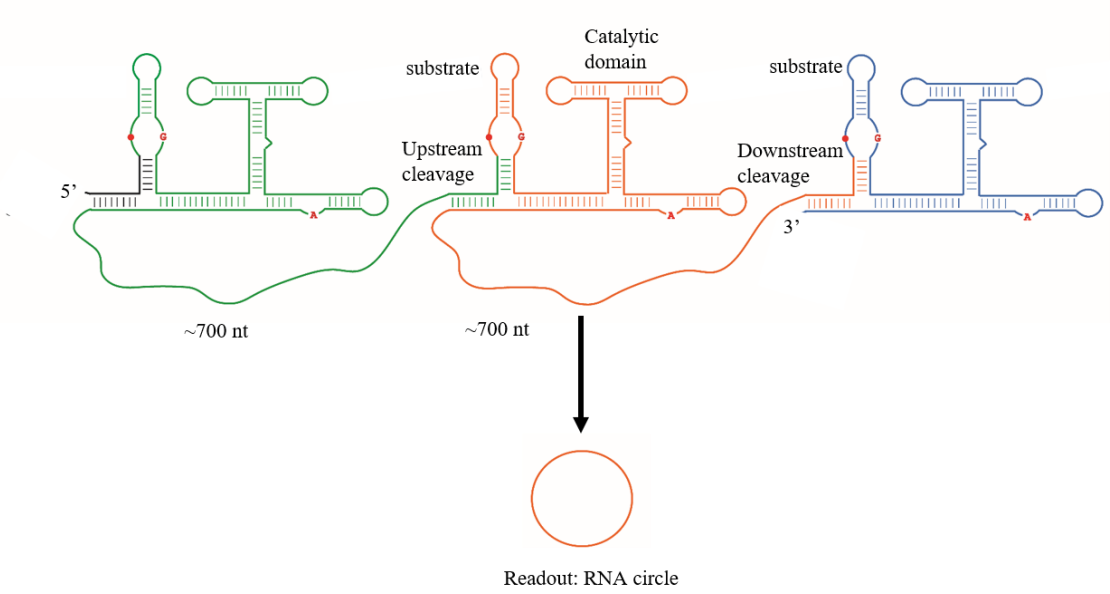

Figure 3.26 Model to study the role of VS ribozymes in rolling-circle replication *in vitro*. The construct consists of three VS ribozyme motifs. The orange molecule flanked by two other ribozymes is capable of generating a circular RNA through cleavage of two substrates shown in orange and blue, and ligation of the one in orange. By making mutations that inhibit either cis or trans activity and assaying for the production of circular RNA, it would be possible to define the interactions between substrates and catalytic domains that are responsible for producing RNA circles, the product of VS catalysis in biology.

RNA SEQUENCES USED IN THE EXPERIMENTS DISCUSSED IN THIS CHAPTER ARE LISTED IN MATERIALS AND METHODS (7.9.1, CHAPTER 7).

REFERENCES

1. Andersen, A. A. and Collins, R. A., 2000. Rearrangement of a stable RNA secondary structure during VS ribozyme catalysis. *Mol. Cell*, 5, 469-478.

2. Andersen, A. A. and Collins, R. A., 2001. Intramolecular secondary structure rearrangement by the kissing interaction of the Neurospora VS ribozyme. *Proc. Natl. Acad. Sci. U. S. A.*, 98, 7730-7735.

3. Babitzke, P. and Yanofsky, C., 1993. Reconstitution of *Bacillus subtilis trp* attenuation *in vitro* with TRAP, the *trp* RNA-binding attenuation protein. *Proc. Natl. Acad. Sci. U. S. A.*, 90, 133–137.

4. Beattie, T. L. and Collins, R. A., 1997. Identification of functional domains in the self-cleaving Neurospora VS ribozyme using damage selection. *J. Mol. Biol.*, 267, 830-840.

5. Beattie, T. L., Olive, J. E. and Collins, R. A., 1995. A secondary-structure model for the self-cleaving region of Neurospora VS RNA. *Proc. Natl. Acad. Sci. U. S. A.*, 92, 4686-4690.

6. Bergonzo, C., Hall, K. B. and Cheatham III, T. E., 2015. Stem-loop V of Varkud satellite RNA exhibits characteristics of the Mg^{2+} bound structure in the presence of monovalent ions. *J. Phys. Chem.*, 119, 12355–12364.

7. Bouchard, P. and Legault, P., 2014. A remarkably stable kissing-loop interaction defines substrate recognition by the *Neurospora* Varkud Satellite ribozyme. *RNA*, 20, 1451-1464.

8. Bouchard, P. and Legault, P., 2014. Structural insights into substrate recognition by the *Neurospora* Varkud Satellite ribozyme: Importance of U-turns at the kissing-loop junction. *Biochemistry*, 53, 258–269.

9. Bujnicki, J.M. and Rychlewski, L., 2001. Unusual evolutionary history of the tRNA splicing endonuclease EndA: relationship to the LAGLIDADG and PD-(D/E) XK deoxyribonucleases. *Protein Sci.*, 10, 656–660.

10. Cai, Z. and Tinoco Jr., I., 1996. Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite. *Biochemistry*, 35, 6026-6036.

11. Cepeda-Plaza, M., McGhee, C. and Lu, Y., 2017. Evidence of a General Acid/Base catalysis mechanism in the 8-17 DNAzyme. *Biochemistry*, doi: 10.1021/acs.biochem. 7b01096

12. Cochrane, J. C. and Strobel, S. A., 2008. Catalytic strategies of self-cleaving ribozymes. *Acc. Chem. Res.*, 41, 1027-1035.

13. Costa, M., Walbott, H., Monachello, D., Westhoff, E. and Michel, F., 2016. Crystal structures of a group II intron lariat primed for reverse splicing. *Science*, 354, 6316-6323.

14. Desjardins, G., Bonneau, E., Girard, N., Boisbouvier, J. and Legault, P., 2011. NMR structure of the A730 loop of the *Neurospora* VS ribozyme: insights into the formation of the active site. *Nucleic Acids Res.*, 39, 4427-4437.

15. Duarte, C. M. and Pyle, A. M., 1998. Stepping through an RNA structure: a novel approach to conformational analysis. *J. Mol. Biol.*, 284, 1465–1478.

16. Duarte, C. M., Wadley, L. M. and Pyle, A. M., 2003. RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. *Nucleic Acids Res.*, 31, 4755–4761.

17. Fayat, G., Mayaux, F. J., Sacerdot, C., Fromant, M., Springer, M., Grunberg-Manago, M. and Blanquet, S., 1983. *Escherichia coli* phenylalanyl-tRNA synthetase operon region. Evidence for an attenuation mechanism. Identification of the gene for the ribosomal protein L20. *J. Mol. Biol.*, 171, 239-261.

18. Fedor, M. J., 2000. The structure and function of the hairpin ribozyme. *J. Mol Biol.*, 297, 268-291.

19. Flinders, J. and Dieckmann, T., 2001. A pH controlled conformational switch in the cleavage site of the VS ribozyme substrate RNA. *J. Mol. Biol.*, 308, 665-679.

20. Gilbert, S. D. and Batey, R. T., 2006. Riboswitches: fold and function. *Chem. Biol.*, 13, 805-807.

21. Heldenbrand, H., Janowski, P.A., Giambasu, G., Giese, T. J., Wedekind, J. E. and York, D. M., 2014. Evidence for the role of active site residues in the hairpin ribozyme from molecular simulations along the reaction path. *J. Am. Chem. Soc.*, 136, 7789–7792.

22. Hiley, S. L., Sood, V. D., Fan, J. and Collins, R. A., 2002. 4-thio-U cross-linking identifies the active site of the VS ribozyme. *EMBO J.*, 21, 4691-4698.

23. Höbartner, C. and Micura, R., 2003. Bistable secondary structures of small RNAs and their structural probing by comparative imino proton NMR spectroscopy. *J. Mol. Biol.*, 325, 421–431.

24. Hoffmann, B., Mitchell, G. T., Gendron, P., Major, F., Andersen, A. A., Collins, R. A., and Legault, P., 2003. NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site. *Proc. Natl. Acad. Sci. U.S.A.*, 100, 7003-7008.

25. Jones, F. D. and Strobel, S. A., 2003. Ionization of a critical adenosine residue in the neurospora Varkud Satellite ribozyme active site. *Biochemistry*, 42, 4265-4276.

26. Kath-Schorr, S., Wilson, T. J., Li, N.-S., Lu, J., Piccirilli, J. A. and Lilley, D. M. J., 2012. General acid-base catalysis mediated by nucleobases in the hairpin ribozyme. *J. Am. Chem. Soc.*, 134, 16717–16724.

27. Keating, K. S., Humphris, E. L. and Pyle, A. M., 2011. A new way to see RNA. *Q. Rev. Biophys.*, 44, 433–466.

28. Kennell, J. C., Saville, B. J., Mohr, S., Kuiper, M. T., Sabourin, J. R., Collins, R. A. and Lambowitz, A.M., 1995. The VS catalytic RNA replicates by reverse transcription as a satellite of a retroplasmid. *Genes Dev.*, 9, 294-303.

29. Knitt, D. S., Narlikar, G. J. and Herschlag, D., 1994. Dissection of the role of the conserved G.U pair in group I RNA self-splicing. *Biochemistry*, 33, 13864-13879.

30. Kraut, J., 1977. Serine proteases: structure and mechanism of catalysis. *Annu. Rev. Biochem.*, 46, 331–358.

31. Lafontaine, D. A., Norman, D. G. and Lilley, D. M. J., 2001a. Structure, folding and activity of the VS ribozyme: importance of the 2-3-6 helical junction. *EMBO J.*, 20, 1415-1424.

32. Lafontaine, D. A., Norman, D. G. and Lilley, D. M. J., 2002a. The global structure of the VS ribozyme. *EMBO J.*, 21, 2461-2471.

33. Lafontaine, D. A., Wilson, T. J., Norman, D. G. and Lilley, D. M. J., 2001b. The A730 loop is an important component of the active site of the VS ribozyme. *J. Mol. Biol.*, 312, 663-674..

34. Lafontaine, D. A., Wilson, T. J., Zhao, Z. Y., and Lilley, D. M. J., 2002b. Functional group requirements in the probable active site of the VS ribozyme. *J. Mol. Biol.*, 323, 23-34.

35. Liberman, J. A., Guo, M., Jenkins, J. L., Krucinska, J., Chen, Y., Carey, P. R. and Wedekind, J. E., 2012. A transition-state interaction shifts nucleobase ionization toward neutrality to facilitate small ribozyme catalysis *J. Am. Chem. Soc.*, 134, 16933-16936.

36. Liu, H., Yu, X., Chen, Y., Zhang, J., Wu, B., Zheng, L., Haruehanroengra, P., Wang, R., Li, S., Lin, J., Li, J., Sheng, J., Huang, Z., Ma, J. and Gan, J., 2017. Crystal structure of an RNA-cleaving DNAzyme. *Nat. Commun.*, 8, 2006. doi: 10.1038/s41467-017-02203-x.

37. Liu, Y., Wilson, T. J. and Lilley, D. M. J., 2017. The structure of a nucleolytic ribozyme that employs a catalytic metal ion. *Nat. Chem. Biol.*, 13, 508-514.

38. Liu, Y., Wilson, T. J., McPhee, S. A. and Lilley, D. M. J., 2014. Crystal structure and mechanistic investigation of the twister ribozyme. *Nat. Chem. Biol.*, 10, 739-745.

39. Lodmell, J. S. and Dahlberg, A. E., 1997. A conformational switch in Escherichia coli 16S ribosomal RNA during decoding of messenger RNA. *Science*, 277, 1262–1267.

40. Martick, M. and Scott, W. G., 2006. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell*, 126, 309-320.

41. Michiels, P. J. A., Schouten, C. H. J., Hilbers, C. W., Heus, H. A., 2000. Structure of the ribozyme substrate hairpin of Neurospora VS RNA: a close look at the cleavage site. *RNA*, 6, 1821-1832.

42. Micura, R. and Höbartner, C., 2003. On secondary structure rearrangements and equilibria of small RNAs. *Chembiochem.*, 4, 984-990.

43. Mir, A. and Golden, B. L., 2016. Two active site divalent ions in the crystal structure of the hammerhead ribozyme bound to a transition state analogue. *Biochemistry*, 55, 633-636.

44. Mir, A., Chen, J., Robinson, K., Lendy, E., Goodman, J., Neau, D. and Golden, B. L., 2015. Two divalent metal ions and conformational changes play roles in the hammerhead ribozyme cleavage reaction. *Biochemistry*, 54, 6369-6381.

45. Nagaswamy, U., Voss, N., Zhang, Z. and Fox, G. E., 2000. Database of non-canonical base pairs found in known RNA structures. *Nucleic Acids Res.*, 28, 375-376.

46. Nguyen, T. H. D., Galej, W. P., Fica, S. M., Lin, P.-C., Newman, A. J. and Nagai, K., 2016. CryoEM structures of two spliceosomal complexes: starter and dessert at the spliceosome feast. *Curr. Opin. Struct. Biol.*, 36, 48-57.

47. Ochieng, P. O., White, N. A., Feig, M. and Hoogstraten, C. G., 2016. Intrinsic Base-Pair Rearrangement in the Hairpin Ribozyme Directs RNA Conformational Sampling and Tertiary Interface Formation *J. Phys. Chem. B*, 120, 10885-10898.

48. Pereira, M. J. B., Nikolova, E. N., Hiley, S. L., Jaikaran, D., Collins, R. A. and Walter, N. G., 2008. Single VS ribozyme molecules reveal dynamic and hierarchical folding toward catalysis. *J. Mol. Biol.*, 382, 496-509.

49. Pinard, R., Hampel, K. J., Heckman, J. E., Lambert, D., Chan, P. A. Major, F. and Burke, J, M., 2001. Functional involvement of G8 in the hairpin ribozyme cleavage mechanism. *EMBO J.*, 20, 6334–6442.

50 Poon, A. H. L, Olive, J. E., McLaren, M. and Collins, R. A., 2006. Identification of separate structural features that affect rate and cation concentration dependence of self-cleavage by the Neurospora VS ribozyme. *Biochemistry*, 45, 13394-13400.

51. Putzer, H., Gendron, N., Grunberg-Manago, M., 1992. Coordinate expression of the two threonyl-tRNA synthetase genes in *Bacillus subtilis*: Control by transcriptional antitermination involving a conserved regulatory sequence. *EMBO J.*, 11,3117–3127.

52. Rastogi, T. and Collins, R. A., 1998. Smaller, faster ribozymes reveal the catalytic core of Neurospora VS RNA. *J. Mol. Biol.*, 277, 215-224.

53. Rastogi, T., Beattie, T. L., Olive, J. E. and Collins, R. A., 1996. A long-range pseudoknot is required for activity of the Neurospora VS ribozyme. *EMBO J.*, 15, 2820-2825.

54. Ren, A., Košutić, M., Rajashankar, K. R., Frener, M., Santner, T., Westhof, E., Micura, R. and Patel, D. J., 2014. In-line alignment and Mg^{2+} coordination at the cleavage site of the env22 twister ribozyme. *Nat. Commun.*, 5:5534. doi: 10.1038/ncomms6534.

55. Ren, A., Micura, R. and Patel, D. J., 2017. Structure-based mechanistic insights into catalysis by small self-cleaving ribozymes. *Curr. Opin. Chem. Biol.*, 41, 71–83.

56. Rupert, P. B. and Ferré-D'Amaré, A. R., 2001. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, 410, 780-786.

57. Rupert, P. B., Massey, A. P., Sigurdsson, S. T. and Ferré-D'Amaré, A. R., 2002. Transition state stabilization by a catalytic RNA. *Science*, 298, 1421-1424.

58. Saville, B. J. and Collins, R. A., 1990. A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell*, 61, 685-696.

59. Savva, R., McAuley-Hecht, K., Brown, T. and Pearl, L., 1995. The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature*, 373, 487-493.

60. Silverman, S. K., Zheng, M., Wu, M., Tinoco Jr., I. and Cech, T. R., 1999. Quantifying the energetic interplay of RNA tertiary and secondary structure interactions. *RNA*, 5, 1665-1674.

61. Sood, V. D. and Collins, R. A., 2002. Identification of the catalytic subdomain of the VS ribozyme and evidence for remarkable sequence tolerance in the active site loop. *J. Mol. Biol.*, 320, 443-454.

62. Szewczak, A. A., Moore, P. B., Chang, Y. L. and Wool, I. G., 1993. The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc. Natl Acad. Sci. U. S. A.*, 90, 9581–9585.

63. Tinsley, R. A. and Walter, N. G., 2006. Pyrrolo-C as a fluorescent probe for monitoring RNA secondary structure formation. *RNA*, 12, 522-529.

64. Tucker B. J. and Breaker, R. R., 2005. Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.*, 15, 342-348.

65. Vitagliano, L., Merlino, A. Zagari, A. and Mazzarella, L., 2002. Reversible substrate-induced domain motions in ribonuclease A. *Proteins*, 46, 97-104.

66. Von Ahsen, U., 1998. Translational fidelity: Error-prone versus hyperaccurate ribosomes. *Chem. Biol.*, 5, R3–R6.

67. Walter, N. G., Harris, D. A.,Pereira, M. J. B. and Rueda, D., 2002. In the fluorescent spotlight: global and local conformational changes of small catalytic RNAs. *Biopolymers*, 61, 224-242.

68. Wilson, T. J. and Lilley, D. M. J., 2011. Do the hairpin and VS ribozymes share a common catalytic mechanism based on general acid-base catalysis? A critical assessment of available experimental data. *RNA*, 17, 213–221.

69. Wilson, T. J., Li, N.-S., Lu, J., Frederiksen, J. K. Piccirilli, J. A. and Lilley, D. M. J., 2010. Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme. *Proc. Natl. Acad. Sci. U. S. A.*, 107, 11751–11756.

70. Wilson, T. J., McLeod, A. C. and Lilley, D. M. J., 2007. A guanine nucleobase important for catalysis by the VS ribozyme. *EMBO J.*, 26, 2489-2500.

71. Wimberly,B., Varani,G. and Tinoco, I. Jr., 1993. The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry*, 32, 1078–1087.

72. Wu, M. and Tinoco Jr., I., 1998. RNA folding causes secondary structure rearrangement. *Proc. Natl. Acad. Sci. U.S.A.*, 95, 11555-11560.

73. Xue, Y., Gracia, B., Herschlag, D., Russel, R., Al-Hashimi, H. M., 2016. Visualizing the formation of an RNA folding intermediate through a fast highly modular secondary structure switch. *Nat. Commun.*, 7, ncomms11768.

74. Zamel, R. and Collins, R. A., 2002. Rearrangement of substrate secondary structure facilitates binding to the Neurospora VS ribozyme. *J. Mol. Biol.*, 324, 903-915.

75. Zamel, R., Poon, A., Jiakaran, D., Andersen, A., Olive, J., De Abreu, D. and Collins, R. A., 2004. Exceptionally fast self-cleavage by a *Neurospora* Varkud satellite ribozyme. *Proc. Natl. Acad. Sci. U. S. A.*, 101, 1467-1472.

76. Zhao, C., Rajashankar, K. R., Marcia, M. and Pyle, A. M., 2015. Crystal structure of group II intron domain 1 reveals a template for RNA assembly. *Nat. Chem. Biol.*, 11, 967-972.

77. Zhao, L. and Xia, T., 2009. Probing RNA conformational dynamics and heterogeneity using femtosecond time-resolved fluorescence spectroscopy. *Methods*, 49, 128-135.

78. Zheng, L., Mairhofer, E., Teplova, M., Zhang, Y., Ma, J., Patel, D. J., Micura, R., Ren, A., 2017. Structure-based insights into self-cleavage by a four-way junctional twister-sister ribozyme. *Nat. Commun.*, 8:1180. doi: 10.1038/s41467-017-01276-y.

Chapter - 4

SMOOTH ACQUISITION OF CATALYTIC FUNCTION IN RIBOZYMES THROUGH INTERSECTION OF NEUTRAL NETWORKS: SETTING THE STAGE

4.1 INTRODUCTION

4.1.1 Growing diversity of endonucleolytic ribozymes warrant investigations into their evolutionary histories

The last four years have witnessed the discoveries of four new classes of endonucleolytic ribozymes that have spurned a renewed interest in the structural and mechanistic investigations of ribozymes (Ren et al., 2017). Structurally and functionally distinct ribonuclease motifs made of RNA are usually found in different biological contexts yet reveal variations on the same theme of nucleobase catalysis in their respective active sites (section 1.2.3, Figure 1.9). This might reflect the lack of chemical diversity in the catalytic repertoire of RNA. This lack of chemical diversity is strongly reflected in the similarities in the active site architectures in the VS, hairpin and hammerhead ribozymes (Figure 3.22) that have led to speculations that these dissimilar RNAs converged to a common catalytic architecture from distant regions in sequence space (section 3.3.4, Suslov et al., 2015). Although these extant ribonuclease ribozymes may have evolved fairly recently in evolutionary history, RNA cleavage was likely an important function of ribozymes in the RNA World (section 1.2.1), with possible roles in primitive gene regulation in RNA-based genomes. The observation that some modern-day ribozymes like the *glmS*, and more speculatively CPEB3 and CLEC2 regulate genetic information in prokaryotes and eukaryotes, respectively (section 1.2.5) support the plausibility of such roles of functional RNA in primitive biology. In

addition to gene regulation, RNA cleavage coupled with ligation could have produced increased complexities in RNA structure and function, as outlined by the accretion model for molecular evolution (Petrov et al., 2015; Manrubia and Briones, 2007). Increased ribozyme complexity as a response to changing environmental conditions, required the emergence of new three-dimensional structures from previously functional folds that carried out similar or different functions (Schmidt, 1999; Talini et al., 2011). The emergence of novel functional folds from pre-existing folds has been often observed for protein enzymes which unlike RNA, benefit from the existence of a more robust platform for analyzing evolutionary lineages (Curtis and Bartel, 2005; Meier and Özbek, 2007).

Contemporary ribozymes exhibit varying degrees of structural complexity in the organization of their global structure and stringency in nucleotide requirements at specific positions in their sequences. This is apparent on comparing three endonucleolytic ribozymes discussed previously in the context of their active site architectures (section 3.3.4), the hammerhead (HH), hairpin (HP) and Varkud satellite (VS). The hammerhead self-cleaving motif is the smallest at ~ 50 nt, while the full-length hairpin ribozyme can extend to ~100 nt (Lilley and Eckstein, 2008). The full-length VS ribozyme is ~200 nt long but a truncated version exhibiting uncompromised cleavage activity contains about 150 nucleotides (Lilley and Eckstein, 2008). While we lack any knowledge about the evolutionary origins of these ribozymes, the modular increase in complexity is discernible. The global structure of the hammerhead is organized by a single three-way junction, whereas the hairpin uses a four-way junction (Lilley and Eckstein, 2008). Seven independently folding subdomains of the VS ribozyme are assembled using three three-way junctions. The active site of the hammerhead is contained within an unpaired bulge at

its only three-way junction and requires minimal structural reorganization during RNA folding to assemble (Lilley and Eckstein, 2008).

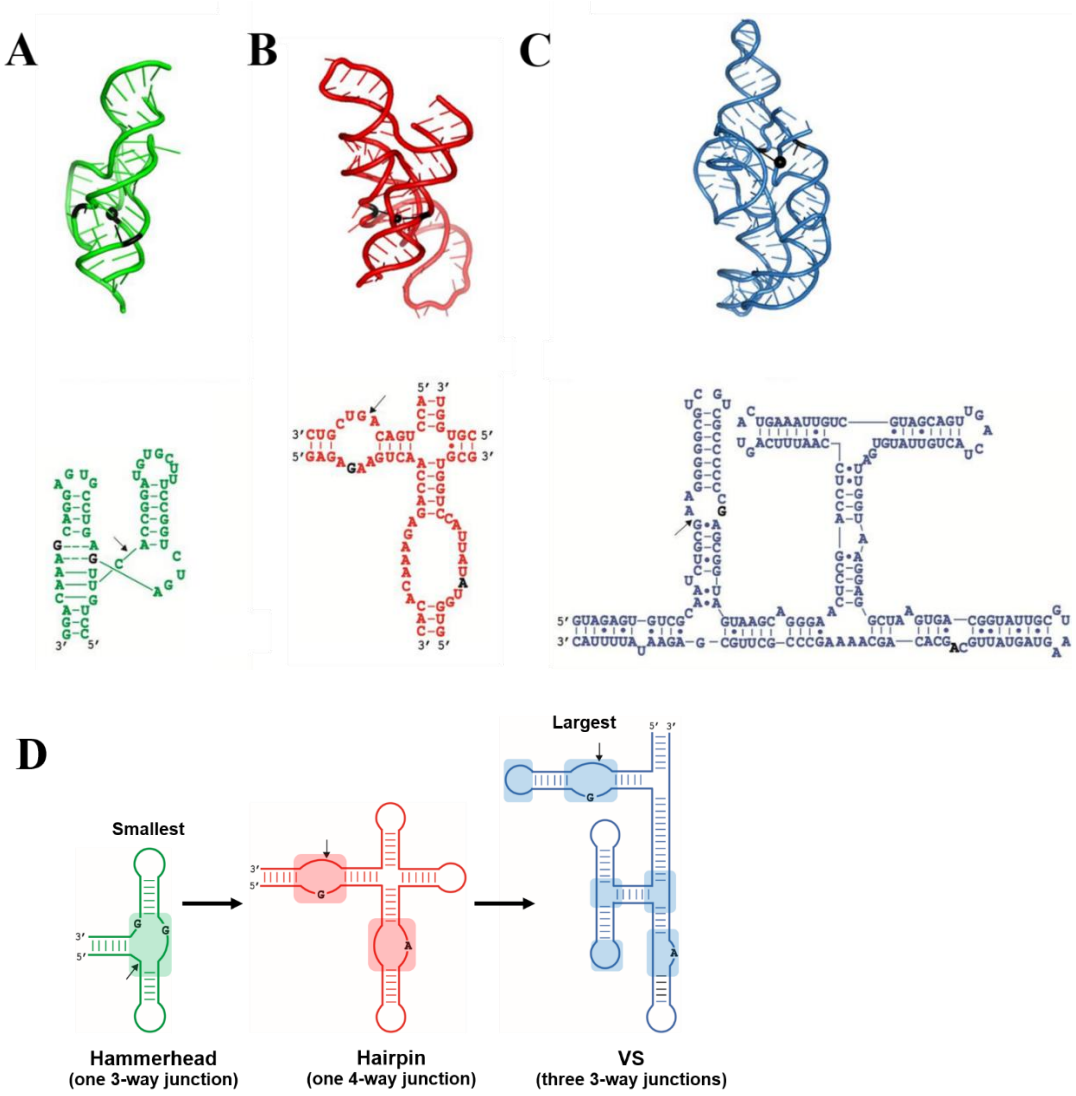

Figure 4.1 The hammerhead, hairpin and VS ribozymes. **A.** Hammerhead. **B.** Hairpin. **C.** VS. Catalytic nucleotides are shown in black and cleavage sites are indicated by black arrows. **D.** Increasing structural complexity from hammerhead to VS. Secondary structures are oriented in a similar topology for comparison. Catalytic nucleotides are shown in black and cleavage sites are indicated by black arrows. Highlighted regions represent conserved structural features in these ribozymes.

The hairpin and VS ribozymes assemble their active sites during substrate docking, which brings into close proximity, catalytic residues that are distant from each other in the secondary structures of the ribozymes (Lilley and Eckstein, 2008). For hairpin catalysis, this involves interactions between two internal loops that house the catalytic nucleobases (Rupert and Ferré-D'Amaré, 2001). In the VS ribozyme, additional tertiary interactions help the substrate dock into a catalytic cleft that brings together the two internal loops, thereby assembling the active site (Rastogi et al., 1996). Thus, structural strategies used to assemble the active sites vary in complexity: from a self-contained ribonuclease motif in the hammerhead to a modular separation of substrate and catalytic domains in the hairpin and VS, with the VS exhibiting added layers of complexity for substrate recognition and binding.

It is possible to estimate the structural complexity of these RNA sequences by representing each position as bits of information (Schneider et al., 1986). The information content required to define the secondary structure of each ribozyme is calculated as follows (Carothers, et al., 2004). For a Watson-Crick base-pair, information content is 2 bits, if additional base-pairs such as G-U and A-C are considered, information at that base-pair position reduces to 1 bit. Two bits of information content are assigned to an invariant position, while a nucleotide that can be either of two bases is assigned 1 bit of information. Unconstrained positions are considered to possess no information, hence assigned 0 bits. Information content in standard hammerhead sequences is about 40-55 bits; a hairpin ribozyme sequence holds about 50-65 bits of information. The VS ribozyme is about four times the size of the hammerhead and twice that of the hairpin and contains within its secondary structure about 110-130 bits of information, excluding the information contained in specific structural motifs that are largely independent of their nucleotide compositions but necessary for folding. This approach is over-simplistic because it does not consider the effect

of nucleotides at the unconstrained positions on RNA folding and does not take into consideration different degrees of stability offered to stems by base-pairs at different positions in the stem. However, it illustrates the principles underlying informational complexity in RNA.

Evolution is perhaps the most prominent example of increasing informational complexity. Studying evolution at the molecular level provides information about incremental changes in the genotype that are often unobservable at the organism phenotype level. Functional RNA embodies both genotype and phenotype. The same sequence of nucleotides (genotype) determines structural (three-dimensional shape) and functional (ligand-binding, catalysis, etc.) phenotypes in RNA, unlike proteins where the corresponding mRNA sequence is the genotype, but phenotype is represented by the properties of the protein molecule. The ease in correlating genotype and phenotype in a single system and the availability of fairly robust computational methods to determine minimum energy secondary structures of ribonucleotide sequences made RNA the system of choice for early excursions into the study of molecular evolution (Cowperthwaite and Meyers, 2007). This choice is relevant to the RNA World, where RNA likely served as both the genotype and phenotype (section 1.2.1), and also in general, as the gene is the smallest and most fundamental unit of evolution. The early 1990s witnessed enormous increase in computing power, resulting in the development of powerful programs for calculating the most probable secondary structure profiles for RNA sequences (Schuster et al., 1994; Huynen et al., 1996; Fontana and Schuster, 1998). Using minimum free energy (mfe) structures as phenotypes, these studies explored the variations of RNA secondary structure (distinguished by their mfes) with variations in its primary structure (sequence) (Cowperthwaite and Meyers, 2007). Some important results pertaining to molecular evolution were obtained using these computational approaches (section 4.1.4); however, with the rise of molecular biology and increasing knowledge of aptamers and

ribozymes, RNA phenotypes could include function. This made RNA more relevant as a model for molecular evolution (Pitt and Ferré-D'Amaré, 2010; Jiménez, et al., 2013).

4.1.2 Sequence space and fitness landscape

Evolution was long regarded as a product of random perturbations to the nucleotide sequence acted upon by local forces of natural selection. While this idea is useful in explaining the diversity of life, it does not offer a framework for theoretical analysis of molecular evolution. As described by Schultes et al., sequence space of a molecule that describes a set of all possible sequence combinations corresponding to a certain length was considered to be "a grab-bag of possibilities" (Schultes et al., 2009). In such a sequence space consisting of all possible sequences, evolved molecular structures were considered to be 'delicately adjusted' by natural selection. Consequently, most mutations to these evolved sequences would be catastrophic to the function of the molecule and perhaps to the organism as whole. Calculations of time estimates for evolutionary transformations in the framework of this paradigm revealed that the probability of identifying a functional sequence in a largely non-functional sequence space was extremely low, and the number of such relevant sequences were too few for meaningful sampling by random mutations. This is referred to as Salisbury's paradox, which is captured by his pithy statement, "if life really depends on each gene being as unique as it appears to be, then it is too unique to come into being by chance mutations." (Salisbury, 1969).

A possible solution to this paradox was provided by the 'Neutral Theory' of molecular evolution by Kimura (Kimura, 1968). Kimura demonstrated that the mutational frequencies to protein sequences are too high to be evolutionarily stable unless a large fraction of these mutations are 'neutral' to selective pressures. This initially heretical theory was expanded to the molecular realm by Smith with the introduction of the concept of sequence space as an organized matrix of

molecules where two neighboring sequences can be interconverted by a single point mutation (Smith, 1970). For an RNA space (RNA consists of 4 types of monomers), if *N* is the length of a sequence and *f* is the fraction of immediate neighbors that are as active (or almost as active) as the parent sequence, then as long as the product of *f* and (4-1) *N*, is greater than 1 (i.e. a sequence has an active neighbor at every step), functional RNA will form a network of meaningful sequences providing feasible paths for evolution by natural selection to occur. The word game used by Smith where a four-letter source word can be transformed into another word with an entirely different set of letters by single-letter changes illustrates this:

WORD-WOR*E*-*G*ORE-GO*N*E-**G*E*NE**.

This is an example of a neutral path. Since the English alphabet has 26 letters, there are many more letter combinations that are gibberish than are meaningful. This is similar to protein sequences that can be made of ~20 amino acid residues; there can be 4^{20} distinct 'four-letter words' in a polypeptide that has four residues. The corresponding RNA sequence space is much smaller with only 4^4 (256) distinct sequences. This makes RNA sequence spaces more tractable to analysis. The concept of a sequence space as outlined here is a platonic construct; not all sequences in a hypothetical space are realized in nature, but they do contain unique information, much like the number system. This concept of sequence space in the light of the neutral theory brings molecular evolution from the domain of 'lucky shots in the dark', to illuminated trajectories that underlie these shots. The modern concept of sequence space is a mathematical one that can be represented by a coordinate system and is subject to statistical correlations (Schultes et al., 2009).

Evolution by natural selection involves random mutational drifts across the vast expanse of sequence space directed by non-random selection pressure. This is expressed in terms of fitness, which can be correlated to a structural or functional phenotype. Fitness landscapes are

mathematical constructs developed to describe the relation between genotype and phenotype (Wright, 1932). Metaphorically, these describe the explorations of genotypes across the sequence surface (usually denoted by the x-y plane) in the direction of greater fitness, usually denoted by peaks on the surface (z-axis). Evolution, therefore, is an adaptive walk on a fitness landscape or diffusion across the landscape with a bias toward climbing fitness peaks (Figure 4.2A). Although these metaphors might imply teleology, peak-climbing is just a visual representation of adaptation to the environment through blind tinkering by the forces of natural selection, not a journey with a destined goal (Jacob, 1977). RNA fitness landscapes are represented in three-dimensional Euclidean space for convenience; however, in reality these are hyper-dimensional constructs, with the number of dimensions increasing with the number of nucleotides in the sequence. For example, the sequence space for an 85-mer has 255 dimensions and $>10^{51}$ sequence possibilities (Schultes, 2009). According to the neutral theory, evolutionary dynamics on a fitness landscape requires at least one functional variant in the immediate neighborhood of each sequence; however, the number of non-functional variants far outnumber the functional sequences (Smith, 1970). This can be caused by global misfolding of an active structure through sequence change, local disruption of important nucleotides leading to loss of function, or both, further lowering the number of functional sequences in the vicinity of a sequence. Intuitively, this suggests that fitness landscapes would be rugged with large regions of inactivity interspersed with active peaks. The driving force behind evolutionary processes is therefore the local fitness differential between planes and peaks in the fitness landscape.

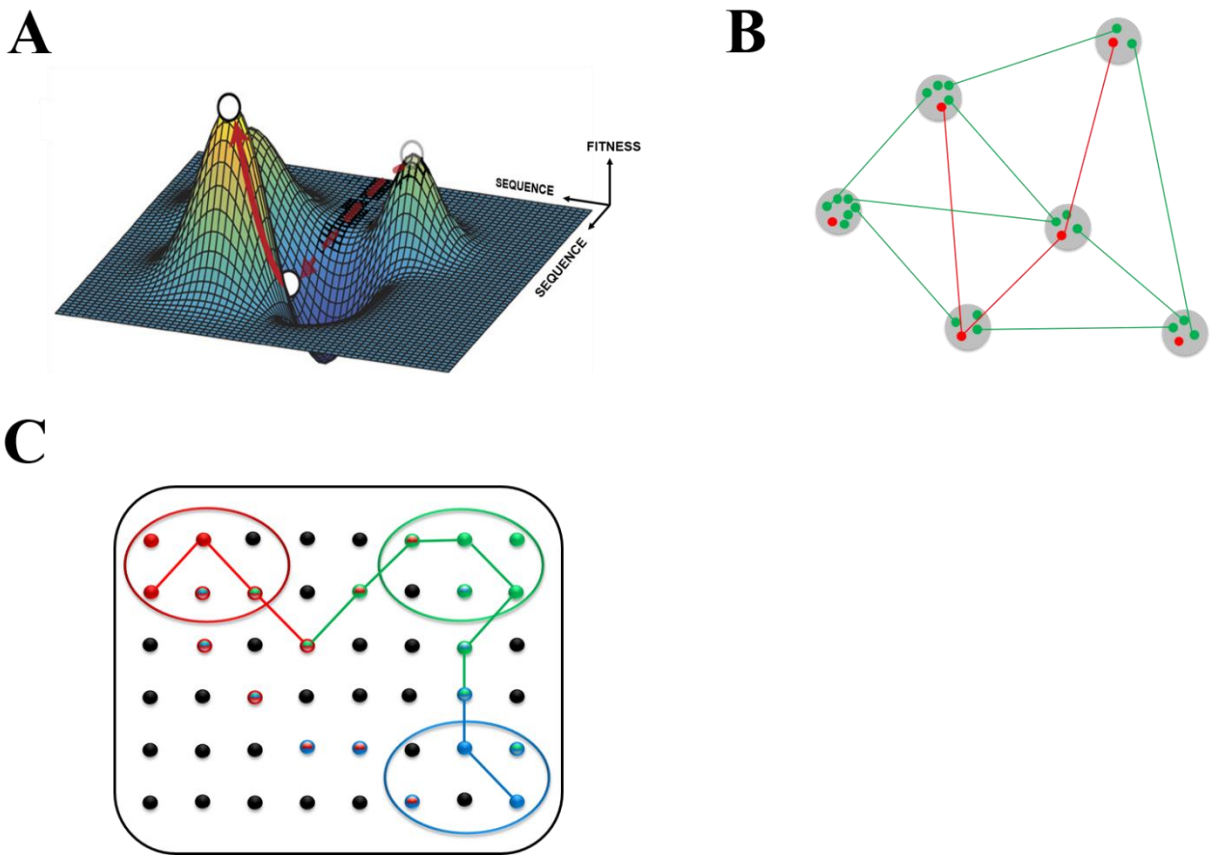

Figure 4.2 Sequence space and fitness landscape. A. A hypothetical fitness landscape. Peaks with local and global maxima represent sequences of high fitness. Evolutionary trajectories are shown by red arrows: dashed and bold arrows indicate journeys to local and global fitness peaks, respectively. (Adapted from Steinberg and Ostermeier, 2016). **B.** Part of an experimental fitness landscape of a GTP-binding aptamer. Clusters of sequences have the same functional phenotype. Red dots indicate the most active sequences in each cluster, green dots indicate sub-optimal sequences with the same function (Jiménez, et al., 2013). **C.** Two-dimensional representation of sequence space containing three primary phenotypes (red, blue and green dots). Clusters corresponding to these phenotypes are separated in sequence space by intervening inactive sequences (black dots), nevertheless distinct phenotypes are connected via sequences that exhibit phenotype plasticity (one genotype-two phenotypes). These bifunctional sequences are shown as dual-colored dots. Black lines indicate neutral paths.

4.1.3 Robustness and plasticity are central to intersecting neutral networks

A ‘peak’ in a fitness landscape is usually a population of related sequences that are separated by single nucleotide mutations (Jiménez et al., 2013, Cowperthwaite and Meyers, 2007). Individual peaks correspond to a specific phenotype which requires a ‘many to one’ mapping of genotype to phenotype at these isolated locations on the landscape (Figure 4.2B, C). The rugged topography of fitness landscapes is a direct result of ‘phenotypic robustness’ of RNA. Robustness can be defined as the invariance of phenotype in the face of environmental or mutational perturbation (Wagner, 2005, Wagner, 2012), the latter being more relevant to our discussion. Robustness is a fundamental concept in evolution as it explains the neutral theory from a molecular perspective. The construction of an RNA molecule involves only four types of nucleotides and since RNA structures largely consists of base-paired helices, these nucleotides vary co-dependently. Watson-Crick base-pairing rules dictate the presence of A or G opposite U and G opposite C, and A-U (or G-U) and G-C pairs are generally interchangeable within a base-paired helix. These features of RNA structure make its sequence space robust. In other words, a wide range of mutations are tolerated in its genotypic landscape as long as they preserve essential structural features of that phenotype. Phenotypic robustness is widely observed in functional RNAs, especially ribozymes, where the primary features dictating function are their secondary structure and the catalytic nucleotides. Most self-cleaving ribozymes with average lengths of 80-150 nucleotides, contain <10 conserved nucleotides that do not tolerate mutations. Robustness in RNA might have been selected by evolutionary forces early on in the RNA World, where RNA was the only functional molecule embodying as both genotype and phenotype. Ribozyme-based replication was most likely error-prone (Joyce, 2002), therefore the gradual accumulation of mutations across generations would have put a heavy error-burden on functional sequences.

Evolution of robust RNA sequences would have reduced the detrimental effects of mutations during replication/polymerization. Therefore, RNA sequences would evolve toward clustered regions of sequence space surrounding fitness peaks (Wagner, 2005, Wagner, 2012). Robustness allows neighboring sequences to possess the same phenotype as the parent sequence, thereby giving rise to 'islands of activity' in the fitness landscape surrounded by a 'sea of inactive sequences' (Schultes et al., 2009).

Since evolution requires acquisition of novel phenotypes, a central problem of molecular evolution is understanding the mechanism by which populations sample sequence space to encounter new phenotypes. Random mutations are the vehicles for this exploration; however, it is difficult for a population to access other isolated phenotype peaks usually located at a great distance and separated by inactive sequences (Cowperthwaite and Meyers, 2007; Jiménez et al., 2013). While robustness results in a smooth local landscape near a fitness peak, it produces a rugged landscape globally. This might suggest that robustness runs counter to evolutionary progress as it resists change. But evolution is a process that has a conservative requirement in addition to its progressive trajectory. During a population's search for new phenotypes, the existing, well-adapted phenotype must be preserved till it encounters a new fitness peak. Robustness plays a vital role in this regard. Since a peak is not a single sequence but a collection of sequences of similar fitness (a result of robustness), each of these sequences are able to diffuse across the landscape in search of new peaks (Huynen et al., 1996; Jiménez et al., 2013). Systems capable of supporting their primary function in multiple configurations that are accessible through mutations have greater flexibility and degrees of freedom to acquire other functions (Wagner, 2005).

Theoretical, computational and more recently experimental results suggest the existence of pathways connecting isolated fitness peaks that consist of active sequences differing from their immediate neighbor by a single point mutation (Huynen et al., 1996; Fontana and Schuster, 1998; Jiménez et al., 2013). Due to the hyper-dimensionality of sequence space, these paths are parts of networks that permeate the fitness landscape, providing connectivity between otherwise isolated fitness peaks (Schultes et al., 2009). This connection matrix is referred to as a neutral network. A neutral network can therefore be defined as a mutationally connected set of genotypes that produce the same phenotype (Cowperthwaite and Meyers, 2007). Since mutations do not alter the phenotype on a neutral network, how do sequences acquire new phenotypes? Early simulations of the evolution of RNA shapes with changes in sequence, predicted the existence of multiple points where two neutral networks intersect. Intersections allow genotypes to jump to a new phenotype network leading to an apparent phenotypic discontinuity (Schuster et al., 1994; Huynen et al., 1996; Fontana and Schuster, 1998). This phenomenon, referred to as 'punctuated equilibria' has been observed in the evolution of RNA (Huynen et al., 1996; Ancel and Fontana, 2000), proteins (Chan and Bornberg-Bern, 2002), digital organisms (Wilke et al., 2001), microorganisms (Burch and Chao, 1999) and at the organismic level as inferred from transition fossils. The success of evolutionary excursions across a fitness landscape is therefore dependent on the existence of sequences at which two neutral networks intersect. Since an intersection sequence lies on the neutral networks of two distinct phenotypes, it needs to satisfy fitness requirements for both. This requires the intersection sequence to embody a dual phenotype, which introduces the concept of plasticity to molecular evolution.

Functional RNA sequences are known to sample multiple conformations but usually only one of those conformations is active. This ability of RNA to adopt multiple folds is sometimes

referred to as conformational plasticity (Ancel and Fontana, 2000; Talini et al., 2011; Lau and Ferré-D'Amaré, 2016). Variations in conformation may include rapid changes in spatial orientation of certain segments of the structure (Marek et al., 2011), local changes in secondary structure (observed in the VS ribozyme substrate - chapter 3), or in extremely rare cases, extensive reorganization of secondary structures resulting in two distinct folds (Lau and Ferré-D'Amaré, 2016). In even rarer cases both folds can possess well-defined functions, leading to the emergence of an additional functional phenotype in an already functional genotype (Lau and Ferré-D'Amaré, 2016). Such a sequence exhibits both structural and functional plasticity. Functional RNA like endonucleolytic ribozymes usually need to fold into distinct tertiary structures to catalyze distinct reactions, therefore functional and structural plasticity are correlated. However, examples of catalytic promiscuity in the group I intron and other *in vitro* selected ribozymes involve minor changes in the active sites without reorganization of their global fold (Forconi and Herschlag, 2005). In the context of intersecting neutral networks, sequences at the points of intersection are required to exhibit both conformational and functional plasticity to be viable. Plasticity presents exciting opportunities for biomolecular structures to acquire novel features without abandoning original traits (Figure 4.2C), but this diversity comes at a cost of activity as each conformer populates a fraction of the total population.

In addition to genotype diversification through mutations, functional plasticity, producing multiple phenotypes from one genotype, provides additional raw material to drive evolution (Schultes and Bartel, 2000; Meier and Özbek, 2007). Robustness determines the existence of neutral networks and their extent of coverage over a fitness landscape and plasticity determines the points where these networks intersect. Simply put, robustness maintains the existing function

of a sequence thereby increasing the probability of it encountering new functions and plasticity allows functional cross-over, enabling molecular evolution.

4.1.4 Illustrations from computation and experiment

The concept of the evolutionary continuum requires a detailed understanding of the relation between genotype and phenotype. This necessitates the determination of fitness landscapes of real molecules. The first fitness landscapes were constructed with RNA structure as a proxy for fitness under the premise that secondary structure dictates three-dimensional fold which in turn determines function of the molecule. A simulation of RNA molecules undergoing cycles of mutation and replication provided the first glimpses of a sequence-structure map for a biologically-relevant molecule (Schuster et al., 1994; Huynen et al., 1996; Fontana and Schuster, 1998). These simulations provided early evidence for the neutral theory and demonstrated that robustness and plasticity were intrinsic to the success of any evolutionary process. The genotype-phenotype maps obtained from these simulations revealed that there are more sequences than there are distinct structures and certain structures that contain stable motifs like stem-loops occur more frequently. This degeneracy is evidence for robustness. These maps also revealed points of phenotypic discontinuity where sequences adopted new shapes without losing fitness after extended neutral drifts. These sequences were found to fold into both existing and new structures, highlighting the role of plasticity in evolutionary transitions (Huynen et al., 1996; Fontana and Schuster, 1998). Since secondary structure was the only metric for fitness in these computational studies, networks connecting genotypes with similar fitness (neutral networks) were numerous (Huynen et al., 1996). However, secondary structure only provides the scaffold for activity in functional RNAs like aptamers and ribozymes. Features like local misfolding and structures of the active site and tertiary

interactions determine much of their function, therefore it is expected that neutral connections between fitness peaks would not be as extensive when fitness is measured not by shape but by biochemical function.

With the advent of high-throughput sequencing techniques, it was possible to 'read' sequence libraries that cover the entirety or a major portion of a molecule's sequence space (Jiménez et al., 2013; Pitt and Ferré-D'Amaré, 2010). This major technical advance allowed the determination of experimental fitness landscapes of functional RNAs for the first time. These studies involved two *in vitro* selected molecules – the class II RNA ligase ribozyme (54 nt) (Pitt and Ferré-D'Amaré, 2010) and a GTP-binding aptamer (24 nt) (Jiménez et al., 2013). The success of these experiments also highlights the convenience of using RNA as a model for studying molecular evolution. While it was possible to cover the entire sequence space of a 24 nt RNA aptamer (Jiménez et al., 2013), a corresponding experiment to study fitness at the organism level, even for the smallest possible genome, would require synthesis of at least 10^{148} sequences (Athvale et al., 2016) (note: there are 10^{86} elementary particles in the observable universe). Calculating fitness landscapes for proteins, though not as absurdly difficult, is still largely prohibitive considering that the complete coverage of an 18-mer polypeptide would require about a kilogram of material (Athvale et al., 2016).

Results from these two experiments were consistent with the outcomes of simulations on the evolution of RNA secondary structures. Fitness landscapes for a single function were found to contain isolated peaks separated by inactive sequences, however, these active clusters contained multiple sequences of similar fitness that could drift to distant locations on the landscape using neutral networks connecting two peaks (Pitt and Ferré-D'Amaré, 2010; Jiménez et al., 2013). The existence of multiple isolated peaks for the same function suggests that sequences showing great

variations and possibly adopting distinct structures could carry out similar functions. The existence of ribozymes with distinct secondary structures but identical function has been reported, where two classes of kinasing ribozymes were evolved from a parent aminoacylating ribozyme (Curtis and Bartel, 2005). While these studies have concentrated on a single RNA, a landmark study by Schultes and Bartel demonstrated that two distinct catalytic motifs, a natural ribonuclease ribozyme (HDV ribozyme) and an artificial ligase ribozyme could be interconverted by single or double mutational steps (Schultes and Bartel, 2000). This study demonstrated that the neutral networks (or at least one neutral path for each network) in RNA approached each other closely and intersected at a sequence that supported both cleavage and ligation reactions (albeit at very low rates in this case). The intersection sequence could be threaded through the secondary structures of both ribozymes and likely adopted two distinct catalytically active folds. This presents an extreme case of plasticity, never observed before, but is essential for successful acquisition of new function through walks along neutral networks. Exclusive nuclease and ligase phenotypes were accessible through mutations to the intersection sequence, although this initial transition was not smooth due to the low activity of the intersection sequence. Subsequent mutational drifts provided smooth access to wild-type nuclease and ligase sequences underscoring the importance of intersecting neutral networks on the acquisition of catalytic function.

Bifunctional sequences have been observed during *in vitro* selection of other functional RNAs. Some of them adopt distinct secondary structures while others retain their global folds but accommodate multiple reactions in their active sites. Examples of the latter include an aptamer sequence that bind both GMP and L-arginine (Connell and Yarus, 1994), a ligase that also catalyzes a cleavage reaction at a second active site in the same fold (Landweber and Pokrovskaya, 1999) and the natural Tetrahymena group I intron ribozyme that catalyzes the hydrolysis of

aminoacyl esters (Piccirilli et al., 1992) and nucleophilic attack on phosphate monoesters (Zaug and Cech, 1986), in addition to self-splicing. Examples of a single RNA sequence adopting two distinct secondary structures to carry out distinct functions were restricted to *in vitro* selected sequences and include an aptamer selected against the GluR2 AMPA receptor that adopts two thermodynamically inconvertible alternative structures (Huang et al., 2009). Bifunctional sequences adopting two distinct structures are presumably central to the emergence of new function in the course *in vitro* selection experiments. The emergence of structurally dissimilar kinase ribozymes from a parent aminoacylase ribozyme involved virtually neutral point mutations (Curtis and Bartel, 2005). An FAD-binding aptamer could be converted to a sequence that binds GMP exclusively via a few mutational steps in which each variant retains ligand-binding capabilities (Held et al., 2003). FAD-binding properties were gradually lost as newer genotypes gained structural features that favored GMP-binding. Intermediate sequences in these neutral paths appeared to bind both FAD and GMP ligands, clearly illustrating the principles of robustness, neutrality and plasticity in the evolution of a new function. A ^{6S}G synthase ribozyme could be obtained from a ^{4S}U synthase ribozyme using a similar approach (Lau and Unrau, 2004).

The diversity of protein folds are most likely results of modifications to pre-existing functional folds rather than products of *de novo* evolution. Cysteine-rich domains of *Hydra* nematocyst wall proteins assume two alternate folds linked by a bridge-sequence that adopts both folds (Meier and Özbek, 2007). Enzymes like sheep bowfly carboxyesterase can be converted to an organophosphate hydrolase by a single point mutation (Newcomb et al., 1997). Human serum paraoxonase (PON1) that primarily hydrolyses lactones could be evolved in lab to hydrolase synthetic organophosphates (Aharoni et al., 2005). Due to the greater diversity of residues in proteins compared to RNA, their sequence space is expected to be more extensive, and fitness

landscapes encompassing fractions of this sequence space are expected to be more rugged. In the absence of experimental data, we do not have a clear understanding of protein fitness landscapes and the structure of neutral networks pervading them.

Computational and experimental efforts over the last two decades have illuminated important principles underlying molecular evolution. However, the evolutionary paths that connect biologically-relevant functional RNA have not been outlined. For example, our understanding of the possible mechanisms for the emergence of endonucleolytic ribozyme activity and the ease of access to distinct catalytic motifs involved in RNA cleavage is limited. Since different classes of endonucleolytic ribozymes catalyze site-specific RNA cleavage on different substrates, it is reasonable to expect that their neutral networks might approach each other closely; however, the possibility that the neutral networks of naturally occurring endonucleolytic ribozymes might intersect remains untested. It is also unclear how larger and more complex scaffolds could have emerged in these ribozymes presumably from simpler self-cleaving motifs. These questions have profound significance for the appearance of new RNA folds and the ease with which catalytic motifs can interconvert or be accessed by divergence from a common ancestor.

We have considered three endonucleolytic ribozymes, the hammerhead, hairpin and VS for our studies. These ribozymes possess varying degrees of complexity (section 4.1.1), yet present similar catalytic strategies for RNA cleavage (section 3.3.4). Investigations into evolutionary connections between these ribozymes warrant the elucidation of their neutral networks. Nodes connecting these distinct phenotypes would correspond to the points of intersection in their respective neutral networks. Plasticity at these intersections enable functional cross-overs, thereby providing smooth access to different catalytic functions. Therefore, we sought to identify sequences that lie at the intersections of these ribozymes taken in pairs. The approach taken was a

hybrid of rational-design, computational assessment of designed secondary structures and experimental validation for catalytic function. The process is detailed in the next section.

4.2 RESULTS

4.2.1 Identification of minimal versions of the hairpin and VS ribozymes

The difference in sizes of wild-type hammerhead (HH) (~50 nt), hairpin (HP) (~100) and VS (~200 nt) sequences makes designing bifunctional intersection sequences for each pair difficult. We decided to obtain shorter versions of the hairpin and VS ribozymes that would be comparable in length to the hammerhead ribozyme. Visual inspection of the secondary structures of the hairpin and VS ribozymes suggests two operations to simplify acquisition of this intersection sequence: (1) contraction of the ribozymes by deletion of nonessential, peripheral elements to create VS and hairpin ribozyme constructs of similar size (Figure 4.3A, B) and (2) circular permutations to arrange the active site elements (cleavage site dinucleotide, catalytic G, and catalytic A) in the same sequence order (Figure 4.3C, D) Justification for the former operation comes from analysis of the primary, secondary, and tertiary structures of ribosomal RNA from archaea, bacteria, and eukarya, which reveal evolutionary expansion of rRNA through recent addition of helical branches (junctions) to ancestral RNA trunks while maintaining structural integrity of the ancestral core (Petrov et al., 2015). Smaller ribozymes like the group I intron are thought to have acquired additional domains for robust folding (Lilley and Eckstein, 2008). From this perspective, three-way junction, J_{345} in the VS ribozyme could have emerged by expansion of an ancestral VS ribozyme structure through insertion of helices 4 and 5 for better folding/substrate binding (Figure 4.3B) (Rastogi et al., 1996). Helix 7 and J_{127}, dispensable for cleavage, is most likely important for regulating cleavage-ligation and cis-trans docking equilibria (Jones et al.,

2001). This helix could have similarly emerged by nucleotide accretion (Figure 4.3B). Further application of this principle eliminated nonessential sequences from the four-way junction (helices C and D) in the hairpin ribozyme (Fedor, 2000) (Figure 4.3A) yielding hypothetical ancestral RNAs of similar topologies. These operations were guided by reported sequences of hairpin (Figure 4.4A, B) (Sarguel et al., 1995; Butcher et al., 1995; Komatsu et al., 1995; Komatsu et al., 1996) and VS ribozymes (Figure 4.4C) (Sood and Collins, 2002). Biochemical and biophysical studies, phylogenetic analysis, secondary structure analysis, and *in vitro* selection support a minimal consensus sequence for the hairpin ribozyme ((Sarguel et al., 1995; Butcher et al., 1995; Komatsu et al., 1995; Komatsu et al., 1996). For the VS ribozyme, replacing J_{345} and helices 4 and 5 with a GAAA tetraloop generates a minimal construct, Δ345 (Figure 4.4C) that catalyzes the site-specific cleavage of its substrate three orders of magnitude faster than the corresponding uncatalyzed reaction (Sood and Collins, 2002). Reducing the length of the linker sequence and deleting the base-paired helix under the internal cleavage loop in Δ345 (Δ345_short) do not affect cleavage (Figures 4.4C, 4.5A, B). The reduction in cleavage activity compared to a wild-type sequence is in part due to weakened substrate binding in the absence of stem-loop 5 that forms a kissing loop interaction with the substrate hairpin to assist in its docking (Figure 4.3B) (Rastogi et al., 1996; Anderson and Collins, 2001; Zamel and Collins, 2002). In addition to catalyzing site-specific cleavage of hairpin and VS substrates exclusively, activity of these hairpin and VS minimal constructs depend on the respective catalytic nucleotides (Lilley and Eckstein, 2008) (Figure 4.5C, G).

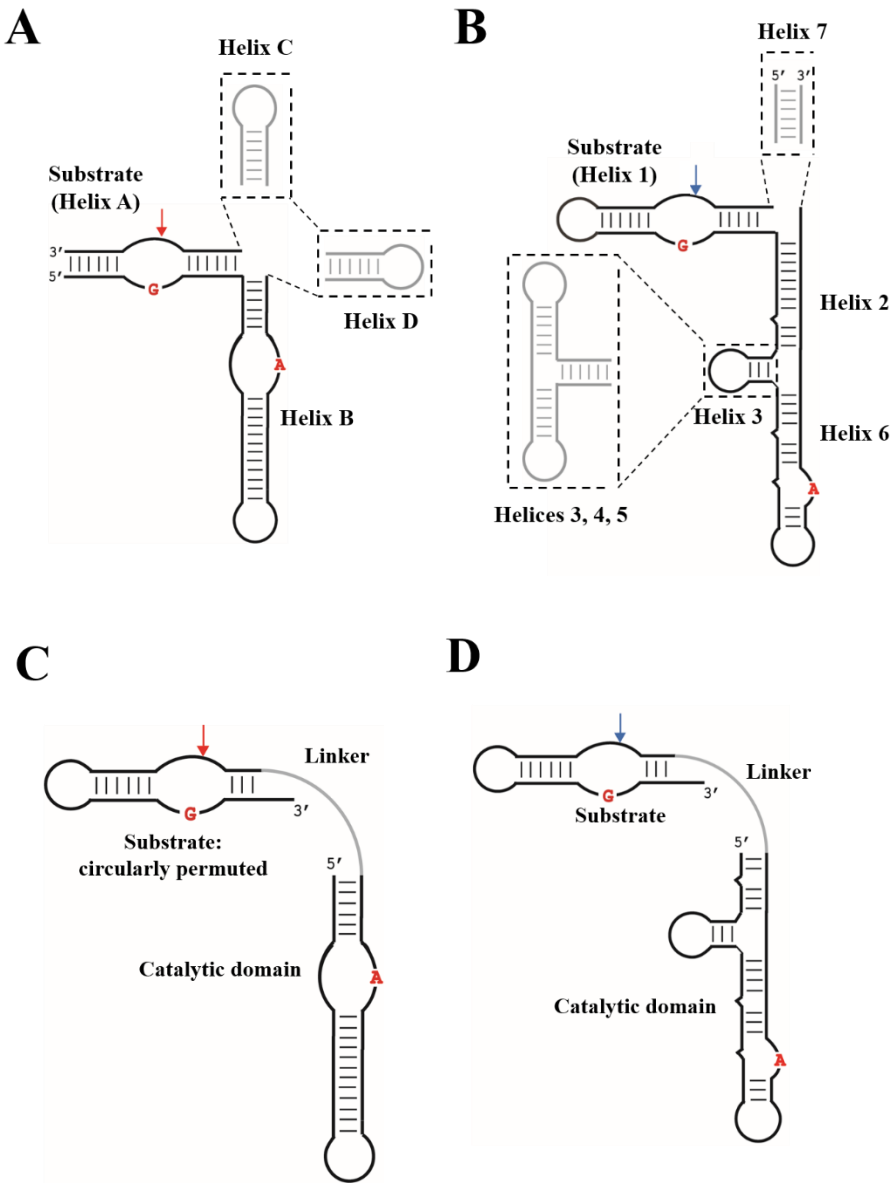

Figure 4.3 Operations to simplify identification of an intersection sequence between VS and hairpin ribozymes. **A.** Deletion of non-essential helices C and D in the wild-type hairpin ribozyme. Cleavage site is indicated by an arrow, and catalytic nucleotides shown in red. **B.** Substitution of helices 3-5 and junction 3-4-5 by a short stem-loop, and deletion of helix 7 and junction 1-2-7 in the wild-type VS ribozyme. Coloring scheme as in A. **C.** Circular permutation of the hairpin ribozyme to match the topology of the VS secondary structure and link substrate and catalytic domains. Cleavage site is indicated by an arrow, and catalytic nucleotides shown in red. **D.** Circular permutation of the VS ribozyme to link substrate and catalytic domains. Cleavage site is indicated by an arrow, and catalytic nucleotides are shown in red.

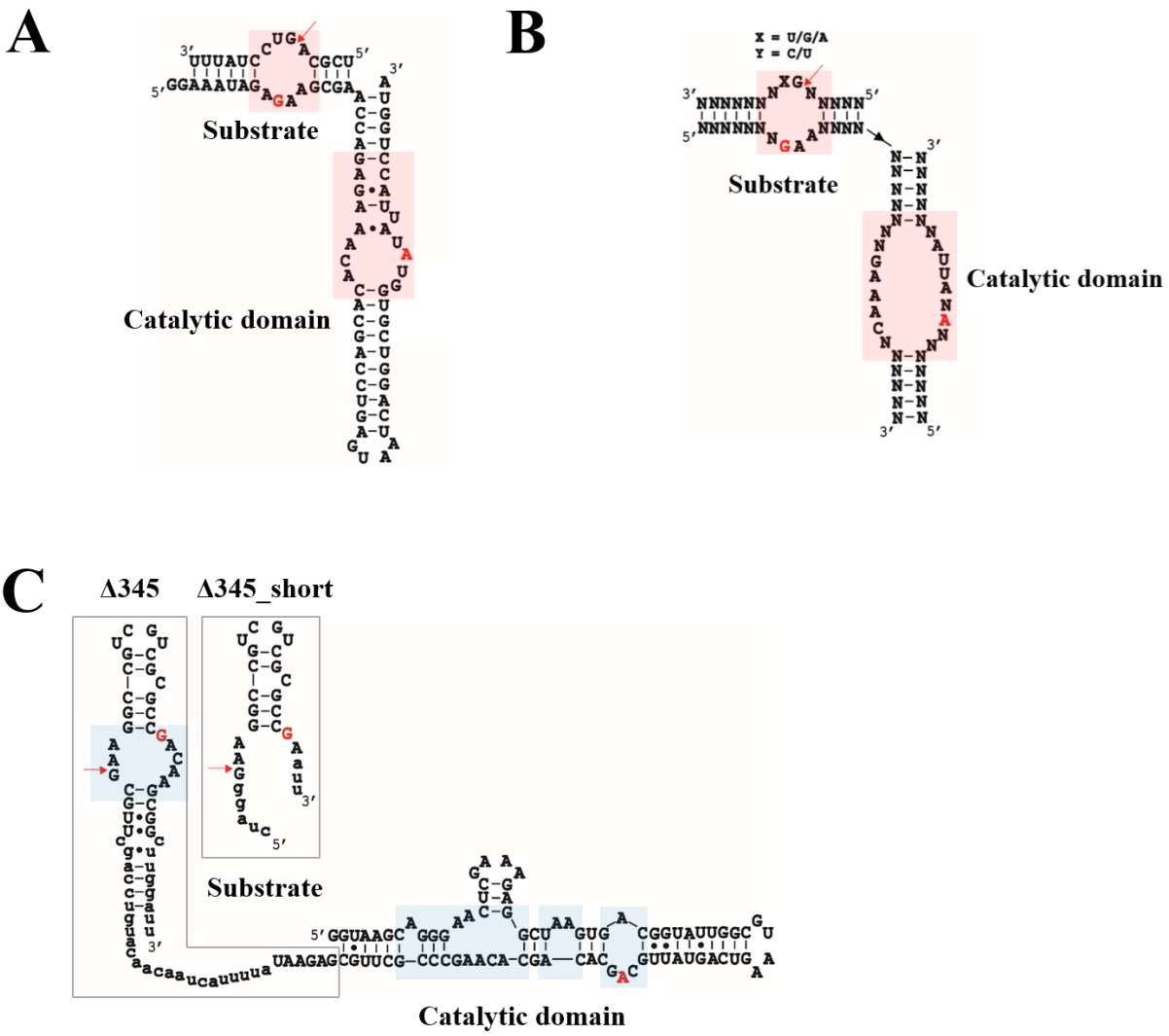

Figure 4.4 Secondary structures of prototype ribozyme sequences used as starting points for rational design of the intersection sequence and consensus sequence of a minimal version of the hairpin ribozyme. A. Secondary structure of a minimal hairpin ribozyme sequence used for rationally designing bifunctional sequences (Butcher et al., 1995). **B.** Minimal consensus sequence for the hairpin ribozyme (Fedor, 2000). **C.** Secondary structure of a minimal VS ribozyme sequence used for rationally designing bifunctional sequences (Sood and Collins, 2002). Cleavage site is indicated by an arrow, and catalytic nucleotides are shown in red. Conserved secondary structure features are highlighted by colored boxes (red for hairpin and blue for VS).

We took a modular approach to design derivatives of these prototype ribozymes. Since catalytic domains of both wild-type ribozymes are capable of cleaving their respective substrate sequences *in trans*, we decided to test the cleavage activities of the catalytic domains of these prototype ribozymes by physically separating them from their substrate domains. Despite cleaving substrates *in trans* (Figure 4.5F), the wild-type hairpin and VS ribozymes differ in the nature of tertiary interactions that help in substrate binding. The hairpin ribozyme binds the substrate through strand annealing, where the substrate strand, forms Watson-Crick base-pairs with the complimentary strand (Fedor, 2000). In addition to this secondary interaction, tertiary interactions between the internal bulges of the substrate and helix B assist in substrate docking (Rupert and Ferré-D'Amaré, 2001). Substrate binding in the VS ribozyme occurs through tertiary interactions between the catalytic domain and substrate that includes a kissing-loop interaction between the hairpins of the substrate stem-loop and stem-loop 5 (Rastogi et al., 1996).

The trans construct derived from the prototypical hairpin ribozyme, HPt cannot bind the substrate through strand annealing due to the absence of a strand overhang in the catalytic domain and its complimentary strand in the substrate. Similar to the hairpin, the absence of helices 4 and 5, and junction 3-4-5 in the VS trans construct, Δ345 precludes the kissing-loop interaction, thus weakening interactions that stabilize substrate docking. These suboptimal features of the trans constructs likely result in slow trans cleavage (Figure 4.5D, F). Interestingly, trans reaction between a VS substrate and the entire ribozyme construct, Δ345 (catalytic domain + substrate) was unsuccessful (Figure 4.5E). This could be due to competitive cis docking of the substrate attached to the catalytic domain (Pereira et al., 2008).

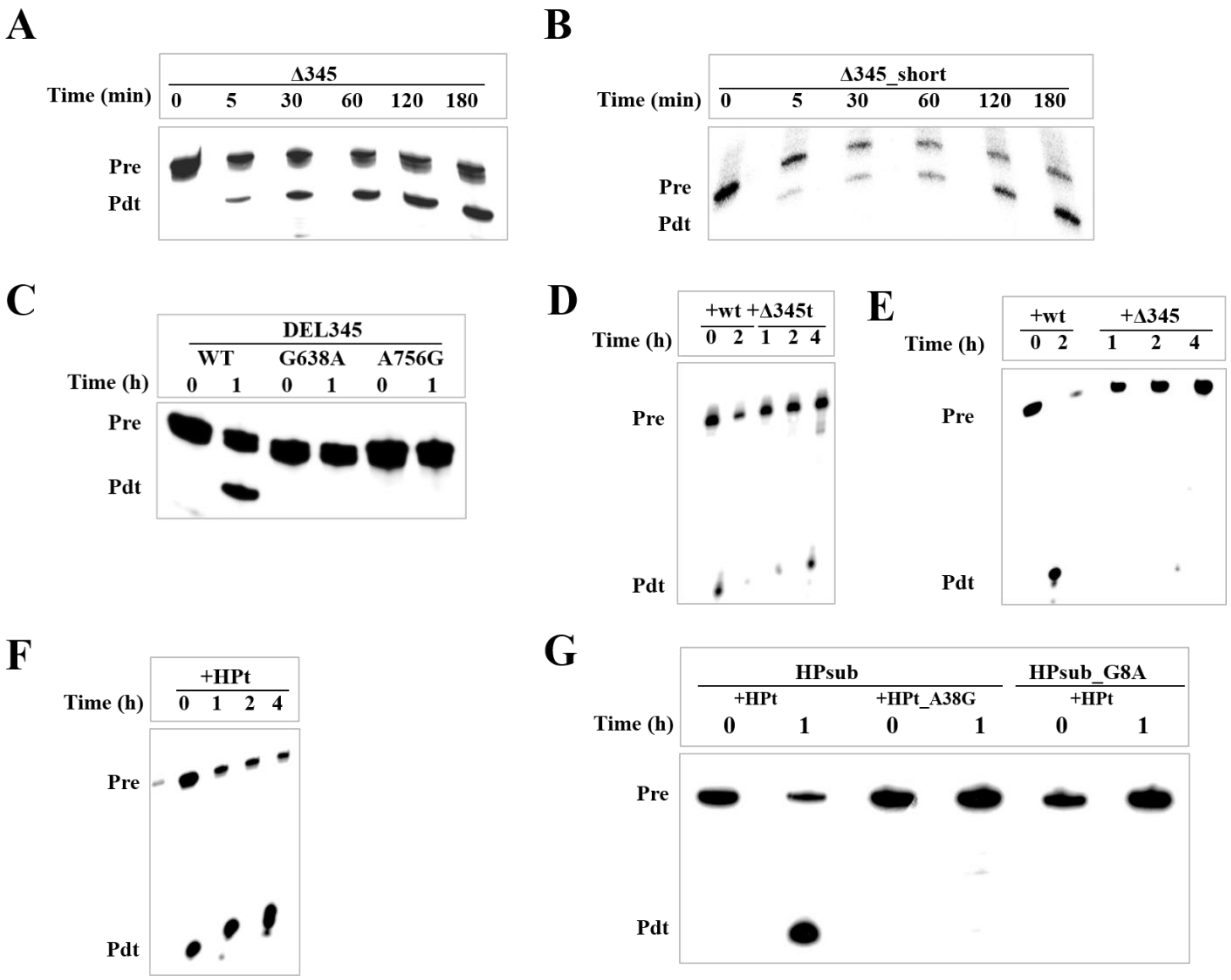

Figure 4.5 Cleavage activities of prototype VS and hairpin ribozymes *in cis* and *in trans*. **A.** Cis cleavage of Δ345. **B.** Cis cleavage of Δ345_short. **C.** Mutating the catalytic guanine to adenine (G638A) and adenine to guanine (A756G) abolishes cleavage in Δ345. **D.** The catalytic domain (Δ345t) of the prototype VS ribozyme (Δ345) catalyzes cleavage of a VS substrate *in trans*. **E.** The intact prototype VS ribozyme (Δ345) is inactive toward trans cleavage. wt indicates wild-type VS ribozyme used as a positive control. **F.** The catalytic domain (HPt) of the prototype hairpin ribozyme catalyzes cleavage of a hairpin substrate *in trans*. **G.** Mutating the catalytic guanine to adenine (G8A) in the substrate and adenine to guanine (A756G) in the catalytic domain abolishes cleavage in the hairpin ribozyme.

With an active trans VS ribozyme in hand, we identified nonessential sequences by sequential truncations. Three sets of truncations (trunc1-3) consisting of the deletion of three base-pairs for each set were performed on the trans construct, Δ345t (Figure 4.6). Truncating helix 2 by three base-pairs (trunc1) and helix 6 by three base-pairs at a position proximal to the catalytically important A730 loop (trunc2) reduce both cis and trans cleavage. Deletion of three base-pairs distant from the A730 loop (trunc3), enhances cleavage compared to full-length constructs, Δ345 and Δ345t. This is consistent with similar truncation studies on the wild-type sequence (Lafontaine et al., 2001).

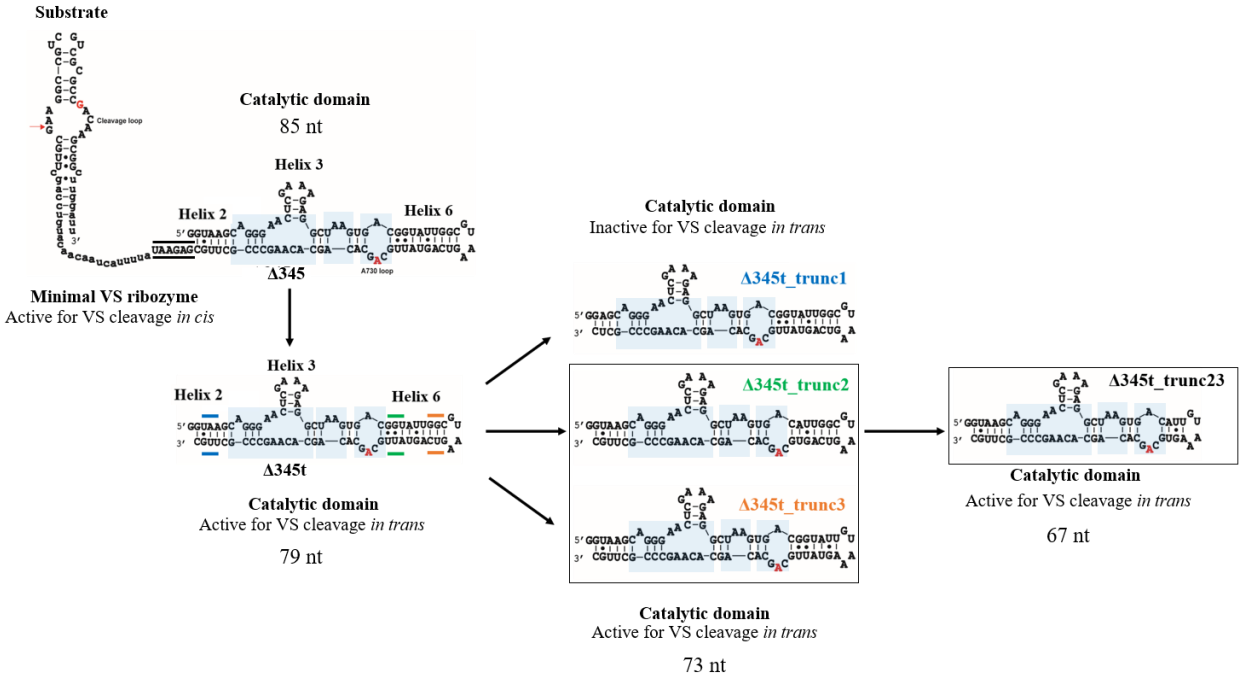

Figure 4.6 Truncation studies on prototype VS ribozyme, Δ345 *in trans*. Three base-pair truncations, trunc1, 2, 3 are indicated by colored bars. Black bar indicates single stranded region of the minimal VS ribozyme that was deleted in the trans construct, Δ345t. Light blue boxes highlight conserved secondary structure features in the VS catalytic domain. Active sequences are shown in boxes,

4.2.2 Designing a bifunctional sequence that cleaves both hairpin and VS substrates *in trans*

Truncation constructs of Δ345t were considered as starting points for the rational design of a bifunctional sequence that supports hairpin and VS cleavage *in trans*. Informed by the secondary structure features of these VS constructs and the consensus sequence for the hairpin ribozyme (Figure 4.4B), we designed sequences that satisfied the following requirements:

1. It should be possible to thread the sequence through secondary structures of both hairpin and VS prototype ribozymes.
2. The intersection sequence must be catalytically active for both ribozyme substrates. Stringency imposed by conformational plasticity, lowers the fraction of molecules populating a certain functional fold and results in reduced activity corresponding to each ribozyme. Nevertheless, the rates of cleavage for either reaction must be sufficiently above that of uncatalyzed RNA cleavage.

The first requirement is thermodynamic. The sequences must fold into two distinct structures of comparable thermodynamic stabilities to be considered a 'bifunctional sequence'. Although the primary requirement is dual catalytic activity (requirement 2), ribozyme function can be usually directly correlated to its three-dimensional structure (Marek et al., 2011). This makes requirement 1 important for initial screening. Due to the absence of folding algorithms that predict tertiary structure, we have used secondary structure as a proxy for global fold. Sequences were initially screened *in silico* using mFOLD (Zuker, 2003) by evaluating their stabilities toward the hairpin and VS folds. mFOLD-predicted secondary structures of the catalytic domains of prototype ribozymes were used as references for fold assignments. Sequences that were predicted to misfold or predominantly adopt only one of the two ribozyme structures were discarded, and those with

comparable stabilities for both ribozyme structures were synthesized by *in vitro* transcription and tested for trans cleavage (Figure 4.7).

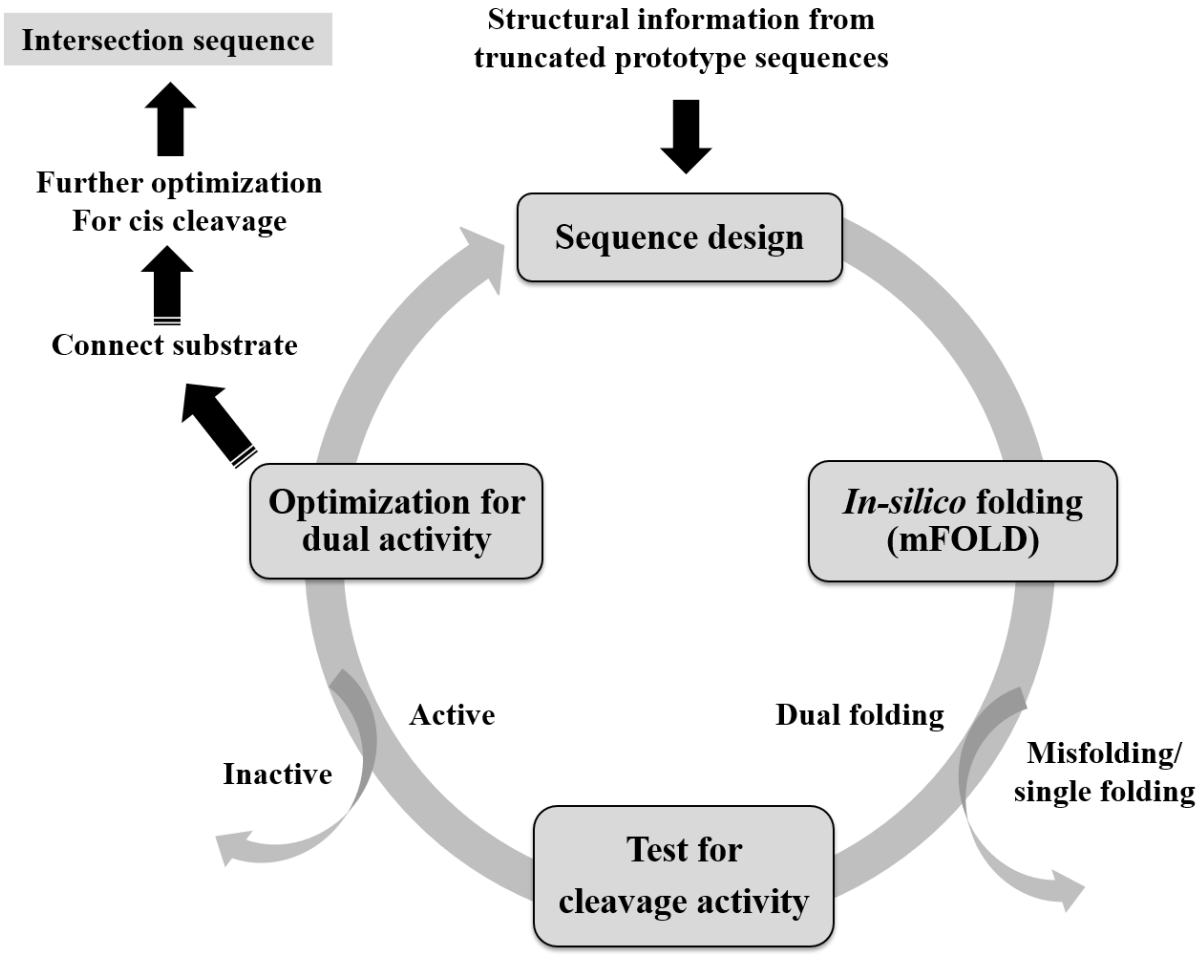

Figure 4.7 Flow diagram for the rational design of an intersection sequence. The process begins with designing sequences that can potentially be threaded through the secondary structures corresponding to the catalytic domains of both hairpin and VS ribozymes. These sequences are screened by *in silico* folding algorithms. Selected sequences are tested for *in vitro* cleavage. Sequences active for cleavage were optimized to impart dual activity. These trans-active catalytic domain sequences were fused to hairpin and VS substrates and tested for cis cleavage. Further optimization leads to a bifunctional intersection sequence for cis cleavage.

Along similar lines, careful inspection of the prototype VS substrate sequence (Figure 4.8A) and the prototype and consensus sequences of the hairpin substrate (Figure 4.8C, D), which consists of a conserved internal bulge (referred to as the internal cleavage loop as it contains the cleavage site) flanked by base-paired helices, enabled us to design substrates for the VS and hairpin ribozymes that closely resembled each other. These will be referred to as VSsub and HPsub, respectively (Figure 4.8B, E). These substrates differed only in the sequences of their conserved internal bulges. Both hairpin and VS substrates obtained from rational design were cleaved by corresponding wild-type ribozymes, but not by the other (Figure 4.9).

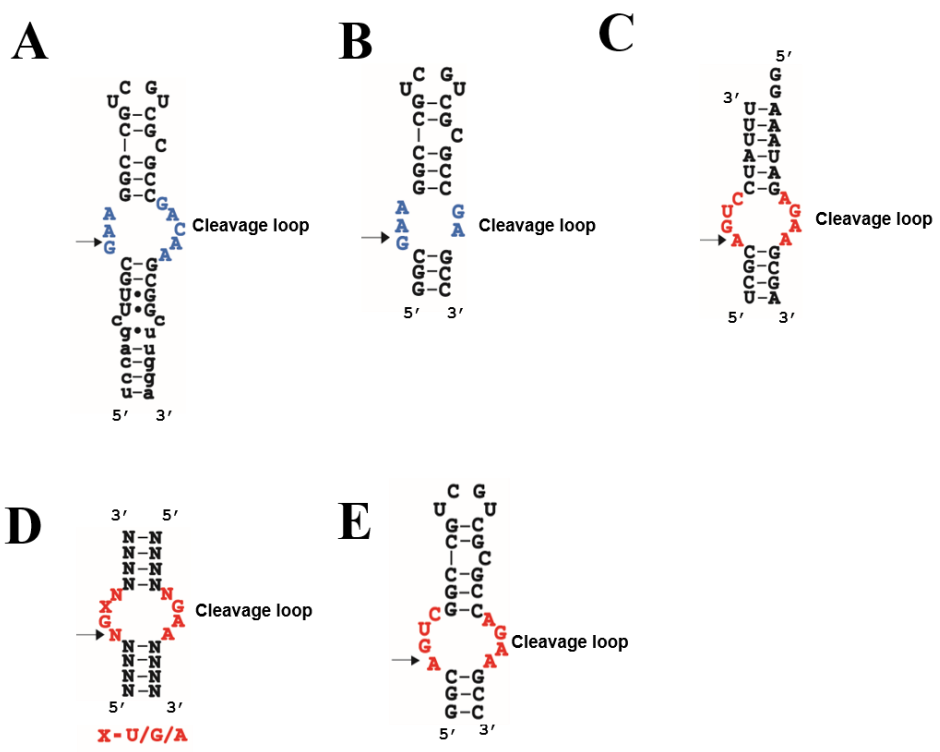

Figure 4.8 Substrate sequences from prototype VS and hairpin ribozymes and substrates used in this work. A. Substrate sequence from prototype VS ribozyme, Δ345 (figure 4.4C). Cleavage loop nucleotides shown in blue. **B.** VS substrate used in this work (see construction of cis-ribozymes in chapter 5). Cleavage loop nucleotides shown in blue. **C.** Substrate sequence from prototype hairpin ribozyme (figure 4.4A). Cleavage loop nucleotides shown in red. **D.** Consensus sequence for hairpin substrate (figure 4.4B). Cleavage loop nucleotides shown in red. **E.** Hairpin substrate used in this work (see construction of cis-ribozymes in chapter 5). Cleavage loop nucleotides shown in red.

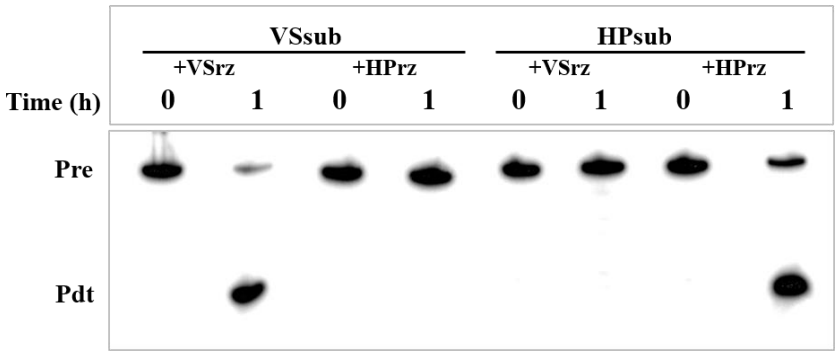

Figure 4.9 Confirming cleavage specificities of VS and hairpin substrates used in this work. VS substrate, VSsub is cleaved by wild-type VS ribozyme (VSrz) but not by wild-type hairpin ribozyme (HPrz). Hairpin substrate, HPsub is cleaved by wild-type HP ribozyme (HPrz) but not by wild-type VS ribozyme (VSrz).

This workflow yielded several sequences that could be approximately threaded through secondary structures of both hairpin and VS ribozyme catalytic domains. However, most sequences folded into more stable hairpin structures (as revealed by mFOLD) and consequently cleaved only hairpin substrates. This could be attributed to a greater number of Watson-Crick base-pairs in the hairpin folds compared to that in the corresponding VS folds. Sequences derived from Δ345t_trunc123 (trans_123) and Δ345t_trunc23 (trans_23 and trans_23mut6) demonstrated trans activity toward the rationally designed hairpin substrate (Figures 4.10A-C, 4.11A, C), but was inactive toward the corresponding VS substrate (Figures 4.10A-C, 4.11B, D). The construct trans_23mut6 differs from trans_23 by six point mutations that preferentially destabilize the hairpin-fold, without affecting the stability of its corresponding VS-fold (Figure 4.10B, C). Construct trans_3 (Figure 4.10D), the design of which was based on Δ345t_trunc3, showed overall reduced activity toward HPsub (Figure 4.11E). However, in addition to cleaving HPsub, a variant of this sequence (trans_3mut6) (Figure 4.10E) also cleaved VSsub, albeit rather inefficiently

(Figure 4.11E, F). Trans_3mut6 contains the same mutations as trans_23mut6. Thus, trans_3mut6 was identified as a single sequence that could cleave both hairpin and VS substrates *in trans*.

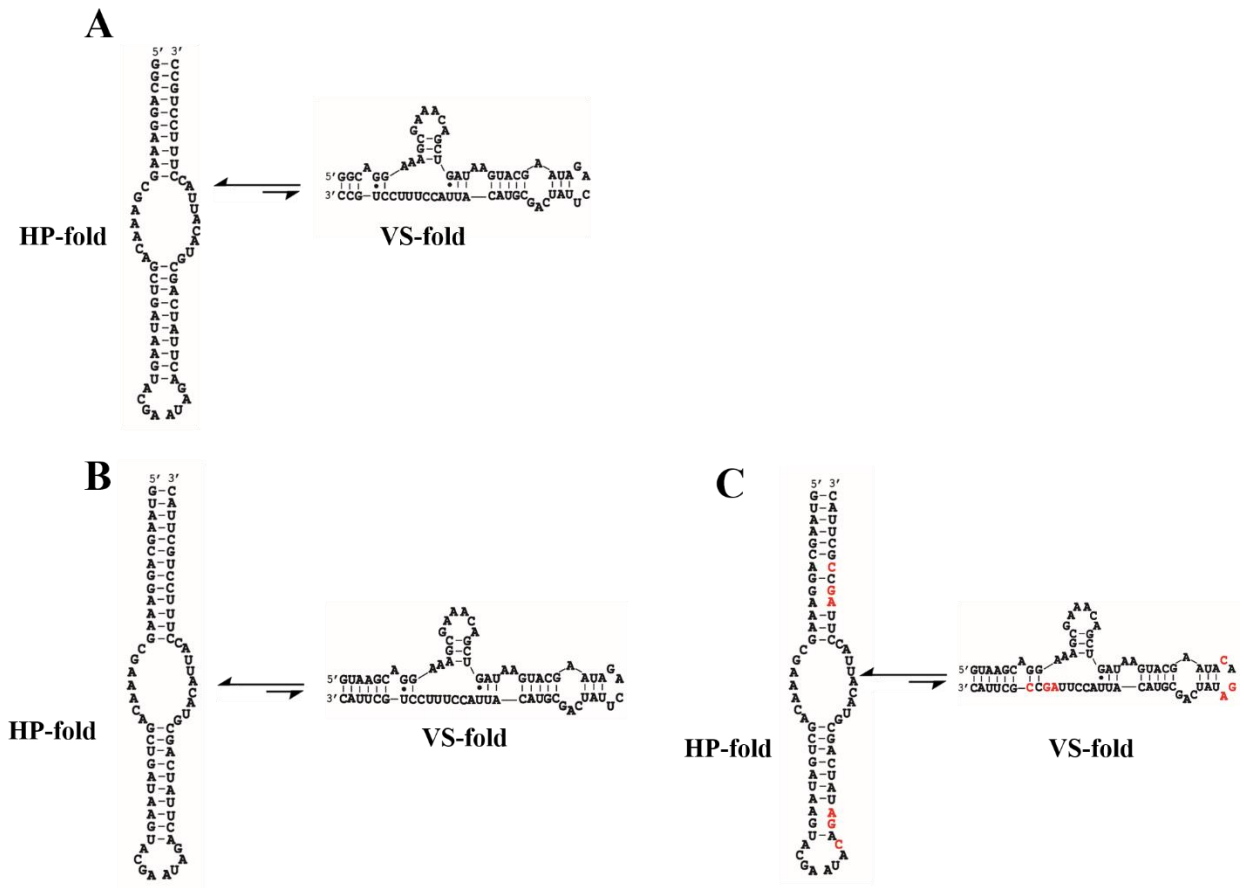

Figure 4.10 Toward a bifunctional catalytic domain sequence. A. Trans_123: putative hairpin and VS folds. **B.** Trans_23: putative hairpin and VS folds. **C.** Trans_23mut6: putative hairpin and VS folds. This sequence is obtained by mutations (in red) to trans_23.

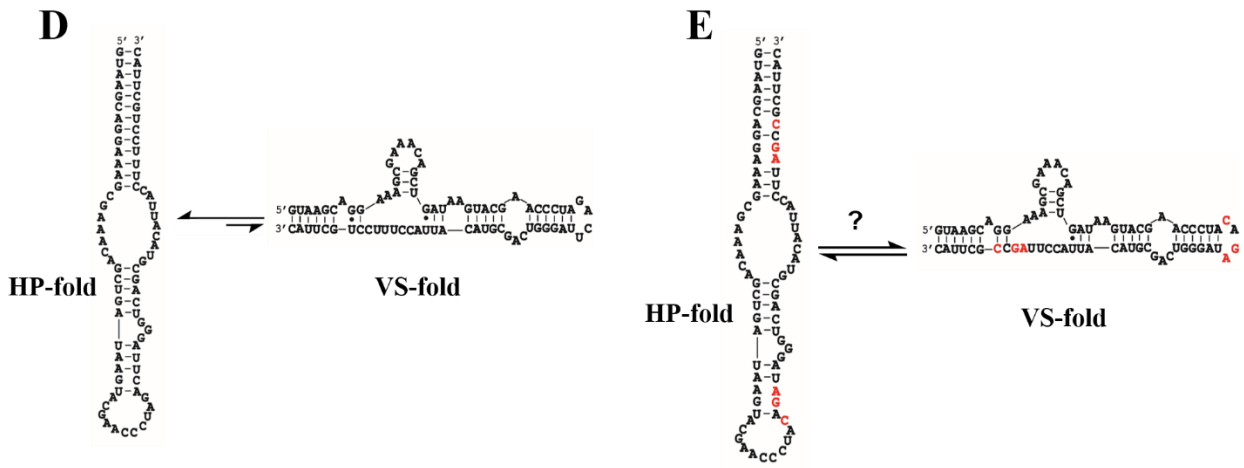

Figure 4.10 Toward a bifunctional catalytic domain sequence (continued). D. Trans_3: putative hairpin and VS folds. **E.** Trans_3mut6: putative hairpin and VS folds. This sequence is obtained by mutations (in red) to trans_3. Equilibrium arrow does not indicate the relative population of two folds.

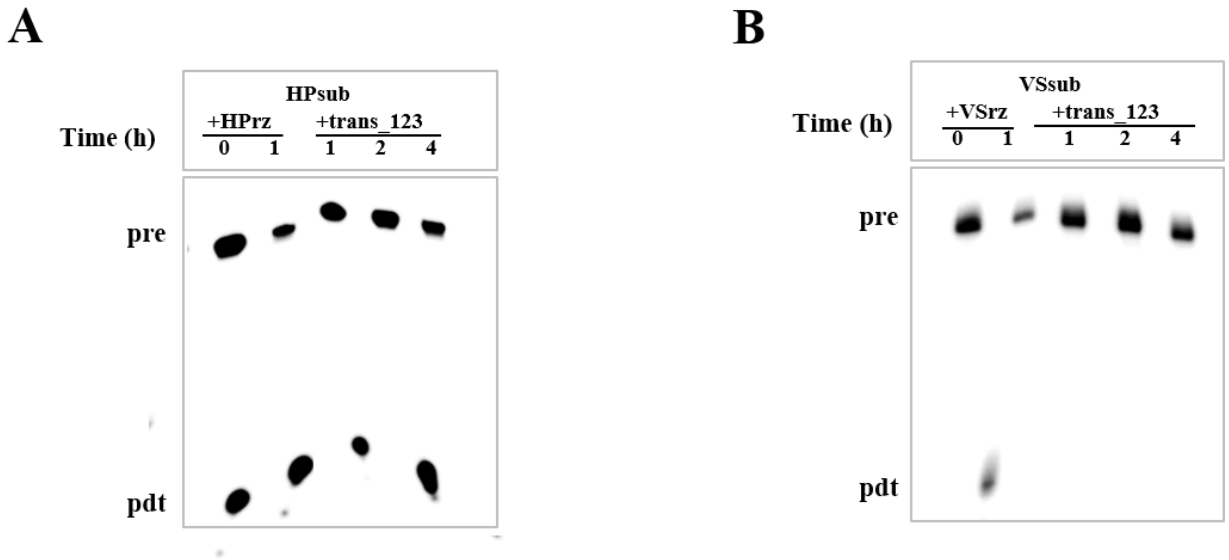

Figure 4.11 Testing hairpin and VS cleavage activities of sequences in figure 4.10.
A. Trans_123 cleaves hairpin substrate. **B.** Trans_123 does not cleave VS substrate.

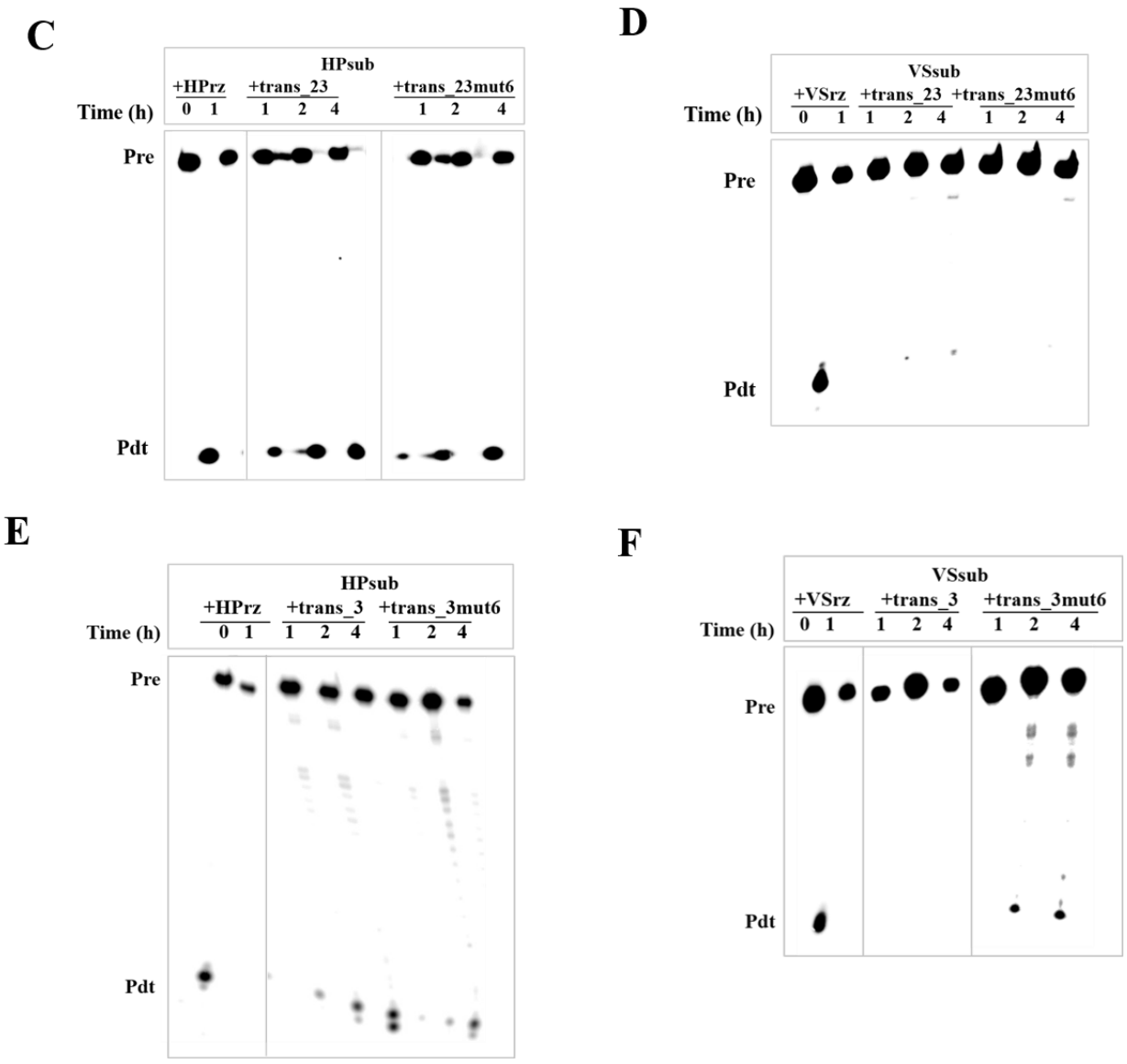

Figure 4.11 Testing hairpin and VS cleavage activities of sequences in figure 4.10 (continued). **C.** Trans_23 and trans_23mut6 cleave hairpin substrate. **D.** Trans_23 and trans_23mut6 do not cleave VS substrate. **E.** Trans_3 and trans_3mut6 cleave hairpin substrate with low efficiency. **F.** Trans_3 does not cleave VS substrate, but trans_3mut6 cleaves VS substrate.

4.3 DISCUSSION

We have reverse engineered a catalytic sequence (trans_3mut6) from wild-type catalytic domains of the hairpin and VS ribozymes that is capable of cleaving both hairpin and VS substrates *in trans*. This bifunctional sequence can be threaded through the secondary structures of both catalytic domains without disrupting the identities of essential nucleotides in both folds. The lower efficiencies of trans cleavage by trans_3mut6, compared to other related constructs like trans_123 or trans_23/23mut6 could be due to the conformational restrictions imposed by this dual phenotype. Despite its low activity, trans_3mut6 is to our knowledge, the first sequence that is capable of cleaving substrates that correspond to two natural ribozymes. In order to create a self-contained catalytic system, the substrate and catalytic domains should be part of the same RNA, as observed in wild-type ribozymes. Therefore, the next step was be to test cleavage activities of the bifunctional sequence *in cis* when fused to hairpin and VS substrates individually. Since, the current bifunctional sequence is an inefficient ribozyme operating at rates that are just above background cleavage, we needed to design robust cis-cleaving ribozymes, created with a catalytic domain sequence that cleaves both hairpin and VS substrates *in cis* when fused with it. The immediate objective was to identify neutral paths that connects hairpin-only and VS-only phenotypes through single step mutations via the bifunctional 'intersection sequence'. The ultimate objective was to identify points of intersections in the neutral networks of endonucleolytic ribozymes.

RNA SEQUENCES USED IN THE EXPERIMENTS DISCUSSED IN THIS CHAPTER ARE LISTED IN MATERIALS AND METHODS (7.9.2, CHAPTER 7).

REFERENCES

1. Aharoni, A., Gaidukov, G., Khersonsky, O., Gould, S. M., Roodveldt, C. and Tawfik, D., 2005. The 'evolvability' of promiscuous protein functions. *Nat. Genet.*, 37, 73–76.

2. Ancel, L. and Fontana, W., 2000. Plasticity, modularity and evolvability in RNA. *J. Exp. Zool.*, 288, 242-283.

3. Andersen, A. A. and Collins, R. A., 2001. Intramolecular secondary structure rearrangement by the kissing interaction of the Neurospora VS ribozyme. *Proc. Natl. Acad. Sci. U. S. A.*, 98, 7730-7735.

4. Athvale, S. A., Spicer, B. and Chen, I. A., 2017. Experimental fitness landscapes to understand the molecular evolution of RNA-based life. *Curr. Opin. Chem. Biol.*, 22, 35-39.

5. Burch, C. L. and Chao, L., 1999. Evolution by small steps and rugged landscapes in the RNA virus φ6. *Genetics*, 151, 921-927.

6. Butcher, S. E., Heckman, J. E. and Burke, J. M., 1995. Reconstitution of hairpin ribozyme activity following separation of functional domains. *J. Biol. Chem.*, 270, 29648-29651.

7. Carothers, J. M., Oestreich, S. C. Davis, J. H. and Szostak, J. W., 2004. Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc.*, 126, 5130-5137.

8. Chan, H. S. and Bornberg-Bern, E., 2002. Perspectives on protein evolution from simple exact models. *Appl. Bioinf.*, 1, 121-144.

9. Cochrane, J. C. and Strobel, S. A., 2008. Catalytic strategies of self-cleaving ribozymes. *Acc. Chem. Res*., 41, 1027-1035.

10, Connell, G. J. and Yarus, M., 1994. RNAs with dual specificity and dual RNAs with similar specificity. *Science*, 264, 1137-1141.

11. Cowperthwaite, M. C. and Meyers, L. A., 2007. How mutational networks shape evolution: lessons from RNA models. *Annu. Rev. Ecol. Evol. Syst*., 38, 203-230.

12. Curtis, E. A. and Bartel, D. P., 2005. New catalytic structures from an existing ribozyme. *Nat. Struct. Mol. Biol.*, 12, 994-1000.

13. Fedor, M. J., 2000. The structure and function of the hairpin ribozyme. *J. Mol Biol.*, 297, 268-291.

14, Fontana, W. and Schuster, P., 1998. Continuity in evolution: On the nature of transitions. *Science*, 280, 1451-1455.

15. Forconi, M. and Herschlag, D., 2005. Promiscuous catalysis by the Tetrahymena group I ribozyme. *J. Am. Chem. Soc.*, 127, 6160-6161.

16. Held, D. M., Greathouse, S. T., Agrawal, A. and Burke, D. H., 2003. Evolutionary landscapes for the acquisition of new ligand recognition by RNA aptamers. *J. Mol. Evol.*, 57, 299-308.

17. Huang, Z., Pei, W., Han, Y., Jayaseelan, S., Shekhtman, A., Shi, H. and Niu, L., 2009. One RNA aptamer sequence, two structures: a collaborating pair that inhibits AMPA receptors. *Nucleic Acids Res.*, 37, 4022-4032.

18. Huynen, M. A., Stadler, P. F. and Fontana, W., 1996. Smoothness within ruggedness: The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. U. S. A.*, 93, 397-401.

19. Jacob, F., 1977. Evolution and tinkering. *Science*, 196, 1161-1166.

20. Jiménez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R., and Chen, I. A, 2013. Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proc. Natl. Acad. Sci. U.S.A.*, 10, 14984-14989.

21. Jones, F. D., Ryder, S. P. and Strobel, S. A., 2001. An efficient ligation reaction promoted by a Varkud Satellite ribozyme with extended 5'- and 3'-termini. *Nucleic Acids Res.*, 29, 5115-5120.

22. Joyce, G. F., 2002. The antiquity of RNA-based evolution. *Nature*, 418, 214-221.

23. Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature*, 217, 624–626.

24. Komatsu, Y., Kanzaki, I. and Ohtsuka, E., 1996. Enhanced folding of hairpin ribozymes with replaced domains. *Biochemistry*, 35, 9815-9820.

25. Komatsu, Y., Kanzaki, I., Koizumi, M. and Ohtsuka, E., 1995. Modification of primary structures of hairpin ribozymes for probing active conformations. *J. Mol. Biol.*, 252-296-304.

26. Lafontaine, D. A., Wilson, T. J., Norman, D. G. and Lilley, D. M. J., 2001. The A730 loop is an important component of the active site of the VS ribozyme. *J. Mol. Biol.*, 312, 663-674.

27. Landweber, L. F. and Pokrovskaya, I. D., 1999. Emergence of a dual-catalytic RNA with metal-specific cleavage and ligase activities: The spandrels of RNA evolution. *Proc. Natl. Acad. Sci. U. S. A.*, 96, 173-178.

28. Lau, M. W. L. and Ferré-D'Amaré, A. R., 2016. Many activities, One structure: Functional plasticity of ribozyme folds. *Molecules*, 21, 1570-1577.

29. Lau, M. W. L., Cadieux, K. E. C. and Unrau, P. J., 2004. Isolation of fast purine nucleotide synthase ribozymes. *J. Am. Chem. Soc.*, 126, 15686–15893.

30. Lilley, D. M. J. and Eckstein F. 2008. *Ribozymes and RNA Catalysis*. RSC publishing.

31. Newcomb, R.D., Campbell, P. M., Ollis, D. L., Cheah, E., Russell, R. J. and Oakeshott, J. G., 1997. A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. *Proc. Natl. Acad. Sci. U. S. A.* 94, 7464–7468.

32. Manrubia, S. C. and Briones, C., 2007. Modular evolution and increase of functional complexity in replicating RNA molecules. *RNA*, 13, 97-107.

33. Marek, M. S., Johnson-Buck, A. and Walter, N. G., 2011. The shape-shifting quasispecies of RNA: one sequence, many functional folds. *Phys. Chem. Chem. Phys.*, 13, 11524–11537.

34. Meier, S. and Özbek, S., 2007. A biological cosmos of parallel universes: does protein structural plasticity facilitate evolution? *BioEssays*, 29, 1095-1104.

35. Pereira, M. J. B., Nikolova, E. N., Hiley, S. L., Jaikaran, D., Collins, R. A. and Walter, N. G., 2008. Single VS ribozyme molecules reveal dynamic and hierarchical folding toward catalysis. *J. Mol. Biol.*, 382, 496-509.

36. Petrov, A. S., Gulen, B., Norris, A. M., Kovacs, N. A., Bernier, C. R., Lanier, K. A., Fox, G. E., Harvey, S. C., Wartell, R. M., Hud, N. V. and Williams, L. D., 2005. History of the ribosome and the origin of translation. *Proc. Natl. Acad. Sci. U. S. A.*, 112, 15396-15401.

37. Piccirilli, J. A., McConnell, T. S., Zaug, A. J., Noller, H. F., Cech, T. R., 1992. Aminoacyl esterase activity of the Tetrahymena ribozyme. *Science*, 256, 1420-1424.

38. Pitt, J. N. and Ferré-D'Amaré, A. R., 2010. Rapid construction of empirical RNA fitness landscapes. *Science*, 330, 376-378.

39. Rastogi, T., Beattie, T. L., Olive, J. E. and Collins, R. A., 1996. A long-range pseudoknot is required for activity of the Neurospora VS ribozyme. *EMBO J.*, 15, 2820-2825.

40. Ren, A., Micura, R. and Patel, D. J., 2017. Structure-based mechanistic insights into catalysis by small self-cleaving ribozymes. *Curr. Opin. Chem. Biol.*, 41, 71–83.

41. Rupert, P. B. and Ferré-D'Amaré, A. R., 2001. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, 410, 780-786.

42. Salisbury, F. B., 1969. Natural selection and the complexity of the gene. *Nature*, 224, 342–343.

43. Sargueil, B., Pecchia, D. B. and Burke, J. M., 1995. An improved version of the hairpin ribozyme functions as a ribonucleoprotein complex. *Biochemistry*, 34, 7739-7748.

44. Schmidt, F. J., 1999. Ribozyme-Why so many, why so few? *Mol. Cells*, 9, 459-463.

45. Schneider, T. D., Stormo, G. D., Gold, L. and Ehrenfeucht, A., 1986. Information content of binding sites on nucleotide sequences. *J. Mol. Biol.*, 188, 415-431.

46. Schultes, E. A. and Bartel, D. P., 2000. One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. *Science*, 289, 448-452.

47. Schultes, E. A., Hraber, P. T. and LaBean, T. H., 2009. No molecule is an island: Molecular evolution and the study of sequence space. *Algorithmic bioprocesses*, Natural Computing Series, Springer-Verlag Berlin Heidelberg, 675-704.

48. Schuster, P., Fontana, W., Stadler, P. F. and Hofacker, I. L., 1994. From sequences to shape and back: a case study in RNA secondary structures. *Proc. R. Soc. Lond. B*, 255, 279-284.

49. Smith, J. M., 1970. Natural selection and the concept of protein space. *Nature*, 225, 563–564.

50. Sood, V. D. and Collins, R. A., 2002. Identification of the catalytic subdomain of the VS ribozyme and evidence for remarkable sequence tolerance in the active site loop. *J. Mol. Biol.*, 320, 443-454.

51. Steinberg, B. and Ostermeier, M., 2016. Environmental changes bridge evolutionary valleys. *Sci. Adv.*, 2 e1500921 DOI: 10.1126/sciadv.1500921.

52. Suslov, N. B., DasGupta, S., Huang, H., Fuller, J. R., Lilley, D. M. J., Rice, P. A. and Piccirilli, J. A., 2015. Crystal structure of the Varkud satellite ribozyme. *Nat. Chem. Biol.*, 11, 840-846.

53. Talini, G., Branciamore, S. and Gallori, E., 2011. Ribozymes: Flexible molecular devices at work. *Biochimie*, 93, 1998-2005.

54. Wagner, A., 2005. Robustness, evolvability, and neutrality. *FEBS Lett.*, 579, 1772-1778.

55. Wagner, A., 2012. The role of robustness in phenotypic adaptation and innovation. *Proc. R. Soc. B*, 279, 1249-1258.

56. Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E. and Adami, C., 2001. Compensatory mutations cause excess of antagonistic epistasis in RNA secondary structure folding. *BMC Evol. Biol.*, 3:3.

56. Wright, S., 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. VI Int. Congr. Genet*., 1, 356–366.

58. Zamel, R. and Collins, R. A., 2002. Rearrangement of substrate secondary structure facilitates binding to the Neurospora VS ribozyme. *J. Mol. Biol.*, 324, 903-915.

59. Zaug, A. J. and Cech, T. R., 1986. The Tetrahymena intervening sequence ribonucleic acid enzyme is a phosphotransferase and an acid phosphatase. *Biochemistry*, 25, 4478-4482.

60. Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31, 3406-3415.

Chapter - 5

SMOOTH ACQUISITION OF CATALYTIC FUNCTION IN RIBOZYMES THROUGH INTERSECTION OF NEUTRAL NETWORKS: EXPLORING NEUTRAL NETWORKS

5.1 RESULTS

5.1.1 Designing a bifunctional sequence that cleaves both hairpin and VS substrates *in cis*

The single catalytic sequence (trans_3mut6, chapter 4) designed to cleave both hairpin and VS substrates *in trans* exhibited slow cleavage even at high millimolar concentrations of Mg^{2+} (200 mM). This could be attributed to the folding restrictions imposed by dual function, which result in a reduced fraction populating each ribozyme fold. Moreover, this sequence might be sufficiently pliable to adopt other non-functional folds in addition to the folds corresponding to the hairpin and VS ribozymes. Wild-type functional RNAs such as ribozymes have evolved sequences that minimize misfolding, but sequences designed to be plastic are more likely to populate multiple tertiary structures. Additionally. the absence of domains in the bifunctional sequence that facilitate substrate binding in wild-type VS ribozyme likely contributes to the inefficient cleavage of the VS substrate. To ameliorate inefficient substrate binding, we connected trans_3mut6 with VSsub and HPsub via linker sequences (Figure 5.1). This operation likely increases catalytically productive encounters between the substrates and the bifunctional catalytic sequence.

Trans_3mut6 was found to be inactive toward cis cleavage of a VS substrate, regardless of the lengths of the linker; however, hairpin substrates containing a base-paired stem under the cleavage loop and connected to the catalytic domain via a 4 nt linker (short1-HP), were cleaved by trans_3mut6 (Figure 5.2). This observation is consistent with the base-pairing requirements in

the hairpin substrate flanking the internal cleavage loop (Fedor, 2000). This is further illustrated by the inactivity of short2-HP toward cleavage by trans_3mut6 (Figure 5.2B).

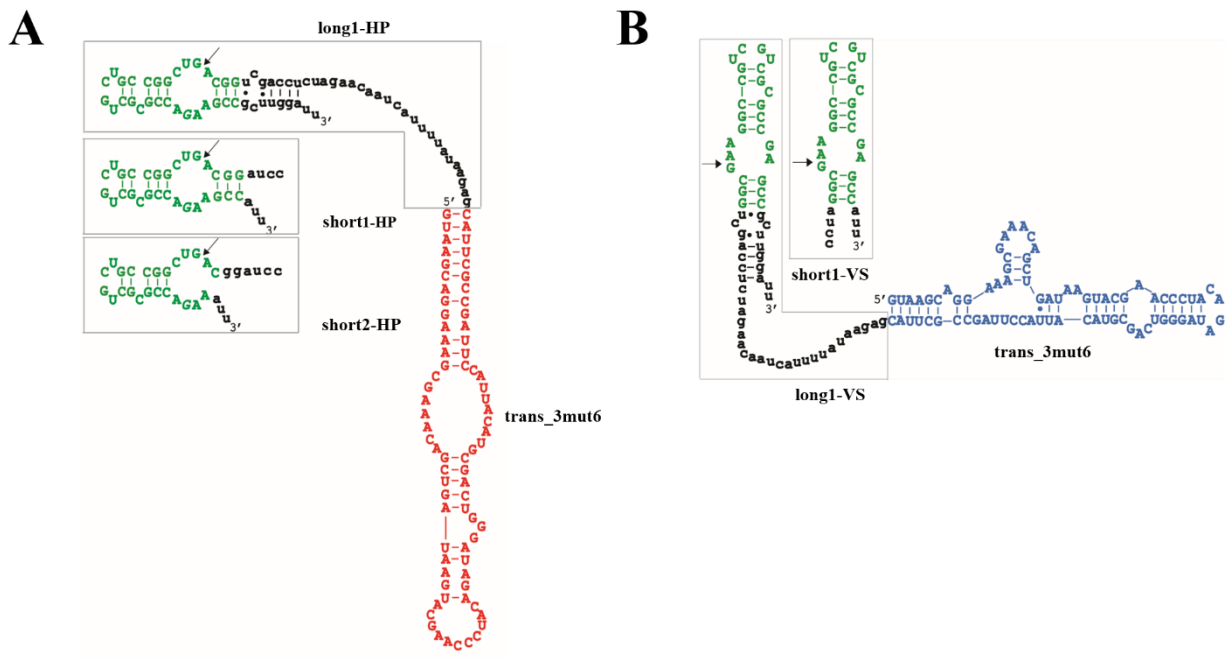

Figure 5.1 Creating cis-ribozyme constructs by fusing hairpin and VS substrates to bifunctional sequence, trans_3mut6. A. Cis hairpin ribozyme constructs created by connecting the hairpin substrate to trans_3mut6 (hairpin fold shown in red) via linkers of different lengths. Long1-HP contains a 23 nt linker sequence followed by a short base-paired helix (6 bp, 1 bulge) in addition to the HPsub sequence and a 3 nt overhang at the 3' terminus. Short1-HP contains a 4 nt linker sequence followed by the HPsub sequence and a 3 nt overhang at the 3' terminus. Short2-HP contains a 6 nt linker sequence followed a variant of HPsub that lacks a stem under the internal cleavage loop. The 3' end contains in a 3 nt overhang. **B.** Cis VS ribozyme constructs created by connecting the VS substrate to trans_3mut6 (VS fold shown in blue) via linkers of different lengths. Long1-VS contains a 23 nt linker sequence followed by a short base-paired helix (6 bp, 1 bulge) in addition to the VSsub sequence and a 3 nt overhang at the 3' terminus. Short1-VS contains a 4 nt linker sequence followed by the VSsub sequence and a 3 nt overhang at the 3' terminus.

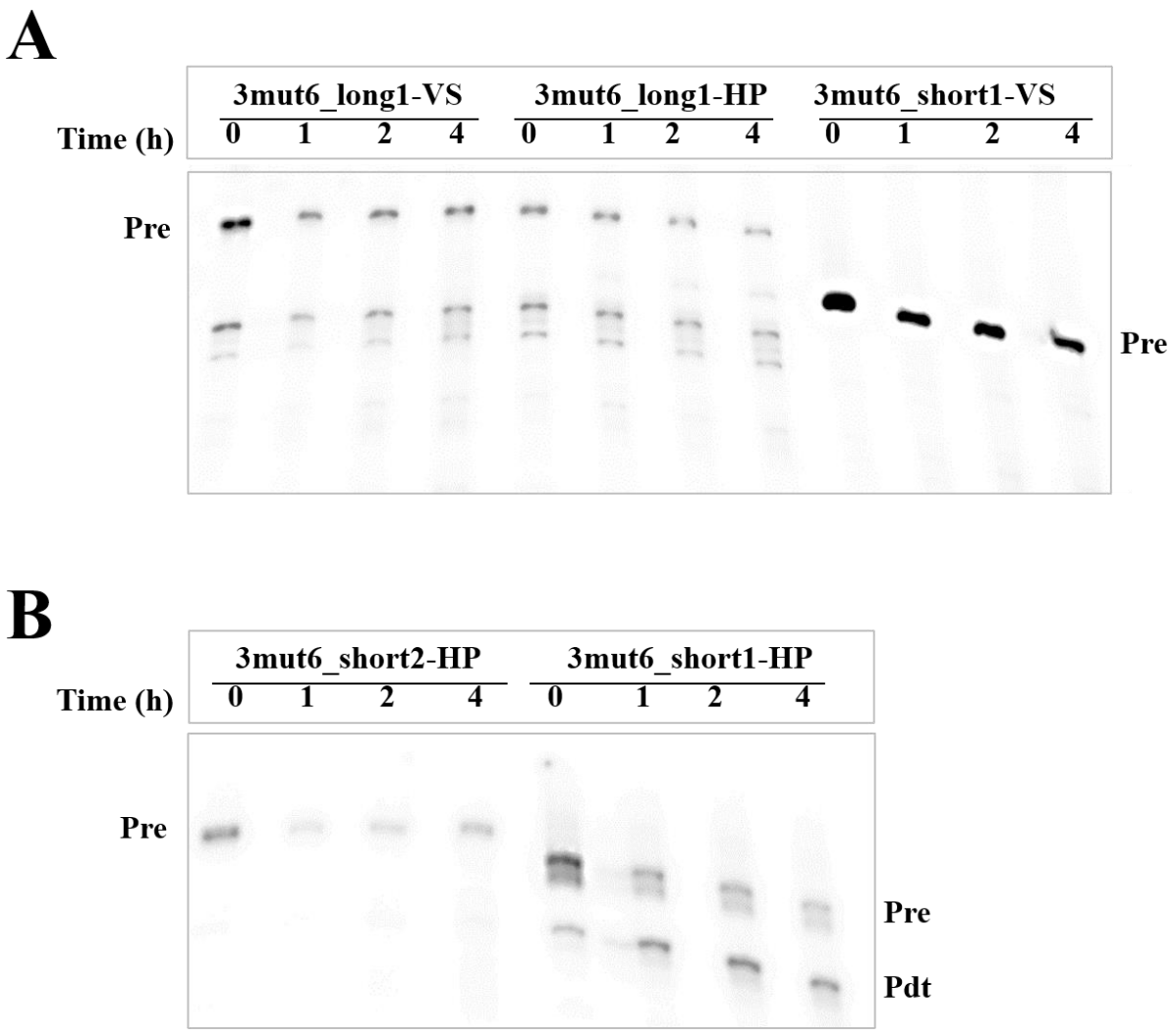

Figure 5.2 Testing cis cleavage of fusion constructs of trans_3mut6 with hairpin and VS substrates connected by different linkers. A. Trans_3mut6 is inactive for VS cleavage *in cis*, regardless of the lengths of linker sequences. Trans_3mut6 does not cleave HPsub when connected with linker, long1. **B.** Trans_3mut6 does not cleave HPsub linked with linker short2; however, shows cis cleavage when tested with HPsub connected by linker short1.

Since the degree of flexibility between the substrate and catalytic domains is determined by the linker (Sargueil et al., 1995; Komatsu et al., 1995; Komatsu et al., 1996; Zamel et al., 2004), we replaced these linkers (Figure 5.1) with a series of linkers ranging from 7 nt to 17 nt for VSsub (Figure 5.3). In line with our goal of designing a single sequence that cleaves both substrates *in*

cis, we tested similar linkers (7 nt to 11 nt) for cis cleavage of HPsub (Figure 5.4). All linkers consist of a CCUA tetranucleotide followed by a polyA chain of varying lengths. Unfortunately, none of these trials generated a catalytically active cis-ribozyme that could cleave hairpin and VS substrates (data not shown).

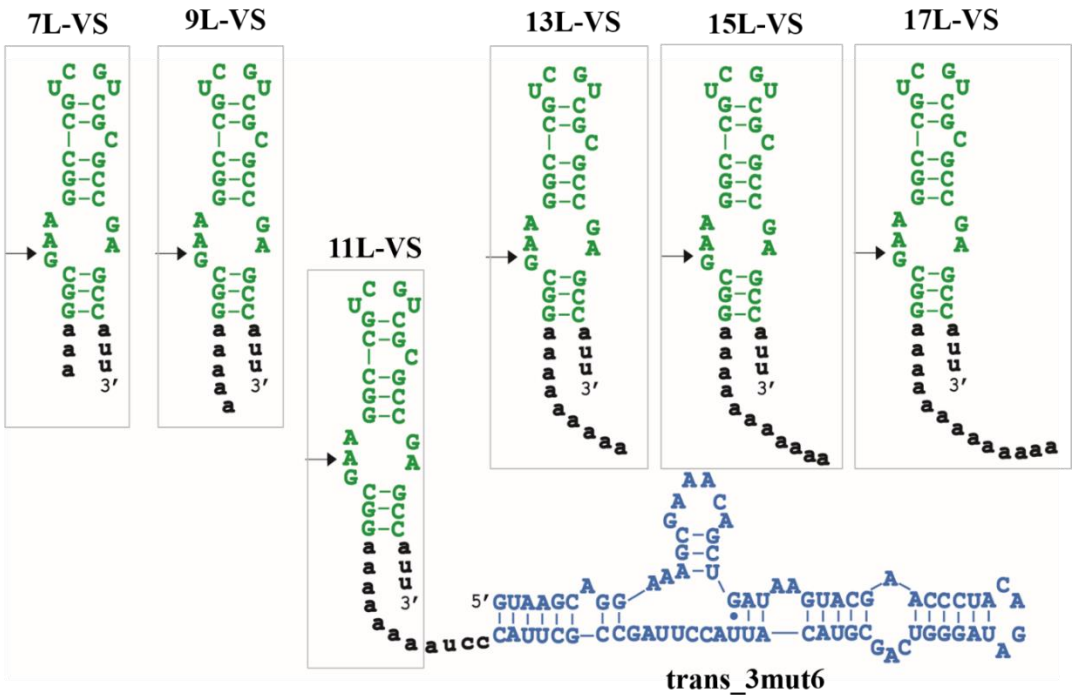

Figure 5.3 Creating cis-ribozyme constructs by fusing VS substrate to trans_3mut6 via linkers of different lengths. None of these cis constructs demonstrated robust cleavage.

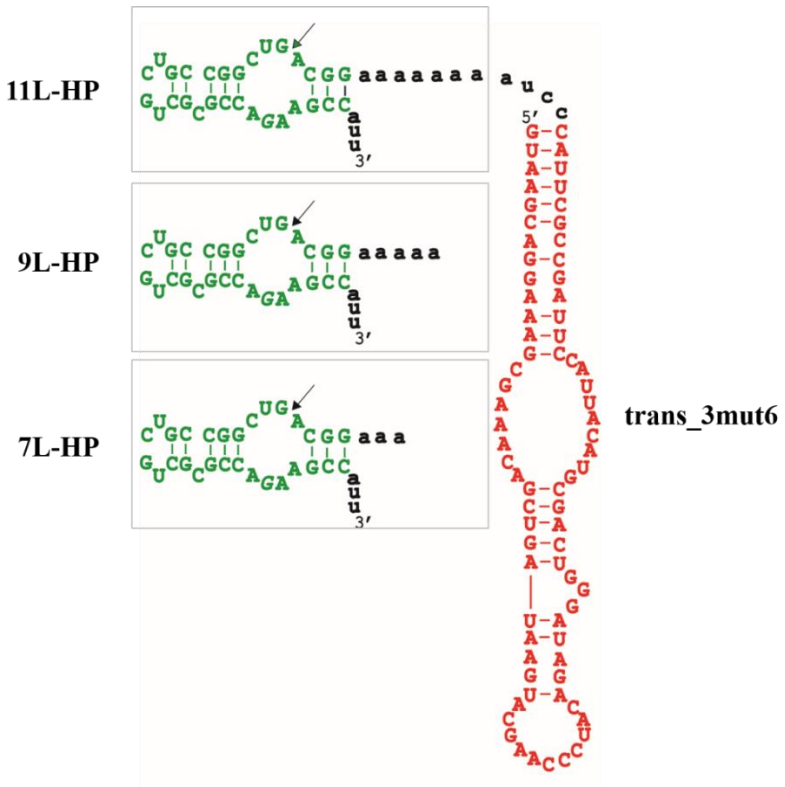

Figure 5.4 Creating cis-ribozyme constructs by fusing hairpin substrate to trans_3mut6 via linkers of different lengths. None of these cis constructs demonstrated robust cleavage.

These results prompted us to test another catalytic sequence that was originally designed for dual cleavage of hairpin and VS substrates but cleaved only hairpin substrates *in trans* with a cleavage rate of 0.012/min (compared to a VS cleavage rate of 0.00016/min). This sequence, trans_23mut6 (Figure 4.10C) differs from trans_3mut6 by 3 base-pairs (Figure 4.10E). We created fusion constructs with trans_23mut6 and hairpin and VS substrates taken separately and connected them by linkers of two different lengths (Figure 5.5). Both hairpin and VS cis constructs with longer linkers underwent site-specific cleavage (Figure 5.6).

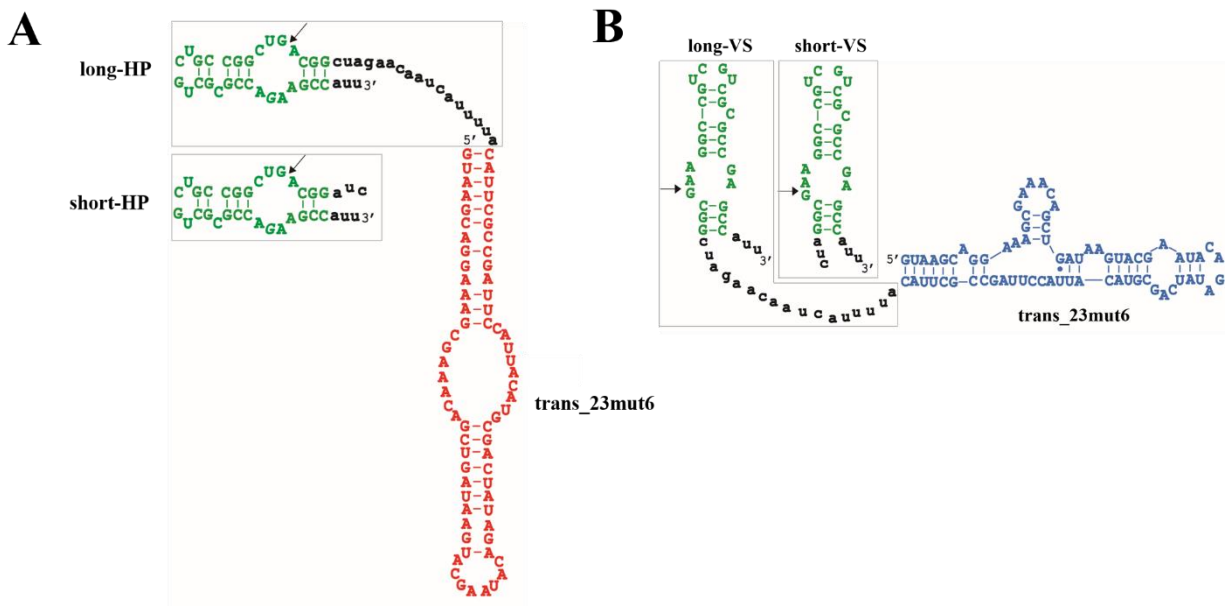

Figure 5.5 Creating cis-ribozyme constructs by fusing hairpin and VS substrates to trans_23mut6. A. Trans_23mut6 was connected to HPsub via a 17 nt linker (long-HP) or a 3 nt linker (short-HP). **B.** Trans_23mut6 was connected to VSsub via a 17 nt linker (long-VS) or a 3 nt linker (short-VS).

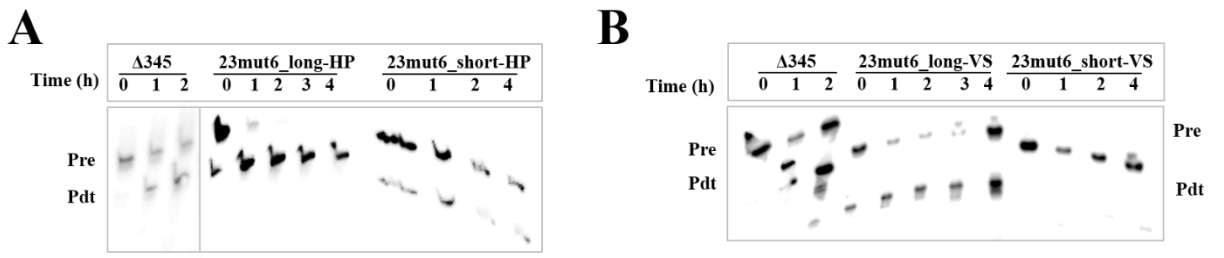

Figure 5.6 Testing cis cleavage activity of trans_23mut6 toward hairpin and VS substrates. Cis-ribozyme constructs comprising trans_23mut6 connected to hairpin (**A**) or VS substrates (**B**) via 17 nt linkers show self-cleavage. Cis-ribozymes with HP or VS substrates connected via 3 nt linkers fail to cleave. Prototype minimal VS ribozyme Δ345 used as positive control.

Therefore, trans_23mut6 represents a catalytic domain sequence that cleaves both hairpin and VS substrates *in cis*. Further, the cleavage is faster and more robust than hairpin or VS cleavage *in trans*. In the framework of molecular evolution (chapter 4), such a functionally plastic sequence corresponds to the point of intersection of two neutral networks representing two distinct functions.

5.1.2 Kinetics of hairpin and VS cleavage *in cis* by HP_INT1_VS

Ribozyme function usually provides a direct readout of their structure (Marek et al., 2011). Since the trans_23mut6 sequence can be threaded through the secondary structures of both hairpin and VS catalytic domains and carries out site-specific cleavage of hairpin and VS substrates, it is expected that trans_23mut6 populates functional conformations of both ribozymes (Figure 5.7). The sequences of these cis-cleaving hairpin and VS ribozymes are identical except in the internal cleavage loops of their substrate domains that contain their respective cleavage sites in addition to other conserved residues that define the identities of the two ribozymes. For simplicity, the bifunctional trans_23mut6 will be referred to as HP_INT1_VS (INT as in intersection sequence), and cis-ribozyme constructs with HP_INT1_VS as the catalytic domain fused to hairpin and VS substrates will be referred to HP_INT1_VS-HP and HP_INT1_VS-VS, respectively.

HP_INT1_VS cleaves both hairpin and VS substrates at rates that are three-orders of magnitude higher (Figure 5.8A-D) compared to corresponding uncatalyzed RNA cleavage under the same reaction conditions (Figure 5.9). Cleavage rates for the hairpin and VS cis-ribozymes created with HP_INT1_VS as their catalytic domain are 0.036±0.003/min and 0.030±0.005/min, respectively (Table 5.1). The 'hairpin' and 'VS' secondary structures adopted by HP_INT1_VS have comparable free energies of -16.5 kcal/mol and -15.7 kcal/mol respectively, as obtained from mFOLD (Zuker, 2003). This supports the structural plasticity and dual function of HP_INT1_VS.

Wild-type hairpin and VS ribozymes accelerate cleavage by utilizing conserved guanine and adenine nucleobases as catalytic moieties and mutating either leads to substantial reduction in cleavage rates (Lilley, 2004). Mutating these residues in the context of HP_INT1_VS-HP and HP_INT1_VS-VS resulted in a modest rate reduction (Figure 5.8E, F). The attenuated dependence on catalytic residues in these minimal ribozymes is expected in the background of low activity, but the large extent of attenuation probably indicates that these sequences use other catalytic strategies for RNA cleavage. The diminished roles of catalytic residues likely contribute to lower cleavage rates intrinsic to HP_INT1_VS. However, it is reasonable to assume that losing dependence on catalytic nucleotides is essential to free the catalytic domain sequence from the structural and functional restraints of either functional fold. Reduction in cleavage rates due to unforeseen local structural perturbations caused by these mutations cannot be discounted.

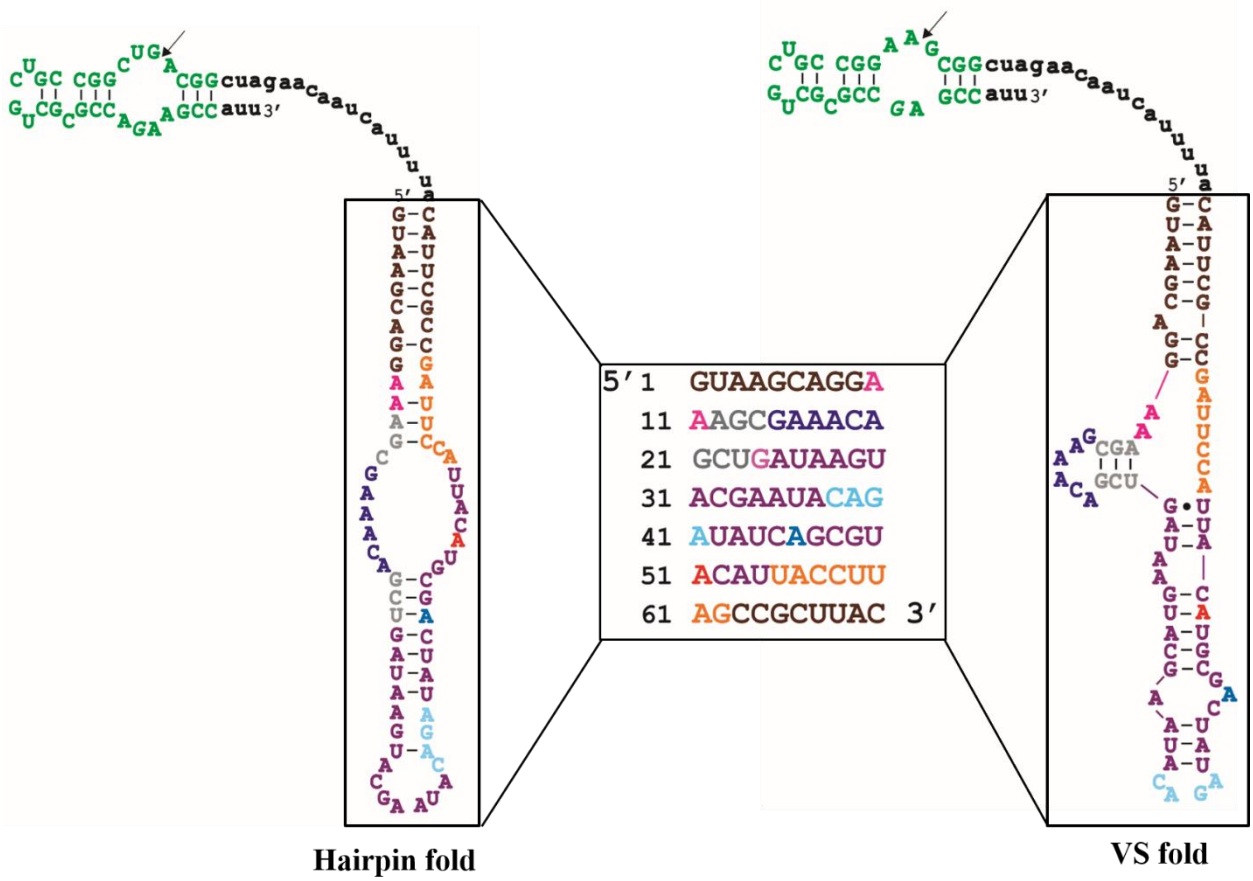

Figure 5.7 Sequence of the bifunctional catalytic domain (HP_INT1_VS) threaded through secondary structures of both hairpin and VS catalytic domains. HP_INT1_VS can assume the secondary structures of both hairpin and VS catalytic domains. These alternate secondary structures have comparable thermodynamic stabilities according to mFOLD predictions. The sequence is color coded according to sub-domains of the VS secondary structure and secondary structure of the hairpin fold reflects the same color code.

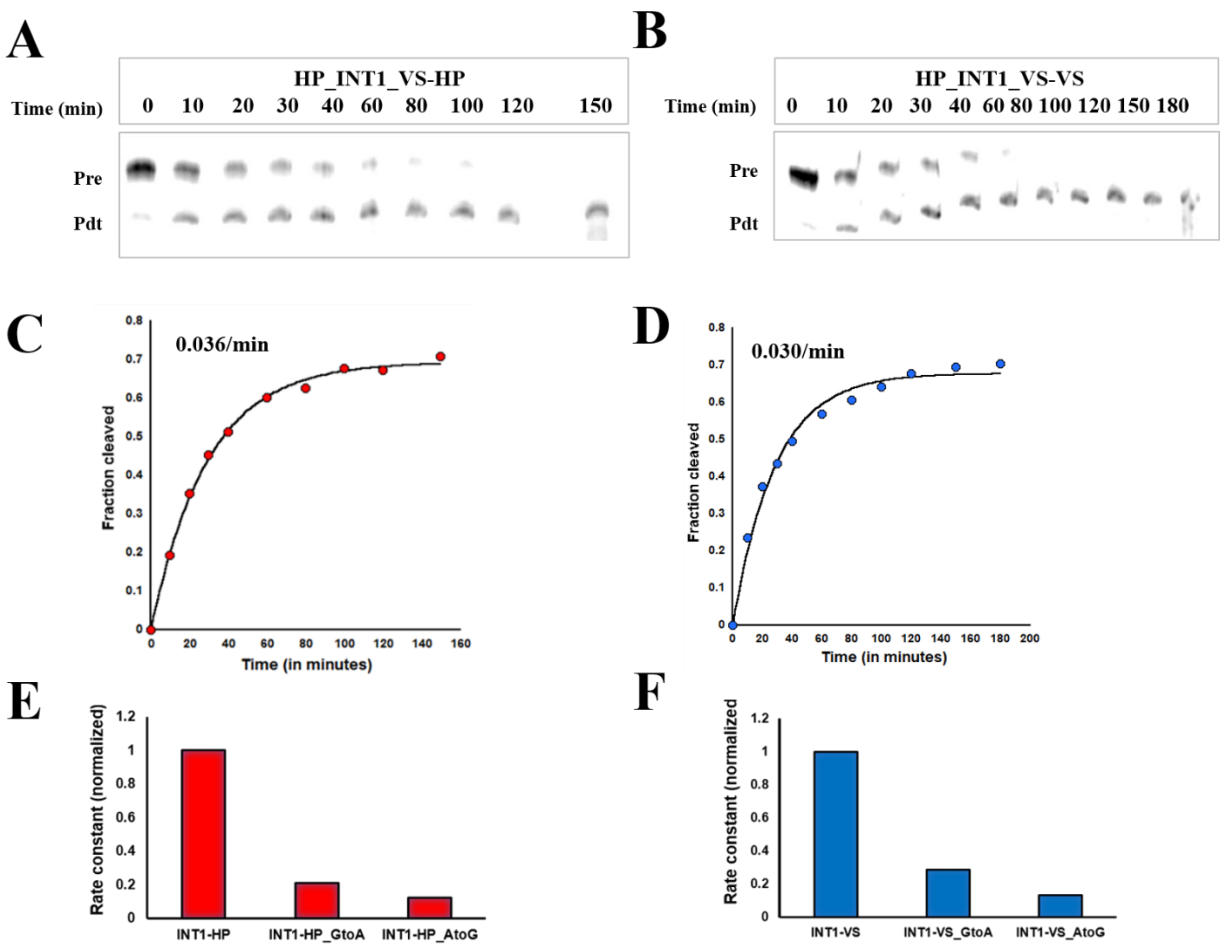

Figure 5.8 HP_INT1_VS cleaves hairpin and VS substrates *in cis*. A. Cis cleavage of HP_INT1_VS-HP. **B.** Cis cleavage of HP_INT1_VS-VS. **C.** Fraction cleaved plotted against time for cis cleavage of HP_INT1_VS-HP. **D.** Fraction cleaved plotted against time for cis cleavage of HP_INT1_VS-VS. **E.** Mutating G and A nucleotides that mediate catalysis in wild-type hairpin sequence, reduces cleavage rates by 4.5 and 8-fold, respectively in HP_INT1_VS-HP. **F.** Mutating G and A nucleotides that mediate catalysis in wild-type VS sequence, reduces cleavage rates by 3.3 and 7.5-fold, respectively in HP_INT1_VS-VS.

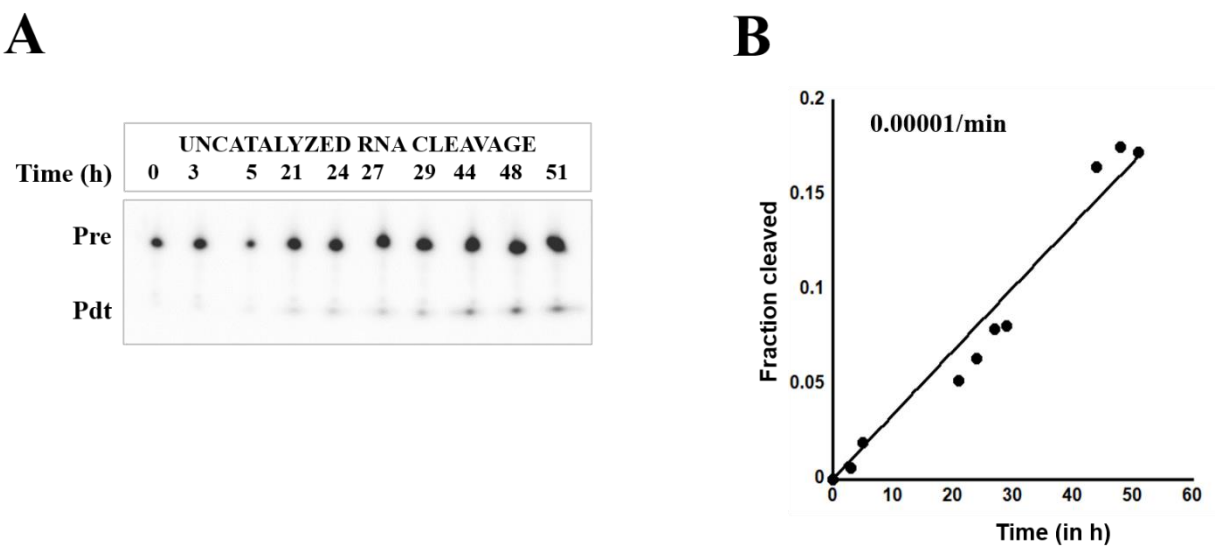

Figure 5.9 Uncatalyzed RNA cleavage. A. Aliquots run on a dPAGE gel at different time points. **B.** Time course for uncatalyzed RNA cleavage. Rate of cleavage is ~0.00001/min. Refer to chapter 7 (section 7.8) for experimental details.

5.1.3 Functional interconversion between hairpin and VS ribozymes through 'neutral' mutational drifts

A bifunctional sequence can drift to exclusively single-function phenotypes via mutational perturbations to their sequence that alter relative stabilities of the two functional folds. Additionally, mutations to nucleotides in the active site can disrupt its local conformation and inhibit catalysis. While this does not affect global folds corresponding to either ribozyme, mutations to essential nucleotides of one of the folds can render that fold inactive. Therefore, it is possible to eliminate functions of either ribozyme fold, by preferentially destabilizing either structure or disrupting the active site of either ribozyme without any change in global conformation. Mutations to the catalytic domain, HP_INT1_VS predicted to stabilize the hairpin fold (and simultaneously destabilize the VS fold) generated sequences that retain their cleavage activity toward the hairpin substrate but gradually eliminated VS cleavage (Figure 5.10). Conversely, mutations to HP_INT1_VS predicted to preferentially stabilize the VS fold (and destabilize the hairpin fold) or disrupt the active site of the hairpin catalytic domain (at conserved residues in loop B) produced sequences that retain cleavage activity toward VS substrate but eliminated hairpin cleavage (Figure 5.11).

Mutational drift from HP_INT1_VS to HP-only sequences (Figure 5.10). The hairpin fold of HP_INT1_VS contains multiple unpaired nucleotides within largely base-paired stems. Mutations (^{HP4}C, ^{HP3}U and ^{HP2}C) that lead to the creation of additional base-pairs stabilize the hairpin fold. Mutations ^{HP3}U and ^{HP2}C are not expected to change the structure of the VS fold as they are located in the terminal loop in this fold. This is consistent with the decrease in VS cleavage that is accompanied by a gradual increase in hairpin cleavage. However, mutation, ^{HP4}C has a dual effect of creating a new base-pair in the hairpin secondary structure and disrupting a base-pair in

the VS secondary structure. This mutation abolishes VS activity but preserves hairpin function. Mutation ^{HP1}G is directed at a cytidine that is part of a loop in both structures and does not affect the structural stabilities of either fold. Mutations ^{HP5}U, ^{HP6}C, ^{HP7}A, preferentially disfavor the VS structure by disrupting the stems flanking the catalytically important internal bulge in helix 6 (A730-loop) (Lafontaine et al., 2001). In the alternate hairpin structure, these mutations lie in the terminal loop hence are expected to be innocuous toward activity. Collectively, these seven mutations shift the conformational equilibrium toward the hairpin fold, which is reflected in the decrease in cleavage rates for VS substrates along the mutational trajectory (Figure 5.12A). Alternate mutational paths support dual function, underscoring the robustness of mutational space (Figure 5.12B, D). Since, each sequence variant retains cleavage activity, the mutational drift from a dual function phenotype to a single function phenotype can be considered neutral.

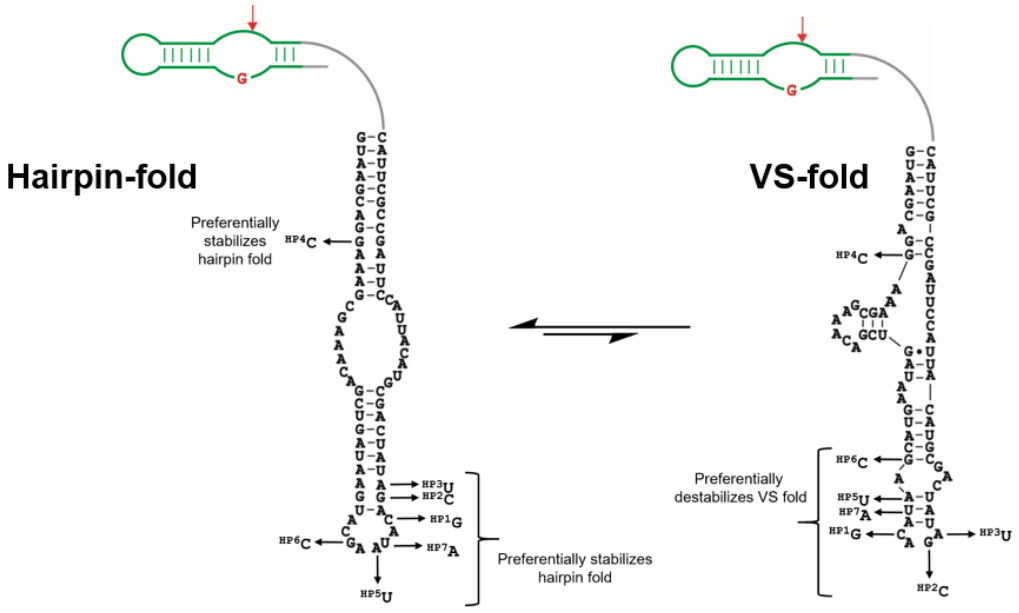

Figure 5.10 Mutational drift from bifunctional HP_INT1_VS to exclusive hairpin function. Directed point mutations to HP_INT1_VS stabilize the hairpin fold, while the corresponding VS fold is relatively unaffected or destabilized. This converts the bifunctional cis-ribozyme to an exclusive hairpin ribozyme. The arrow indicates equilibrium between the hairpin and VS catalytic domains.

Mutational drift from HP_INT1_VS to VS-only sequences (Figure 5.11). Mutation ^{INT2}addC provides additional stabilization to the VS fold by introducing a new C-G base-pair to helix 2 and introduces a one-nucleotide bulge in the corresponding hairpin fold. Both hairpin and VS functions are retained as result. Subsequently, we introduced directed deletions ($^{VS1}\Delta$, $^{VS2}\Delta$, $^{VS4}\Delta$, $^{VS6}\Delta$) to conserved nucleotides that reside in loop B of HP_INT1_VS-HP, which promptly eliminate hairpin function (hairpin function is abolished after $^{VS1}\Delta$). $^{VS1}\Delta$ and $^{VS2}\Delta$ involve deletions of terminal loop nucleotides in the corresponding VS fold, thus are expected to preserve its cleavage activity. Deletions $^{VS4}\Delta$ and $^{VS6}\Delta$ further disrupt the hairpin active site and result in a weakened stem with two bulged nucleotides (A and C) proximal to catalytic A730 loop in the VS fold. Subsequent deletions ($^{VS3}\Delta$ and $^{VS5}\Delta$) of these bulged nucleotides restore a fully-paired stem in the VS fold, without affecting the secondary structure of corresponding hairpin fold. Since, these mutations serve to disrupt the active site of the hairpin fold without any global effects, there is likely no shift in the conformational equilibrium between the two folds. Therefore, this mutational drift eliminates hairpin activity without enriching VS activity (Figure 5.12A). This is in contrast to the corresponding drift toward exclusive hairpin function.

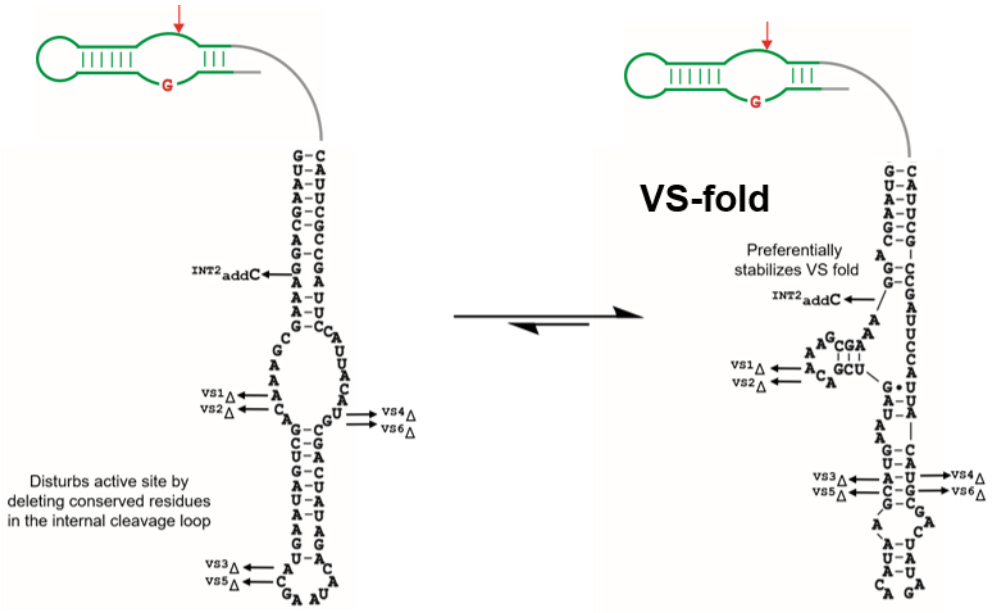

Figure 5.11 Mutational drift from bifunctional HP_INT1_VS to exclusive VS function. Directed point mutations to HP_INT1_VS disrupt the active site of hairpin fold, while the corresponding VS fold is relatively unaffected or stabilized. This converts the bifunctional cis-ribozyme to an exclusive VS ribozyme. The arrow indicates equilibrium between the hairpin and VS catalytic functions.

Exclusive hairpin or VS function can be found within a few mutations of the intersection sequence, HP_INT1-VS. Although distinct structural and functional phenotypes are considered to be isolated from each other in the vast sequence space, the existence of neutral paths between these phenotypes enable smooth acquisition of both phenotypes from sequences that exhibit structural and functional plasticity. HP_INT1_VS, a bifunctional sequence, encounters exclusive hairpin or VS function along its neutral network in only a few mutational steps. An 'HP-only' phenotype can be accessed through four point mutations from HP_INT1_VS, whereas a 'VS-only' sequence is just 2 mutations from it (Figure 5.12C). Initial mutations to HP_INT1_VS that stabilizes the hairpin fold over the VS fold (^{HP3}U and ^{HP2}C), create sequence variants with increases catalytic activities toward hairpin substrates and a reduced reactivity toward VS substrates consistent with the expected shift in the equilibrium between the two folds favoring the hairpin structure. Mutation

^{HP4}C provides an additional base-pair to the hairpin secondary structure but destabilizes the essential junction J_{236} in the secondary structure of the VS (see chapter 2). This leads to an abrupt elimination of VS activity. Mutation $^{VS1}\Delta$ promptly abolishes hairpin activity as it involves deleting a conserved residue in the active site of the hairpin fold, maintaining the integrity of the VS fold. Predicted secondary structures corresponding to hairpin and VS fold (Figure 5.7) are consistent with the effects of directed mutations on the catalytic functions of HP_INT1_VS.

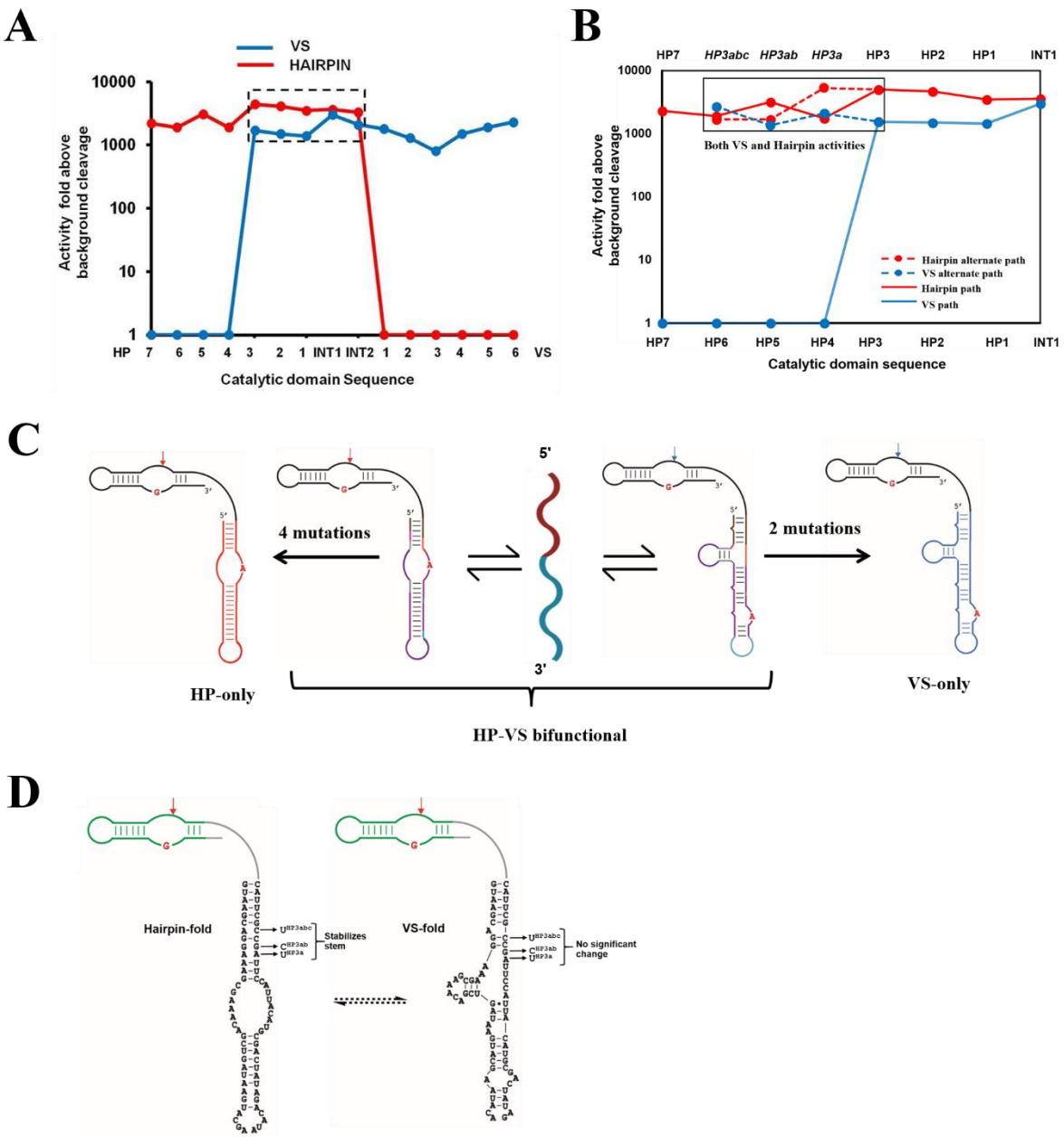

Figure 5.12 Smooth access to hairpin and VS functions from HP_INT1_VS. A. Neutral paths corresponding to hairpin and VS functions intersect across multiple sequences. **B.** Certain mutations preserve neutrality in the RNA sequence space. Mutations to HP3 that destabilize the VS fold completely abolishes VS activity (HP4 to HP7, in A). However, other mutations can lead bifunctional sequences that have comparable VS and hairpin activities - HP3a, HP3b, HP3c shown on the x-axis on top (in italics; corresponding paths are shown in dotted lines). Dual activities of these sequences illustrate alternative neutral paths that can diverge from one point in a neutral network. **C.** Distinct hairpin and VS phenotypes can be accessed through smooth mutational paths diverging from bifunctional, HP_INT1_VS. **D.** Description of point mutations that preserve dual activity discussed in B.

5.1.4 Hairpin and VS substrates are connected by neutral paths

The chemical step of catalysis is preceded by substrate docking, mediated by tertiary interactions between the substrate and catalytic domains of the VS and hairpin ribozymes (Wilson and Lilley, 2011). Since the intersection sequence, HP_INT1_VS cleaves both VS and hairpin substrates, we characterized possible mutational pathways that connect the two substrate domains. As the substrates have identical stems and differ only in the sequences of their conserved bulges, we designed several five-step pathways consisting of substrate variants with mutations in their conserved bulges that includes transitions, transversions, point additions and deletions (Figure 5.13). Each step introduces a single point mutation to the substrate to provide a smooth transition from one substrate to another across a neutral path, where each mutation preserves some features of the source sequence and acquires new features of the target. Wild-type hairpin and VS ribozymes have a specificity for cleavage loop sequences and especially the nucleotide immediately downstream of the cleavage site (Fedor, 2000, Lilley, 2004). This requirement is relaxed in the context of a bifunctional catalytic domain, and mutations to the cleavage site of the hairpin substrate reduce cleavage rates by ~ 3-fold. Active substrate variants that differ from their immediate neighboring sequences by a single mutation (Hamming distance = 1) connect the catalytic domain variants favoring either hairpin or VS cleavage. This provides a smooth transition between complete cis-ribozymes (Figure 5.14). Encouraged by the success in designing neutral paths consisting of substrate variants that connect hairpin and VS functionality between HP_INT1_VS-HP and HP_INT1_VS-VS, we created cis-ribozyme constructs by fusing each of the 15 catalytic domain sequences with 7 substrate sequences that include the canonical hairpin and VS sequences, in addition to the 5 most active intermediates (S1-S5) connecting them, to

generate 105 cis cleaving ribozymes. These constructs set the stage for interrogating the landscape that borders VS and hairpin functions.

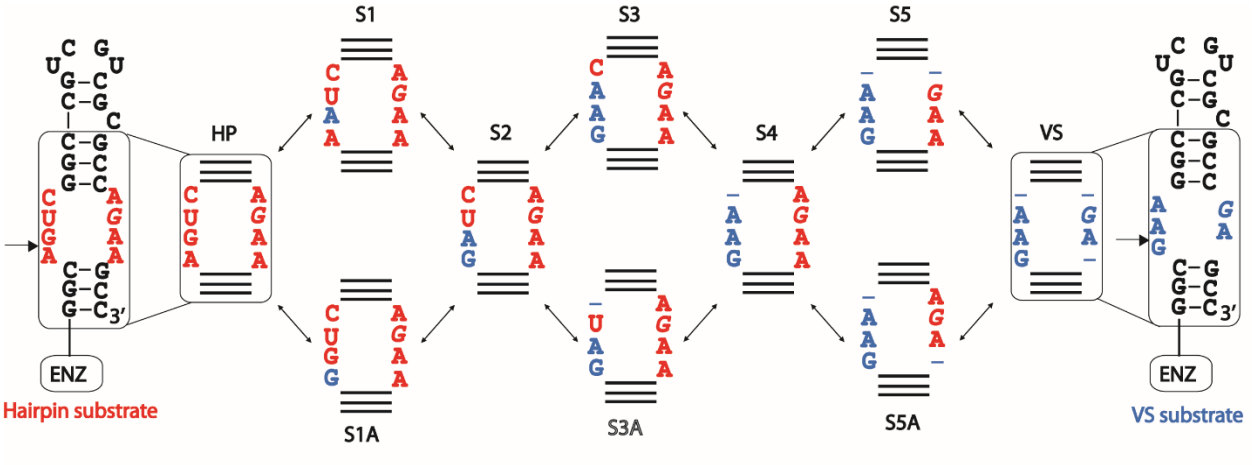

Figure 5.13 Mutational paths connect hairpin and VS substrates. Hairpin and VS substrates can be interconverted by step-wise single point mutations. S3 represents a central intermediate in this five-step mutational walk, whereas S1, S2 and S4, S5 approximately resembles the canonical hairpin and VS substrates respectively. Nucleotides that resemble a hairpin cleavage loop are shown in red and those resembling a VS cleavage loop are shown in blue. ENZ indicates catalytic domain linked to these substrate variants.

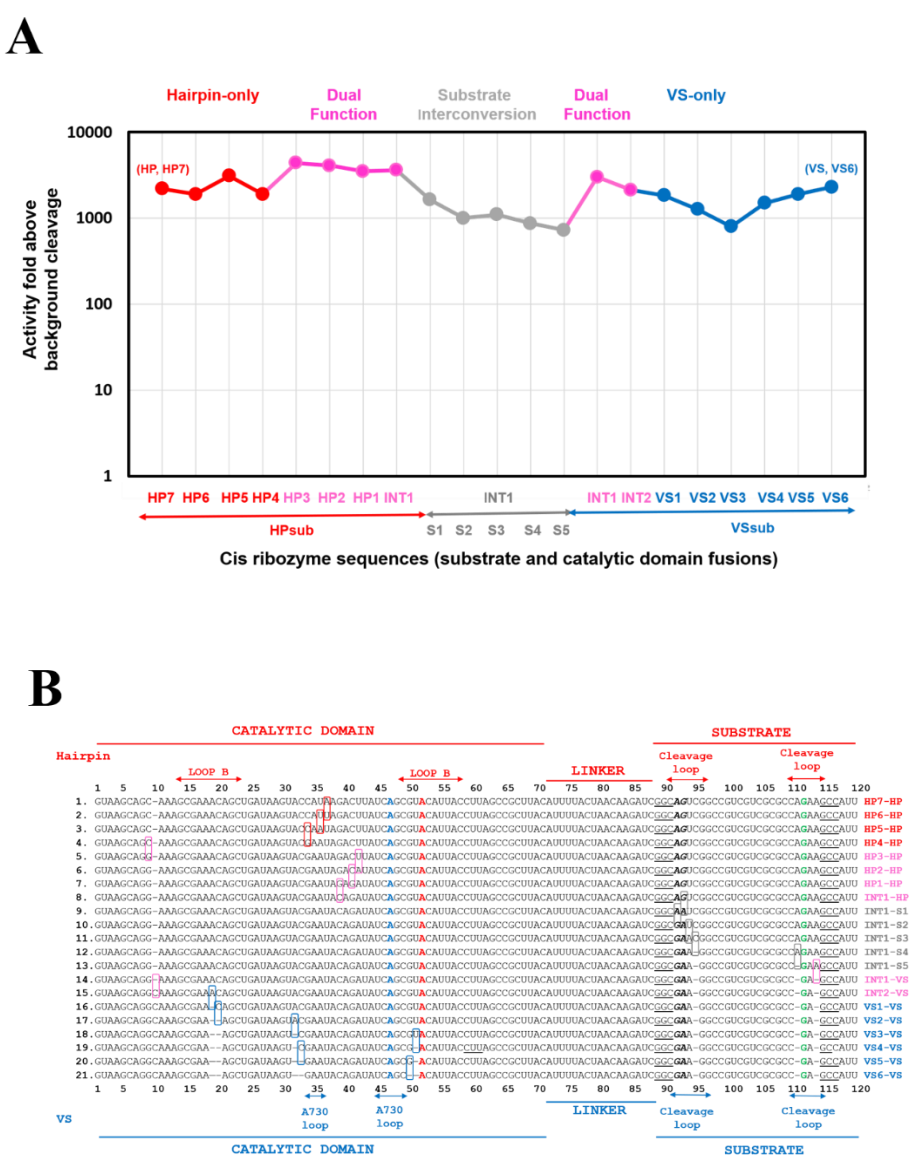

Figure 5.14 Interconversion of complete minimal forms of the hairpin and VS ribozymes. A. Complete cis-cleaving minimal forms of the hairpin and VS ribozymes can be interconverted by single-step mutations to the substrate and catalytic domains. Such a mutational path is evolutionarily feasible as each sequence in this path is active for site-specific RNA cleavage. Red dots represent sequences that exclusively favor hairpin cleavage, blue dots represent sequences that exclusively favor VS cleavage and pink dots represent sequences that exhibit dual VS and hairpin activities. Grey dots represent substrate sequence variants fused to HP_INT1_VS. **B.** Sequence alignment for the intermediates in the neutral path in (A) with important domains labeled. Putative catalytic residues that are important in wild-type sequences of the hairpin and VS ribozymes are colored: catalytic G is shown in green for both phenotypes, catalytic A in the VS fold is shown in blue and hairpin fold in red. Point mutations that separate each sequence are highlighted by colored boxes that represents the predominant phenotype (coloring scheme as in A).

5.1.5 Local fitness landscape at the intersection of hairpin and VS function

We measured cleavage rates for all 105 (15 catalytic domain variants X 7 substrate variants) cis-ribozyme constructs. Collectively, these results define a local fitness landscape that reveal the smoothness of the connection between the hairpin and VS ribozymes networks (Figure 5.15). In agreement with the general observation that functional RNA sequences are found in clusters surrounded by non-functional sequences in their relevant fitness landscapes (Jiménez et al., 2013), active cis-cleaving ribozyme constructs populate the vicinity of the intersection sequence that is active for the cleavage of both VS and hairpin substrates as well as each of the five substrate variants to different degrees. Catalytic compatibility between catalytic domains and substrate variants becomes weaker on moving up (toward VS-like catalytic domains) or down (toward hairpin-like catalytic domain) from the intersection sequence. This is likely due to decrease in functional plasticity of the catalytic domains that is required to catalyze cleavage of substrate variants with sequence features intermediate to the canonical VS and hairpin substrates. Mutants one, two and three mutations away from the intersection sequence toward hairpin function (HP1, HP2, HP3) that involve preferential stabilization of the hairpin fold without disturbing their alternate VS fold, show gradual increase in hairpin cleavage accompanied by decrease in VS cleavage. However, mutations HP4-HP7 specifically destabilize the VS fold, which is reflected in the immediate abolition of VS cleavage, while still supporting hairpin function. A single mutation to the intersection sequence, INT1 yields a variant INT2 that preserves dual function; however, further mutations involving deletions to certain nucleotides that constitute the conserved loop B in the hairpin fold but are part of a non-essential hexaloop in the VS fold, abrogate hairpin activity while preserving the ability to cleave the VS substrate. The loss of either VS or hairpin function most likely leads to gain in specificity toward the cognate substrate of the active ribozyme. The

intersection sequence (INT1) and the catalytic domain sequences around it (HP1-HP3 and INT2) support functional interconversion between the VS and the hairpin ribozymes, revealing points of intersection of several neutral networks corresponding to the VS and hairpin ribozymes, and establish a smooth connection between two distinct endonucleolytic motifs. Cleavage rates of all sequences involved in mapping this landscape are listed in tables 5.1 and 5.2.

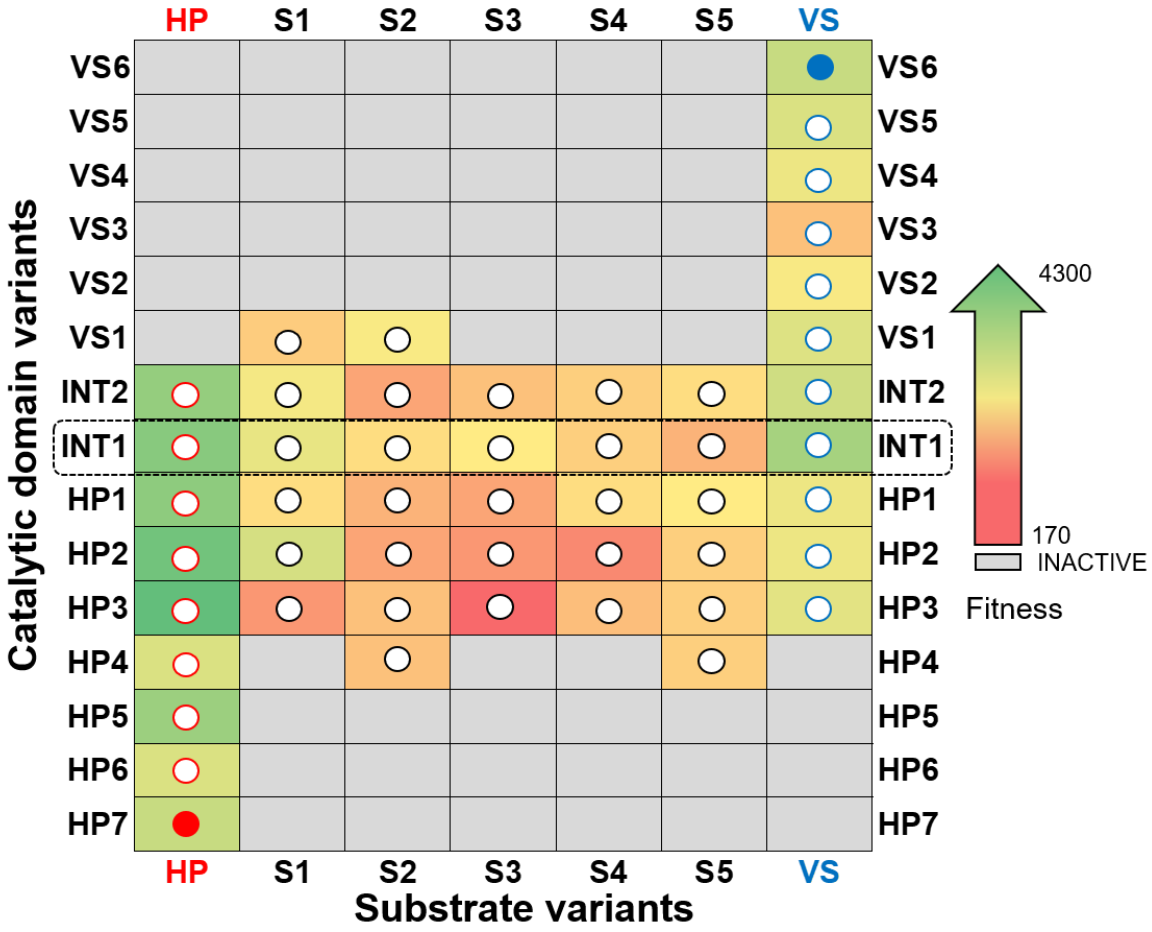

Figure 5.15 Local fitness landscape at the intersection of hairpin and VS functions.
Site-specific RNA cleavage activity (fitness) plotted against sequence variants of the substrate (x-axis) and catalytic domains (y-axis). The two-dimensional heat map represents intersection of functional landscapes of the hairpin and VS ribozymes, where functional interconversion is brought about through neutral pathways involving point mutants of the substrate and catalytic domains. The landscape is divided into two distinct regions: the VS-only (top) and HP-only (bottom) spaces, separated by sequences that are catalytically active for the cleavage of both hairpin and VS substrates including their sequence variants. The paths between sequences with exclusive hairpin activity (co-ordinates (HP, HP7) shown as a red dot) and exclusive VS activity (co-ordinates (VS, VS6) shown as a blue dot) are neutral. Grey squares indicate non-functional sequences. Cis ribozyme with INT1 as their catalytic domain are highlighted by a box.

Sequence	Hairpin activity	VS activity
Sequences involved in hairpin-VS functional interconversion (Figures 5.12)		
	Mean	***Mean***
HP7	0.023±0.002	Inactive
HP6	0.019±0.002	Inactive
HP5	0.0315±0.0007	Inactive
HP4	0.019±0.002	Inactive
HP3	0.043±0.006	0.017±0.001
HP2	0.041±0.007	0.015±0.001
HP1	0.035±0.004	0.014±0.002
INT1	0.036±0.003	0.030±0.005
INT2	0.033±0.006	0.021±0.004
VS1	Inactive	0.018±0.002
VS2	Inactive	0.013±0.002
VS3	Inactive	0.008±0.001
VS4	Inactive	0.015±0.003
VS5	Inactive	0.019±0.003
VS6	Inactive	0.023±0.004
Catalytic domain variants the HP functional space that retain dual function (Figure 5.12B)		
HP3a	0.057±0.003	0.030±0.009
HP3ab	0.017±0.003	0.014±0.004
HP3abc	0.017±0.001	0.027±0.003
Catalytic residue mutants (Figure 5.8 E, F)		
INT1-VS_GtoA	-	0.009±0.002
INT1-VS_AtoG	-	0.004±0.001
INT1-HP_GtoA	0.008±0.002	-
INT1-HP_AtoG	0.0045±0.0021	-
Substrate variants with lower activities (Figure 5.13)		
INT1-S1a (GGUC-AGAA)	0.012±0.003	0.012±0.003
INT1-S3a (GAU-AGAA)	0.00019±0.00001	0.00019±0.00001
INT1-S5a (GAA-AGA)	0.0049±0.0003	0.0049±0.0003

Table 5.1 Cleavage rate constants (min^{-1}) of RNA sequences in the exploration of intersecting neutral networks of hairpin and VS ribozymes

Sequence	Activity	Sequence	Activity
	Mean		***Mean***
HP7-S1	Inactive	INT2-S1	0.027±0.010
HP7-S2	Inactive	INT2-S2	0.014±0.003
HP7-S3	Inactive	INT2-S3	0.007±0.003
HP7-S4	Inactive	INT2-S4	0.018±0.004
HP7-S5	Inactive	INT2-S5	0.021±0.006
HP6-S1	Inactive	VS1-S1	0.009±0.004
HP6-S2	Inactive	VS1-S2	0.012±0.003
HP6-S3	Inactive	VS1-S3	Inactive
HP6-S4	Inactive	VS1-S4	Inactive
HP6-S5	Inactive	VS1-S5	Inactive
HP5-S1	Inactive	VS2-S1	Inactive
HP5-S2	Inactive	VS2-S2	Inactive
HP5-S3	Inactive	VS2-S3	Inactive
HP5-S4	Inactive	VS2-S4	Inactive
HP5-S5	Inactive	VS2-S5	Inactive
HP4-S1	Inactive	VS3-S1	Inactive
HP4-S2	0.008±0.001	VS3-S2	Inactive
HP4-S3	Inactive	VS3-S3	Inactive
HP4-S4	Inactive	VS3-S4	Inactive
HP4-S5	0.009±0.003	VS3-S5	Inactive
HP3-S1	0.005±0.003	VS4-S1	Inactive
HP3-S2	0.008±0.001	VS4-S2	Inactive
HP3-S3	0.0016±0.0009	VS4-S3	Inactive
HP3-S4	0.0078±0.0005	VS4-S4	Inactive
HP3-S5	0.010±0.002	VS4-S5	Inactive
HP2-S1	0.020±0.005	VS5-S1	Inactive
HP2-S2	0.005±0.002	VS5-S2	Inactive
HP2-S3	0.0050±0.0003	VS5-S3	Inactive
HP2-S4	0.003±0.002	VS5-S4	Inactive
HP2-S5	0.009±0.001	VS5-S5	Inactive
HP1-S1	0.009±0.002	VS6-S1	Inactive
HP1-S2	0.0058±0.0007	VS6-S2	Inactive
HP1-S3	0.006±0.001	VS6-S3	Inactive
HP1-S4	0.010±0.003	VS6-S4	Inactive
HP1-S5	0.0110±0.0006	VS6-S5	Inactive
INT1-S1	0.016±0.002		
INT1-S2	0.010±0.003		
INT1-S3	0.011±0.002		
INT1-S4	0.008±0.002		
INT1-S5	0.0073±0.0005		

Table 5.2 Cleavage rate constants (min^{-1}) of RNA sequences in the exploration of intersecting neutral networks of hairpin and VS ribozymes: figure 5.15.

5.1.6 Bifunctional minimal sequences can access wild-type features by mutation and domain accretion

VS and hairpin ribozyme sequences found in biology contain domains that are absent in the minimal forms used in this work, which in addition to facilitating substrate binding, assist the sequences to fold into single functional folds. The domains eliminated from the wild-type hairpin ribozyme, however, do not contribute to cleavage (Fedor, 2000). The effects of evolved catalytic capabilities and escaping the restraints of dual function most likely allow wild-type ribozymes to catalyze RNA cleavage faster with rate accelerations of 10^5 or more (Lilley and Eckstein, 2008). Simple mutational operations involving point mutations to non-essential hairpin loops, deletions of unpaired nucleotides in single stranded regions and base pairs in stems, and pairwise base pair mutations in stems, all of which commonly occur in evolution, convert the intersection sequence (HP_INT1_VS) to minimal forms of the VS ribozyme that exclusively cleave VS substrate exclusively with increased rates (Figure 5.16). Addition of domains (stem-loops) to the catalytic domain of the minimal VS construct yields sequences that resemble the wild-type ribozyme with enhanced cleavage activities (Figure 5.17). During the transition from cis-cleaving systems to trans systems, we added junction 3-4-5 to the minimal catalytic domain of the VS ribozyme in the form of two sequence variants, one with a wild-type helix 4 but a truncated helix 5 and the other with wild-type helix 5 and a truncated helix 4 (Figure 5.16). This enables direct comparison to well-studied trans VS ribozyme constructs that cleave VS substrates with rates of ~0.7/min (Guo and Collins, 1995). Interestingly, the catalytic domain sequence with helix 5 truncated to just three base-pairs is inactive for trans cleavage; however, a variant with full-length helix 5 but a truncated helix 4 exhibits robust trans cleavage. These observations are consistent with the role of stem-loop 5 in the binding and activation of the VS substrate via the kissing-loop interaction (Rastogi et al.,

1996; Lafontaine et al., 2002). Restoring wild-type domain 3-4-5 further increases trans cleavage. Addition of helix 7, a domain that attenuates cleavage by favoring the reverse reaction, ligation, exhibits slightly reduced cleavage rates. However, this sequence is about 2-fold faster than the wild-type sequence (Figure 5.17), which might reflect the effect of a circular permutation in this sequence that differentiates it from a canonical wild-type catalytic domain.

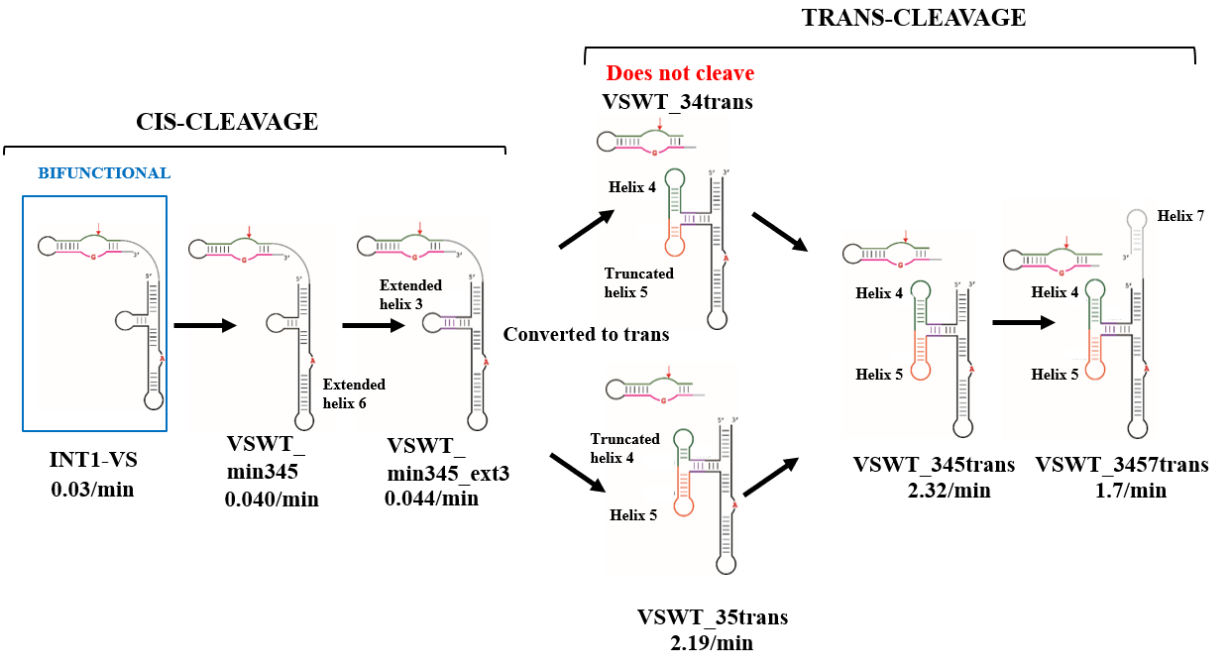

Figure 5.16 Addition of nucleotides can restore wild-type activity to a bifunctional VS ribozyme. Catalytic activity of HP_INT1_VS-VS can be improved by nucleotide additions and mutations to helices 3 and 6, thereby making its catalytic core resemble the wild-type sequence. Converting the catalytic domain into a trans ribozyme (VSWT_min345_ext3 to VSWT_35trans) enables comparison to the extensively studied trans VS ribozyme constructs. Further nucleotide additions in form of extensions to helix 4 and addition of helix 7 preserve robust trans cleavage. VSWT_3457trans resembles a full-length, wild-type catalytic domain and is related to it by a circular permutation.

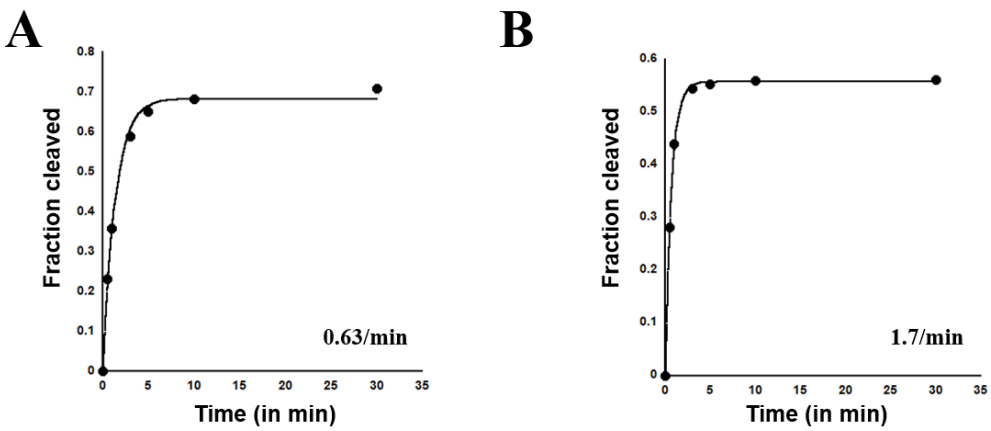

Figure 5.17 Kinetics of trans cleavage by wild-type VS ribozyme (VSrz) and VSWT_3457trans. A. Fraction cleaved plotted against time for trans cleavage by VSrz. **B.** Fraction cleaved plotted against time for trans cleavage by VSWT_3457trans.

Similar operations convert the bifunctional catalytic domain, HP_INT1_VS into an efficient trans-acting hairpin ribozyme, HPTransRz (Figures 5.18). This was accomplished primarily through complimentary mutations and deletions to the base-paired regions of the hairpin fold of HP_INT1_VS.

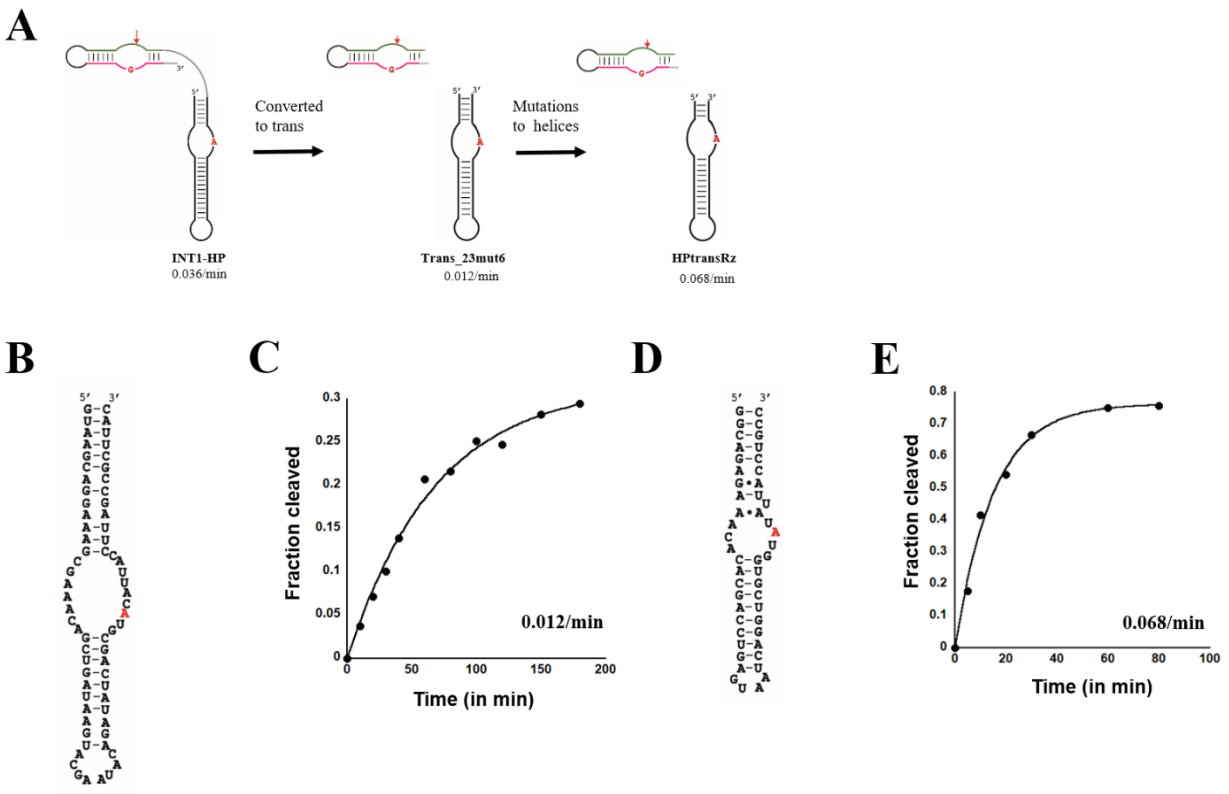

Figure 5.18 Mutations to the stems in HP_INT1_VS enhance its hairpin trans cleavage activity. A. From a bifunctional cis cleaving hairpin ribozyme to trans cleaving ribozyme with enhanced activity. **B.** Secondary structure corresponding to the hairpin fold of the intersection sequence, HP_INT1_VS (see chapter 4). **C.** Fraction cleaved plotted against time for trans cleavage of a hairpin substrate by HP_INT1_VS. **D.** Secondary structure of a trans-acting hairpin ribozyme, HPTransRz that resembles the catalytic domain (HPt) of the prototype hairpin ribozyme sequence (see chapter 4). **E.** Fraction cleaved plotted against time for hairpin trans cleavage by HPTransRz.

These operations mirror ‘evolution by domain accretion’, a phenomenon observed in several functional RNAs (Bokov and Steinberg, 2009; Fox et al., 1977; Petrov et al., 2015; Lilley and Eckstein, 2008). Neutral paths incorporating these changes in a step-wise manner can be easily envisioned that connect minimal forms of VS and hairpin ribozymes to their respective wild-type sequences, thereby establishing a link between two naturally occurring, apparently dissimilar ribozymes and unequivocally demonstrate that their neutral networks intersect (Figure 5.19). This defines a pathway by which two distinct ribonuclease ribozymes could have emerged by evolutionary divergence from a common ancestor.

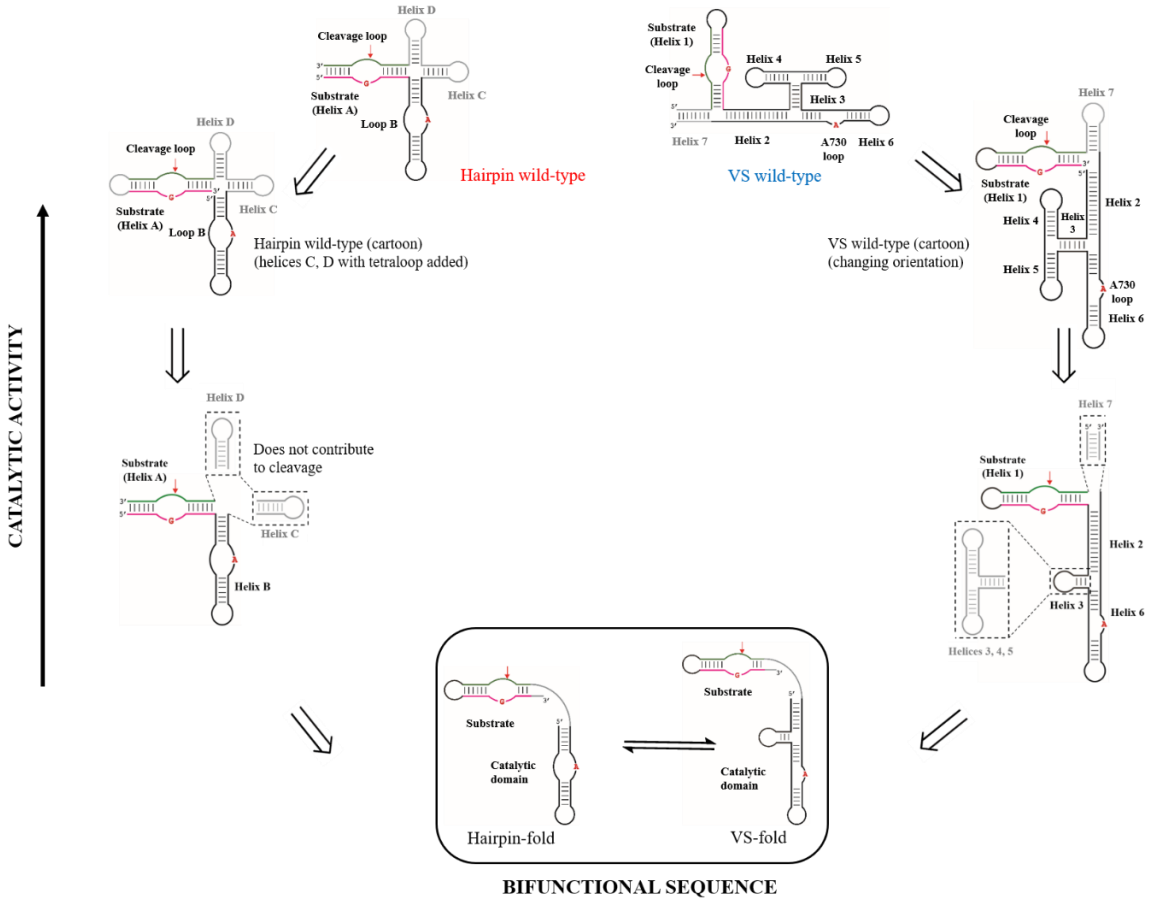

Figure 5.19 Emergence of distinct wild-type hairpin and VS ribozymes from structurally and functionally plastic catalytic ancestor sequences. Bifunctional sequences exhibiting conformational plasticity can evolve to wild-type hairpin of VS ribozyme sequences by mutations and domain accretion producing ribozymes with well-defined tertiary folds and more efficient catalytic abilities.

5.1.7 Sequence variants of the hammerhead self-cleaving motif catalyze cleavage of a hairpin substrate *in cis*

The hammerhead ribozyme is a self-contained endonucleolytic motif. It is the smallest and simplest of all known natural ribozymes. Sequence variants of this motif have emerged from *in vitro* selection experiments designed to evolve self-cleaving ribozymes (Conaty et al., 1999; Tang and Breaker, 2000; Salehi-Ashtiani and Szostak, 2001; Popović et al., 2015), leading to the speculation that the hammerhead constitutes the simplest solution for site-specific ribonuclease activity that can be created with an RNA framework. Nucleotides essential for cleavage activity are found within the core of the ribozyme located at a three-way junction. Folding random 100 nt RNA sequences in a computational grid has revealed that ~1.6 X 10^{10} molecules must be searched to encounter a hammerhead motif with 50% probability (Knight et al., 2005). This result is consistent with the emergence of the hammerhead motif from selection experiments that usually involve 10^{13}-10^{15} distinct molecules. Informed by the consensus sequence and secondary structure of the hammerhead (Figure 5.20A) and hairpin (Figure 5.20B) ribozymes, we designed a single sequence, HH_INT_HP that could be threaded through the secondary structure of the complete hammerhead ribozyme and the catalytic domain of the hairpin ribozyme (Figure 5.21). This sequence catalyzes autocatalytic hammerhead cleavage and simultaneously catalyzes the cleavage of a hairpin substrate *in cis* when the two domains are connected by a linker (Figure 5.22). The approach used to achieve this was similar to that used for designing a dual HP/VS catalytic sequence (Figure 4.7). The cis-cleaving construct (HH_INT_HP-HP) consisting of HH_INT_HP and HPsub, connected via a 17 nt linker, generates two cleavage products. These correspond to cleavage of the hairpin substrate as HH_INT_HP assumes the functional fold of the hairpin catalytic domain, and the self-cleavage of the hammerhead motif, because this sequence is able to

fold into the hammerhead structure, independent of sequence context. This is consistent with the use of hammerhead ribozyme sequences at the 3' ends of transcripts to generate homogeneous ends during *in vitro* transcription (Price et al., 1995).

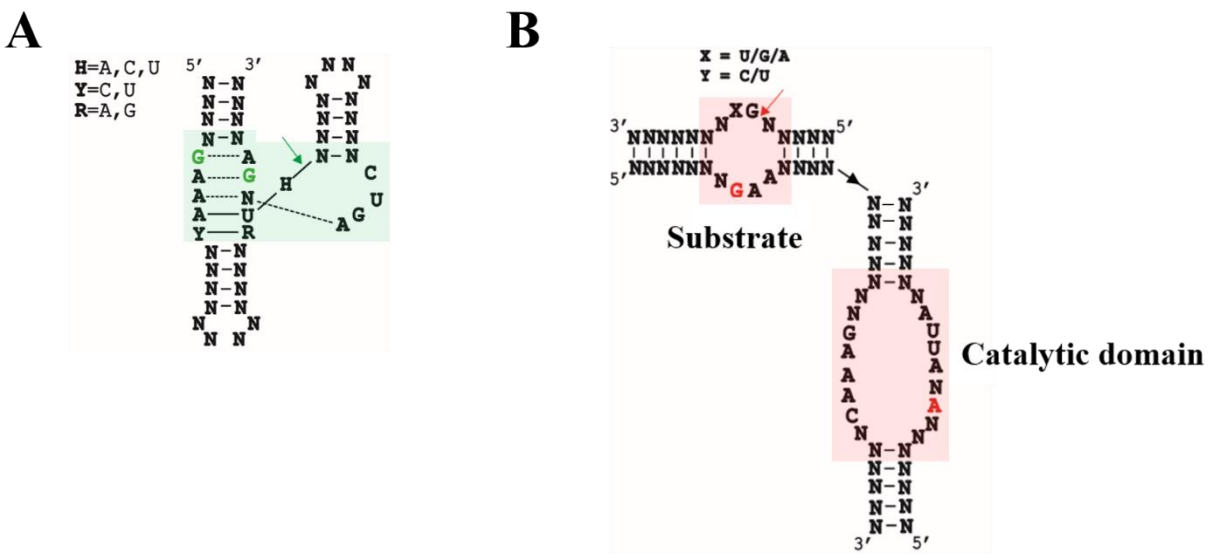

Figure 5.20 Consensus sequence for the hammerhead and hairpin ribozymes.
A. Consensus sequence for the hammerhead ribozyme. Catalytic nucleotides are shown in green, cleavage site is indicated by a green arrow. **B.** Consensus sequence for the hairpin ribozyme. Catalytic nucleotides are shown in red, cleavage site is indicated by a red arrow. Conserved structural features are highlighted by colored boxes (green for hammerhead and red for hairpin).

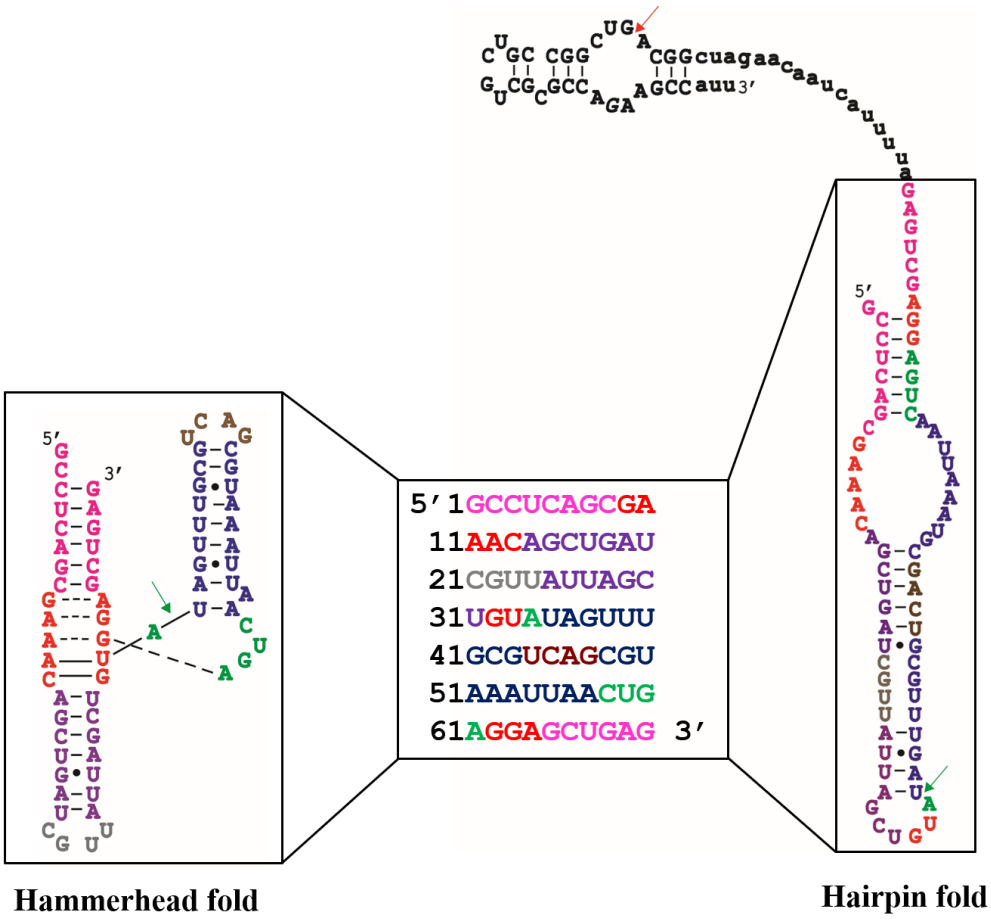

Figure 5.21 Bifunctional sequence (HH_INT_HP) that can be threaded through secondary structures of both the hammerhead self-cleaving ribozyme and the catalytic domain of the hairpin ribozyme. HH_INT_HP can assume the secondary structures of both hammerhead ribozyme and the catalytic domain of the hairpin ribozyme. These alternate secondary structures have comparable thermodynamic stabilities. The sequence is color coded according to domains of the hammerhead secondary structure and secondary structure of the hairpin fold reflects the same color code. The hammerhead and hairpin cleavage sites are denoted by green and red arrows respectively. Since HH_INT_HP can assume the functional folds of the hammerhead self-cleaving motif and the catalytic domain of the hairpin ribozyme simultaneously, the fusion of HH_INT_HP with HPsub, results in two cleavage products generated from hammerhead and hairpin cleavage; the respective cleavage sites in the same molecule are indicated by green and red arrows, respectively.

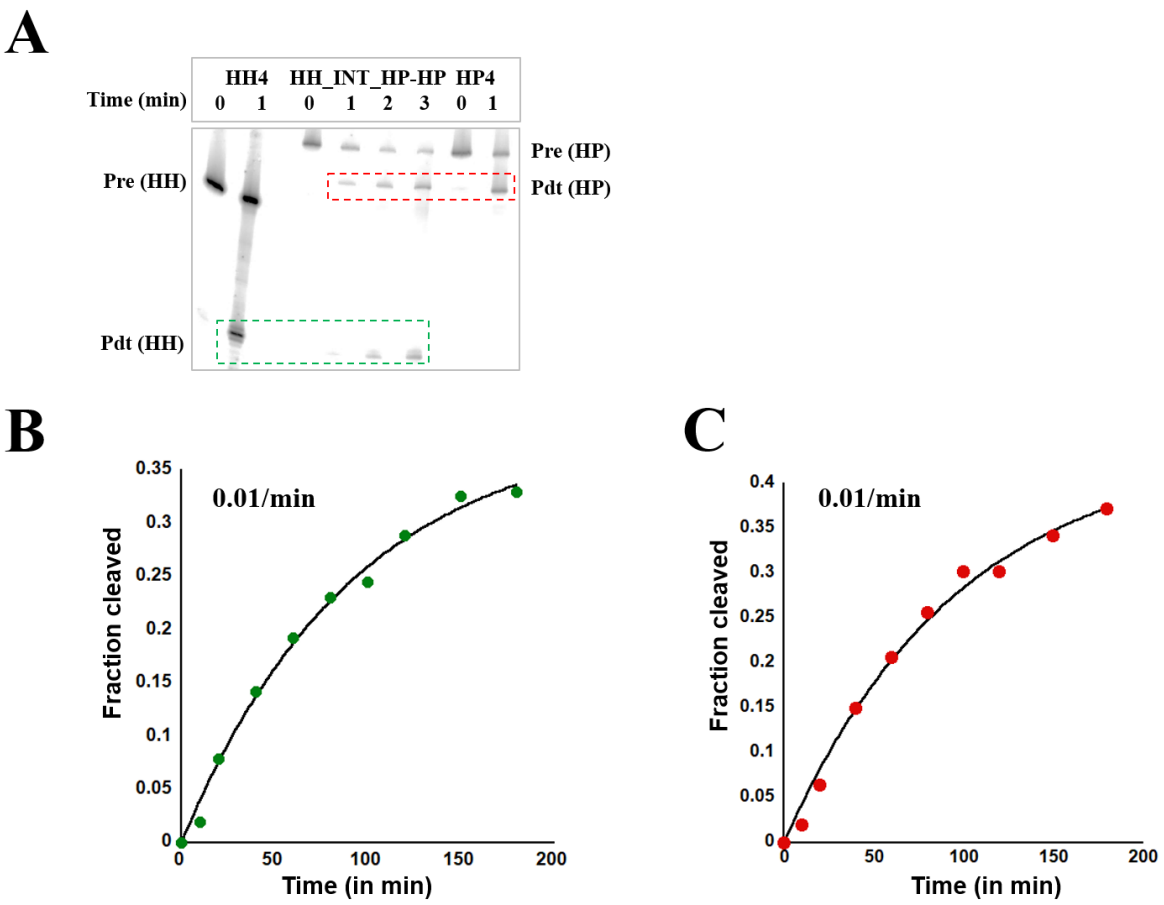

Figure 5.22 HH_INT_HP catalyzes autocatalytic hammerhead cleavage and the cleavage of a hairpin substrate *in cis*. A. Dual hammerhead and hairpin cleavage activity of HH_INT_HP. Hairpin and hammerhead cleavage products are highlighted by red and green boxes. Respectively. **B.** Fraction cleaved plotted against time for cis cleavage of HH_INT_HP. **D.** Fraction cleaved plotted against time for cis cleavage of HH_INT_HP-HP.

5.1.8 Functional interconversion between hammerhead and hairpin ribozymes through 'neutral' mutational drifts

HH_INT_HP catalyzes 'hammerhead' self-cleavage and cleaves a hairpin substrate *in cis*. Its dual function generates cleavage products from both hairpin and hammerhead cleavage (Figure 5.22A). Mutations to this bifunctional sequence (Figures 5.23, 5.24) can eliminate the activity of either fold, which is accompanied by an increase in the activity of the other fold (Figure 5.25A, B). The first set of point mutations to HH_INT_HP preferentially destabilizes the hairpin-fold by disrupting the base paired helix immediately under the conserved internal bulge. Disrupting the A-U base-pair distant to the bulge (^{HH1}A), preserves a four base-paired helix under the internal bulge thereby preserving hairpin activity. However, further mutations to base-pairs proximal to the internal bulge destabilize the stem likely resulting in local misfolding. This eliminates hairpin activity without affecting hammerhead activity since the mutated nucleotides are located in a loop region of the hammerhead fold of HH_INT_HP, hence largely interchangeable. The reduction and eventual elimination of hairpin activity is accompanied by an increase in hammerhead activity possibly reflecting a shift in conformational equilibrium in favor of the hammerhead fold. Similarly, a second set of four point mutations to nucleotides that correspond to the other loop domain of the hammerhead fold of HH_INT_HP, preferentially stabilize the hairpin fold of HH_INT_HP by introducing additional base-pairs in its stem. This likely shifts the equilibrium in favor of the more stable hairpin fold, resulting in an approximate increase in hairpin activity across all four mutations while completely abrogating hammerhead cleavage. This toggling between hammerhead and hairpin function in response to point mutations present a molecular switch in the mutational path between the two ribozymes and could have potential synthetic biology applications in designing genetic circuits where point mutations can regulate cleavage at distant

locations. This switching can also be induced by changing reaction conditions (200 mM Mg^{2+} to 4 M NH_4^+) to favor hairpin cleavage (Figure 5.25D, E). This could have implications for evolutionary divergence resulting from a differential response to changing environmental conditions.

HP_INT1_VS exhibited an attenuated dependence on the catalytic nucleotides in its hairpin and VS fold for cleavage of their respective substrates (Figure 5.8E, F). In contrast, HH_INT_HP self-cleavage (hammerhead function) was abolished when either catalytic guanine ('G8' or 'G12' in wild-type sequences) was mutated to an adenine (Figure 5.26A). Mutations to catalytic nucleotides in HH4 (exclusive hammerhead function) resulted in a similar loss of hammerhead function (Figure 5.26B). This could suggest that the hammerhead fold of the bifunctional sequence retains the catalytic mechanism of the wild-type ribozyme further underscoring its robustness as a catalytic motif.

Therefore, HH_INT_HP represents a point of intersection of neutral paths corresponding to hammerhead and hairpin function. Paralleling the effect of mutations to a hairpin-VS intersection sequence (HP_INT1_VS), exclusive hammerhead or hairpin functions were encountered only after one (HP-only) or two (HH-only) point mutations (Figure 5.25). Much like the hairpin-VS dual activity profile, the hammerhead-hairpin dual activity profile further highlights the importance of phenotypic plasticity exhibited by intersection sequences (HP_INT1_VS and HH_INT_HP) in the acquisition of distinct functionality. Table 5.2 lists the cleavage rates for hammerhead and hairpin cleavage for the sequences in figure 5.25B.

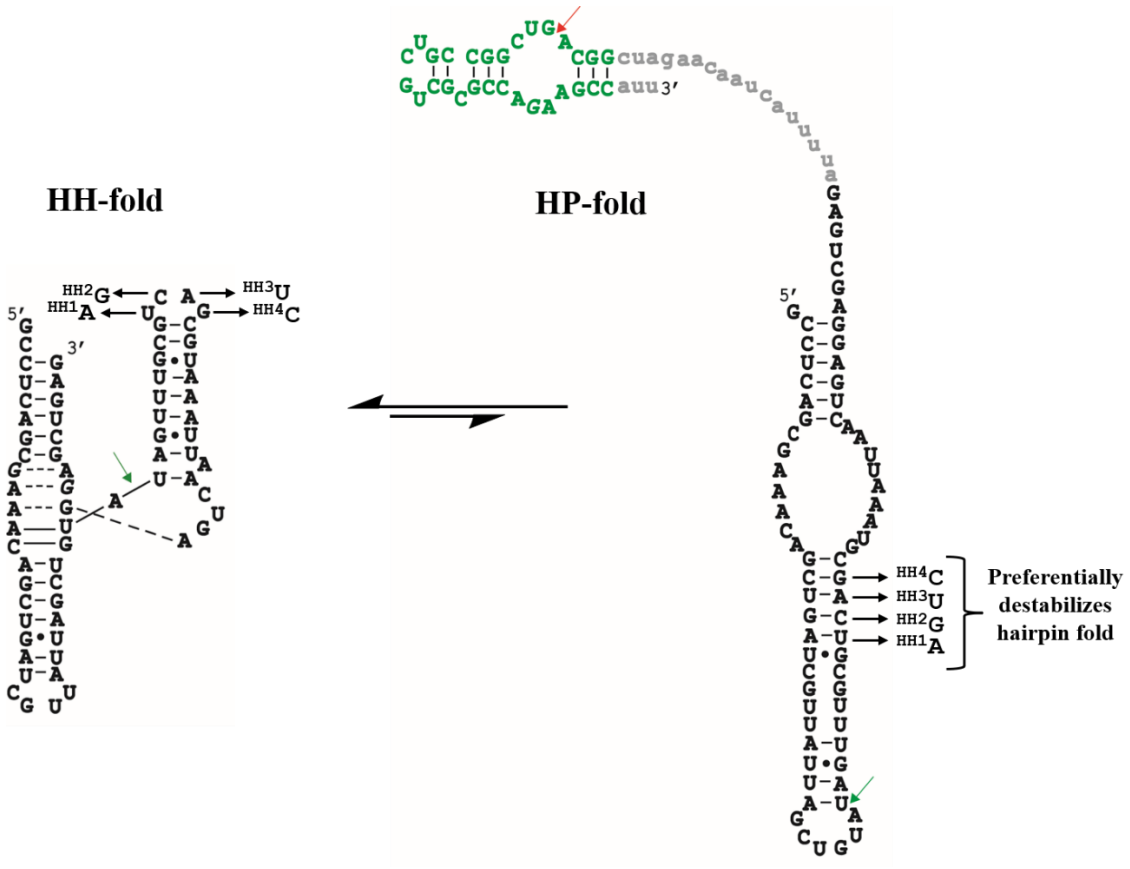

Figure 5.23 Mutational drift from bifunctional HH_INT_HP to exclusive hammerhead function. Directed point mutations to HH_INT_HP destabilize the hairpin fold, while the corresponding hammerhead fold is relatively unaffected. This converts the bifunctional sequence to an exclusive hammerhead ribozyme. Equilibrium arrow represents equilibrium between the hammerhead self-cleaving motif and the catalytic domain of the hairpin ribozyme.

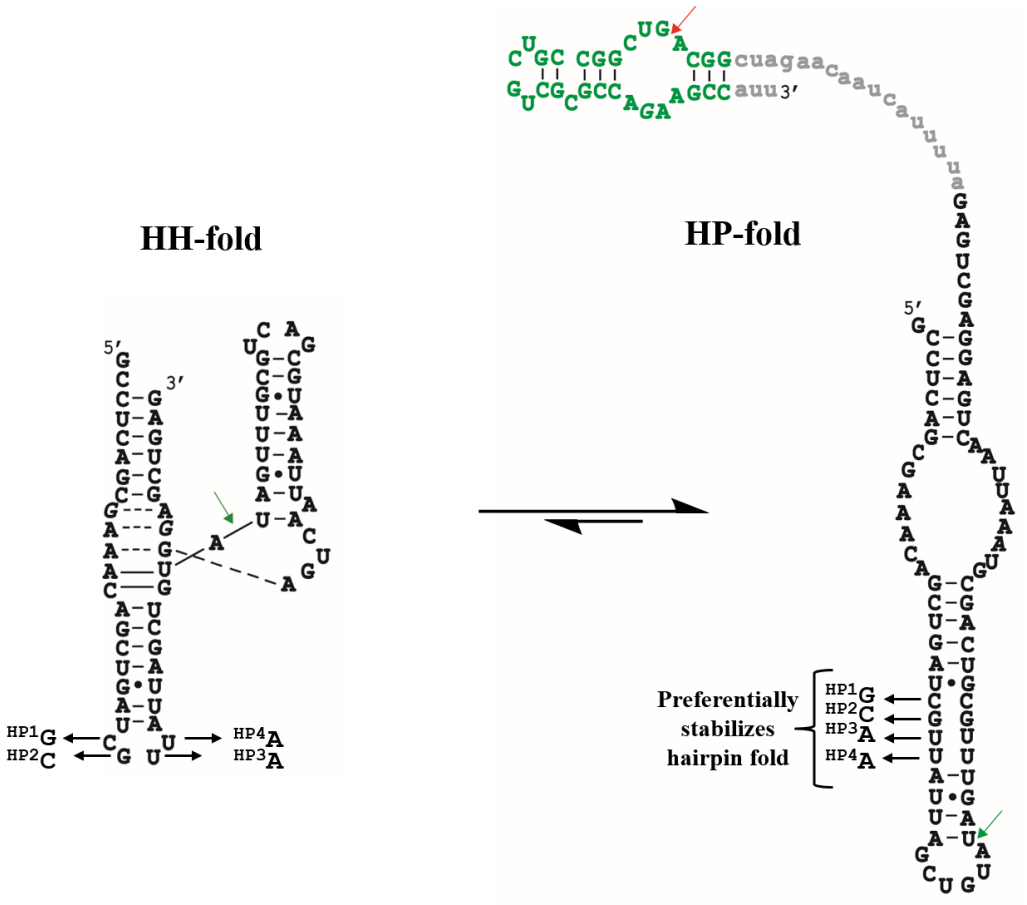

Figure 5.24 Mutational drift from bifunctional HH_INT_HP to exclusive hairpin function. Directed point mutations to HH_INT_HP stabilize the hairpin fold, while the corresponding hammerhead fold is relatively unaffected. This converts the bifunctional sequence to a hairpin ribozyme catalytic domain. Equilibrium arrow represents equilibrium between the hammerhead self-cleaving motif and the catalytic domain of the hairpin ribozyme.

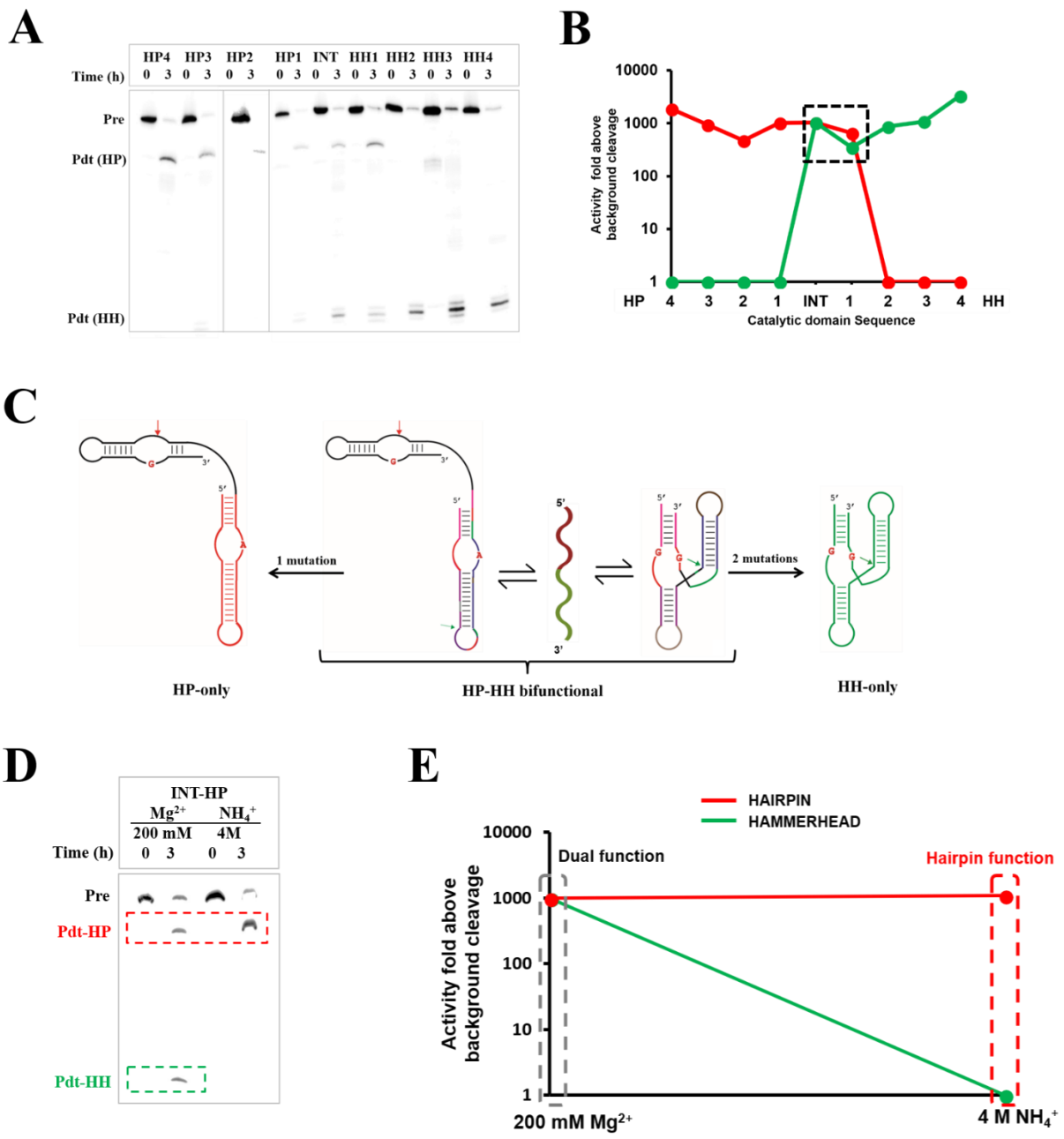

Figure 5.25 Smooth access to hairpin and hammerhead functions from HH_INT_HP. A. Mutations to the intersection sequence, HH_INT_HP, gradually eliminate either hammerhead (left of INT) or hairpin (right of INT). **B.** Neutral paths corresponding to hairpin and hammerhead functions intersect across two sequences. **C.** Distinct hairpin and hammerhead phenotypes can be accessed through smooth mutational paths diverging from HH_INT_HP that supports both function. **D.** Replacing 200 mM Mg^{2+} with 4 M NH_4^+ eliminates hammerhead cleavage but retains hairpin cleavage. **E.** Rates of hairpin cleavage in presence of 200 mM Mg^{2+} and 4 M NH_4^+ are comparable.

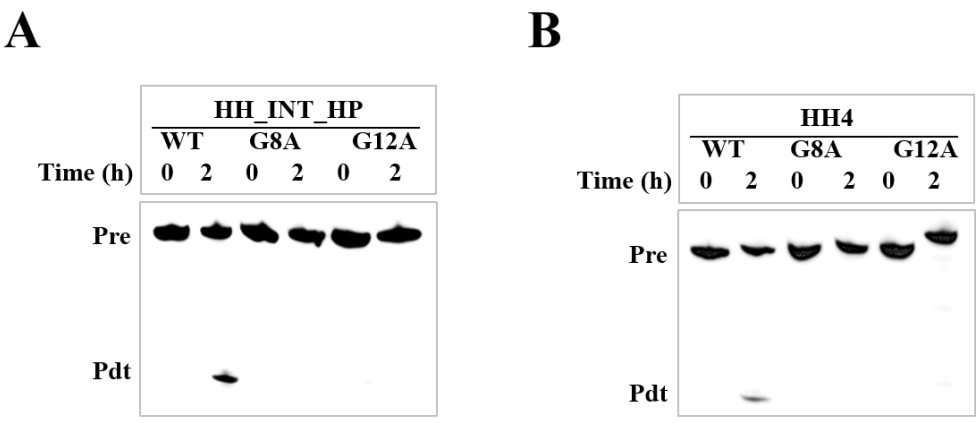

Figure 5.26 Mutating catalytic nucleotides in HH_INT_HP and HH4 abolish hammerhead activity. **A.** G8A and G12A mutations to bifunctional HH_INT_HP abolish hammerhead self-cleavage. **B.** G8A and G12A mutations to hammerhead sequence, HH4 abolish hammerhead self-cleavage.

Sequence	Hairpin activity	Hammerhead activity	VS activity
Sequences involved in intersecting neutral networks in hammerhead-hairpin interconversion (Figure 5.25)			
	Mean	***Mean***	***Mean***
HP4	0.019±0.002	Inactive	-
HP3	0.0094±0.0003	Inactive	-
HP2	0.0047±0.0006	Inactive	-
HP1	0.010±0.001	Inactive	-
INT	0.010±0.003	0.010±0.003	Inactive
HH1	0.0065±0.0026	0.0035±0.0009	-
HH2	Inactive	0.00867±0.00006	-
HH3	Inactive	0.011±0.003	-
HH4	Inactive	0.033±0.007	-
INTtoINT1_1	0.0263±0.011	Inactive	Inactive
INTtoINT1_2	0.003±0.001	Inactive	Inactive
INTtoINT1_3	0.039±0.010	Inactive	Inactive

Table 5.3 Cleavage rate constants (min^{-1}) of RNA sequences in the exploration of intersecting neutral networks of hairpin and hammerhead ribozymes (figure 5.25A, B) and transition between bifunctional sequences, HH_INT_HP and HP_INT1_VS (figure 5.28).

5.1.9 Increase in structural complexity in ribozymes can occur across intersecting neutral networks

Increase in complexity is one of the most important consequences of evolution by natural selection. Functional RNAs usually exhibit greater activity when their informational content is increased (Carothers, et al., 2004; Irene Chen and Mark Ditzler, personal communication). A classic example of this was provided by Szostak and co-workers, where they showed that GTP-binding RNA aptamers that contained 10 additional bits of information compared to their parent RNA bound GTP with 10-fold greater affinity. However, this additional complexity decreased the abundance of those sequences by ~1000-fold in the selected pool of active molecules (Carothers, et al., 2004). A similar relationship between informational complexity, reactivity and abundance was observed with the artificial ligase ribozyme (Carothers, et al., 2004). These observations correlate increase in complexity to an increase in functional capability and suggests that the emergence of a functional RNA becomes less probable with increased activity and complexity. Together, these results highlight the importance of identifying plausible evolutionary paths that lead to the emergence of more complex ribozymes from simpler precursors.

Our results provide a framework to understand the emergence of complexity in functional RNA (Figure 5.27). Simple self-cleaving phenotypes can interface with catalytic sequences that are active for hairpin cleavage through neutral mutational drifts, via catalytic sequences that support both hammerhead and hairpin functions (HH_INT_HP). This sequence represents the intersection between the neutral networks of the hammerhead and hairpin ribozymes (Figure 5.25B), thereby providing a step-up in terms of complexity. The bulwark of dual function enables mutational exploration of sequence space. Consequently, further mutations to this sequence restricted only to non-conserved residues can convert this sequence to a new phenotype that

preserves only its hairpin activity but gains VS activity (HP_INT1_VS (Figure 5.28, table 5.3). This sequence represents the intersection between the neutral networks of the hairpin and VS ribozymes (Figure 5.12A) and provides access to additional complexity. Intermediates with dual function (HH_INT_HP and HP_INT1_VS) can bridge phenotypes with distinct structures and functions and enhance the capacity of functional RNA to explore its evolutionary sequence space

Mutational drifts from intersection sequences HH_INT_HP and HP_INT1_VS lead to sequences with exclusive HH, HP or VS functions (Figures 5.12A, 5.25B, 5.27). Additional mutations to these monofunctional sequences and acquisition of entire domains through nucleotide accretion can eventually yield more complex and catalytically-active wild-type sequences (Figure 5.19). It is to be noted that although the increasing order for phenotype complexity is HH<HP<VS, bifunctional sequences, HH_INT_HP and HP_INT1_VS are intrinsically more complex than their immediate monofunctional neighbors due to the stringency on nucleotide identities enforced by the requirements of dual function. This suggests that the emergence of such bifunctional sequences are rare events; however, variants of simpler catalytic motifs such as the hammerhead could have served as templates that produced such sequences through ‘neutral’ mutations.

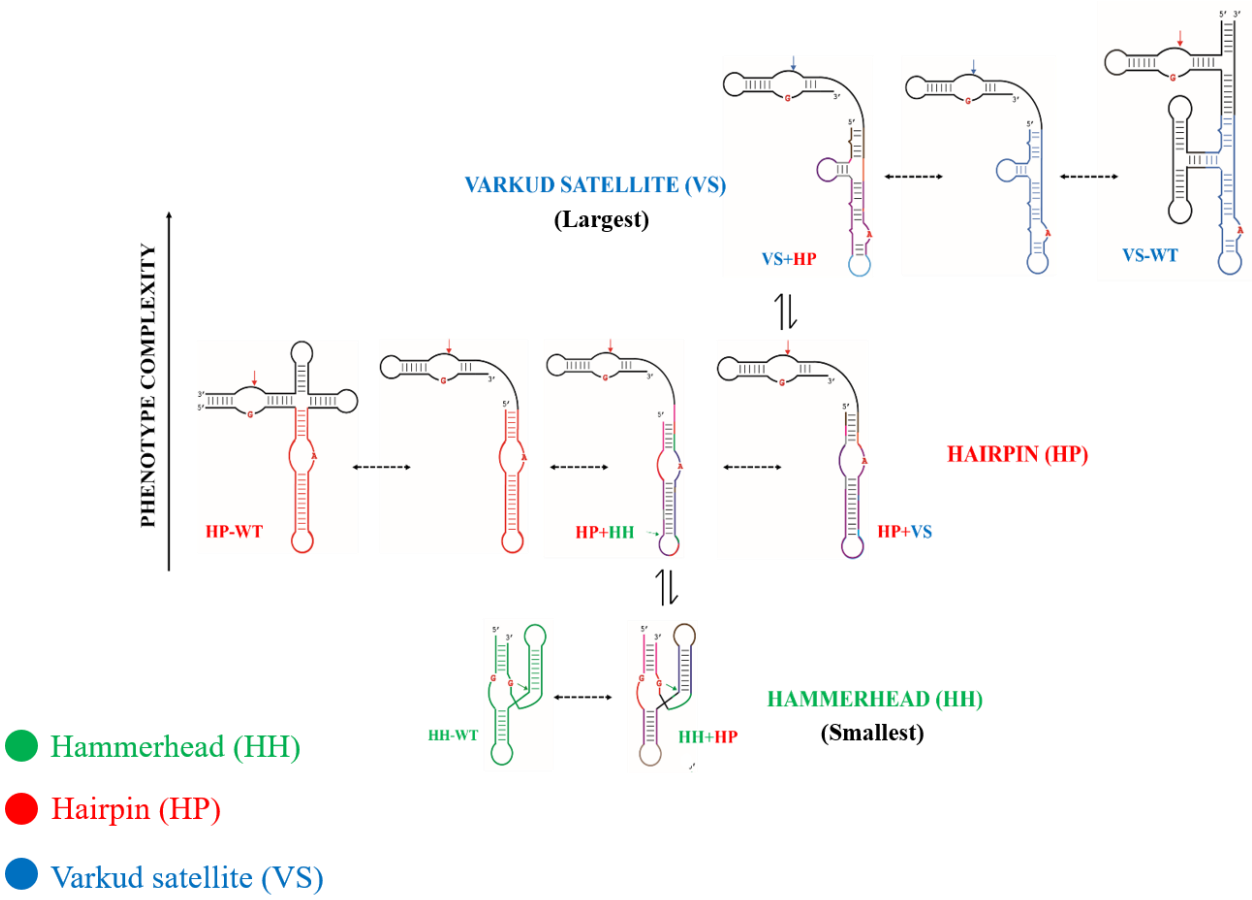

Figure 5.27 Smooth transition to more complex phenotypes. Simple self-cleaving motifs such as the hammerhead ribozyme can gain more complex structural features via intersection sequences connecting it to more complex phenotypes.

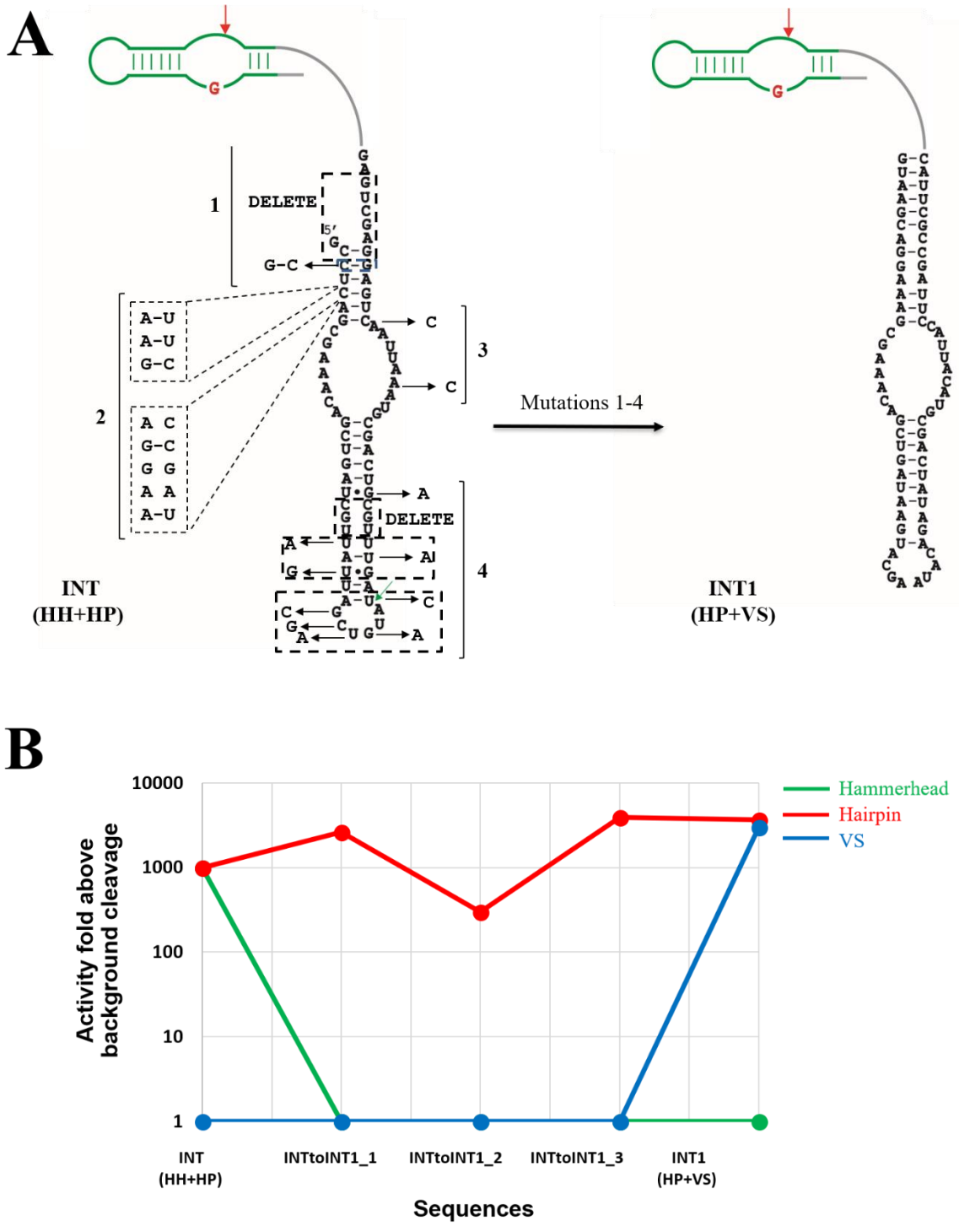

Figure 5.28 Mutational path between bifunctional sequences: HH_INT_HP and HP_INT1_VS. A. Sets of mutations (mutations 1-4) to non-conserved regions of bifunctional HH_INT_HP yield bifunctional HP_INT_VS. Intermediates in this path retain hairpin activity but lose the hammerhead activity of the parent HH_INT_HP. In this exercise hairpin function bridges hammerhead and VS functions.

5.2 DISCUSSION

Molecular evolution of RNA is shrouded in historical uncertainty. Unlike DNA or protein, RNA does not survive the onslaught of time, thus 'ancient' RNA cannot be retrieved from fossilized remains. In contrast DNA is routinely extracted from fossil organisms in efforts to place them in the tree of life (Poiner and Stankiewicz, 1999). 80-million-year-old proteins have been successfully extracted from dinosaur fossils (Schroeter et al., 2017). This allows comparison of these ancient proteins to contemporary proteins sequences, thereby generating a consensus sequence that is ultimately expressed recombinantly in the lab. Not only does this exercise help 'resurrect' these primitive proteins and study their structure/function but their homologies with modern proteins provide clues about evolutionary pathways that connect these resurrected proteins to their 'descendants' thriving in extant biology. Evolutionary investigations of functional RNA do not afford these tools, hence the primary way to understand the evolutionary events that led to the emergence of different RNA functions, is through reconstruction of plausible pathways that are consistent with the current framework of molecular evolution.

We have studied three RNAs that catalyze site-specific RNA cleavage-the hammerhead, hairpin and VS ribozymes. Although these ribozymes employ similar catalytic strategies and share similarities in the architecture of their active sites, they have distinct structural motifs that bolster their catalytic apparatuses and differ in overall complexity. In the light of these differences, it is generally assumed that these molecules evolved independently and are isolated in functional space. Several experimental findings led us to inspect the plausible evolutionary histories of these RNAs.

1. Two distinct RNA functions can be connected via intersecting neutral networks (Schultes and Bartel, 2000) through the existence of an intersection sequence that supports both functions.

2. Emergence of novel functional folds and new function in RNA aptamers and ribozymes (Held et al., 2003; Jaeger et al., 1999; Curtis and Bartel, 2005; Lau and Unrau, 2009) from pre-existing functional sequences.

3. Recurring emergence of the hammerhead ribozyme from random RNA pools when selected for RNA cleavage (Conaty et al., 1999; Tang and Breaker, 2000; Salehi-Ashtiani and Szostak, 2001; Popović et al., 2015). This suggests that the hammerhead most likely presents the simplest RNA-based solution for catalyzing RNA cleavage. The robustness of the hammerhead motif and its repeated selection highlight the possibility that this motif could have evolved multiple times in evolutionary history simply from statistical forces and due to the constraints imposed by limited chemical variability in RNA catalysts (Salehi-Ashtiani and Szostak, 2001).

Collectively, these results could help us formulate a working hypothesis about the evolution of more complex endonucleolytic ribozymes from relatively simpler motifs and the diversification of endonucleolytic function. Reconstructing neutral pathways connecting hammerhead, hairpin and VS function provides a point of entry into this discussion. The facile emergence of hammerhead motifs from random RNA sequences under selection pressure and the self-contained organization of the active fold (in contrast to the hairpin and VS ribozymes that consists of distinct substrate and catalytic domains that are brought together by tertiary interactions) make it an attractive candidate for a primordial self-cleaving ribozyme. Since the active site of the hammerhead is contained in a largely conserved 13 nt three-way junction, the robustness of the peripheral helix sequences can presumably allow the exploration of mutational space. Minor perturbations are generally not expected to alter the global functional fold of the ribozyme, however, contain ‘seeds of future innovation’ (Wagner, 2005). As long as the hammerhead motif maintains the integrity of its catalytic core and overall fold, sequence variants

of the motif could acquire additional catalytic functions as they encounter novel phenotypes. Conservation of the active site during the acquisition of new functions is essential for evolutionary progress; random mutations without any adaptive trajectory (widely accepted before the formulation of the neutral theory) could be catastrophic for catalytic function. A new functional fold is achieved only after the accumulation of sufficient mutational perturbations to the original sequence, leading to the sudden emergence of a new phenotype (punctuated equilibrium). This requires structural plasticity in these intermediate sequences in addition to bifunctionality.

We have demonstrated that sequence variants of the cis-cleaving hammerhead motif can adopt functional folds corresponding to the catalytic domain of the hairpin ribozyme and sequence variants of the catalytic domain of the hairpin ribozyme can in turn adopt a VS functional fold. Functional plasticity allows for the existence of points of intersection in neutral networks corresponding to ribozymes of varying structural complexities such as the hammerhead and hairpin, and the hairpin and VS, thereby enabling an increase in structural complexity. Our work demonstrates the existence of neutral paths toward increasing structural complexity in catalytic RNA. Bifunctional sequences in these paths act as nodes providing smooth transition to a more complex structural phenotype (Figure 5.27).

We identified these points of intersection (considering the hyper-dimensionality of sequence space, multiple points of intersection exist), thereby experimentally demonstrating functional interconversion between natural ribozymes for the first time,. Bifunctional sequences at the intersections of these neutral networks have smooth access to distinct catalytic functions via neutral drifts along well-defined mutational trajectories. Such intersection sequences would therefore provide nodes for the diversification of RNA function. Ancestral sequences capable of dual function would diverge from these nodes into distinct evolutionary trajectories resulting in

the emergence of new function. After encountering a new function, these sequences could acquire additional structural modules to reduce misfolding and evolve more robust chemical mechanisms for more efficient catalysis. Interestingly, the secondary structures corresponding to the intersection sequences corresponding to each ribozyme pair identified in this work most likely employ distinct sets of nucleotides to create their respective active sites. This allows for functional divergence and the emergence of new catalytic functions through mutations without compromising fitness. The parallels in catalytic strategies used by endonucleolytic ribozymes and the striking similarities in the active site architecture of the hammerhead, hairpin and VS ribozymes probably reflects local evolutionary convergence due to the lack of chemical diversity in catalytic systems created with RNA.

In the context of our work, a simple hammerhead-like self-cleaving motif could have evolved novel capabilities through adaptive walks across a fitness landscape that selects for the ability to cleave RNA (Figure 5.29). Through ribozyme-catalyzed ligation presumably viable in the RNA World, these self-cleaving sequences would associate with short but unique RNA sequences producing new self-cleaving ribozymes of increased complexities. Interestingly, most contemporary endonucleolytic ribozymes catalyze ligation in addition to cleavage (except the HDV ribozyme) (Lilley and Eckstein, 2008). In fact, full-length hammerhead, hairpin and VS ribozymes are better ligases than nucleases (De la Pena, 2016; Fedor, 2000; Jones et al., 2001). This presents an attractive model for the evolution of complexity in endonucleolytic ribozymes by modular accretion of smaller, independently-folding sequences (Schmidt, 1999; Manrubia and Briones, 2007). Modular increase in complexity has been observed in the evolution of ribosomal RNA (Fox et al., 1977), group I intron (van der Horst, 1991) and RNase P (Schmidt, 1999).

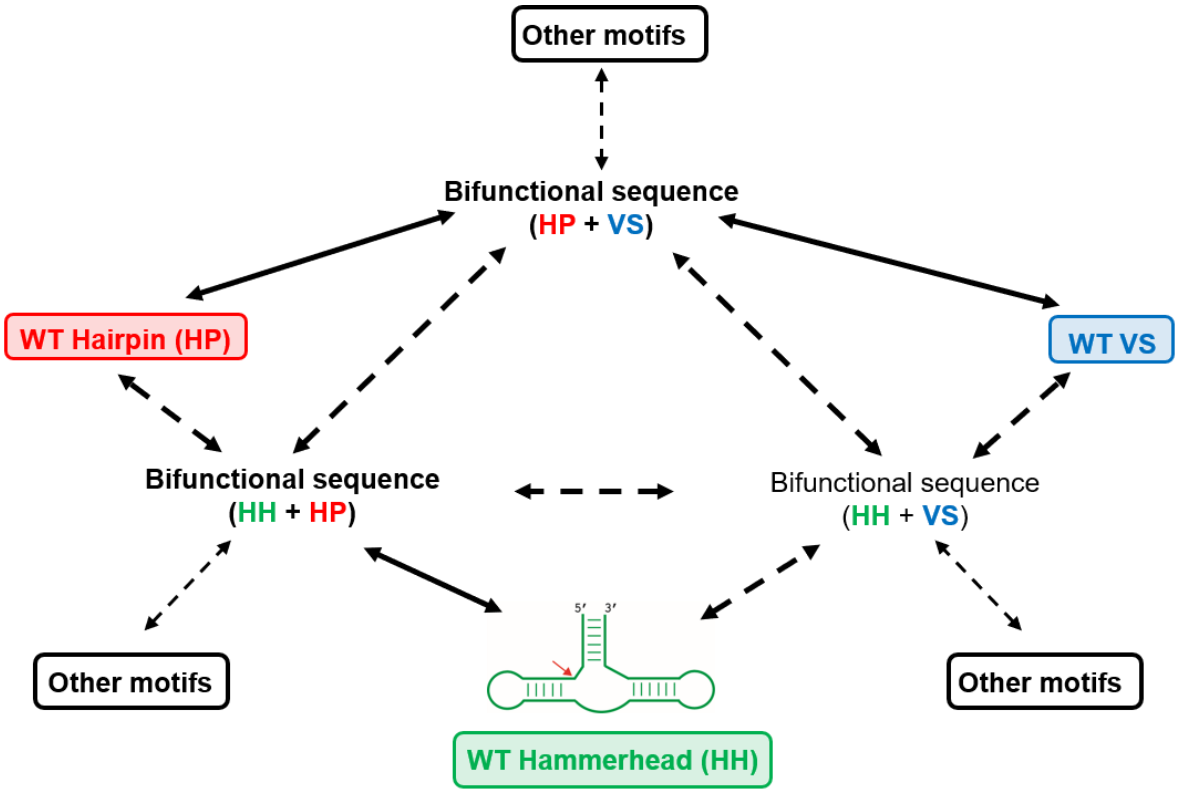

Figure 5.29 Divergence of ribonuclease function from the hammerhead self-cleaving motif. A hypothetical evolutionary network outlining probable routes (indicated by double-headed arrows) to the emergence of distinct endonucleolytic ribozyme motifs from a simple hammerhead self-cleaving ribozyme. Solid, bold arrows indicate paths outline in this work. Dashed, bold arrows indicate paths that could be outlined conceptually (see section 5.3-future directions). Dashed, light arrows indicate possible undiscovered routes to other endonucleolytic ribozyme motifs.

Such a model allows smaller, pre-folded sequences to assemble into a larger functional structure without the need for step-by-step polymerization, an intrinsically sluggish and error-prone process. Within this conceptual framework, the hammerhead or related self-cleaving motifs could be the catalytic core around which new ribozymes could have evolved. The emergence of hammerhead ribozyme variants from a pool of random RNA sequences during the laboratory evolution of self-cleaving ribozyme motifs in quasi-prebiotic conditions lends further credence to this hypothesis (Popović et al., 2015). It is possible that the first endonucleolytic ribozymes were indiscriminate toward RNA cleavage and substrate-specificity arose later in their evolutionary histories. Promiscuity in ancestral endonucleolytic ribozymes afforded the opportunity for multiple combinations of catalytic domain sequences and substrates to emerge, but perhaps only the optimal combinations gave rise to catalytic precursors to the fully-evolved modern-day ribozymes. This involves a compatibility search between catalytic domains and substrate sequences, similar to the one illustrated for the hairpin and VS ribozymes (Figure 5.15).

The robust folding properties of the hammerhead make it suitable for conjugation with other functional RNAs like aptamers to generate novel functionalities (Tang and Breaker, 1997). In addition to its ability to fold independent of its sequence context, hammerhead sequences can be accessed in segments through non-enzymatic template-directed polymerization of activated ribonucleotides (Prywes et al., 2016) and in their entirety through ribozyme-catalyzed transcription of its RNA template (Wochner et al., 2011). Conceptualizing a simple and evolvable catalytic sequence (or more likely multiple related variants all connected by neutral paths), drastically lowers the uncertainty of encountering unique catalytic ribonuclease sequence in the expansive sequence space of functional RNA.

In an effort to create intersection sequences for delineating intersecting neutral paths leading to functional cross-overs, we eliminated peripheral domains from wild-type ribozymes. This has led to the identification of fundamental structural requirements of these ribozymes. Studying the minimal catalytic sequences with the layers of sophistication stripped away might help highlight some of the fundamental similarities between ribozymes that appear different in their wild-type contexts. This in turn provide insights into the *de novo* design of catalytic systems based on RNA.

While studying minimal sequences generate useful information regarding the construction of catalytic cores of ribozymes, it has been realized on multiple occasions that although these peripheral sequences do not contain catalytic information, they contain information about RNA folding. However, the evolutionary potential embedded in these sequences has hardly been appreciated. Our work highlights the importance of these peripheral sequences in enabling functional RNAs to escape existing functional folds and encounter novel structure and function. Robustness in peripheral sequences provides the freedom to explore vast expanses of sequence space without compromising existing activity, consequently facilitating functional 'leaps' across intersecting neutral networks. The result is a 'small step' for the molecule but a 'giant leap' for evolution.

It is to be noted that the existence of intersecting neutral networks between any two extant ribozymes does not necessarily establish any evolutionary relationship; however, suggests a plausible mechanism through which they could be related. Similarly, successful identification of bifunctional sequences with activities corresponding to the hammerhead and other endonucleolytic ribozymes does not affirm the direct emergence of these motifs from the hammerhead, but simply suggests that it is possible for a relatively simple hammerhead-like motif to produce the catalytic

diversity/complexity in endonucleolytic ribozymes through gradual mutational drifts, consistent with the current framework of molecular evolution.

5.3 FUTURE DIRECTIONS

We have identified neutral pathways between the hammerhead, hairpin and VS ribozymes and outlined routes from bifunctional sequences to wild-type functionalities. This has been directly illustrated for HP_INT1_VS; however, a similar exercise should be possible for HH_INT_HP (Figure 5.27). We are working on identifying a dual HH-VS sequence through the rational design approach outlined here. We have identified catalytic sequences that can be presumably threaded through secondary structures of the hammerhead and VS ribozymes, but exhibit either HH or VS cleavage but not both. Further optimizations are required for obtaining a bifunctional sequence for HH and VS functions. Once this is achieved, the next step would be to connect these three bifunctional sequences (HH+HP, HP+VS, HH+VS) through neutral paths. Success would reinforce the robustness of the connectivity between these distinct ribonuclease motifs. Other endonucleolytic ribozymes could be introduced in our workflow to test the generality of this approach and more broadly, the existence of neutral networks connecting RNA-cleaving ribozymes.

This work is meant to initiate deeper investigations into the evolutionary relationships between endonucleolytic ribozymes. We present bifunctional catalytic sequences as plausible intermediates in the diversification of catalytic function via the intersection of neutral networks. However, our approach has relied on rational design and directed point mutations to explore the local sequence spaces of each ribozyme and is therefore limited. High-throughput sampling of

sequence space promises to provide much more information about the fitness landscapes of these RNAs.

A randomized library of catalytic sequences for each of the three ribozymes (usually 97% wild-type nucleotide and 1% each of the other three for each position or different extents of doping) would be generated. Current technical capabilities for sequencing allows screening of nearly all possible triple mutants of the catalytic domain sequences used in this work (Kobori and Yokobayashi, 2016). This library would be challenged for cis cleavage of both hairpin and VS substrates and both active and inactive molecules would be identified by high-throughput sequencing. These results will provide activity profiles for each nucleotide position in each of the three ribozymes, thereby generating an approximate fitness matrix for each ribozyme and substrate combination. It is expected that multiple neutral networks connecting sequences with the highest fitness (fitness peaks) would intersect in the fitness landscape. Pathways connecting fitness peaks corresponding to different substrates would provide multiple routes for functional interconversion and access to distinct phenotypes.

Mutagenized libraries could be prepared for the other self-cleaving ribozymes to generate functional sequence maps that could guide rational design of possible bifunctional sequences that are active for hammerhead cleavage in addition to their native activity. Successful design would allow us to explore mutational space and identify intersecting neutral networks between the hammerhead and other endonucleolytic ribozymes, further supporting the possibility of catalytic divergence from a hammerhead-like sequence.

RNA SEQUENCES USED IN THE EXPERIMENTS DISCUSSED IN THIS CHAPTER ARE LISTED IN MATERIALS AND METHODS (7.9.3, CHAPTER 7).

REFERENCES

1. Adami, C., Ofria, C. and Collier, T. C., 2000. Evolution of biological complexity. *Proc. Natl. Acad. Sci. U. S. A.*, 97, 4463-4468.
2. Butcher, S. E., Heckman, J. E. and Burke, J. M., 1995. Reconstitution of hairpin ribozyme activity following separation of functional domains. 270, 29648-29651.
3. Carothers, J. M., Oestreich, S. C. Davis, J. H. and Szostak, J. W., 2004. Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc.*, 126, 5130-5137.
4. Conaty, J., Hendry, P. and Lockett, T., 1999. Selected classes of minimized hammerhead ribozyme have very high cleavage rates at low Mg^{2+} concentration. *Nucleic Acids Res.*, 27, 2400-2407.
5. De la Peña, M., García-Robles, I. and Cervera, A., 2017. The hammerhead ribozyme: A long history for a short RNA. *Molecules*, 22, 78-90.
6. Fedor, M. J., 2000. The structure and function of the hairpin ribozyme. *J. Mol Biol.*, 297, 268-291.
7. Fox, G. E., Magrum, L. J., Balch, W. E., Wolfe, R. S. and Woese, C. R., 1977. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc. Natl. Acad. Sci. U. S. A.*, 74, 4537-4541.
8. Guo, H. C. T. and Collins, R. A., 1995. An efficient *trans*-cleavage of a stem-loop RNA substrate by a ribozyme derived from *Neurospora* VS RNA. *EMBO J.*, 14, 368-376.
9. Jiménez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R., and Chen, I. A, 2013. Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proc. Natl. Acad. Sci. U. S. A.*, 10, 14984-14989.
10. Jones, F. D. and Strobel, S. A., 2003. Ionization of a critical adenosine residue in the neurospora Varkud Satellite ribozyme active site. *Biochemistry*, 42, 4265-4276.
11. Knight, R., De Sterck, H., Markel, R., Smit, S., Oshmyansky, A., and Yarus, M., 2005. Abundance of correctly folded RNA motifs ion sequence space, calculated on computational grids. *Nucleic Acids Res.*, 33, 5924-5935.
12. Kobori, S. and Yokobayashi, Y., 2016. High-throughput mutational analysis of a twister ribozyme. *Angew. Chem Int. Ed. Engl.*, 55, 10354-10357.
13. Komatsu, Y., Kanzaki, I. and Ohtsuka, E., 1996. Enhanced Folding of Hairpin Ribozymes with Replaced Domains. *Biochemistry*, 35, 9815-9820.
14. Komatsu, Y., Kanzaki, I., Koizumi, M. and Ohtsuka, E., 1995. Modification of primary structures of hairpin ribozymes for probing active conformations. *J. Mol. Biol.*, 252, 296-304.
15. Lafontaine, D. A., Norman, D. G. and Lilley, D. M. J., 2002. The global structure of the VS ribozyme. *EMBO J.*, 21, 2461-2471.

16. Lafontaine, D. A., Wilson, T. J., Norman, D. G. and Lilley, D. M. J., 2001. The A730 loop is an important component of the active site of the VS ribozyme. *J. Mol. Biol.*, 312, 663-674.

17. Lilley, D. M. J., 2004. The Varkud satellite ribozyme. *RNA*, 10, 151-158.

18. Manrubia, S. C. and Briones, C., 2007. Modular evolution and increase of functional complexity in replicating RNA molecules. *RNA*, 13, 97-107.

19. Marek, M. S., Johnson-Buck, A. and Walter, N. G., 2011. The shape-shifting quasispecies of RNA: one sequence, many functional folds. *Phys. Chem. Chem. Phys.*, 13, 11524–11537.

20. Poiner, H. N. and Stankiewicz, B. A., 1999. Protein preservation and DNA retrieval from ancient tissues. *Proc. Natl. Acad. Sci. U. S. A.*, 96, 8426-8431.

21. Popović, M., Fliss, P. S., Ditzler, M. A., 2015. *In vitro* evolution of distinct self-cleaving ribozymes in diverse environments. *Nucleic Acids Res.*, 43, 7070-7082.

22. Price, S. R., Ito, N., Oubridge, C., Avis, J. M. and Nagai, K., 1995. Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J. Mol. Biol.*, 249, 398-408.

23. Prywes, N., Blain, J, C., Del Frate, F. and Szostak, J. W., 2016. Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *eLife*, 5. e17756.

24. Rastogi, T., Beattie, T. L., Olive, J. E. and Collins, R. A., 1996. A long-range pseudoknot is required for activity of the Neurospora VS ribozyme. *EMBO J.*, 15, 2820-2825.

25. Salehi-Ashtiani, K. and Szostak, J. W., 2001. *In vitro* evolution suggests multiple origins for the hammerhead ribozyme. *Nature*, 414, 82-84.

26. Sargueil, B., Pecchia, D. B. and Burke, J. M., 1995. An improved version of the hairpin ribozyme functions as a ribonucleoprotein complex. *Biochemistry*, 34, 7739-7748.

27. Schmidt, F. J., 1999. Ribozyme-Why so many, why so few? *Mol. Cells*, 9, 459-463.

28. Schroeter, E. R., DeHart, C. J., Cleland, T. P., Zheng, W, Thomas, P. M., Kelleher, N. L., Bern, M. and Schweitzer, M. H., 2017. Expansion for the *Brachylophosaurus canadensis* collagen I sequence and additional evidence of the preservation of cretaceous protein. *J. Proteome Res.*, 16, 920–932.

29. Tang, J. and Breaker, R. R., 1997. Rational design of allosteric ribozymes. *Chem. Biol.*, 4, 453-459.

30. Tang, J. and Breaker, R. R., 2000. Structural diversity of self-cleaving ribozymes. *Proc. Natl. Acad. Sci. U. S. A.*, 97, 5784-5789.

31. Van der Horst, G., Christian, A. and Inoue, T., 1991. Reconstitution of a group I intron self-splicing reaction with an activator RNA. *Proc. Natl. Acad. Sci. U. S. A.*, 88, 184-188.

32. Wagner, A., 2005. Robustness, evolvability, and neutrality. *FEBS Lett.*, 579, 1772-1778.

33. Wilson, T. J. and Lilley, D. M. J., 2011. Do the hairpin and VS ribozymes share a common catalytic mechanism based on general acid-base catalysis? A critical assessment of available experimental data. *RNA*, 17, 213–221.

34. Wochner, A., Attwater, J., Coulson, A. and Holliger, P., 2011. Ribozyme-catalyzed transcription of an active ribozyme. *Science*, 332, 209-212.

35. Zamel, R., Poon, A., Jaikaran, D., Andersen, A., Olive, J.,De Abreu, D. and Collins, R. A., 2004. Exceptionally fast self-cleavage by a Neurospora Varkud satellite ribozyme. *Proc. Natl. Acad. Sci. U. S. A.*, 101, 1467-1472.

36. Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31, 3406-3415.

Chapter - 6

CONCLUSIONS

RNA catalysis has come a long way from being a subject of mere curiosity three decades ago to a field of active research that has resulted in a deep understanding of the structural principles that guide the construction of functional RNA molecules. Rapid advances in the structural biology of RNA, especially x-ray crystallography have been a major driving force behind successful investigations of RNA catalysis. Here, we have reported the first (and only) crystal structures of the largest endonucleolytic ribozyme found in biology, the Varkud satellite (VS) ribozyme. These structures capture different functional states of the ribozyme and thus provide insights into various aspects of folding and catalysis.

The VS ribozyme is more efficient for trans cleavage than cleavage *in cis*. This could reflect its function in its biological context. The crystal structures capture the VS active site created *in trans* by the substrate and catalytic domains from two separate RNA molecules. The structures reveal three-way junctions as the predominant structural motif that organizes individual helices into the catalytically-active fold of the ribozyme.

The VS ribozyme is constructed in a modular fashion, with independently-folding domains that could have been added to a smaller catalytic core by evolution through nucleotide accretion. The modularity in the organization of the functional fold extends to substrate docking and remodeling. Our results suggest that the register shift in the substrate-helix physically couples the loop-loop kissing interaction to rearrangements in the internal cleavage loop of the substrate that harbors the cleavage site. Kissing loop interactions occur ubiquitously in structured RNAs and many RNA stem-loops are ‘interrupted’ by internal bulges. Coupling register shifts to kissing interactions could provide a general mechanism by which RNAs form tertiary interactions with

internal bulges.

Crystal structures of constructs containing mutations to either catalytic residues provided the first glimpse into the catalytically-competent active site of the VS ribozyme. Our results provide a structural framework for the chemical step of catalysis and implicate G638 and A756 in general acid-base catalysis. In addition, new catalytically-relevant interactions revealed by the high-resolution structures of the active site, promise to stimulate further investigations into the catalytic mechanism of the VS ribozyme. Comparing the VS active site with that of other endonucleolytic ribozymes such as the hairpin and hammerhead revealed common structural and catalytic themes, which extend to RNA-cleaving enzymes made of DNA and protein.

The increasing structural and catalytic diversity of endonucleolytic ribozymes present fundamental questions regarding the evolutionary origins of distinct catalytic functions. Distinct ribonuclease motifs could have emerged independently or alternatively could have originated from simpler self-cleaving motifs. Through the rational design of bifunctional ribozyme sequences, we have established connections between distinct endonucleolytic motifs that correspond to VS, hairpin and hammerhead functions. Furthermore, we have demonstrated that it is possible to access distinct catalytic functions from these bifunctional sequences by neutral mutational drifts. Our work establishes the first examples of intersecting neutral networks between naturally-occurring RNAs and demonstrates smooth acquisition of distinct catalytic functions in RNA along these neutral networks. Using the hammerhead, hairpin and VS ribozymes, we have delineated plausible pathways toward increased complexity (hammerhead to hairpin to VS). Considering the informational simplicity and the ease of evolvability of the hammerhead ribozyme, this or similar self-cleaving motifs could have been instrumental in the diversification of ribonuclease function in RNA.

Beyond their importance in modern biological systems, ribozymes are important model systems in the quest to understand biological catalysis and are central players in the 'RNA World hypothesis'. Uncovering novel features of extant RNA enzymes promises to reveal fundamental principles of biocatalysis and illuminate capabilities of early life forms. Enzymes are intrinsic to life and RNA enzymes could have been the hands that rocked its cradle.

Chapter - 7

MATERIALS AND METHODS

7.1 RNA synthesis for downstream applications

7.1.1 Generating DNA templates for *in vitro* transcription

DNA templates used to generate sequences for downstream experiments were synthesized by polymerase chain reaction (PCR) of gBlock® gene fragments (for crystallization constructs in chapters 2 and 3, and sequences longer than 200 nucleotides in chapters 4 and 5), Ultramers® (for sequences between 100 and 200 nucleotides in chapters 4 and 5) and simple oligos (for sequences under 100 nucleotides in chapters 4 and 5) using appropriate primers (all DNA sequences were purchased from Integrated DNA Technologies-IDT). An annealing temperature of 55°C was used for all PCR protocols. For generating crystallization constructs, the PCR amplified sequences were subcloned into EcoRI and XBaI sites of pUC19 vector (New England Biolabs) under a T7 promoter. The recombinant plasmids were sequenced to confirm the presence of the insert of interest. Templates for *in vitro* transcription were generated by PCR amplification of the desired fragment using suitable primers. The first two nucleotides of all reverse primers were designed to contain a 2'OCH_3 modification to reduce transcriptional heterogeneity at the 3' end of the RNA synthesized (Kao et al., 2001). All PCR products were purified using a Wizard® SV gel and PCR cleanup system (Promega).

7.1.2 *In vitro* transcription

DNA templates were transcribed by T7 RNA Polymerase (50 μg/mL) (expressed and purified in-house) in a reaction buffer containing 40 mM Tris-HCl (pH 8), 2 mM spermidine, 10 mM NaCl, 25 mM $MgCl_2$, 10 mM Dithiothreitol (DTT), 30 U/mL RNase inhibitor murine (NEB), 2.5 U/mL thermostable inorganic pyrophosphatase (TIPPase) (NEB), 4 mM of NTPs (ATP, CTP, GTP, CTP stock solutions made from powder purchased from Sigma) each, and 30-60 pmol/mL DNA template, for 2 h at 37°C. Reaction was quenched by adding 20 U/mL DNase I (RNase-free DNase, Promega) and incubating for 30 min at 37°C. Transcribed RNA was P/C/I extracted 2-3 times with equal volumes of phenol-chloroform-isoamyl alcohol solution (phenol solution buffered with 0.1 M citrate at pH 4.3 was purchased from Sigma). For RNA sequences that cleave almost completely during transcription, lower concentrations of $MgCl_2$ (5 mM) and shorter reaction times (30 minutes for transcription and 15 minutes for DNase I digestion) were used.

7.1.3 Purification of transcribed RNA

For RNA used for crystallization trials, P/C/I extracted RNA was loaded onto a NAP-10 column (GE Healthcare) pre-equilibrated with gel filtration buffer (for VS constructs in chapter 2 and 3) containing 10 mM Tris (pH 7), 25 mM KCl, 5 mM $MgCl_2$. RNA was eluted with 1.5 mL gel filtration buffer and loaded onto HiLoad 16/60 superdex 200 pg size exclusion column in an AKTAxpress® filtration system (GE healthcare). All purifications were performed at 4°C. Fractions corresponding to purified RNA dimer were collected and concentrated to 10-20 mg/mL using an Amicon Ultra-15® column (30 kDa molecular weight cut-off), aliquoted into small fractions and stored at -80°C. The entire process using native size exclusion chromatography

(nSEC) avoided any denaturation-renaturation steps that could lead to RNA aggregation and undesired conformational heterogeneity detrimental to crystallization.

For RNA used for cleavage assays and other biochemical experiments, P/C/I extracted RNA was mixed with 9M urea and gel loading dye (90% (v/v) formamide, 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol) in 1:1 ratio and heated to 95^{o}C for 2 min and purified by denaturing polyacrylamide gel electrophoresis (gel mix contained 8 M urea). Desired bands were cut out and RNA was extracted into TEN Buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 50 mM NaCl) by passive elusion at 4 ^{o}C for 16-24 h (slices dissolved in TEN buffer were frozen for 20 min and thawed before overnight elusion for more efficient extraction). Extracted RNA was concentrated and ethanol precipitated, followed by suspension of the dried pellet in doubled distilled water.

Short RNA sequences with lengths under 35 nucleotides (VS and hairpin substrates in chapters 4 and 5) were purchased from IDT and purified by denaturing polyacrylamide gel electrophoresis and identical downstream steps were followed.

7.2 5'-radiolabeling of RNA sequences

In vitro transcribed RNA sequences contain 5’ triphosphates that need to be removed in order to be radiolabeled. This is done by treating them with shrimp alkaline phosphatase (SAP, ThermoFisher Scientific) (~1 U/50 pmol RNA to be labeled) after incubation at 90°C for 2 min, followed by cooling for 3 minutes (on ice or room temperature). The reaction mixture was incubated for 30 min at 37°C followed by heat inactivation of the enzyme at 65°C for 10 min.

Phosphatase-treated and commercially synthesized RNA were 5'-labeled by T4 polynucleotide kinase (T4 PNK) in a proprietary reaction buffer (NEB) in the presence of 25-50

pmol γ^{32}-ATP (Perkin-Elmer) for 30 min at 37°C. Reactions were quenched with gel loading dye (90% (v/v) formamide, 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol) and loaded onto a denaturing gel. The gel was exposed to an autoradiography film (Kodak) for 30 to 60 s or more if required in a dark room. The film was developed in an X-ray developer (Alphatek AX200), the developed film was aligned with the gel using the markers and bands corresponding to full-length labeled RNA were cut out and soaked in TEN buffer. Followed by a quick freeze-thaw cycle, labeled RNA was extracted into the buffer through overnight passive elusion and ethanol-precipitated in the presence of 2-3 μL 20 U/μL RNase-free glycogen (Thermo-Fisher) for formation of a visible pellet. The pellet was dried and suspended in double-distilled water, preferably just before using it for downstream applications.

7.3 Light scattering

VS ribozyme monomer and dimer fractions were passed through 0.2 μm cutoff centrifugal filter units (Millipore) and equilibrated to RT. Light scattering measurements were performed with DynaPro spectrometer (Wyatt). Data analysis was performed using native DynaPro software package.

7.4 Determination of equilibrium constant for dimerization

FPLC purified monomeric RNA was concentrated in the range of 0.01-80 μM. RNA was allowed to equilibrate at 37°C for 30 min. Samples corresponding to each concentration were loaded onto Tricorn high performance (10/300) superdex 200 column (GE Healthcare). Chromatogram peaks were integrated using the default HPLC software (Waters). Data was fitted

to a two-state model in Matlab: θ = [monomer]/ (K_D + [monomer]); where θ = fraction dimerized and K_D = dimerization equilibrium constant.

7.5 X-ray crystallography

7.5.1 Crystallization

Frozen aliquots (section 7.1.3), were rapidly thawed and diluted to 7.5 mg/mL. To decrease the number of nucleation events (Chayen, 2009), RNA was passed over 0.2 μm centrifugal filter units (Millipore). Mosquito liquid handling robot (TTP Labtech) was used to set up high-throughput (HT) hanging drop vapor diffusion crystallization screens with drops containing 0.1 μL RNA and 0.1 μL crystallization screen buffer. Crystal Screen I/II (Hampton), Index Screen I/II (Hampton), Natrix Screen I/II (Hampton) and RNA crystallization screen (Sigma) were used for initial crystallization trials. HT- crystallization trays were stored at 23-25°C and 4°C. Crystal 'hits' were observed in 100 mM bis-Tris pH 6.5, 2.0 M ammonium sulfate (Index I). Crystals appeared and grew to full size within 2-3 days. Crystallization was scaled up to 2 μL drops (1 μL RNA + 1 μL crystallization buffer) in VDX 24-well plates containing 500 μL crystallization buffer. VSx_G638A and VSx_C634 crystals were grown in 100 mM sodium cacodylate pH 6.8, 2.0 M ammonium sulfate and VSx_A756G crystals were grown in 100mM sodium citrate pH 5.8, 4% 1, 3-butanediol and 2.0 M ammonium acetate. Iridium-derivatized crystals were prepared by soaking native VSx_G638A_tGU crystals in a stabilizing solution supplemented with ~80 mM iridium hexamine (Keel et al., 2007).

7.5.2 Data collection and processing

Crystals were screened at the Advanced Photon Source (APS) GM/CA-CAT and SBC-CAT. Data sets were collected as APS NE-CAT. A single wavelength anomalous diffraction experiment was performed on the VSx_G638A_tGU crystals soaked in a stabilizing solution supplemented with ~ 80 mM iridium hexamine (Keel et al., 2007). Clear diffraction was observed to a resolution of at least 4.5 Å. We indexed, integrated and scaled the data using HKL2000/3000 (Otwinowski and Minor, 1997), identified heavy atom sites and calculated phases with SHELXD (Schneider and Sheldrick, 2002) / AutoSol (Adams et al., 2010), and carried out refinement swith Phenix/ERRASER (Adams et al., 2010; Chou et al., 2011). Sixteen iridium sites were found using SHELXD (Schneider and Sheldrick, 2002), aided by the hkl2map GUI_ENREF_32 (Pape and Schneider, 2004). SAD phasing was performed using AutoSol (Adams et al., 2010) with the four highest occupancy heteroatom sites from SHELXD as input parameters. Initial density modification and model building was done by AutoBuild (Adams et al., 2010). Ellipsoidal truncation and anisotropic scaling were applied to the native dataset prior to refinement using the Diffraction Anisotropy Server (Strong et al., 2006). Native datasets from VSx_G638A and VSx_C634 were indexed, integrated and scaled using HKL2000/3000 (Otwinowski and Minor, 1997) and (https://rapd.nec.aps.anl.gov/rapd). VSx_A756G datasets were integrated and scaled by RAPD (https://rapd.nec.aps.anl.gov/rapd). Phases of the VSx_A756G and VSx_C634 datasets were obtained by molecular replacement using VSx_G638A as search model using Phaser (Adams et al., 2010). Model building for both structures was completed with COOT (Emsley and Cowtan, 2004) with aid of RCrane (Keating and Pyle, 2010). Refinement was carried out with the Phenix/ERRASER pipeline (Adams et al., 2010; Chou et al., 2011). Metal ions were assigned on

the basis of coordination distance and temperature factors (Harding, 2006; Stahley et al., 2007). The reliability factors (R-factors) of the final models are listed in table 7.1.

Construct	R_{free} (%)	R_{work} (%)
VSx_G638A	17.7	21.5
VSx_A756G	23.5	26.9
VSx_C634	21.6	24.6

Table 7.1 R-factors of final crystallographic model of VS ribozyme constructs.

7.6 Small-angle x-ray scattering (SAXS)

SAXS experiments were conducted in the SIBYLS beamline at the Advanced Light Source synchrotron as previously described (Hura et al., 2009). The samples were purified by nSEC and prepared as described in 7.1.3. For each experiment, concentrated samples (1.5 mg/ml, 2.0 mg/ml (two wells), 2.5 mg/ml, 3.0 mg/ml, 3.5 mg/ml, 4.0 mg/ml, 4.5 mg/ml in 25 µl were placed in a 96-well plate. SAXS data were collected continuously with Q ranging from 0.012 – 0.324 with exposures of 0.5, 1. 6 seconds. Buffer (10 mM Tris (pH 7.5), 25 mM KCl and 5 mM $MgCl_2$) blanks were performed both before and after each sample exposure and subtracted from the sample signal. Within each concentration, each buffer-subtracted exposure was checked for radiation damage and any oversaturated points were removed before being averaged together. The final experimental scattering curve was calculated by scaling the averaged datasets for each concentration to the highest concentration (4.5 mg/ml) dataset and merging with ALMERGE (Franke et al., (2012), extrapolating to infinite dilution. SAXS curves were calculated from the crystal structure atomic coordinates and fit to the experimental data using the FOXS (Schneidman-Duhovny et al., 2013). SAXS data have been deposited in the SASBDB under accession code SASDAC9.

7.7 Structure analysis and video preparation

Helical axes for all stems were computed using Curves+ (Lavery et al., 2009). The interhelical angles were computed with vector methods in MatLab (MathWorks). Solvent accessible surface area was calculated using POPs (Cavallo et al., 2013) and PDBePISA (http://www.ebi.ac.uk/pdbe/pisa/). Tertiary interactions were validated with Assemble (Jossinet et al., 2010). Torsion angles and the nature of sugar puckers were calculated using Amigos II (Wadley and Pyle, 2004). All figures were made in Pymol (Schrodinger) (LLC) and edited in Illustrator (Adobe).

The movie depicting the transition from an undocked to a docked VS ribozyme substrate was generated using the Morph function in PyMol (Schrodinger) (LLC). The coordinates from the NMR structure of the isolated, undocked substrate (PDB ID: 1HWQ) (Flinder et al., 2001) and the docked substrate from our crystal structure of the full-length RNA (PDB ID: 5V3I) were superimposed using the ALIGN function in PyMol (Schrodinger) (LLC). For the docked state, helices 2–7 were not shown for clarity. The movie was compiled using the eMovie plugin.

7.8 Ribozyme cleavage kinetics

Cis cleavage assays. Radiolabeled RNA in ddH_2O was heated at 70°C for 2 minutes in the presence of 25 mM KCl, followed by incubation on ice for 3 minutes and at room temperature for 5 minutes. Tris-HCl (pH 8) and spermidine were added to final concentrations of 50 mM and 2 mM, respectively, and incubated at 37°C for 15 minutes. $MgCl_2$ was added to initiate cleavage and the reaction was incubated at 37°C. Aliquots at different time points were withdrawn from the reaction and quenched with stop dye (90% (v/v) formamide, 0.05% (w/v) xylene cyanol, 0.05%

(w/v) bromophenol blue) and frozen on dry ice. Reaction end times were chosen according to their expected rates.

Trans cleavage assays. Protocol 1. This protocol was used for slow trans cleavage (chapter 4) was identical to that of cis cleavage assays. Only the substrate was radiolabeled and both ribozyme and labeled substrate were incubated at 70°C for 2 minutes in the presence of 25 mM KCl in the first step of RNA refolding.

Protocol 2. This protocol was used for fast trans cleavage (chapter 5, section 5.1.6). Labeled RNA and unlabeled ribozyme were incubated separately at 70°C for 2 minutes in the presence of 10 mM Mg^{2+} (for VS trans cleavage) or 100 mM Mg^{2+} (for hairpin trans cleavage), 25 mM K^{+}, 25 mM Tris (pH 8) and 2 mM spermidine, cooled on ice for 3 minutes followed by incubation in RT for 5 minutes. Reaction was initiated by mixing ribozyme and labeled substrate and the reaction was incubated at RT (for VS trans cleavage) or 37°C (for hairpin trans cleavage). Aliquots were withdrawn at appropriate intervals, quenched in stop dye and frozen on dry ice. Ribozymes concentrations were 0.5 µM and 10 µM for VS and hairpin cleavage respectively.

Measuring uncatalyzed background RNA cleavage. DNA mimicking the sequence of the 5' strand of the substrate (containing the cleavage site) with a single ribonucleotide just upstream of the cleavage site was used for measuring the kinetics of uncatalyzed RNA cleavage. An all-RNA sequence was not used to avoid non-specific cleavage across the entire substrate sequence at high Mg^{2+} concentrations used in the reaction that complicates calculations. A DNA sequence with a single ribonucleotide yields a clean single product band after cleavage that can be easily quantified. Two such ‘substrate’ sequences were designed that differed only in their desired cleavage sites reflecting the natural difference in canonical cleavage sites in the VS and hairpin

ribozymes (GpA for VS and ApG for hairpin). Substrate sequences were incubated at 37°C for 48 hours. The sequences used are as follows (bold and italicized nucleotides indicate canonical cleavage sites, r***G*** or r***A*** are ribonucleotides):

VSsub_oneribo_ss: GGCr***GA***AGGCCGTC

HPsub_oneribo_ss: GGCr***AG***TCGGCCGTC

Aliquots were loaded on to a 20% dPAGE for separating precursor and product bands. The gels were exposed to PhosphorImager screens (Amersham Biosciences) overnight or more if required and scanned on a Typhoon Trio imager (GE Healthcare) in 'Storage phosphor' mode. Bands were quantified in ImageQuant TL with automated rolling ball algorithm for background subtraction. Fraction cleaved was calculated as the ratio of the intensities of the product and precursor bands. In gels that contain other bands resulting from non-specific cleavage, fraction cleaved was estimated by dividing the intensity of the product band by the total intensity in the gel. Fraction cleaved was plotted against reaction time points and data was fitted to a first-order exponential in Kaleidagraph (Synergy Software). Composite data were plotted in Excel (Microsoft) or Origin 8/2017 (OriginLab).

7.9 RNA sequences used in this work

7.9.1 Crystallization constructs and control sequences of the VS ribozyme (chapters 2, 3)

CONSTRUCT	SEQUENCE	DESCRIPTION	PURPOSE
VSx_G638A_ tGU	GGCGCUGUGUCGCAAUCUGCGAAGGGCGUCGUCGG CCCAAGCGGUAGUAAGCAGGGAACUCACCUCCAAU GAAACACAUUGUCGUAGCAGUUGACUACUGUUAU GUGAUUGGUAGAGGCUAAGUGACGGGUGUGGCGU AAGCCACGUCCGCAGCACAGCACAAGCCCGCUUGC GAGAUUACAGCGC	Contains two tandem GUs in helix 6 to create iridium binding site. Green: Ir-binding site	SAD phasing.
VSx_G638A	GGCGCUGUGUCGCAAUCUGCGAAGGGCGUCGUCGG CCCAAGCGGUAGUAAGCAGGGAACUCACCUCCAAU GAAACACAUUGUCGUAGCAGUUGACUACUGUUAU GUGAUUGGUAGAGGCUAAGUGACGGUAUUGGCGU AAGCCAAUACCGCAGCACAGCACAAGCCCGCUUGC GAGAUUACAGCGC	Contains shortened helix 4 with AAACA pentaloop and G638A mutation. Red, underlined: catalytic nucleotides	Native data collection. SAXS studies.
VSx_A756G	GGCGCUGUGUCGCAAUCUGCGAAGGGCGUCGUCGG CCCGAGCGGUAGUAAGCAGGGAACUCACCUCCAAU GAAACACAUUGUCGUAGCAGUUGACUACUGUUAU GUGAUUGGUAGAGGCUAAGUGACGGUAUUGGCGU AAGCCAAUACCGCGGCACAGCACAAGCCCGCUUGC GAGAUUACAGCGC	Contains shortened helix 4 with AAACA pentaloop and A756G mutation. *Solved by MR of VSx_G638A.*	Structure of active site with wild-type G638.
VSx_A756C	GGCGCUGUGUCGCAAUCUGCGAAGGGCGUCGUCGG CCCGAGCGGUAGUAAGCAGGGAACUCACCUCCAAU GAAACACAUUGUCGUAGCAGUUGACUACUGUUAU GUGAUUGGUAGAGGCUAAGUGACGGUAUUGGCGU AAGCCAAUACCGCCGCACAGCACAAGCCCGCUUGC GAGAUUACAGCGC	Contains shortened helix 4 with AAACA pentaloop and A756G mutation. **Not solved.**	Structure of active site with wild-type G638.
VSx_C634	GGCGCUGUGUCGCAAUCUGCGAAGGGCGUCGUCGC CCCAAGCGGUAGUAAGCAGGGAACUCACCUCCAAU GAAACACAUUGUCGUAGCAGUUGACUACUGUUAU GUGAUUGGUAGAGGCUAAGUGACGGUAUUGGCGU AAGCCAAUACCGCAGCACAGCACAAGCCCGCUUGC GAGAUUACAGCGC	Contains shortened helix 4 with AAACA pentaloop and G638A mutation. Contains wild-type helix 1b. *Solved by MR of VSx_G638A.*	Structure of wild-type helix 1b.
VS_WT	GGCGCUGUGUCGCAAUCUGCGAAGGGCGUCGUCGG CCCGAGCGGUAGUAAGCAGGGAACUCACCUCCAAU UUCAGUACUGAAAUUGUCGUAGCAGUUGACUACU GUUAUGUGAUUGGUAGAGGCUAAGUGACGGUAUU GGCGUAAGCCAAUACCGCAGCACAGCACAAGCCCG CUUGCGAGAUUACAGCGC	Wild-type helix 4	Control.
VS_G638A	GGCGCUGUGUCGCAAUCUGCGAAGGGCGUCGUCGG CCCAAGCGGUAGUAAGCAGGGAACUCACCUCCAAU UUCAGUACUGAAAUUGUCGUAGCAGUUGACUACU GUUAUGUGAUUGGUAGAGGCUAAGUGACGGUAUU GGCGUAAGCCAAUACCGCAGCACAGCACAAGCCCG CUUGCGAGAUUACAGCGC	Contains WT helix 4 and mutation of the putative general base guanine (G638) to adenine.	Control.
VS_G638	GGCGCUGUGUCGCAAUCUGCGAAGGGCGUCGUCGG CCCGAGCGGUAGUAAGCAGGGAACUCACCUCCAAU GAAACACAUUGUCGUAGCAGUUGACUACUGUUAU GUGAUUGGUAGAGGCUAAGUGACGGUAUUGGCGU AAGCCAAUACCGCAGCACAGCACAAGCCCGCUUGC GAGAUUACAGCGC	Same as crystallization construct VSx_G638A but with WT putative general base guanine.	Control.

CONSTRUCT	SEQUENCE	DESCRIPTION	PURPOSE
VS_G638_Δ1aΔ7	GGAAUCUGCGAAGGGCGUCGUCGGCCC**G**AGCGGU AGUAAGCAGGGAACUCACCUCCAAUUUCAGUACUG AAAUUGUCGUAGCAGUUGACUACUGUUAUGUGAU UGGUAGAGGCUAAGUGACGGUAUUGGCGUAAGCC AAUACCGCAGCAC**A**GCACAAGCCCGCUUGCGAGAU UACAGCGC	Lacks helices 1a and 7, hence junction 1-2-7	Testing the role of nucleotides upstream of cleavage site in dimerization
VS_G638_Δ1aΔ1aΔ7	GGCGGUAGUAAGCAGGGAACUCACCUCCAAUUUCA GUACUGAAAUUGUCGUAGCAGUUGACUACUGUUA UGUGAUUGGUAGAGGCUAAGUGACGGUAUUGGCG UAAGCCAAUACCGC**A**GCACAGCACAAGCCCGCUUG CGAGAUUACAGCGC	Lacks helices 1a, 1b, hence junction 1-2-7 and stem-loop 1	Testing the role of nucleotides upstream of cleavage site, and nucleotides in the substrate stem-loop in dimerization

7.9.2 Toward a trans-active bifunctional sequence for hairpin and VS cleavage (chapter 4)

CONSTRUCT	SEQUENCE
HPrz (wt)/HPt	GACCAGAGAAACACACGACCUGAGUAAUCAGGUCGUGGU**A**UAUUACCUGGUA
VSrz (wt)	GGCGGUAGUAAGCAGGGAACUCACCUCCAAUUUCAGUACUGAAAUUGUCGUAGCAGUUG ACUACUGUUAUGUGAUUGGUAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGC **A**GCACAGCACAAGCCCGCUUGCGAGAAU
Δ345	GGUAAGCAGGGAACUCGAAAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGC**A**G CACAGCACAAGCCCGCUUGCGAGAAUauuuuacuaacaagaucuccagcuuGC***GA***AGGCCGUCGUCGC GCC**G**ACUAAGCGGcuuggaauu (yellow: trunc1, blue: trunc2, green: trunc3)
Δ345trunc1	GGGCAGGGAACUCGAAAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGC**A**GCAC AGCACAAGCCCGCCGAGAAUauuuuacuaacaagaucuccagcuuGC***GA***AGGCCGUCGUCGCGCC**G**AC UAAGCGGcuuggaauu
Δ345trunc2	GGUAAGCAGGGAACUCGAAAGAGGCUAAGUGACAUUGGCGUAAGUCAGUGC**A**GCACAGC ACAAGCCCGCUUGCGAGAAUauuuuacuaacaagaucuccagcuuGC***GA***AGGCCGUCGUCGCGCC**G**AC UAAGCGGcuuggaauu
Δ345trunc3	GGUAAGCAGGGAACUCGAAAGAGGCUAAGUGACGGUAUUGUAAAGUAUUGC**A**GCACAGC ACAAGCCCGCUUGCGAGAAUauuuuacuaacaagaucuccagcuuGC***GA***AGGCCGUCGUCGCGCC**G**AC UAAGCGGcuuggaauu
Δ345_G638A	GGUAAGCAGGGAACUCGAAAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGC**A**G CACAGCACAAGCCCGCUUGCGAGAAUauuuuacuaacaagaucuccagcuuGC***GA***AGGCCGUCGUCGC GCC**A**ACUAAGCGGcuuggaauu
Δ345_A756G	GGUAAGCAGGGAACUCGAAAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGC**G** GCACAGCACAAGCCCGCUUGCGAGAAUauuuuacuaacaagaucuccagcuuGC***GA***AGGCCGUCGUCG CGCC**G**ACUAAGCGGcuuggaauu
Δ345_short	GGUAAGCAGGGAACUCGAAAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGC**A**G CACAGCACAAGCCCGCUUGCcuagg***GA***AGGCCGUCGUCGCGCC**G**Aauu
Δ345t	GGUAAGCAGGGAACUCGAAAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGCAG CACAGCACAAGCCCGCUUGC (yellow: trunc1, blue: trunc2, green: trunc3)
Δ345t_trunc1	GGGCAGGGAACUCGAAAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGCAGCAC AGCACAAGCCCGCC
Δ345t_trunc2	GGUAAGCAGGGAACUCGAAAGAGGCUAAGUGACAUUGGCGUAAGUCAGUGCAGCACAGC ACAAGCCCGCUUGC
Δ345t_trunc3	GGUAAGCAGGGAACUCGAAAGAGGCUAAGUGACGGUAUUGUAAAGUAUUGCAGCACAGC ACAAGCCCGCUUGC
Δ345t_trunc23	GGUAAGCAGGGAACUCGAAAGAGGCUAAGUGACAUUGUAAAGUGCAGCACAGCACAAGC CCGCUUGC
Δ345t_trunc123	GGGCAGGGAACUCGAAAGAGGCUAAGUGACAUUGUAAAGUGCAGCACAGCACAAGCCCG CC
Trans_123	GGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCUUUC CUGCC
Trans_23	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU UUCCUGCUUAC
Trans_23mut6	GGUAACAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU UAGCCGCUUAC (red letters indicate point mutations)
Trans_3	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUAGACUUAGGGUCAGCGUACA UUACCUUUCCUGCUUAC
Trans_3mut6	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAGCGUACA UUACCUUAGCCGCUUAC (red letters indicate point mutations)
HPsub	GGC***AG***UCGGCCGUCGUCGCGCCA**G**AAGCCauu
VSsub	GGC***GA***AGGCCGUCGUCGCGCC**G**AGCCauu
Bold: catalytic nucleotides, Bold, italicized: cleavage site nucleotides, lowercase letters indicate linker 3'-end nucleotides.	

7.9.3 Sequences in chapter 5

7.9.3.1 Toward a cis-active bifunctional sequence for hairpin and VS cleavage

CONSTRUCT	SEQUENCE
Trans_3mut6_long1-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACgagaauauuuuacuaacaagaucuccagcuGGC***GA***AGGCCG UCGUCGCGCCGAGCCgcuuggauu
Trans_3mut6_short1-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACccuaGGC***GA***AGGCCGUCGUCGCGCCGAGCCauu
Trans-3mut6_ long1-HP	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACgagaauauuuuacuaacaagaucuccagcuGGC***AG***UCGGCC GUCGUCGCGCCAGAAGCCgcuuggauu
Trans-3mut6_ short1-HP	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACccuaGGC***AG***UCGGCCGUCGUCGCGCCAGAAGCC auu
Trans_3mut6_ short2-HP	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACccuaggc***AG***UCGGCCGUCGUCGCGCCAGAAauu
Trans_3mut6_ 7L-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACccua*aaa*GGC***GA***AGGCCGUCGUCGCGCCGAGCCa uu
Trans_3mut6_ 9L-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACccua*aaaaa*GGC***GA***AGGCCGUCGUCGCGCCGAGC Cauu
Trans_3mut6_ 11L-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACccua*aaaaaaa*GGC***GA***AGGCCGUCGUCGCGCCGA GCCauu
Trans_3mut6_ 13L-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACccua*aaaaaaaaa*GGC***GA***AGGCCGUCGUCGCGCCG AGCCauu
Trans_3mut6 _15L-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACccua*aaaaaaaaaaa*GGC***GA***AGGCCGUCGUCGCGCC GAGCCauu
Trans_3mut6_17L-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACccua*aaaaaaaaaaaaa*GGC***GA***AGGCCGUCGUCGCG CCGAGCCauu
Trans_3mut6_ 7L-HP	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACccua*aaa*GGC***AG***UCGGCCGUCGUCGCGCCAGAA GCCauu
Trans_3mut6 _9L-HP	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACccua*aaaaa*GGC***AG***UCGGCCGUCGUCGCGCCAGA AGCCauu
Trans_3mut6 _11L-HP	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAACCCUACAGAUAGGGUCAG CGUACAUUACCUUAGCCGCUUACccua*aaaaaaa*GGC***AG***UCGGCCGUCGUCGCGCCAG AAGCCauu
Trans_23mut6 _long-VS (INT1-VS)	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACA UUACCUUAGCCGCUUACauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCCGAGCC auu
Trans_23mut6 _short-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACA UUACCUUAGCCGCUUACcuaGGC***GA***AGGCCGUCGUCGCGCCGAGCCauu
Trans_23mut6_ long-HP (INT1-HP)	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACA UUACCUUAGCCGCUUACauuuuacuaacaagaucGGC***AG***UCGGCCGUCGUCGCGCCAGAA GCCauu
Trans_23mut6_ short-HP	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACA UUACCUUAGCCGCUUACcuaGGC***AG***UCGGCCGUCGUCGCGCCAGAAGCCauu

Bold, italicized: cleavage site nucleotides, lowercase letters indicate linker or 3'-end nucleotides, underlined: base-paired 5' and 3' end of substrate.

7.9.3.2 Hairpin-VS dual activity

CONSTRUCT	SEQUENCE
Sequences involved in local fitness landscape for hairpin-VS interconversion (Figure 5.15)	
HP7	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACCAUAAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***AG***UCGGCCGUCGUCGCGCCAGAAGCCauu
HP7-S1	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACCAUAAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***AA***UCGGCCGUCGUCGCGCCAGAAGCCauu
HP7-S2	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACCAUAAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***UCGGCCGUCGUCGCGCCAGAAGCCauu
HP7-S3	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACCAUAAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***ACGGCCGUCGUCGCGCCAGAAGCCauu
HP7-S4	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACCAUAAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCCAGAAGCCauu
HP7-S5	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACCAUAAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCCGAAGCCauu
HP7-VS	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCCGAGCCauu
HP6	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACCAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***AG***UCGGCCGUCGUCGCGCCAGAAGCCauu
HP6-S1	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACCAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***AA***UCGGCCGUCGUCGCGCCAGAAGCCauu
HP6-S2	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACCAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***UCGGCCGUCGUCGCGCCAGAAGCCauu
HP6-S3	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACCAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***ACGGCCGUCGUCGCGCCAGAAGCCauu
HP6-S4	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACCAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCCAGAAGCCauu
HP6-S5	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACCAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCCGAAGCCauu
HP6-VS	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCCGAGCCauu
HP5	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***AG***UCGGCCGUCGUCGCGCCAGAAGCCauu
HP5-S1	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***AA***UCGGCCGUCGUCGCGCCAGAAGCCauu
HP5-S2	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***UCGGCCGUCGUCGCGCCAGAAGCCauu
HP5-S3	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***ACGGCCGUCGUCGCGCCAGAAGCCauu
HP5-S4	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCCAGAAGCCauu
HP5-S5	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCCGAAGCCauu
HP5-VS	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAUUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCCGAGCCauu
HP4	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***AG***UCGGCCGUCGUCGCGCCAGAAGCCauu
HP4-S1	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***AA***UCGGCCGUCGUCGCGCCAGAAGCCauu
HP4-S2	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***UCGGCCGUCGUCGCGCCAGAAGCCauu
HP4-S3	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***ACGGCCGUCGUCGCGCCAGAAGCCauu
HP4-S4	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCCAGAAGCCauu
HP4-S5	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCCGAAGCCauu

HP4-VS	GUAAGCAGCAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCGAGCCauu
HP-3	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
HP3-S1	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCAAUCGGCCGUCGUCGCGCCAGAAGCCauu
HP3-S2	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAUCGGCCGUCGUCGCGCCAGAAGCCauu
HP3-S3	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAACGGCCGUCGUCGCGCCAGAAGCCauu
HP3-S4	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCAGAAGCCauu
HP3-S5	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCGAAGCCauu
HP3-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCGAGCCauu
HP2	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
HP2-S1	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCAAUCGGCCGUCGUCGCGCCAGAAGCCauu
HP2-S2	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAUCGGCCGUCGUCGCGCCAGAAGCCauu
HP2-S3	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAACGGCCGUCGUCGCGCCAGAAGCCauu
HP2-S4	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCAGAAGCCauu
HP2-S5	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCGAAGCCauu
HP2-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCGAGCCauu
HP1	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
HP1-S1	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCAAUCGGCCGUCGUCGCGCCAGAAGCCauu
HP1-S2	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAUCGGCCGUCGUCGCGCCAGAAGCCauu
HP1-S3	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAACGGCCGUCGUCGCGCCAGAAGCCauu
HP1-S4	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCAGAAGCCauu
HP1-S5	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCGAAGCCauu
HP1-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCGAGCCauu
INT1-HP	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
INT1-S1	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCAAUCGGCCGUCGUCGCGCCAGAAGCCauu
INT1-S2	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAUCGGCCGUCGUCGCGCCAGAAGCCauu
INT1-S3	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAACGGCCGUCGUCGCGCCAGAAGCCauu
INT1-S4	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCAGAAGCCauu
INT1-S5	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCGAAGCCauu
INT1-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCGAGCCauu
INT2-HP	GUAAGCAGGCAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACC
	UUAGCCGCUUACauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
INT2-S1	GUAAGCAGGCAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACC
	UUAGCCGCUUACauuuuacuaacaagaucGGCAAUCGGCCGUCGUCGCGCCAGAAGCCauu

INT2-S2	GUAAGCAGGCAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACC
	UUAGCCGCUUACauuuuacuaacaagaucGGC*GA*UCGGCCGUCGUCGCGCCAGAAGCCauu
INT2-S3	GUAAGCAGGCAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACC
	UUAGCCGCUUACauuuuacuaacaagaucGGC*GA*ACGGCCGUCGUCGCGCCAGAAGCCauu
INT2-S4	GUAAGCAGGCAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACC
	UUAGCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCAGAAGCCauu
INT2-S5	GUAAGCAGGCAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACC
	UUAGCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAAGCCauu
INT2-VS	GUAAGCAGGCAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACC
	UUAGCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAGCCauu
VS1	GUAAGCAGGCAAAGCGAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAGCCauu
VS1-S1	GUAAGCAGGCAAAGCGAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC*AA*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS1-S2	GUAAGCAGGCAAAGCGAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC*GA*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS1-S3	GUAAGCAGGCAAAGCGAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC*GA*ACGGCCGUCGUCGCGCCAGAAGCCauu
VS1-S4	GUAAGCAGGCAAAGCGAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCAGAAGCCauu
VS1-S5	GUAAGCAGGCAAAGCGAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAAGCCauu
VS1-HP	GUAAGCAGGCAAAGCGAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU
	UAGCCGCUUACauuuuacuaacaagaucGGC*AG*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS2	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCUU
	AGCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAGCCauu
VS2-S1	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCUU
	AGCCGCUUACauuuuacuaacaagaucGGC*AA*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS2-S2	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCUU
	AGCCGCUUACauuuuacuaacaagaucGGC*GA*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS2-S3	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCUU
	AGCCGCUUACauuuuacuaacaagaucGGC*GA*ACGGCCGUCGUCGCGCCAGAAGCCauu
VS2-S4	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCUU
	AGCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCAGAAGCCauu
VS2-S5	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCUU
	AGCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAAGCCauu
VS2-HP	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCUU
	AGCCGCUUACauuuuacuaacaagaucGGC*AG*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS3	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGUACAUUACCUUA
	GCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAGCCauu
VS3-S1	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGUACAUUACCUUA
	GCCGCUUACauuuuacuaacaagaucGGC*AA*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS3-S2	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGUACAUUACCUUA
	GCCGCUUACauuuuacuaacaagaucGGC*GA*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS3-S3	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGUACAUUACCUUA
	GCCGCUUACauuuuacuaacaagaucGGC*GA*ACGGCCGUCGUCGCGCCAGAAGCCauu
VS3-S4	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGUACAUUACCUUA
	GCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCAGAAGCCauu
VS3-S5	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGUACAUUACCUUA
	GCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAAGCCauu
VS3-HP	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGUACAUUACCUUA
	GCCGCUUACauuuuacuaacaagaucGGC*AG*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS4	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGACAUUACCUUAG
	CCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAGCCauu
VS4-S1	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGACAUUACCUUAG
	CCGCUUACauuuuacuaacaagaucGGC*AA*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS4-S2	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGACAUUACCUUAG
	CCGCUUACauuuuacuaacaagaucGGC*GA*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS4-S3	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGACAUUACCUUAG
	CCGCUUACauuuuacuaacaagaucGGC*GA*ACGGCCGUCGUCGCGCCAGAAGCCauu
VS4-S4	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGACAUUACCUUAG
	CCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCAGAAGCCauu

VS4-S5	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGACAUUACCUUAG CCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAAGCCauu
VS4-HP	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUCGAAUACAGAUAUCAGCGACAUUACCUUAG CCGCUUACauuuuacuaacaagaucGGC*AG*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS5	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCGACAUUACCUUAGC CGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAGCCauu
VS5-S1	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCGACAUUACCUUAGC CGCUUACauuuuacuaacaagaucGGC*AA*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS5-S2	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCGACAUUACCUUAGC CGCUUACauuuuacuaacaagaucGGC*GA*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS5-S3	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCGACAUUACCUUAGC CGCUUACauuuuacuaacaagaucGGC*GA*ACGGCCGUCGUCGCGCCAGAAGCCauu
VS5-S4	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCGACAUUACCUUAGC CGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCAGAAGCCauu
VS5-S5	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCGACAUUACCUUAGC CGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAAGCCauu
VS5-HP	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCGACAUUACCUUAGC CGCUUACauuuuacuaacaagaucGGC*AG*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS6	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCACAUUACCUUAGCC GCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAGCCauu
VS6-S1	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCACAUUACCUUAGCC GCUUACauuuuacuaacaagaucGGC*AA*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS6-S2	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCACAUUACCUUAGCC GCUUACauuuuacuaacaagaucGGC*GA*UCGGCCGUCGUCGCGCCAGAAGCCauu
VS6-S3	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCACAUUACCUUAGCC GCUUACauuuuacuaacaagaucGGC*GA*ACGGCCGUCGUCGCGCCAGAAGCCauu
VS6-S4	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCACAUUACCUUAGCC GCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCAGAAGCCauu
VS6-S5	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCACAUUACCUUAGCC GCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAAGCCauu
VS6-HP	GUAAGCAGGCAAAGCGAAAGCUGAUAAGUGAAUACAGAUAUCAGCACAUUACCUUAGCC GCUUACauuuuacuaacaagaucGGC*AG*UCGGCCGUCGUCGCGCCAGAAGCCauu
Catalytic domain variants the HP functional space that retain dual function (Figure 5.12B)	
HP3a	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU UUGCCGCUUACauuuuacuaacaagaucGGC*AG*UCGGCCGUCGUCGCGCCAGAAGCCauu
HP3a-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU UUGCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAGCCauu
HP3ab	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU UUCCCGCUUACauuuuacuaacaagaucGGC*AG*UCGGCCGUCGUCGCGCCAGAAGCCauu
HP3ab-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU UUCCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAGCCauu
HP3abc	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU UUCCUGCUUACauuuuacuaacaagaucGGC*AG*UCGGCCGUCGUCGCGCCAGAAGCCauu
HP3abc-VS	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUAGACUUAUCAGCGUACAUUACCU UUCCUGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAGCCauu
Catalytic residue mutants (Figure 5.8 E, F)	
INT1-VS_GtoA	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU UAGCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCAAGCCauu
INT1-VS_AtoG	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCGGCGUACAUUACCU UAGCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCGAGCCauu
INT1-HP_GtoA	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU UAGCCGCUUACauuuuacuaacaagaucGGC*AG*UCGGCCGUCGUCGCGCCAAAAGCCauu
INT1-HP_AtoG	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUGCAUUACCU UAGCCGCUUACauuuuacuaacaagaucGGC*AG*UCGGCCGUCGUCGCGCCAGAAGCCauu
Substrate variants with lower activities (Figure 5.13)	
INT1-S1a	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU UAGCCGCUUACauuuuacuaacaagaucGGC*GG*AGGCCGUCGUCGCGCCGAGCCauu
INT1-S3a	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU UAGCCGCUUACauuuuacuaacaagaucGGC*GA*UGGCCGUCGUCGCGCCAGAAGCCauu
INT1-S5a	GUAAGCAGGAAAGCGAAACAGCUGAUAAGUACGAAUACAGAUAUCAGCGUACAUUACCU UAGCCGCUUACauuuuacuaacaagaucGGC*GA*AGGCCGUCGUCGCGCCAGAGCCauu
Hairpin catalytic A: red, VS catalytic A: blue	

7.9.3.3 From intersection sequence, HP_INT1_VS to wild-type activity

CONSTRUCT	SEQUENCE
VSWT_min345	GUAAGCAGGGAACUCGAAAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGC**A**GCACAGCACAAGCCCGCUUGCauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCC**G**AGCCauu
VSWT_min345_ext3	GUAAGCAGGGAACUCACCUCGAAAUUGGUAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGC**A**GCACAGCACAAGCCCGCUUGCauuuuacuaacaagaucGGC***GA***AGGCCGUCGUCGCGCC**G**AGCCauu
VSWT_34trans	GUAAGCAGGGAACUCACCUCCAAUUUCAGUACUGAAAUUGUCGUAUGACUUAUGUGAUUGGUAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGC**A**GCACAGCACAAGCCCGCUUGC
VSWT_35trans	GUAAGCAGGGAACUCACCUCCAAGUACUUGUCGUAGCAGUUGACUACUGUUAUGUGAUUGGUAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGC**A**GCACAGCACAAGCCCGCUUGC
VSWT_345trans	GUAAGCAGGGAACUCACCUCCAAUUUCAGUACUGAAAUUGUCGUAGCAGUUGACUACUGUUAUGUGAUUGGUAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGC**A**GCACAGCACAAGCCCGCUUGC
VSWT_3457trans	GUAAGCAGGGAACUCACCUCCAAUUUCAGUACUGAAAUUGUCGUAGCAGUUGACUACUGUUAUGUGAUUGGUAGAGGCUAAGUGACGGUAUUGGCGUAAGUCAGUAUUGC**A**GCACAGCACAAGCCCGCUUGCGAGAAUAUUUUAC*GAAA*GUAGAGUGUCGC
HPTransRz	GGCAGAGAAACACACGACCUGAGUAAUCAGGUCGUGGU**A**CAUUACCUGCC
Purple: helix 3, green: helix 4, orange: helix 5, grey: helix 7, bold: catalytic nucleotides, bold italics: cleavage site nucleotides	

7.9.3.4 Hammerhead-hairpin dual activity and connecting bifunctional sequences, INT to INT1

CONSTRUCT	SEQUENCE
HH4	GCCUCAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGAGUCCGUAAAUUAAC UGAGGAGCUGAG
HH4-HP	GCCUCAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGAGUCCGUAAAUUAAC UGAGGAGCUGAGauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
HH3	GCCUCAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGAGUGCGUAAAUUAAC UGAGGAGCUGAG
HH3-HP	GCCUCAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGAGUGCGUAAAUUAAC UGAGGAGCUGAGauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
HH2	GCCUCAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGAGAGCGUAAAUUAAC UGAGGAGCUGAG
HH2-HP	GCCUCAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGAGAGCGUAAAUUAAC UGAGGAGCUGAGauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
HH1	GCCUCAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGACAGCGUAAAUUAAC UGAGGAGCUGAG
HH1-HP	GCCUCAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGACAGCGUAAAUUAAC UGAGGAGCUGAGauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
INT-HH	GCCUCAGCGAAACAGCUGAUCGUAAUUAGCUGUAUAGUUUGCGUCAGCGUAAAUUAAC UGAGGAGCUGAG\
INT-HP	GCCUCAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGUCAGCGUAAAUUAAC UGAGGAGCUGAGauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
HP1	GCCUCAGCGAAACAGCUGAUCGUAAUUAGCUGUAUAGUUUGCGUCAGCGUAAAUUAAC UGAGGAGCUGAGauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
HP1-HH	GCCUCAGCGAAACAGCUGAUCGUAAUUAGCUGUAUAGUUUGCGUCAGCGUAAAUUAAC UGAGGAGCUGAG
HP2	GCCUCAGCGAAACAGCUGAUCGAAAUUAGCUGUAUAGUUUGCGUCAGCGUAAAUUAAC UGAGGAGCUGAGauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
HP2-HH	GCCUCAGCGAAACAGCUGAUCGAAAUUAGCUGUAUAGUUUGCGUCAGCGUAAAUUAAC UGAGGAGCUGAG
HP3	GCCUCAGCGAAACAGCUGAUCCAAAUUAGCUGUAUAGUUUGCGUCAGCGUAAAUUAACU GAGGAGCUGAGauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
HP3_HH	GCCUCAGCGAAACAGCUGAUCCAAAUUAGCUGUAUAGUUUGCGUCAGCGUAAAUUAACU GAGGAGCUGAG
HP4	GCCUCAGCGAAACAGCUGAUGCAAAUUAGCUGUAUAGUUUGCGUCAGCGUAAAUUAAC UGAGGAGCUGAGauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
HP4_HH	GCCUCAGCGAAACAGCUGAUGCAAAUUAGCUGUAUAGUUUGCGUCAGCGUAAAUUAAC UGAGGAGCUGAG
INTtoINT1_1-HP	GUCAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGUCAGCGUAAAUUAACUG ACauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCauu
INTtoINT1_1-VS	GUCAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGUCAGCGUAAAUUAACUG ACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCGAGCCauu
INTtoINT1_2-HP	GUAAGCAGGAAAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGUCAGCGUAA AUUAACUUAGCCGCUUACauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCa uu
INTtoINT1_2-VS	GUAAGCAGGAAAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGUCAGCGUAA AUUAACUUAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCGAGCCauu
INTtoINT1_3-HP	GUAAGCAGGAAAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGUCAGCGUAC AUUACCUUAGCCGCUUACauuuuacuaacaagaucGGCAGUCGGCCGUCGUCGCGCCAGAAGCCa uu
INTtoINT1_3-VS	GUAAGCAGGAAAGCGAAACAGCUGAUCGUUAUUAGCUGUAUAGUUUGCGUCAGCGUAC AUUACCUUAGCCGCUUACauuuuacuaacaagaucGGCGAAGGCCGUCGUCGCGCCGAGCCauu
Hairpin catalytic residues: red, Hammerhead catalytic residues: green, VS catalytic residues: purple, bold italics: cleavage site nucleotides – red for hairpin, blue for VS and green for hammerhead, changes connecting INT to INT in orange and grey	

REFERENCES

1. Adams, P. D. *et al.*, 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, 66, 213-221.

2. Cavallo, L., Kleinjung, J. and Fraternali, F., 2003. POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucl. Acids. Res.*, 31, 3364-3366.

3. Chayen, R., 2009. Rigorous filtration for protein crystallography. *J. Appl.Cryst.*, 42, 743-744.

4. Chou, F.-C., Sripakdeevong, P. and Das, R., 2011. Correcting pervasive errors in RNA crystallography with Rosetta. *Nat. Methods*, 2013 10(1), 74–76.

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

6. Flinders, J. and Dieckmann, T., 2001. *J. Mol. Biol.*, 308, 665-679.

7. Franke, D., Kikhney, A. G. and Svergun, D. I., 2012. Automated acquisition and analysis of small angle X-ray scattering data. *Nuc. Inst. Meth. A.*, 689, 52–59.

8. Harding, M. M., 2006. Small revisions to predicted distances around metal sites in proteins. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 678-682.

9. Hura, G. L. *et al.*, 2009. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods*, 6, 606–612.

10. Jossinet, F., Ludwig, T. E. and Westhof, E., 2010. Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics*, 26, 2057-2059.

11. Kao, C., Rudisser, S. and Zheng, M., 2001. A simple and efficient method to transcribe RNAs with reduced 3' heterogeneity. *Methods*, 23, 201-205.

12. Keating, K. S. and Pyle, A. M., 2010. Semiautomated model building for RNA crystallography using a directed rotameric approach. *Proc. Natl. Acad. Sci. U.S.A.*, 107, 8177-8182.

13. Keel, A. Y., Rambo, R. P., Batey, R. T. and Kieft, J. S., 2007. A general strategy to solve the phase problem in RNA crystallography. *Structure*, 15, 761-772.

14. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. and Zakrzewska, K., 2009. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids. Res.*, 37, 5917-5929.

15. LLC, S. The PyMOL Molecular Graphics System, Version 1.2r3pre.

16. Otwinowski, Z. and Minor, W., 1997. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.*, 276, 307-326.

17. Pape, T. and Schneider, T. R., 2004. HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Cryst.*, 37, 843-844.

18. Schneider, T. R. and Sheldrick, G. M., 2002. Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.*, 58, 1772-1779.

19. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. and Sali, A., 2013. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.*, 105, 962–974.

20. Stahley, M. R., Adams, P. L., Wang, J. and Strobel, S. A., 2007. Structural metals in the group I intron: a ribozyme with a multiple metal ion core. *J. Mol. Biol.*, 372, 89-102.

21. Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D. and Eisenberg, D., 2006. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.*, 103, 8060-8065.

22. Wadley, L. M. and Pyle A. M., 2004. The identification of novel RNA structural motifs using COMPADRES: an automated approach to structural discovery. *Nucleic Acids Res.*, 32, 6650-6659.

APPENDICES

Appendix - 1

USING RNA AS A DEVICE: SELECTIVE AND SENSITIVE DETECTION OF LEAD (II) BY THE SPINACH RNA APTAMER

CONTENTS OF THIS CHAPTER HAVE BEEN PUBLISHED AS AN ARTICLE IN CHEMICAL COMMUNICATIONS (***Chem. Commun.*** **2015, 51, 9034-9037**). *ALL MATERIALS OF THE ARTICLE HAVE BEEN ADAPTED WITH THE COPYRIGHT PERMISSION FROM THE ROYAL SOCIETY OF CHEMISTRY.*

A.1 INTRODUCTION

A.1.1 Biomolecular recognition by RNA

Biomolecules, by virtue of their three-dimensional structure, are able to create intricate scaffolds that present specialized surfaces for ligand binding. The extensive network of weak interactions and precisely positioned functional groups in these cavities, referred to as 'binding pockets' provide binding modules that are selective toward their respective ligands in addition to being extremely tight binders. Riboswitches (section 1.2.1) are prominent examples of such recognition modules in nature and are employed by bacteria for gene regulation (Roth and Breaker, 2009). Riboswitches can selectively bind to cations, anions, and neutral and charged small molecules by exploiting their tertiary folds. About a decade prior to the discovery of riboswitches, work from the Szostak, Joyce and Gold labs delivered the first examples of artificial RNA sequences capable of binding specified targets. These RNA sequences were named aptamers (now expanded to DNA) (Ellington and Szostak, 1990; Robertson and Joyce, 1990; Tuerk and Gold, 1990) by Ellington. *In vitro* selection or selective evolution of ligands by exponential enrichment (SELEX), the molecular biology technique (Tuerk and Gold, 1990) used to generate these sequences has resulted in hundreds of aptamers (both DNA and RNA) that have found applications

in therapeutics, drug discovery, synthetic biology and *in vivo* imaging (Berens et al., 2015; Cho et al., 2009; Keefe et al., 2010; Paige et al., 2012). One such aptamer that inhibits vascular endothelial growth factor 165 ($VEGF_{165}$), Pegaptanib (marketed as Macugen®), is an FDA-approved drug for treatment of neovascular age-related macular degeneration (AMD), a leading cause of blindness worldwide (Bell et al., 1999).

A.1.2 Fluorescent RNA aptamers

Tagging biomolecules with fluorescent dyes has been one of the most effective techniques for elucidating biochemical processes *in vivo*. The advent of green fluorescent protein (GFP) and its variants enabled cellular tracking and investigations into the various steps in the flow of genetic information that culminates in protein expression (Gardes and Kaether., 1996). *In vitro* selection has yielded RNA aptamers that fluoresce when they bind their cognate fluorophores. The restricted molecular environment in the binding pockets of these aptamers results in fluorescence activation of otherwise non-fluorescent small molecules. The resurgence of fluorescent RNA aptamers began about a decade after the development of the malachite-green aptamer (Grate and Wilson, 1999), the first of its kind and the blue fluorescent RNA aptamer (BFR). These aptamers fluoresce when they bind malachite green and Hoechst dyes respectively (Sando et al., 2008). The Spinach RNA aptamer, an RNA mimic of GFP that selectively binds an analog of the fluorophore in GFP, DFHBI (3, 5-difluoro-4-hydroxybenzylidene imidazolinone) was developed in the group of Sammie Jaffrey (Paige et al., 2011). Spinach aptamer promised to revolutionize the field, with greater photo-stability and high quantum yields and since its emergence, it has been extensively used for analytical purposes. Spinach coupled to other aptamers (Strack et al., 2014) and hybridization sequences (Bhadra and Ellington, 2014; Akhter and Yokobayashi, 2014) has been used to detect

metabolites and proteins *in vivo* and oligonucleotides *in vitro*. In these Spinach-coupled devices binding of a specific analyte triggers the formation of the active conformation of Spinach, thereby activating DFHBI fluorescence. Spinach has also been used to monitor *in vitro* transcription efficiency in real time (Hofer, et al., 2013). In addition to variants of Spinach like Spinach2 (Strack et al., 2013 and iSpinach (Autour et al., 2016), another green fluorescent RNA, Broccoli (Filonov et al., 2014) and its derivatives Red and Orange Broccoli (Song et al., 2017), have been derived from Spinach using fluorescence-activated cell sorting (FACS). Other independent RNA aptamers that followed Spinach include Mango that binds a thiazole orange derivative (TO1) (Dolgosheina et al., 2014), Corn that binds 3,5-difluoro-4-hydroxybenzylidene imidazolinone-2-oxime (DFHO) and the DIR-aptamer that promiscuously binds both dimethylindole red (DIR) to activate red and oxazole thiazole blue (OTB) to activate blue fluorescence (Tan, et al., 2017) have led to the development of a suite of fluorescent modules with a wide range of spectral properties and stabilities. Structure determination of these aptamers has revealed G-quadruplexes as common binding motifs shared by most (but not all). This has opened doors for further innovation through rational design.

A.1.3 G-quadruplex as a recurring motif in fluorescent RNA aptamers

Fluorophores of fluorescent RNA aptamers share a planar structure made of aromatic heterocycles and charged side chains. While the charged functional groups help mediate stabilizing interactions with the negatively charged RNA backbone, the planer rings stack against RNA bases, often in binding pockets that are created by multiple base-stacks. G-quadruplexes are common nucleic acid motifs that provide efficient stacking platforms for planar molecules (Sen and Gilbert, 1988). These unusual motifs are created by multiple stacks of guanine quartets (four planar

guanines forming hydrogen bonds with their Watson-Crick and Hoogsteen faces) stabilized by metal ions (usually K^+ and Na^+ in biology) (Figure A.1) that are located in the core of the quadruplex assembly (Burge, et al., 2006). A popular line of anti-cancer therapy uses planar aromatics to stabilize these structures in telomeres, in the hope of inducing cell death by inhibiting telomerase, an enzyme that replenishes the ends of telomeres when they shorten after cell division (Neidle, 2010). Although a surprise when initially discovered in Spinach (Huang et al., 2014, Deigan-Warner et al., 2014), in retrospect, a G-quadruplex seems to be an obvious choice to bind planar small molecules like the ligands of these fluorescent RNA aptamers.

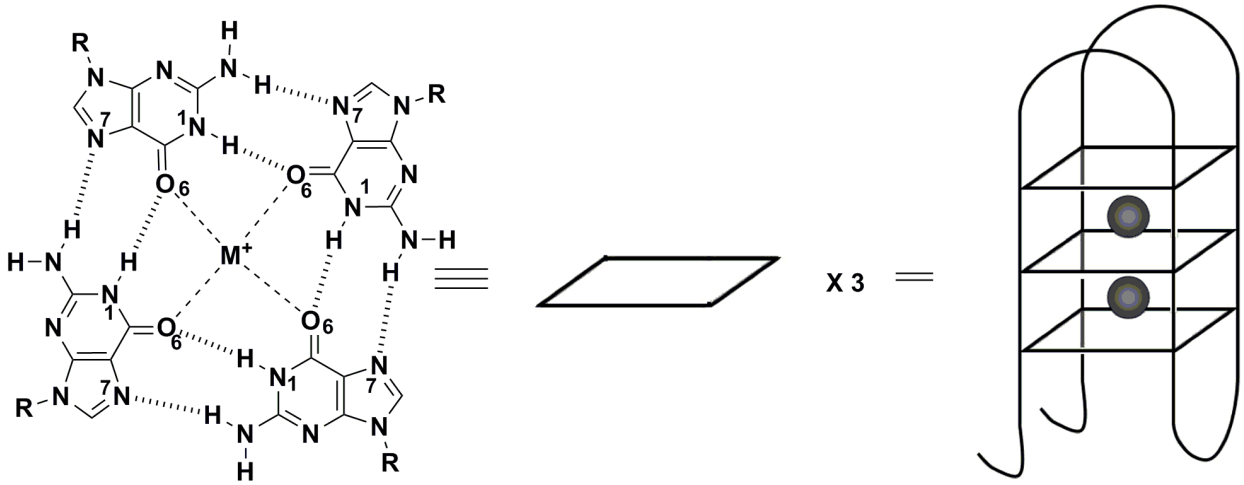

Figure A.1 G-quadruplex is an assembly of stacked guanine quartets. A G-quartet is formed by in-plane interactions between guanines at their Watson-Crick and Hoogsteen faces and is stabilized by electrostatic interactions between a cation (usually K^+ or Na^+) and the carbonyl oxygen atoms of the four guanines. Figure created by Dr. Pradeepkumar P. I., Indian Institute of Technology, Bombay.

While G-rich sequences in both DNA and RNA can fold into quadruplexes, they usually assume distinct topologies. DNA sequences have a propensity to form anti-parallel quadruplexes while RNA sequences form parallel quadruplexes. These distinct topologies are distinguishable by their respective signatures in circular dichroism (CD) spectra (Burge, et al., 2006). A-form nucleic acids (usually RNA) show a positive peak at ~260 nm, whereas B-form nucleic acids (usually DNA), exhibit a positive peak at ~275 nm. In addition, RNA G-quadruplexes show minor and intense negative peaks at 240 nm and 210 nm, respectively. A-form DNA shows a similar high intensity negative peak at 210 nm, accompanied by a strong positive signal at 190 nm, and B-DNA exhibit a low intensity negative peak at 240 nm (Vorlíčková et al., 2012). Presence of high millimolar concentrations of monovalent cations like Na^+ and K^+ induce/stabilize quadruplexes of both topologies; however, Na^+ and K^+ have been observed to favor quadruplexes with antiparallel and parallel topologies respectively (Burge, et al., 2006). Divalent cations like Pb^{2+}, Ca^{2+}, Ba^{2+} and Sr^{2+} can also induce the formation of G-quadruplexes in G-rich DNA sequences (Zavyalova, et al., 2016), with Pb^{2+} inducing particularly compact and stable structures likely due to an optimal combination of ionic radius and charge (Frank et al., 2000). This is illustrated by the large difference in melting temperatures of a G-rich sequence d($GTG_3TAG_3CG_3TTG_2$) in presence of different cations; T_m values of the G-rich sequence in the presence of Pb^{2+}, K^+ and Na^+ were 72°C, 51°C and 42°C, respectively (Frank et al., 2000). Assuming this favorable interaction between DNA G-quadruplex and Pb^{2+} extends to RNA-quadruplexes, minute quantities of Pb^{2+} should induce G-rich RNA sequences to fold into stable parallel quadruplexes. This assumption and the discovery of a K^+-stabilized G-quadruplex as an essential motif in the fluorescent Spinach RNA aptamer inspired us to develop the first RNA-based sensor for Pb^{2+}.

A.1.4 Nucleic acid sensors for lead (II)

Heavy metal toxicity is a serious environmental threat that affects both children and adults. Its effects are more adverse in developing countries in the absence of general awareness and governmental implementation of safety standards. Pb^{2+} concentrations as low as 0.5-1 μM can impair development and cause vascular and nervous system anomalies in children. Higher concentrations can cause hypertension, kidney damage, and infertility, and concentrations in the high range of 5-7 μM can lead to coma and eventual death (Centers for Disease Control and Prevention; National Institutes of Health). More than 500,000 children between the ages 1 and 5 years have unusually high Pb^{2+} concentrations in their blood and about 4 million live in lead-contaminated houses (Centers for Disease Control and Prevention; National Institutes of Health). The economic burden of lead toxicity on society is in the tens of billions of dollars (Centers for Disease Control and Prevention; National Center for Environmental Health). Traditional methods of detecting Pb^{2+} such as atomic absorption spectroscopy (AAS), atomic emission spectroscopy (AES) and inductively coupled plasma-mass spectrometry (ICP-MS) involve sophisticated and expensive equipments and are not suitable for on-site detection (Wang et al., 2013) Multi-electrode electrochemical sensors are portable but can be expensive and require considerable upkeep (Kim et al., 2012). Development of small molecule-based chemosensors like derivatized anthracene moieties and larger scaffolds made of organic macrocycles like calixarenes often have low selectivity and sensitivity though the latter can be improved by optimizing their design (Kim et al., 2012). Biomolecular sensors generally promise better selectivity due to the presence of more intricate scaffolds for ligand recognition. Modified peptides and biomimetic polymers have been used as selective sensors for $Pb^{2+,}$ presenting an array of polymer-based devices for selective and sensitive detection (Kim et al., 2012).

The last decade witnessed the development of a variety of DNA-based Pb^{2+}sensors (Figure A.2). Most of these sensors fall broadly into two classes based on their mechanisms of action- first, Pb^{2+}-dependent DNAzyme-catalyzed RNA cleavage (Yu et al., 2003; Lan et al., 2010; Zhang et al., 2011; Li et al., 2012; Gao et al., 2013; Li et al., 2014) and second, Pb^{2+}-dependent formation of DNA G-quadruplex (Table A.1) (Li et al., 2010a; Li et al., 2010b; Li et al., 2011b; Li et al., 2012; Li et al., 2013a; Guo et al., 2012). The first class of sensors consists of variants of a Pb^{2+}-dependent RNA-cleaving DNAzyme (primarily the 8-17 and GR5 motifs) with covalently modified termini. Modifications usually include fluorophore-quencher pairs that arrest fluorescence in the absence of Pb^{2+} but are separated from each other on Pb^{2+}-induced cleavage, leading to a fluorescence signal (Figure A.2A) (Kim et al., 2012). Other modifications such as nanoparticles use electrochemistry or colorimetry for detection (Li et al., 2013b; Li et al., 2012). The second class exploits the potential of Pb^{2+} to induce formation of G-quadruplexes in single or double-stranded G-rich DNA (Li et al., 2010a; Li et al., 2011b; Li et al., 2012; Li et al., 2013a; Li et al., 2010b; Guo et al., 2012). This class is made of two distinct types of sensors that incorporate different features of the first class of sensors. The first sub-class consists of G-rich sequences (such as the thrombin-binding aptamer) capable of folding into quadruplexes that are decorated with fluorophore and quencher pairs at their 5' and 3' termini. In the absence of Pb^{2+} the fluorophore produces strong signal, which is sequestered as the fluorophore-quencher pair is brought into close proximity by the formation of a quadruplex in the presence of Pb^{2+} (Figure A.2B). The second sub-class comprises G-quadruplex-based DNAzymes including AGRO100 and PS2.M that use a hemin cofactor to catalyze the H_2O_2-mediated oxidation of substrates such as Ampex Ultra red (AUR) and $ABTS^{2-}$/luminol leading to colorimetric or chemiluminiscence-based detection. G-quadruplexes stabilized by K^+ are able to support the hemin cofactors necessary for these reactions.

However, Pb^{2+} binding makes the quadruplex motif more compact, a configuration that is unable to catalyze the DNAzyme reaction (Figure A.2C). Therefore, this sub-class of sensors are primarily 'switch-off' sensors that exhibit marked decrease in signal in the presence of Pb^{2+} (Kim et al., 2012). Due to the requirement for covalent modification or chemical reactions that generate colored/chemiluminescent products, the applications of most DNA-based sensors discussed above are limited to certain favourable conditions of pH and ionic strength. DNA-based electrochemical sensors, most of which use one of the above mechanisms involve elaborate assemblies, usually requiring immobilization of DNA molecules to gold electrodes thereby reducing cost effectiveness (Li et al., 2011b; Gao et al., 2013; Li et al., 2013a; Liu et al., 2010; Zhang et al., 2011). To bypass these drawbacks, we have expanded the repertoire of nucleic acid-based Pb^{2+} sensors by developing the first Pb^{2+} sensor made of RNA.

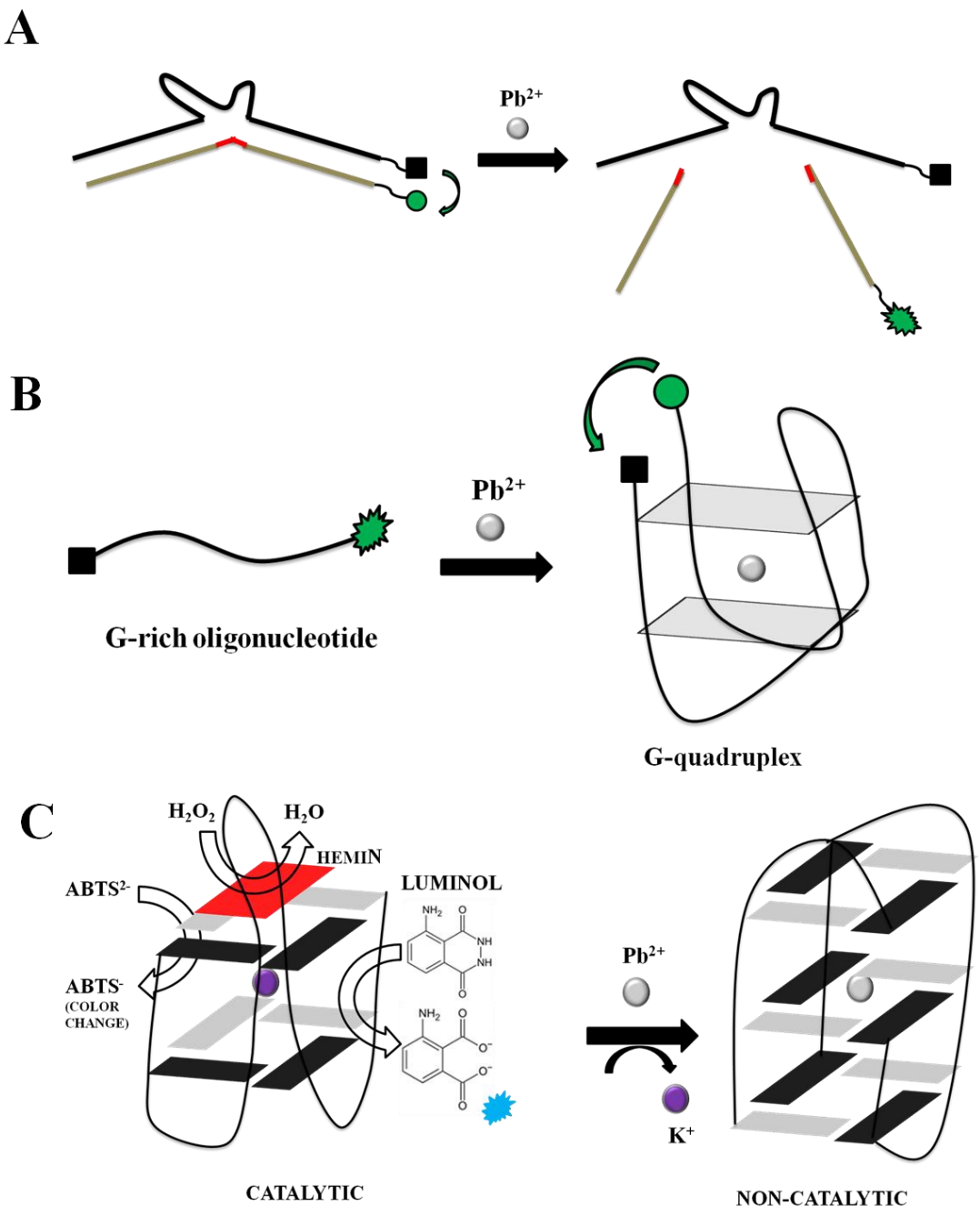

Figure A.2 Different classes of DNA-based Pb^{2+} sensors. A. Pb^{2+}-dependent RNA-cleaving DNAzyme, cleaves RNA substrate in the presence of Pb^{2+} thereby separating fluorophore from quencher. This results in strong fluorescence. **B.** G-rich single-stranded DNA folds into a G-quadruplex in presence of Pb^{2+} bringing the fluorophore quencher together leading to a 'turn off' signal. **C.** A K^+-stabilized G-quadruplex DNAzyme that catalyzes peroxide-mediated oxidation on binding hemin, switches to an inactive conformation when Pb^{2+} displaces K^+. This constitutes a 'turn-off' signal.

A.1.5 Designing the Spinach lead sensor

The Spinach RNA aptamer contains a two-layered G-quadruplex motif that serves as a stacking platform for DFHBI binding (Huang et al., 2014; Deigan-Warner et al., 2014). Formation of this quadruplex motif requires millimolar concentrations of K^+, Na^+, or NH_4^+ (Huang et al., 2014). Flanking the central binding pocket, which includes an essential base-triple in addition to the quadruplex, are two base-paired stems (Figure A.3). The stem just below the quadruplex is essential for its stability; however, this stem can be abridged to just five base-pairs without effecting fluorescence (Huang et al., 2014). We used this truncated Spinach construct for our studies. Spinach functions optimally in 100 mM K^+, and its absence completely abolishes fluorescence of the Spinach-DFHBI complex (Huang et al., 2014). The G-rich region of Spinach that is presumably unstructured in the absence of K^+, folds into a G-quadruplex on encountering Pb^{2+}. This enables fluorophore binding, thereby activating fluorescence signal. We used this Spinach-DFHBI complex *sans* K^+ as a 'switch on' fluorescent sensor (Figure A.3).

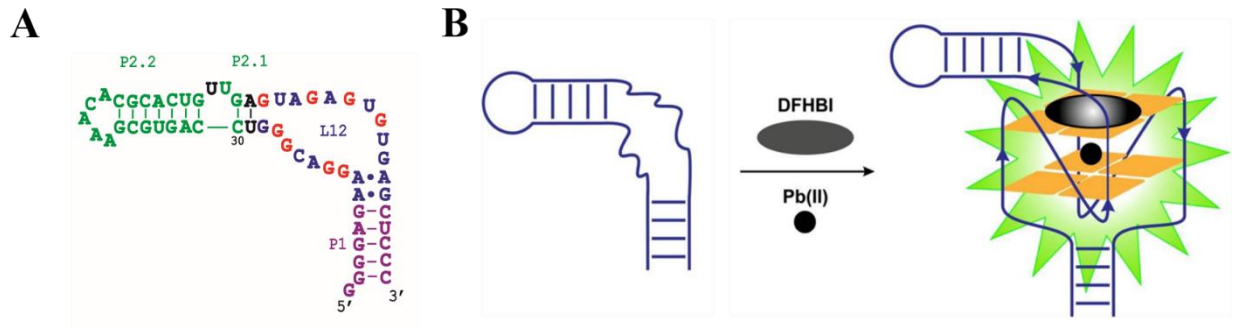

Figure A.3 Secondary structure and schematic illustration of the Spinach sensor. A. Secondary structure of the truncated Spinach RNA construct. Quadruplex guanines are shown in red, other unpaired nucleotides in the fluorophore binding pocket are colored blue, nucleotides involved in an essential base-triple are shown in black. **B.** The unstructured region (shown by curved lines) in the Spinach RNA aptamer folds into a two-layered G-quadruplex in the presence of Pb^{2+} and binds DFHBI fluorophore, activating fluorescence.

A.2 RESULTS

A.2.1 Spinach sensor exhibits high sensitivity and selectivity toward lead (II)

The hallmarks of an efficient sensor are its high sensitivity and selectivity. High sensitivity enables detection of minute quantities of the analyte, which is especially important for detecting Pb^{2+} contamination in real-world samples. Addition of sub-micromolar quantities of Pb^{2+} resulted in a strong fluorescence signal around 501 nm (Figure A.3B). This suggests that Pb^{2+} can support formation of the RNA G-quadruplex in Spinach, thereby allowing RNA to bind the fluorophore and activate its fluorescence. Even at 1 μM concentration of Pb^{2+} we observed a strong fluorescence signal that was visible under a hand-held UV lamp (Figure A.4).

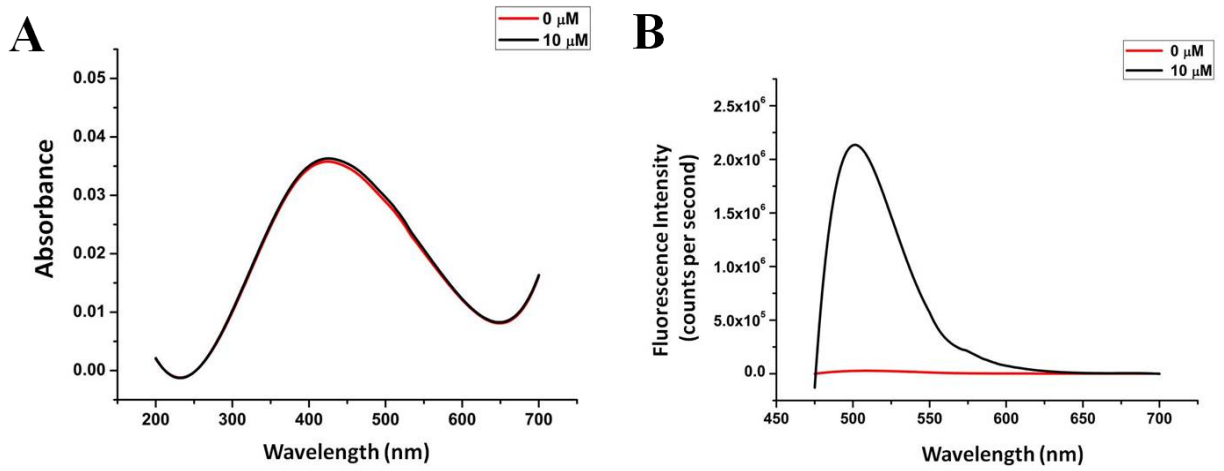

Figure A.4 Absorption (A) and emission (B) spectra for the Spinach sensor in the absence (red) and presence (black) of 10 μM Pb^{2+}. A. Absorption spectra of the sensor are identical in the absence or presence of Pb^{2+} (two spectra are superimposed on each other). **B.** Emission spectra of the sensor indicate that the Spinach sensor shows enhanced fluorescence in presence of Pb^{2+}. Concentrations of RNA and DFHBI were 5 μM and 20 μM respectively.

To test the sensitivity of the Spinach sensor, we measured fluorescence at different concentrations of Pb^{2+} and signal enhancement was observed up to 10 μM (Figure A.5), consistent with increased folding of the RNA into a quadruplex with increased concentrations of Pb^{2+}. Increasing concentrations of Pb^{2+} in the presence of DFHBI (without the Spinach aptamer) showed no fluorescence enhancement, establishing the role of the RNA in fluorescence activation of DFHBI (Figure A.6). The decrease in fluorescence after 10 μM Pb^{2+} might reflect non-specific quenching by Pb^{2+} (Figure A.5A, C) (Yuen et al., 2014). The signal response was linear in the range 5 nM-500 nM Pb^{2+} (inset Figure A.5B) and the detection limit for Pb^{2+} was 6 nM (based on 3σ/slope method, where σ is the standard deviation of blank), which is well below the maximum permissible level for Pb^{2+} concentration in drinking water (72 nM or 15 ppb) (Wang et al., 2013). The linear range and low detection limit make our Spinach sensor suitable for quantitative detection of Pb^{2+} at low to moderate concentrations that are likely to be encountered in real-world samples.

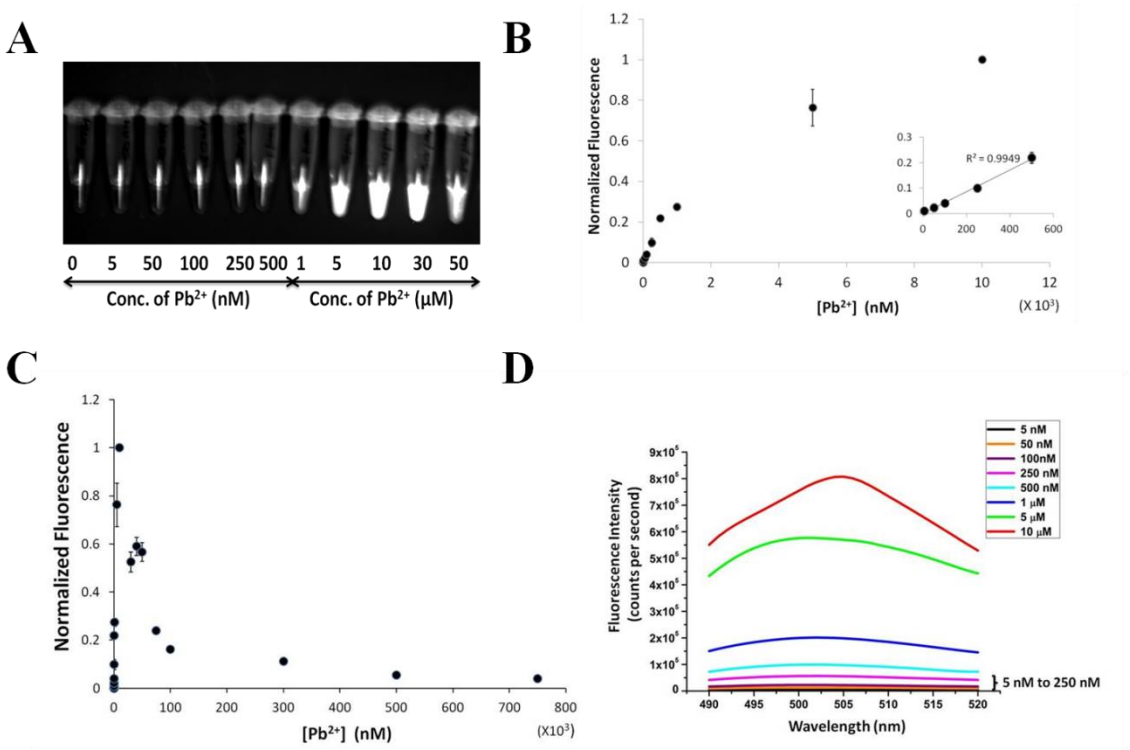

Figure A.5 Fluorescence response of the Spinach sensor to increasing concentrations of Pb^{2+}. **A.** Spinach sensor gives a strong fluorescence signal at Pb^{2+} concentrations ≥ 1 µM as observed under UV light. RNA and DFHBI concentrations were 5 µM and 20 µM respectively. **B.** Spinach fluorescence versus the concentration of Pb^{2+}. Inset: signal increases linearly (R^2=0.9949) with Pb^{2+} concentration in the range of 5 nM-500 nM. Experiments were performed at a pH 7.5 and 5 mM Mg^{2+}. RNA and DFHBI concentrations were 100 nM and 1 µM, respectively. **C.** Complete plot for **B** showing response of Spinach sensor to increasing concentrations of Pb^{2+} (5 nM-750 µM). Fluorescence decreases after 10 µM. **D.** Fluorescence spectra in the region 490 nm–520 nm (excitation wavelength 468 nm). The color code indicates concentrations of Pb^{2+}. RNA and DFHBI concentrations were 5 µM and 20 µM, respectively.

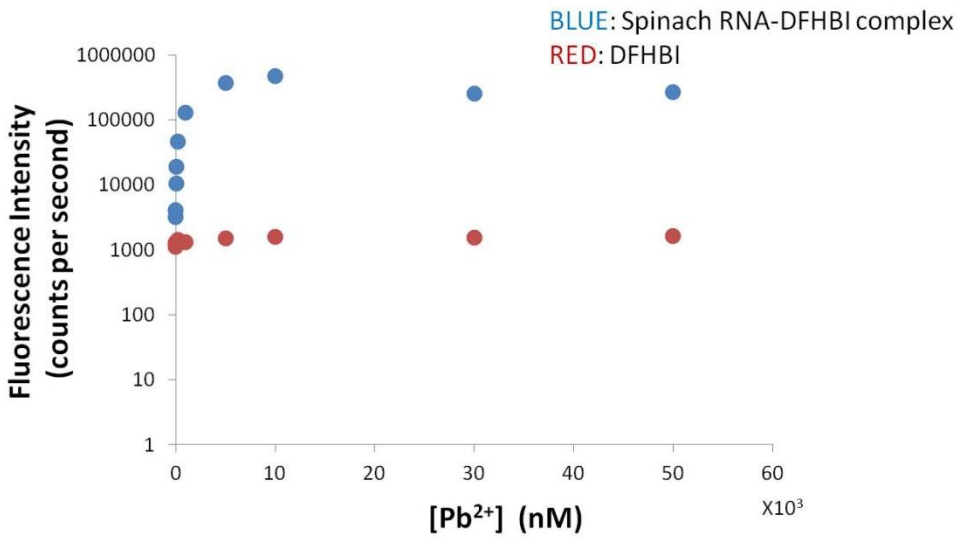

Figure A.6 Fluorescence signal of DFHBI with different concentrations of Pb^{2+} (5 nM-50 µM) in the presence (blue) or absence (red) of spinach RNA. Fluorescence signal remains constant over the wide concentration range of Pb^{2+} in the absence of RNA but shows a sharp rise in its presence. RNA and DFHBI concentrations were 5 µM and 20 µM, respectively.

One of the motivations of using a biomolecular sensor was the expected high selectivity due to reasons outlined above (section A.1). To test selectivity of the Spinach sensor toward Pb^{2+} we tested it with different environmentally relevant cations. Cations known to stabilize G-quadruplexes, Na^{+}, K^{+} and Ca^{2+} (Smirnov and Shafer, 2000; Lee et al., 2007) showed fluorescence signals that were only slightly above background even when present in 30-fold (Na^{+}) and 3-fold (K^{+}, Ca^{2+}) excess over Pb^{2+}. The Spinach sensor was more than 17000-fold selective for Pb^{2+} compared to the most competing ion Ca^{2+} (inset Figure A.7), which to our knowledge makes it the most selective nucleic acid-based sensor for Pb^{2+} (Table A.1). Fluorescence signal due to the presence of Pb^{2+} was virtually unaffected in a metal soup that contained Ag^{+}, Ca^{2+}, Co^{2+}, Cr^{3+}, Cd^{2+}, K^{+}, Na^{+}, Cu^{2+}, Fe^{2+}, Mn^{2+}, Ni^{2+}, Zn^{2+} (Figure A.7). These observations highlight the high selectivity of the sensor and demonstrate its potential utility for analysis of Pb^{2+} in samples containing other metal ions.

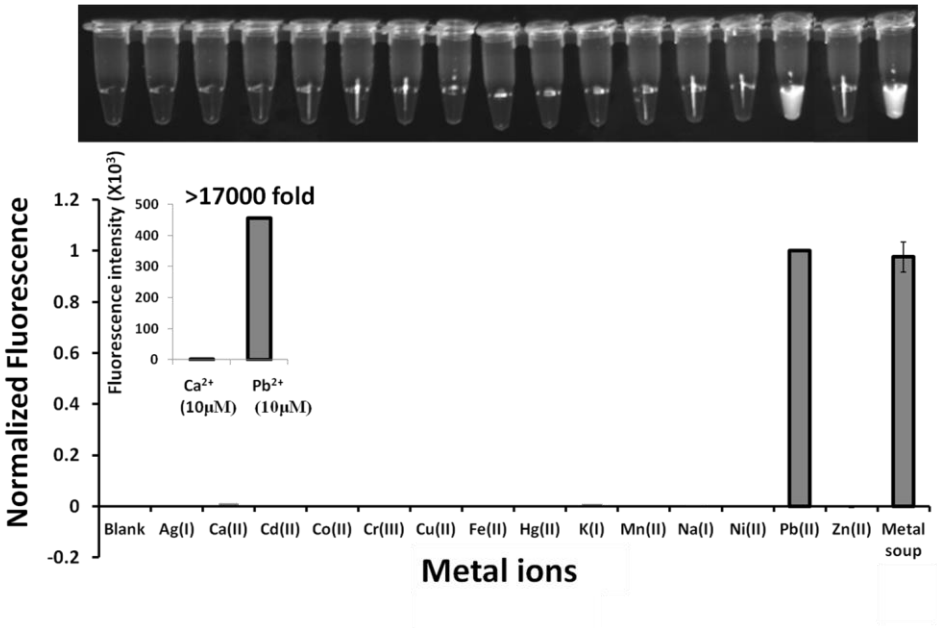

Figure A.7 Spinach sensor detects Pb^{2+} with greater than 17000-fold selectivity relative to the most interfering cation Ca^{2+} (inset). Pb^{2+} was assayed at 10 μM, Ca^{2+} and K^{+} at 30 μM and the remaining metal ions at 300 μM. Experiments were performed at a pH 7.5 and 5 mM Mg^{2+}. RNA and DFHBI concentrations were 5 μM and 20 μM, respectively.

To test the role of Pb^{2+}-induced G-quadruplex formation in the fluorescence signal emitted by the Spinach sensor, we performed circular dichroism (CD) spectroscopy on the Spinach sensor (Spinach RNA-DFHBI complex) in the absence and presence of Pb^{2+}. An enhancement in the 264 nm peak in the CD spectrum of Spinach upon addition of 10 μM Pb^{2+} suggested the formation of a G-quadruplex in presence of Pb^{2+} (Figure A.8A). The relatively low peak enhancement presumably reflects the fact that only about 14% (8 out of the total of 57 nucleotides) of the RNA folds into a G-quadruplex. For further confirmation, we made directed mutations to the guanines in both layers of the quadruplex in order to disrupt its formation. This completely abrogates Pb^{2+}-induced Spinach fluorescence, providing further evidence that Pb^{2+} supports formation of the G-quadruplex motif in Spinach (Figure A.8B).

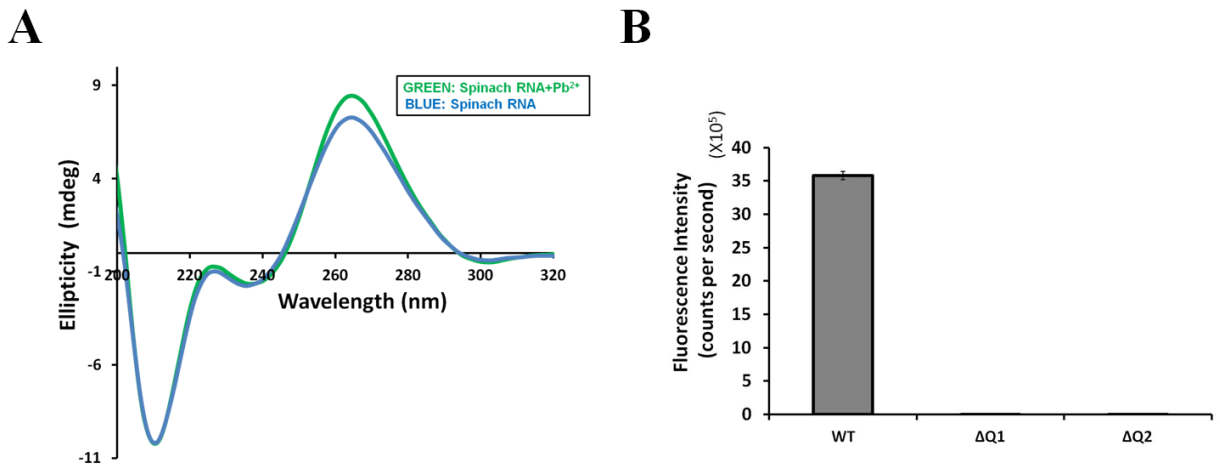

Figure A.8 Spinach sensor signal is due to Pb^{2+}-induced formation of G-quadruplex. A. CD spectra of Spinach-DFHBI complex. A distinct enhancement of the positive peak at 264 nm after addition of 10 μM Pb^{2+} (green) indicates formation of G-quadruplex in Spinach RNA in the presence of Pb^{2+} at this concentration. The spectrum in blue is obtained without Pb^{2+}. Experiments were performed at a pH 7.5 and 5 mM Mg^{2+}. **B.** Wild-type Spinach sensor shows strong fluorescence signal in presence of 10 μM Pb^{2+}, but quadruplex mutants of the sensor, ΔQ1 and ΔQ2 do not show fluorescence over background. RNA and DFHBI concentrations were 5 μM and 20 μM respectively.

A.2.2 Binding studies with the Spinach sensor: metal ions and DFHBI

Spinach RNA has to bind both fluorophore and the quadruplex-stabilizing cation in order to fluoresce. Two cations that activated fluorescence (albeit at high concentrations) in Spinach, in addition to Pb^{2+} were Ca^{2+} and K^{+}. We investigated the nature of cation (Pb^{2+}, Ca^{2+} and K^{+}) binding to Spinach RNA in the presence of the DFHBI fluorophore and vice-versa to understand the nature of dependence of the sensor on its various components. We measured Spinach fluorescence in the presence of increasing concentrations of Pb^{2+}, Ca^{2+} and K^{+} at lower concentrations of RNA (100 nM) and DFHBI (1μM). A sigmoidal increase in fluorescence with increase in Pb^{2+} concentration, is in sharp contrast to the response curves for K^{+} and Ca^{2+} binding that exhibit reduced and no signal enhancements, respectively at low concentrations of RNA and fluorophore (Figure A.9A).

However, at higher concentrations ([RNA] = 5 μM, [DFHBI] = 20 μM), both cations exhibit similar sigmoidal response curves (Figure A.9B, C). The apparent K_D values for metal ion binding to Spinach RNA were 0.0013±0.0004 mM (Hill coefficient n=1.5±0.4), 2.3±1.1 mM (Hill coefficient n=0.7±0.3) and 9.6±0.4 mM (Hill coefficient n=1.6±0.1) (Huang et al., 2014) for Pb^{2+}, Ca^{2+} and K^+, respectively. Thus, Pb^{2+} binds to Spinach with an affinity that is three orders of magnitude greater than that of K^+ or Ca^{2+}, further emphasizing the selectivity of the sensor towards Pb^{2+}.

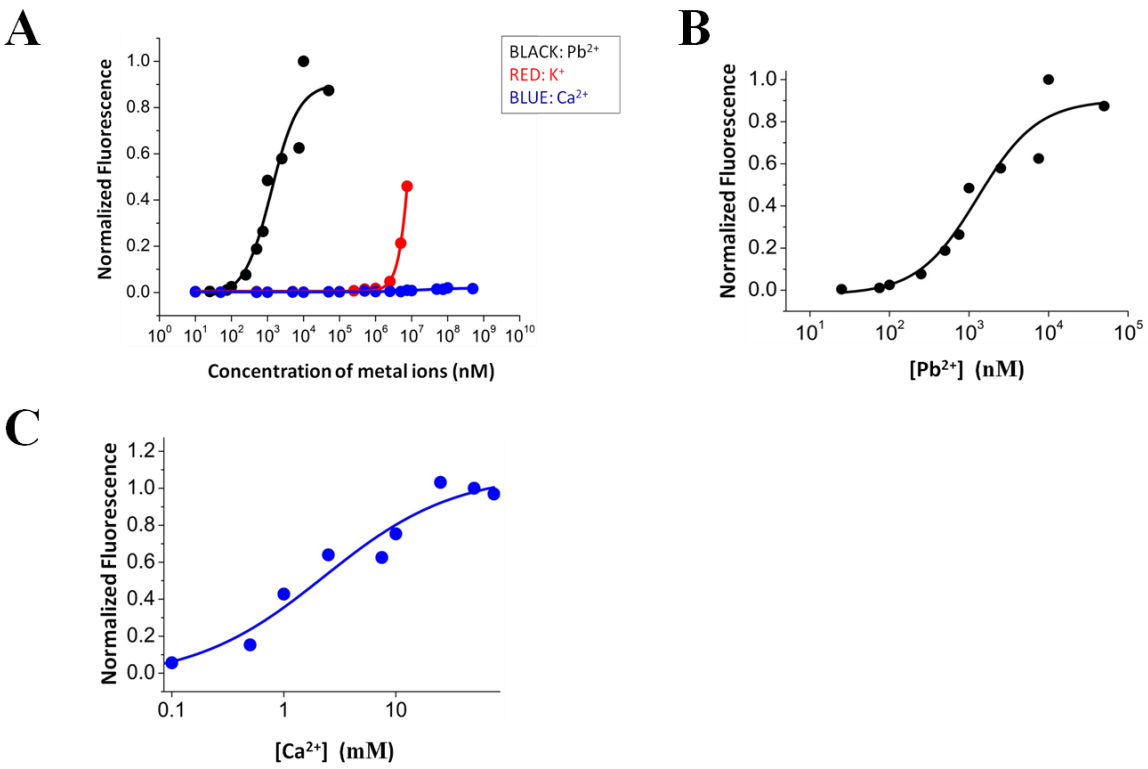

Figure A.9 Concentration dependence of Spinach fluorescence activation by Pb^{2+} and Ca^{2+}. **A.** Concentration dependence at low RNA/DFHBI concentrations. RNA and DFHBI concentrations were 100 nM and 1 μM, respectively. **B.** Pb^{2+} shows a strong affinity towards Spinach RNA with an apparent K_D of 1.29 ± 0.45 μM (n=1.5±0.4). **C.** Ca^{2+} shows a weaker binding affinity with an apparent K_D of 2.29 ± 1.12 mM (n=0.7±0.3). Concentrations of RNA and DFHBI were 100 nM and 1 μM, respectively, for the Pb^{2+} assays and 5 μM and 20 μM, respectively, for Ca^{2+} assays.

In a complimentary experiment, fluorescence was measured at increasing concentrations of DFHBI, in the presence of 10 μM Pb^{2+} (optimal concentration for Pb^{2+}), Ca^{2+} and K^{+}. Unsurprisingly, only Pb^{2+} could activate fluorescence in Spinach at low micromolar concentrations across a range of DFHBI concentrations (0.01-10 μM) (Figure A.10A, B). This observation, in addition to a markedly lower K_D value for Pb^{2+} binding, provides credence to our original hypothesis regarding the tight binding of Pb^{2+} to an RNA G-quadruplex. However, at higher concentrations of cations, a sigmoidal increase in fluorescence similar to that in the presence of Pb^{2+} was observed for Ca^{2+} [+](Figure A.10C) and K^{+} (Huang et al., 2014). K_D values calculated for DFHBI binding in presence of 10 μM Pb^{2+}, 100 mM K^{+} and 50 mM Ca^{2+} were 1.14±0.09 μM, 0.30±0.68 μM (Huang et al., 2014) and 20.45±5.85 μM, respectively.

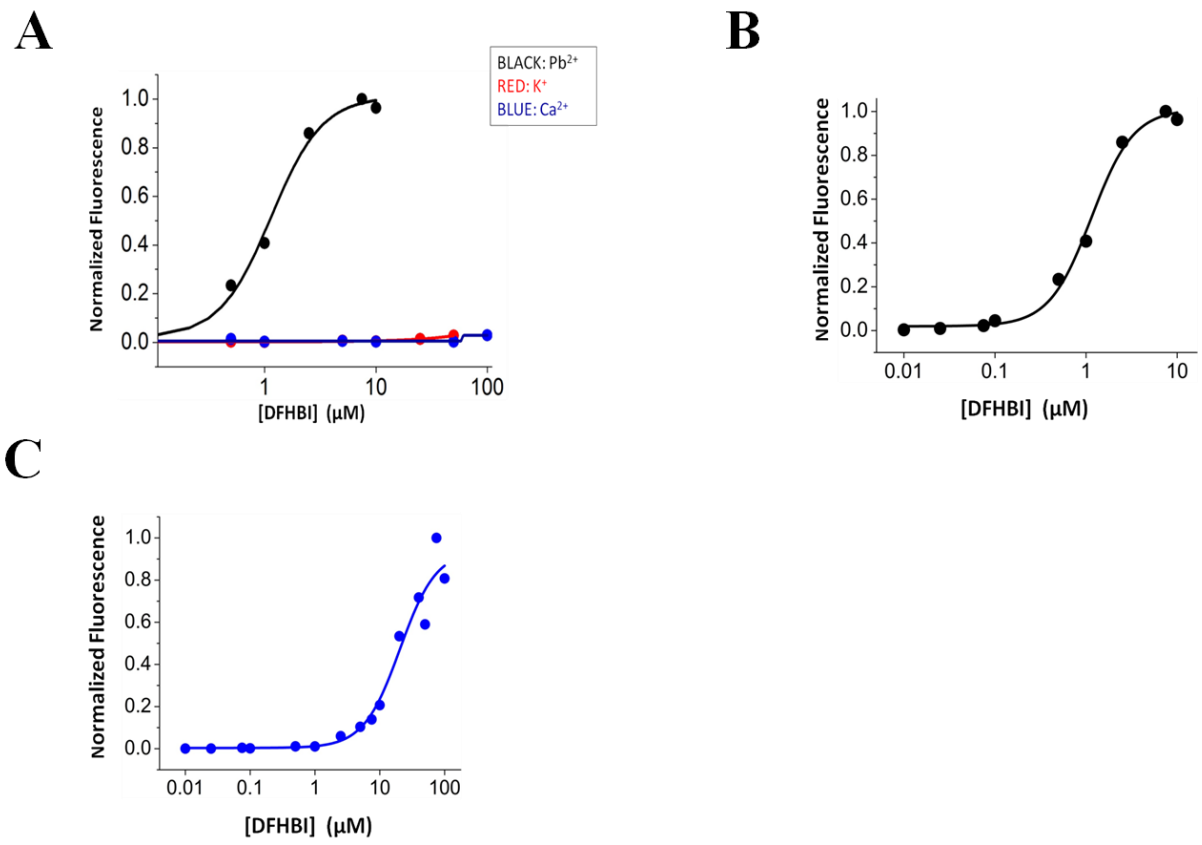

Figure A.10 Dependence of Spinach fluorescence activation on DFHBI concentration in the presence of Pb^{2+} and Ca^{2+}. A. At 10 μM concentrations of metal ions only the Pb^{2+} stabilized sensor has the capacity to activate fluorescence over the DFHBI concentration range tested (black). In contrast, the K^+ and Ca^{2+} stabilized sensors required much higher concentrations of DFHBI (red and blue, respectively). RNA concentration was 30 nM. **B.** Binding curve for DFHBI binding to Spinach in the presence of 10 μM Pb^{2+}. A fit of the data to the Hill equation gave an apparent K_D of 1.14 ± 0.09 μM. **C.** Binding curve for DFHBI binding to Spinach in the presence of 50 mM Ca^{2+}. A fit of the data to the Hill equation gave an apparent K_D of 20.45 ± 5.85 μM. RNA concentration was 30 nM in for both A and B.

A.2.3 Real-world application: detecting lead (II) in tap water

Detection methods based on Pb^{2+}-induced G-quadruplex formation tend to yield low signal to noise ratios due to the presence of high concentrations (in the millimolar range) of metal ions like K^+, Na^+ and Ca^{2+} that induce G-quadruplex formation. High background signals make their use in practical samples potentially problematic (Zhan et al., 2013). To assess this possible limitation in the Spinach sensor, we tested its performance using samples of tap water. First, we examined the sensor's stability in tap water and observed no degradation over the course of the measurement including after incubation at 37ºC overnight in presence of 10 μM Pb^{2+} (Figure A.11A). The stability of Spinach at 37ºC overnight (~16 h) was encouraging as the incubation times for the Spinach sensor fall between 15-30 minutes. The sensor was stable to degradation between pH 3 and pH 11 in the time frame of detection (Figure A.11B) and fluorescence signal was optimal between pH 6 and pH 8 (Figure A.11C).

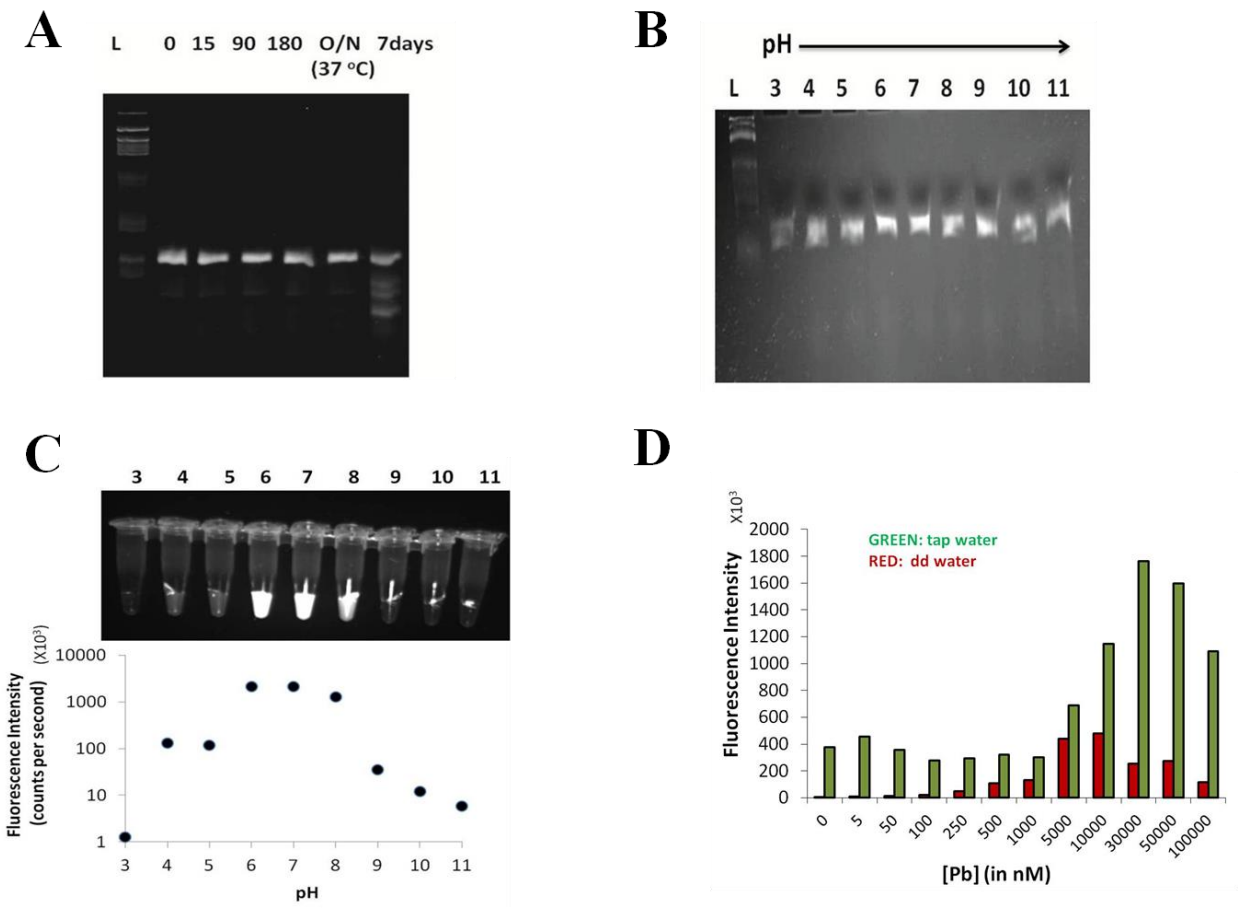

Figure A.11 Applicability of the Spinach sensor in 'real-world' conditions. A. Denaturing polyacrylamide gel electrophoresis (dPAGE) showing the stability of Spinach sensor in tap water in the presence of 10 μM Pb^{2+} incubated at 25ºC unless otherwise mentioned. Tested time points are 15, 90, 180 minutes, overnight (O/N) and 7 days. The RNA aptamer shows minimal cleavage in the presence of 10 μM Pb^{2+} in tap water after incubation at 37ºC overnight. **B.** dPAGE showing the stability of Spinach sensor at different pH conditions (3-11) incubated with 10 μM Pb^{2+} at 25ºC for 60 minutes. L denotes low-range ssRNA ladder. **C.** Activity of spinach sensor was interrogated in a broad range pH 3-11. Spinach sensor shows optimal activity in the pH range 6-8. RNA and DFHBI concentrations were 5 μM and 20 μM, respectively. **D.** Spinach sensor (using [RNA]/[DFHBI] = 5 μM/20 μM) showed high background fluorescence in tap water (green bar) even in the absence of Pb^{2+}, unlike in ddH_2O. The overall dependence of fluorescence signal on Pb^{2+} concentration is, however, similar to that in ddH_2O.

ICP-MS analysis of tap water samples showed that Ca^{2+}, Na^{+} and K^{+} were present in 568 μM (22.7 ppm), 350 μM (8.1 ppm) and 28 μM (1.1 ppm) concentrations, respectively. The absence of significant fluorescence in the presence of K^{+} and Na^{+} at these concentrations (Figure A.7) indicated that the presence of Ca^{2+} caused the observed background signal when the Spinach sensor was used in tap water (Figure A.11D). To identify conditions that improve selectivity for detection of Pb^{2+} over Ca^{2+}, we re-analyzed the DFHBI binding curves (fluorescence signal as a function of increasing DFHBI concentrations) in the presence of a fixed concentration of Pb^{2+} or Ca^{2+} (Figure 1.10). We found that Spinach required significantly higher concentrations of DFHBI to give the same intensity of fluorescence signal in the presence of Ca^{2+} compared Pb^{2+}. For example, in the presence of 50 mM Ca^{2+}, Spinach did not show substantial signal above background when the concentration of DFHBI was below 10 μM, (Figure A.10C) but significant signal was observed in the presence of 10 μM Pb^{2+}, even at DFHBI concentrations as low as 0.5 μM (Figure A.10B). Thus, the background signal due to Ca^{2+} was minimized by lowering the concentration of DFHBI used from 20 μM (used with 5 μM RNA for most assays) to 1 μM (used with 100 nM RNA). These conditions rendered the sensor completely insensitive to the presence of Ca^{2+}in tap water. We spiked tap water with different concentrations of Pb^{2+}. Fluorescence signal increased linearly with Pb^{2+} concentration in the range of 100 nM-5 μM (Figure A.12). These results demonstrate a potential real-world application of the Spinach sensor for detecting Pb^{2+} in tap water.

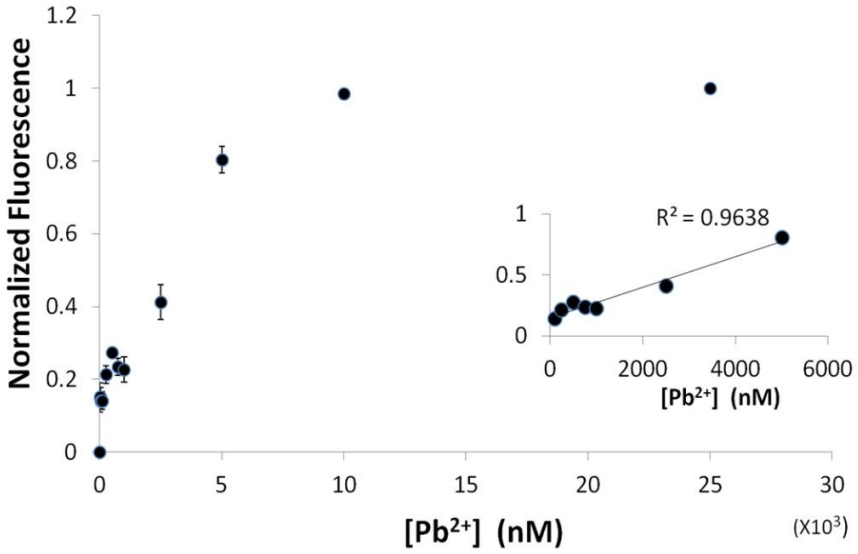

Figure A.12 Spinach sensor detects lead (II) in tap water. The sensor responds linearly (R^2=0.9638) to Pb^{2+} concentrations in the range of 100 nM-5 μM in tap water (inset). RNA and DFHBI concentrations were 100 nM and 1 μM, respectively.

A.3 DISCUSSION AND SUMMARY

We have developed the first RNA-based sensor for Pb^{2+} that expands the applications of the Spinach RNA aptamer. The Spinach lead sensor consists of RNA that can be easily synthesized from inexpensive deoxyoligonucleotide templates. The Spinach fluorophore, DFHBI can be purchased from Lucerna™ (http://www.lucernatechnologies.com/DFHBI-5-mg-p25.html) or synthesized in-house (Huang et al., 2014). The distinct advantages of the Spinach lead sensor can be summarized as follows (See table A.1 for comparison with other nucleic acid based Pb^{2+} sensors):

1. It does not need covalent modification to the nucleic acid precluding additional chemical steps. The Spinach fluorophore can just be added to the RNA for creating the RNA-fluorophore complex.
2. It does not require nanoparticles or quantum dots as part of its assembly, thereby reducing complexity and cost.
3. High quantum yield of the fluorophore produces a bright signal that is fairly stable over a period of days. This allows ‘on-spot ‘visualization using a handheld UV lamp without requiring elaborate laboratory setup.
4. The primary advantage of our Spinach lead sensor is its excellent selectivity (in addition to a high sensitivity). To our knowledge this is the highest among nucleic acid-based lead sensors and is comparable to the best lead detection systems available.

Other potential applications of our sensor could include measuring Pb^{2+} in paint, which contains no interfering G-quadruplex stabilizing cations (Mazumder et al., 2010). In this case, detection must be preceded by hydrochloric acid treatment to bring the lead in solution (as soluble $PbCl_2$). The Spinach sensor could also be used to detect/quantify Pb^{2+} in contaminated cells. The

higher limit for Pb^{2+} in healthy blood serum is 2.4 μM, which falls well within the range of detection for the Spinach sensor. The low millimolar concentrations of Ca^{2+} and K^{+} in blood serum (2.7 mM and 4 mM respectively) are below or almost equal to their respective K_D values and Na^{+} is in high millimolar concentrations. Even in the absence of Pb^{2+}, this would lead to high background fluorescence due to the formation of the G-quadruplex platform in Spinach, triggered by these cations. Blood samples must be pre-treated with chelators like EGTA for Ca^{2+} or crown ethers of different cavity sizes for Na^{+} and K^{+}. The spinach sensor could also be adapted as an 'on-spot' dipstick system. This would require the RNA to be derivatized with biotin for immobilization on a streptavidin-coated surface. The fluorophore needs to be immobilized as well, which could likely be possible by covalent linkage to the RNA at the binding pocket to constitute the primed sensor. This device would be dipped in analyte solution for a certain period of time, followed by incubation at 37°C (to allow for RNA folding). Fluorescence could be visualized by holding it under a hand-held UV lamp or quantified in a plate-reader. However, these are entirely conceptual applications that need to be tested and optimized for effective implementation.

A DNA-version of the Spinach sensor could be a better alternative to the current RNA-based device in terms of greater stability and cost. However, since the fluorophore binding pocket of Spinach RNA uses 2'-OH to mediate stabilizing interactions between RNA and fluorophore (Huang et al., 2014), a 'deoxySpinach' would likely not be active. An effective DNA-Spinach could be obtained by carrying out a fresh *in vitro* selection for DFHBI binding from a random pool of DNA, possibly enriched in guanines to bias the selection toward the emergence of a G-quadruplex-containing aptamer.

Mechanism of action	Detection method	Sensitivity (Detection Limit)	Selectivity (w.r.t. the most interfering cation)	Incubation time	Linear Range	Ref.
Pb^{2+} induced formation of G-quadruplex	**Fluorescence *(Increase)***	**6 nM**	**~ 17000-fold**	**15 min**	**5-500 nM**	**This work**
	Fluorescence *(Decrease)*	1.0 nM	~ 5-fold	30 min	5 nM-1 μM	(a)
	Fluorescence *(Decrease)*	0.4 nM	~ 15-fold	15 min	0.5 – 500 nM	(b)
	Fluorescence *(Increase)*	5 nM	~ 30-fold	2 hours	20 nM-1 μM	(c)
	Fluorescence *(Decrease)*	~4 nM	~ 3-fold	5 min	0-960 nM	(d)
	Fluorescence *(Anisotropy)*	1.0 nM	~ 50-fold	30 min	10 nM-2 μM	(e)
	Electrochemistry *(Increase)*	0.4 nM	~4-fold	10 min	1 nM-1 μM	(f)
	Electrochemistry *(Decrease)*	0.075 nM	~ 5-fold	30 min	0.1 nM-0.1 μM	(g)
	Electrochemistry *(Increase)*	3.3 pM	-	-	0.01 – 160 nM	(h)
	Surface Plasmon resonance (SPR) *(Increase)*	>0.4 nM	-	5-6 hours	-	(i)
Pb^{2+} induced formation of G-quadruplex DNAzyme	Fluorescence *(Increase)*	0.4 nM	~4 fold	2-3 hours	0-1 μM	(j)
	Fluorescence *(Increase)*	0.05 nM	~ 40-fold	2 hours	0-100 nM	(k)
	Colorimetry *(Decrease)*	32 nM	~ 6-fold	1-2 hours	100 nM-10 μM	(l)
	Chemiluminiscence *(Decrease)*	1 nM			10^{-9}-$10^{-6.5}$ M	
	Electrochemistry *(Decrease)*	0.5 nM	-	4 hours	-	(m)
	Surface Plasmon Resonance	~ 0.005 pM 0.1 nM	~ 25-fold	-	-	(n)
	Electrochemistry *(Decrease)*		~ 6-fold			

Table A.1 Summary of nucleic acid-based Pb^{2+} sensors. Sensors that fall into three broad classes but use multiple detection methods are compared in terms of their selectivity, sensitivity, incubation times and concentration range of Pb^{2+} for linear response.

Mechanism of action	Detection method	Sensitivity (Detection Limit)	Selectivity (w.r.t. the most interfering cation)	Incubation time	Linear Range	Ref.
Pb^{2+} DNAzyme catalyzed RNA cleavage	Fluorescence *(Increase)*	3.7 nM	~ 6000-fold	6 min	-	(o)
	Fluorescence *(Increase)*	-	~ 40-fold	15 min	10 nM-4 μM	(p)
	Fluorescence *(Increase)*	~ 3 nM	~ 500-fold	30 min	5-100 nM	(q)
	Electrochemistry *(Decrease)*	0.9 pM	~ 2-fold	1 hour	2 pM-1nM	(r)
	Electrochemistry *(Increase)*	0.6 mM	~ 50-fold	1 hour	1 nM-1 μM	(s)

Table A.1 Summary of nucleic acid-based Pb^{2+} sensors (continued). Sensors that fall into three broad classes but use multiple detection methods are compared in terms of their selectivity, sensitivity, incubation times and concentration range of Pb^{2+} for linear response.

References for Table A.1

(a) Guo, L. et al., 2012. *Biosens. Bioelectron.*, 35, 123-127.
(b) Wang, W., Jin, Y.,Zhao, Y.,Yue, X. and Zhang, C., 2013. *Biosens. Bioelectron.*,41, 137-142.
(c) Li, T., Dong, S. and Wang, E., 2010. *J. Am. Chem. Soc*., 132, 13156-13157.
(d) Zhan, S. et al., 2013. *RSC Adv*., 3, 16962-16966.
(e) Zhang, D., Yin, L., Meng, Z. A., Yu, Z., Guo, L. and Wang, H., 2014. *Anal. Chim, Acta.*, 812, 161- 167.
(f) Li, F., Feng, Y., Zhao, C. and Tang, B., 2011. *Chem. Commun.*, 47, 11909-11911.
(g) Li, F., Yang, L., Chen, M., Li, P. and Tang, B., 2013. *Analyst*, 138, 461-466.
(h) Liu, M. et al., 2010. *Environ. Sci. Technol*., 44, 4241-4246.
(i) Cheng, S. et al. *Biosens. Bioelectron*., 2014, 53, 479-485.
(j) Li, C.-L., Liu, K.-T., Lin, Y.-W. and Chang, H.-T., , 2011. *Anal. Chem.*, 83, 225-23.
(k) Li, C.-L., Huang, C.-C., Chen, W.-H., Chiang, C.-K. and Chang, H-T., 2012. *Analyst*, 137, 5222-5228.
(l) Li, T., Wang, E. and Dong, S., 2010. *Anal. Chem.*, 82, 1515-1520.
(m) Li, F., Yang, L., Chen, M.; Qian, Y.and Tang, B., 2013. *Biosens. Bioelectro*n., 41, 903-906.
(n) Pelossof, G., Tel-Vered, R. and Willner, I., 2012. *Anal. Chem*., 84, 3703-3709.
(o) Lan, T., Furuya, K. and Lu, Y., 2010. *Chem. Commun.*, 46, 3896-3898.
(p) Lu, Y., Liu, J., Li, J., Bruesehoff, P. J., Pavot, C. M.-B. and Brown, A. K., 2003. *Biosens. Bioelectron*., 41, 529-540.
(q) Li, H., Zhang, Q., Cai, Y., Kong, D.-M. and Shen, H.-X., 2012. *Biosens. Bioelectron.*, 34, 159-164.
(r) Gao, A., Tang, C.-X., He, X.-W. and Yin, X.-B., 2013. *Analyst*, 138, 263-268.
(s) Zhang, H., Jiang, B., Xiang, Y., Su, J., Chai, Y. and Yuan, R., 2011. *Biosens. Bioelectron.*, 41, 529-540.

A.4. MATERIALS AND METHODS

A.4.1 Construct preparation, RNA synthesis and purification, and other chemicals

The sequence of the Spinach sensor RNA is: 5'-GGGGAGAAGGACGGGUCCAGUG CGAAACACGCACUGUUGAGUAGAGUGUGAGCUCCC-3'. DNA cloning, synthesis (by *in vitro* transcription) and purification of RNA by FPLC was performed following the procedure outlined in ref. 9a (Huang *et al.*). DFHBI was synthesized by following a previously reported scheme in ref. 9a (Huang *et al.*). Metal salts like $Pb(CH_3COO)_2.3H_2O$ (used as a source of Pb^{2+} for all experiments) were purchased from Sigma-Aldrich (ACS grade).

A.4.2 RNA stability assays

Stability of the spinach sensor was assayed at different conditions relevant to possible real-life contexts. For tap water assays, the sensor was incubated with 10 μM Pb^{2+} at 25ºC in tap water for 15, 90, 180 minutes and 7 days and at 37ºC overnight (~16 h). For pH assays, the sensor was incubated with 10 μM Pb^{2+} at 25ºC for 60 minutes in buffers with pH from 3 to 11. Aliquots were mixed with standard loading dye and loaded on an analytical dPAGE. Gel was stained with ethidium bromide for visualization under UV.

A.4.3 Fluorescence assays

Purified RNA was ethanol precipitated twice and refolded in the presence of a folding buffer containing 10 mM Tris (pH 7.5) and 5 mM Mg^{2+} (heating at 90ºC for two minutes in water followed incubation on ice for 5 minutes; then heating at 50ºC for 30 minutes in presence of 5 mM Mg^{2+} followed incubation on ice for 5 minutes). All experiments were performed in the same buffer conditions. For metal binding assays, fluorophore DFHBI was added to the RNA and

incubated at 37°C for 10 minutes. Concentrations of RNA and DFHBI used for sensitivity and selectivity assays were 5 μM and 20 μM, respectively, unless otherwise mentioned. Different concentrations of Pb^{2+} (for sensitivity and selectivity assays) or other metal ions (for selectivity assays) were added and incubated at 37°C for 15 minutes. For DFHBI binding assays, RNA was taken in 30 nM concentration. For tap water experiments, 100 nM RNA and 1 μM DFHBI was used. Fluorescence emission was measured using a Fluorolog-3 spectrofluorometer equipped with a thermo-controller (Horiba Inc.) at an excitation wavelength of 468 nm and emission range of 490-520 nm with a slit width of 5 nm; results were the average of three consecutive independent measurements and each measurement was performed in triplicate (the data points represent the average of these triplicate measurements). All measurements were done at 25°C. Data were normalized and plotted in Microsoft Excel and Origin.

A.4.4 RNA-metal ion and RNA-fluorophore affinity measurements

Binding affinities of Pb^{2+} and Ca^{2+} to Spinach RNA were determined by measuring the increase in fluorescence as a function of concentration of these cations in the presence of 100 nM RNA and 1 μM DFHBI and 5 μM RNA and 20 μM DFHBI, respectively. Determination of binding affinity of the fluorophore to metal bound RNA was performed in presence of 10 μM Pb^{2+} or 50 mM Ca^{2+} and 30 nM RNA. For each concentration of fluorophore, a background signal for the fluorophore without RNA in the buffer was also measured and subtracted from the signal observed for the fluorophore with RNA. Curves were fitted to the Hill equation: $y= (C_0 + Cmax*x^{n)}/(K_D^n + x^n)$. All measurements were done at 25°C. Data were normalized and plotted in Origin.

A.4.5 Circular dichroism (CD) measurements

Measurements were performed using a JASCO-1500 CD Spectrometer. Samples were prepared same way as in the case of fluorescence experiments. Concentrations of RNA and DFHBI were 5 μM and 20 μM, respectively. Three scans from 180 to 320 nm at 1 nm intervals were accumulated with a scan rate of 100 nm min^{-1} and averaged (Path length=1 mm). All measurements were done at 25°C. Data were plotted in Microsoft Excel.

A.4.6 Mutagenesis of the G-quadruplex region

The following variants of the Spinach sensor RNA sequence were used:

ΔQ1: 5'-GGGGAGAAG**C**ACG**U**GUCCAGUGCGAAACACGCACUGUUGA**C**UAGA**U**

UGUGAGCUCCC-3'

ΔQ2: 5'-GGGGAGAA**C**GAC**U**GGUCCAGUGCGAAACACGCACUGUUGAGUA**C**AG

U**U**UGAGCUCCC-3'

The wild-type sequence (the sequence of the Spinach sensor) used as a positive control was:

WT: 5'-GGGGAGAAGGACGGGUCCAGUGCGAAACACGCACUGUUGAGUAGAGU

GUGAGCUCCC-3'

The G-quadruplex guanines are highlighted in blue and the G mutations are highlighted in red. Fluorescence experiments were performed as indicated above. Concentrations of RNA and DFHBI were 5 μM and 20 μM, respectively. For each construct (wild-type and two mutants), the background signal without Pb^{2+} was subtracted from the signal observed with Pb^{2+}. Data were plotted in Microsoft Excel.

A.4.7 ICP-MS analysis of tap water

Data were collected using an Agilent 7700x inductively coupled plasma mass spectrometer (ICP-MS) using ICP-MS MassHunter version B01.03. Samples were diluted in a 2% HNO_3 matrix and analyzed against an 8 point standard curve (for Pb^{2+}) or 5 point standard curve (for Na^{+}, Ca^{2+}, and K^{+}) over the range from 0.05 ppb to 25 ppb (for Pb) or 0.1 ppm to 1 ppm (for Na^{+}, Ca^{2+}, and K^{+}). For all analyses, the correlation coefficient was > 0.9990 for all analytes of interest. The signal was normalized against a Tb 159 internal standard. Data collection was performed in Spectrum Mode with five replicates per sample and 100 sweeps per replicate. Each sample was analyzed three times for consistency.

REFERENCES

1. Akter, F. and Yokobayashi, Y., 2014. RNA signal amplifier circuit with integrated fluorescence output. *ACS Synth. Biol.*, 4, 655-658.

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

3. Bell, C., Lynam, E., Landfair, D. J., Janjic, N. and Wiles, M. E., 1999. Oligonucleotide NX1838 inhibits VEGF165-mediated cellular responses *in vitro*. *In Vitro Cell. Dev. Biol.-Animal*, 35, 533-542.

4. Berens, C., Groher, F. and Suess, B., 2015. RNA aptamers as genetic control devices: the potential of riboswitches as synthetic elements for regulating gene expression. *Biotechnol J.*, 2, 246-257.

5. Bhadra, S. and Ellington, A. D., 2014. A Spinach molecular beacon triggered by strand displacement. *RNA*, 20, 1183–1194.

6. Cho, E. J., Lee, J. W., and Ellington, A. D., 2009. Applications of aptamers as sensors. *Annu. Rev. Anal.Chem.*, 2, 241-264.

7. Deigan-Warner, K. D., Chen, M. D., Song, W., Strack, R. L., Thorn, A., Jaffrey, S. R. and, Ferré-D'Amaré, A. R., 2014. Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat. Struct. Mol. Biol.*, 21, 658–663.

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

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

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

11. Gao, A., Tang, C.-X., He, X.-W. and Yin, X.-B., 2013. Electrochemiluminescent lead biosensor based on GR-5 lead-dependent DNAzyme for $Ru(phen)_3^{(2+)}$ intercalation and lead recognition. *Analyst*, 138, 263–268.

12. Gerdes, H. H. and Kaether, C., 1996. Green fluorescent protein: applications in cell biology. *FEBS Lett.*, 389, 44-47.

13. Grate, D. and Wilson, C., 1999. Laser-mediated, site-specific inactivation of RNA transcripts. *Proc. Natl. Acad. Sci. U S A.*, 96, 6131-6136.

14. Guo, L., Nie, D., Qiu, C., Zheng, Q., Wu, H., Ye, P., Hao, Y., Fu, F. and Chen, G., 2012. A G-quadruplex based label-free fluorescent biosensor for lead ion. *Biosens. Bioelectron.*, 35, 123–127.

15. Hofer, K. L., Langejurgen, V. and Jäschke, A., 2013. Universal aptamer-based real-time monitoring of enzymatic RNA synthesis. *J. Am. Chem. Soc.*, 135, 13692–13694.

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

17. Keefe, A. D., Pai, S. and Ellington, A., 2010. Aptamers as therapeutics. *Nat. Rev. Drug Discov.*, 9, 537-550.

18. Kim, H. N., Ren, W. X., Kim, J. S. and Yoon, J., 2012. Fluorescent and colorimetric sensors for detection of lead, cadmium, and mercury ions. *Chem. Soc. Rev.*, 41, 3210-3244.

19. Kotch, F. W., Fettinger, J. C. and Davis, J. T., 2000. A lead-filled G-quadruplex: Insight into the G-quartet's selectivity for Pb^{2+} over K^{+}. *Org. Lett.*, 2, 3277–3280.

20. Lan, T., K, Furuya, K. and Lu, Y., 2010. A highly selective lead sensor based on a classic lead DNAzyme. *Chem. Commun.*, 46, 3896-3898.

21. Lee, M. P. H., Parkinson, G. N., Hazel, P. and Neidle, S., 2007. Observation of the coexistence of sodium and calcium ions in a DNA G-quadruplex ion channel. *J. Am. Chem. Soc.*, 129, 10106–10107.

22. Li, C.-L., Huang, C.-C., Chen, W.-H., Chiang, C.-K. and Chang, H.-T., 2012. Peroxidase mimicking DNA–gold nanoparticles for fluorescence detection of the lead ions in blood. *Analyst*, 137, 5222-5228.

23. Li, C.-L., Liu, K.-T., Lin, Y.-W. and Chang, H.-T., 2011a. Fluorescence detection of lead (II) ions through their induced catalytic activity of DNAzymes. *Anal. Chem.*, 83, 225–230.

24. Li, F., Feng, Y., Zhao, C. and Tang, B., 2011b. Crystal violet as a G-quadruplex-selective probe for sensitive amperometric sensing of lead. *Chem. Commun.*, 47, 11909–11911.

25. Li, F., Yang, L., Chen, M., Li, P. and Tang. B., 2013a. A selective amperometric sensing platform for lead based on target-induced strand release. *Analyst*, 138, 461-466.

26. Li, F., Yang, L., Chen, M., Qian, Y. and Tang, B., 2013b. A novel and versatile sensing platform based on HRP-mimicking DNAzyme-catalyzed template-guided deposition of polyaniline. *Biosens. Bioelectron.*, 41, 903–906.

27. Li, H., Zhang, Q., Cai, Y., Kong, D.-M. and Shen, H.-X., 2012. Single-stranded DNAzyme-based Pb^{2+} fluorescent sensor that can work well over a wide temperature range. *Biosens. Bioelectron.*, 34, 159–164.

28. Li, T., Dong, S. and Wang, E., 2010a. A lead (II)-driven DNA molecular device for turn-on fluorescence detection of lead (II) ion with high selectivity and sensitivity. *J. Am. Chem. Soc.*, 132, 13156–13157.

29. Li, T., Wang, E. and Dong, S., 2010b. Lead (II)-induced allosteric G-quadruplex DNAzyme as a colorimetric and chemiluminiscence sensor for highly sensitive and selective Pb^{2}+ detection. *Anal. Chem.*, 82, 1515–1520.

30. Li, W., Yang, Y., Chen, J., Zheng, Q., Wang, Y., Wang, F. and Yu, C., 2014. Detection of lead (II) ions with a DNAzyme and isothermal strand displacement signal amplification. *Biosens. Bioelectron.*, 53, 245–249.

31. Liu, M., Zhao, G., Tang, Y., Yu, Z., Lei, Y., Li, M., Zhang, Y. and Li, D., 2010. A simple, stable and picomole level lead sensor fabricated on DNA-based carbon hybridized TiO_2 nanotube arrays. *Environ. Sci. Technol.*, 44, 4241–4246.

32. Lu, Y., Liu, J., Li, J., Bruesehoff, P. J., Pavot, C. M-B. and Brown, A. K., 2003. New highly sensitive and selective catalytic DNA biosensors for metal ions. *Biosens. Bioelectron.*, 41, 529-540.

33. Mazumdar, D., Liu, J., Lu, G., Zhou, J. and Lu, Y., 2010. *Chem. Commun.*, 46, 1416-1418.

34. Neidle, S., 2010. Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.*, 277, 11118-11125.

35. Paige, J. S., Nguyen-Duc, T., Song, W. and Jaffrey, S. R., 2012. Fluorescence imaging of cellular metabolites with RNA. *Science*, 335, 1194-1196.

36. Paige, J. S., Wu, K. Y. and Jaffrey, S. R., 2011. RNA mimics of green fluorescent protein. *Science*, 333, 642-646.

37. Robertson, D. L. and Joyce, G. F., 1990. Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, 344, 467-468.

38. Sando, S., Narita, A., Hayami, M. and Aoyama, Y., 2008. Transcription monitoring using fused RNA with a dye-binding light-up aptamer as a tag: a blue fluorescent RNA. *Chem Commun.*, 33, 3858-3860.

39. Smirnov, I. and Shafer, R. H., 2000. Lead is unusually effective in sequence-specific folding of DNA. *J. Mol. Biol.*, 2000, 296, 1–5.

40. Song, W., Filonov, G. S., Kim, H., Hirsch, M., Li, X., Moon, J. D. and Jaffrey, S. R., 2017. Imaging RNA polymerase III transcription using a photostable RNA–fluorophore complex *Nat. Chem. Biol.*, 13, 1187–1194.

41. Strack, R. L., Disney, M. D. and Jaffrey, S. R., 2013. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat. Methods*, 12, 1219-1224.

42. Strack, R. L., Song, W. and Jaffrey, S. R., 2014. Using Spinach-based sensors for fluorescence imaging of intracellular metabolites and proteins in living bacteria. *Nat. Prot.*, 2014, 9, 146–155.

43. Tuerk, C. and Gold, L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 249, 505-510.

44. Vorlíčková, M., Kejnovská, I., Sagi, J., Renčiuk, D., Bednářová, K., Motlová, J. and Kypr, J., 2012. Circular dichroism and guanine quadruplexes. *Methods*, 57, 64-75.

45. Wang, W., Jin, Y., Zhao, Y., Yue, X. and Zhang, C., 2013. Single-labeled hairpin probe for highly specific and sensitive detection of lead(II) based on the fluorescence quenching of deoxyguanosine and G-quartet. *Biosens. Bioelectron.*, 41, 137–142.

46. Yuen, L. H., Franzini, R. M., Tan, S. S. and Kool, E. T., 2014. Large-scale detection of metals with a small set of fluorescent DNA-like chemosensors. *J. Am. Chem. Soc.*, 2014, 136, 14576–14582.

47. Zavyalova, E., Tagiltsev, G., Reshetnikov, R., Arutyunyan, A. and Kopylov, A., 2016. Cation coordination alters the conformation of a thrombin-binding G-quadruplex DNA aptamer that affects inhibition of thrombin. *Nucleic Acid Ther.*, 26, 299-308.

48. Zhan, S., Wu, Y., Liu, L., Xing, H., He, L., Zhan, X., Luo, Y. and Zhou, P., 2013. A simple fluorescent assay for lead (II) detection based on lead (II)-stabilized G-quadruplex formation. *RSC Adv.*, 3, 16962–16966.

49. Zhang, H., Jiang, B., Xiang, Y., Su, J., Chai, Y. and Yuan, R., 2011. DNAzyme-based highly sensitive electronic detection of lead via quantum dot-assembled amplification labels. *Biosens. Bioelectron.*, 28, 135–138.

Appendix - 2

ADDITIONAL AUTHOR CONTRIBUTIONS

1. **Chapters 2 and 3: Crystal structure of the VS ribozyme**
    - Analysis of VS ribozyme oligomerization, SAD phasing and solution of the VSx_G638A structure were performed by Dr. Nikolai B. Suslov.
    - Additional analysis of SAXS data was performed by Dr. James R. Fuller.
    - Additional crystallization trials (VSx_G638A) were performed by Dr. Hao Huang

2. **Appendix 1: Lead (II) detection by Spinach Sensor**
    - DFHBI, the fluorophore for Spinach aptamer was synthesized by Dr. Nan-sheng Li.
    - Dr. Sandip Shelke made Figure A.3B and participated in useful discussion at the inception of the project.