

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Ligand-responsive RNA switches: viral translation regulators, therapeutic targets, and tunable building blocks for nanotechnology

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Chemistry

by

Mark A. Boerneke

Committee in charge:

Professor Thomas Hermann, Chair
Professor Michael Gilson
Professor Ulrich Muller
Professor Emmanuel Theodorakis
Professor David Wyles

2016

Copyright

Mark A. Boerneke, 2016

All rights reserved

The Dissertation of Mark A. Boerneke is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2016

DEDICATION

I would like to dedicate my thesis to my wife Renae for her constant love and support. I am so thankful every day to have her in my life.

I would also like to dedicate my thesis to my family. My parents, Lee and Peggy, and my brothers, Jeff and Bradley, have been an incredible source of support from day one. My grandparents James, Joyce, Leroy, and Eunice have also served as a great inspiration to me, and are champions in education.

I would also like to thank my advisor, Prof. Thomas Hermann, for his mentorship and genuine interest in my development as a scientist, an educator, and a person.

I would also like to thank my coworkers in the Hermann Lab for all their help and support in my years at UCSD.

TABLE OF CONTENTS

Signature Page.....	iii
Dedication Page.....	iv
Table of Contents.....	v
List of Figures.....	xi
List of Tables.....	xx
Acknowledgements.....	xxii
Vita.....	xxiv
Abstract of the Dissertation.....	xxv
Chapter 1: Introduction to RNA switches involved in the regulation of viral translation and their therapeutic potential.....	1
Introduction.....	2
Functional non-coding RNA and its potential as a drug target.....	2
Hepatitis C Virus (HCV).....	3
HCV Internal Ribosome Entry Site (IRES).....	5
HCV IRES Domain II.....	10
HCV IRES Subdomain IIa.....	13
RNA Switches in Other Viral IRES Elements.....	14
X-ray crystallography for studying RNA structure and ligand binding.....	16
Fluorescence tools for studying RNA conformational changes, and ligand binding.....	17

Use of RNA Switches as Tunable Building Blocks for Nanotechnology.....	20
Chapter 2: Functional conservation and capture by a common ligand despite structural differences in diverse viral RNA switches.....	21
Abstract.....	22
Introduction.....	22
Results and Discussion.....	26
Guanine Captures an Extended Conformation of the Subdomain IIa Switch.....	26
Domain II-like motifs from other IRES RNAs are unrelated in sequence but fold like the HCV archetype.....	35
Subdomain IIa RNA from the SVV IRES Element Adopts a 90° Bent Fold.....	41
Domain II-like RNAs are interchangeable functional modules despite their sequence dissimilarity.....	48
Discussion of viral switches captured by a guanosine.....	66
Discussion of functional similar switches in five viruses despite sequence dissimilarity.....	72
Conclusions.....	73
Extended discussion of guanine-like molecules.....	74
Materials and Methods.....	114
RNA Preparation.....	114
FRET Folding Experiments.....	114
FRET Compound Screening Experiments.....	115
In vitro translation experiments.....	115
Compound testing for in vitro translation inhibition.....	116

HCV Bicistronic Reporter Mutagenesis.....	116
Crystallization and Data Collection.....	117
Structure Solution and Refinement.....	118
HCV Replicon Mutagenesis.....	119
HCV Replicon Assay.....	119
Acknowledgements.....	124
Chapter 3: A closer look at ligand-responsive RNA mechanical switches and their widespread discovery in diverse viral IRES elements.....	125
Abstract.....	126
Introduction.....	126
Ligand-responsive RNA mechanical switches versus metabolite-sensing riboswitches.....	126
Results and Discussion.....	129
Discovery of the RNA switches in viral IRES elements.....	133
Structure of the ligand-free RNA switches in HCV and SVV.....	136
Structure of a ligand-bound switch.....	140
IIa Switch function.....	141
Significance and applications of the IIa switches.....	144
Conclusions.....	146
Materials and Methods.....	147
In Vitro Translation Experiments.....	147
HCV Bicistronic Reporter Mutagenesis.....	147
Acknowledgements.....	148

Chapter 4: Tools and Techniques for Studying RNA Structure, Folding, and Ligand Binding.....	149
Abstract.....	150
Introduction.....	150
RNA crystallography.....	150
Förster resonance energy transfer (FRET).....	151
Heavy atom containing fluorescent ribonucleoside analog.....	152
Results and Discussion.....	153
RNA crystallography.....	153
Förster resonance energy transfer (FRET).....	161
Monitoring ligand binding at viral RNA switches.....	169
Ribosomal A-site RNA with an incorporated heavy atom containing fluorescent ribonucleoside analog.....	172
Conclusions.....	184
Materials and Methods.....	186
Preparation of RNA constructs for FRET experiments.....	186
FRET titration experiments monitoring folding and ligand binding....	187
Crystallization and Data Collection.....	188
Structure Solution and Refinement.....	189
Acknowledgements.....	190
Chapter 5: Crystal structure-guided design of self-assembling RNA nanotriangles.	191
Abstract.....	192
Introduction.....	192

Results and Discussion.....	196
Design of RNA Nanotriangles.....	196
Self-Assembly of RNA Nanotriangles.....	198
Stability of RNA Nanotriangles.....	201
Effect of Switch-Binding Ligand on RNA Nanotriangle Assembly....	202
Crystal Structure of Small RNA Nanotriangle.....	205
Discussion of Crystal Structures used in Design.....	212
Discussion of Self-Assembling Constructs.....	219
Discussion of Programmable Large Triangle.....	240
Discussion of Triangle Stability.....	246
Discussion of Ligand-Responsive Nanostructures.....	252
Discussion of RNA Nanostructure Crystallography.....	257
Discussion of RNA Small Triangle Atomic Force Microscopy.....	261
Conclusions.....	265
Materials and Methods.....	266
Preparation of RNA.....	266
Gel Electrophoresis.....	266
Thermo- and Chemical Stability.....	267
RNA Assembly and Dissociation in Presence of Switch-Binding Ligand.....	267
Crystallization, Data Collection, and Structure Determination.....	267
Atomic Force Microscopy.....	268
Acknowledgements.....	269

References..... 271

LIST OF FIGURES

Figure 1.1:	Hepatitis C virus (HCV) genome.....	5
Figure 1.2:	General translation initiation mechanisms.....	6
Figure 1.3:	Model of HCV IRES translation initiation with the role of subdomain IIa highlighted.....	8
Figure 1.4:	Structure of HCV IRES subdomain IIa RNA in absence and presence of a binding ligand.....	11
Figure 1.5:	Secondary structure predictions of domain II motifs in viral IRES elements from HCV and other flaviviruses.....	15
Figure 1.6:	RNA switch FRET experiments.....	18
Figure 1.7:	Utility of dual functional selenophene-conjugated uridine analog..	19
Figure 1.8:	Design and characterization of RNA nanostructures.....	20
Figure 2.1:	Structures and ligands of viral internal ribosome entry sites (IRES).....	24
Figure 2.2:	Ligand binding to the HCV IRES.....	28
Figure 2.3:	RNA switch FRET experiments.....	31
Figure 2.4:	HCV FRET experiments with guanine control compounds.....	32
Figure 2.5:	Impact of guanine, adenine, cytosine, uracil and the benzimidazole 1 on IRES-driven translation.....	34
Figure 2.6:	Titration of 3 (guanine) with FRET-labelled HCV IRES IIa RNA carrying an A57U mutation.....	35
Figure 2.7:	Domain II-like RNAs in viral IRES elements.....	37
Figure 2.8:	Structures and sequences of dye-labeled viral IRES subdomain IIa constructs for FRET experiments.....	39
Figure 2.9:	FRET experiments with constructs shown in Figure 2.8.....	40

Figure 2.10:	Structure of the SVV IRES subdomain IIa RNA.....	42
Figure 2.11:	Crystal structure of the SVV subdomain IIa RNA.....	43
Figure 2.12:	Crystal structure of an extended SVV subdomain IIa RNA construct.....	45
Figure 2.13:	Stereo view of a superposition of the subdomain IIa crystal structures from the HCV (red) and SVV (orange/white) IRES IIa elements.....	48
Figure 2.14:	Structures and sequences of HCV IRES domain II chimera constructs.....	50
Figure 2.15:	Functional competence of chimera HCV IRES elements.....	51
Figure 2.16:	HCV IRES domain II mutations' effect on translation efficiency in an <i>in vitro</i> translation assay.....	53
Figure 2.17:	Construction of HCV IRES chimeras.....	54
Figure 2.18:	HCV replicon experiments in Huh-7.5.1 cells.....	56
Figure 2.19:	Effect on translation efficiency of subdomain IIa replacement in the HCV IRES by the corresponding motif from CSFV.....	59
Figure 2.20:	Effect on translation efficiency of subdomain IIa replacement in the HCV IRES by the corresponding motif from AEV.....	61
Figure 2.21:	Effect on translation efficiency of subdomain IIa replacement in the HCV Δ dII + BVDV dII chimera IRES by the predicted corresponding motif from AEV.....	63
Figure 2.22:	Model for the modular domain II architecture in the IRES elements of RNA viruses.....	65
Figure 2.23:	Functional competence of chimera HCV IRES with domain II elements shortened or lengthened by 3 base pairs (-3bp, +3bp) in the basal helix.....	66
Figure 2.24:	Modeling of guanine 3 and quinazolinone 4 by superimposition on the benzimidazole 1 in the ligand binding site of the subdomain IIa complex crystal structure.....	68

Figure 2.25:	HCV IRES central domain secondary structure showing a pseudoknot in domain III and the domain IV hairpin holding the AUG start codon in black.....	69
Figure 2.26:	Proposed mechanism for the regulation of IRES activity by the subdomain IIa switch through an interaction with a G residue from domain IV.....	70
Figure 2.27:	pH dependence of guanine-like scaffolds binding affinity for the HCV IIa target.....	75
Figure 2.28:	Arginine and derivative testing in the FRET assay.....	77
Figure 2.29:	P-1 (guanine) titration with FRET-labelled HCV IRES IIa RNA...	84
Figure 2.30:	P-2 titration with FRET-labelled HCV IRES IIa RNA.....	85
Figure 2.31:	P-3 titration with FRET-labelled HCV IRES IIa RNA.....	86
Figure 2.32:	P-4 titration with FRET-labelled HCV IRES IIa RNA.....	87
Figure 2.33:	P-5 (guanosine) titration with FRET-labelled HCV IRES IIa RNA.....	88
Figure 2.34:	P-6 (acyclovir) titration with FRET-labelled HCV IRES IIa RNA.	89
Figure 2.35:	Q-1 titration with FRET-labelled HCV IRES IIa RNA.....	91
Figure 2.36:	Q-2 titration with FRET-labelled HCV IRES IIa RNA.....	92
Figure 2.37:	Q-3 titration with FRET-labelled HCV IRES IIa RNA.....	93
Figure 2.38:	Q-4 titration with FRET-labelled HCV IRES IIa RNA.....	94
Figure 2.39:	Q-5 titration with FRET-labelled HCV IRES IIa RNA.....	95
Figure 2.40:	Q-6 titration with FRET-labelled HCV IRES IIa RNA.....	96
Figure 2.41:	Q-7 titration with FRET-labelled HCV IRES IIa RNA.....	97
Figure 2.42:	Q-8 titration with FRET-labelled HCV IRES IIa RNA.....	98
Figure 2.43:	Q-9 titration with FRET-labelled HCV IRES IIa RNA.....	99

Figure 2.44:	Q-10 titration with FRET-labelled HCV IRES IIa RNA.....	101
Figure 2.45:	Q-11 titration with FRET-labelled HCV IRES IIa RNA.....	101
Figure 2.46:	Pt-1 (pterin) titration with FRET-labelled HCV IRES IIa RNA.....	103
Figure 2.47:	Pt-2 (neopterin) titration with FRET-labelled HCV IRES IIa RNA.....	104
Figure 2.48:	Pt-3 (folic acid) titration with FRET-labelled HCV IRES IIa RNA	105
Figure 2.49:	Pt-4 (lumazine) titration with FRET-labelled HCV IRES IIa RNA	106
Figure 2.50:	Pt-5 titration with FRET-labelled HCV IRES IIa RNA.....	107
Figure 2.51:	Pt-6 titration with FRET-labelled HCV IRES IIa RNA.....	108
Figure 2.52:	T-1 (tetrahydrobiopterin, THB) titration with FRET-labelled HCV IRES IIa RNA.....	110
Figure 2.53:	T-2 titration with FRET-labelled HCV IRES IIa RNA.....	111
Figure 2.54:	D-1 (L-sepiapterin) titration with FRET-labelled HCV IRES IIa RNA.....	113
Figure 2.55:	D-2 (isoxanthopterin) titration with FRET-labelled HCV IRES IIa RNA.....	114
Figure 3.1:	Comparison of ligand-responsive RNA mechanical switches in contrast to metabolite-sensing riboswitches.....	128
Figure 3.2:	Structure of HCV IRES subdomain IIa RNA switch in absence and presence of binding ligand.....	130
Figure 3.3:	Structures of IRES subdomain IIa RNA switches from SVV and HCV.....	131
Figure 3.4:	Structure of the HCV IRES subdomain IIa RNA switch in complex with ligand 1	132
Figure 3.5:	Functional activity and secondary structures of IRES IIa switches.	136
Figure 3.6:	Mutational analysis of SVV IIa switch.....	139

Figure 3.7:	Consensus secondary structure of domain II analogs from viral IRES elements studied here.....	144
Figure 4.1:	RNA switch FRET experiments.....	152
Figure 4.2:	BVDV IRES subdomain IIa switch RNA crystal constructs.....	154
Figure 4.3:	CSFV IRES subdomain IIa switch RNA crystal constructs.....	155
Figure 4.4:	SVV IRES subdomain IIa switch RNA crystal constructs.....	156
Figure 4.5:	AEV IRES subdomain IIa switch RNA crystal constructs.....	157
Figure 4.6:	Viral IRES subdomain IIa switch RNA crystal constructs.....	158
Figure 4.7:	Secondary structure and folding of viral RNA switches monitored by FRET experiments with cyanine dye-labeled oligonucleotides.	163
Figure 4.8:	Normalized relative FRET signal for the Mg ²⁺ titration of the HCV conventional (●) and modular (○) FRET constructs.....	165
Figure 4.9:	Structures and sequences of dye-labeled viral IRES subdomain IIa constructs for FRET experiments.....	166
Figure 4.10:	Secondary structures of functional viral IRES IIa switches.....	169
Figure 4.11:	Binding of a translation inhibitor ligand 1 to the HCV RNA switch as monitored by FRET experiments with cyanine dye-labeled oligonucleotides.....	171
Figure 4.12:	Utility of dual functional selenophene-conjugated uridine analog..	173
Figure 4.13:	Aminoglycosides which bind the ribosomal decoding site (A-site)	174
Figure 4.14:	Structure of SeU-modified A-site RNA construct.....	175
Figure 4.15:	Electron density map around the SeU-modified A-site RNA construct.....	177
Figure 4.16:	Superimposition of the SeU-modified (white) and non-modified (red; PDB: 1T0D) A-site RNA constructs.....	178
Figure 4.17:	Stacking of neighboring residues G1405 and C1407 on the selenophene-modified U1406.....	179

Figure 4.18:	Superimposition of the SeU-modified (cyan/indigo) and non-modified neomycin-bound (yellow/pink, PDB: 2A04) A-site RNA constructs.....	181
Figure 4.19:	Superimposition of the SeU-modified (cyan/indigo) and non-modified paramomycin-bound (yellow/green, PDB: 1J7T) A-site RNA constructs.....	182
Figure 4.20:	Superimposition of the SeU-modified (cyan/indigo) and non-modified kanamycin A-bound (yellow/brown, PDB: 2ESI) A-site RNA constructs.....	182
Figure 4.21:	Superimposition of the SeU-modified (cyan/indigo) and non-modified tobramycin-bound (yellow/blue, PDB: 1LC4) A-site RNA constructs.....	183
Figure 4.22:	Superimposition of the SeU-modified (cyan/indigo) and non-modified hygromycin B-bound (yellow/blue, PDB: 1HNZ) A-site RNA constructs.....	184
Figure 5.1:	Several significant structural achievements in the history of RNA nanotechnology.....	194
Figure 5.2:	Design of self-assembling RNA nanotriangles.....	198
Figure 5.3:	Nanotriangle assembly and stability. Structures were analyzed by native PAGE in the presence of 2.5mM MgCl ₂	200
Figure 5.4:	Magnesium ion concentration dependence of large nanotriangle self-assembly.....	201
Figure 5.5:	Nanotriangle self-assembly efficiency and dissociation in the presence of ligand.....	203
Figure 5.6:	Large nanotriangle self-assembly efficiency in the presence of a binding ligand and control compounds.....	204
Figure 5.7:	Dissociation of symmetrical and programmable large nanotriangles.....	205
Figure 5.8:	Crystal structure of the self-assembling RNA nanotriangle.....	206
Figure 5.9:	In the intermolecular packing of the crystal structure, the RNA nanotriangle forms face-to-face dimers.....	207

Figure 5.10:	In the extended crystal packing, triangle dimers form layers that stack in alternating orientations.....	208
Figure 5.11:	Stereo view of the crystal structure and electron density map.....	209
Figure 5.12:	One-dimensional twinning disorder of crystal packing layers of the self-assembling RNA nanotriangle.....	210
Figure 5.13:	Crystal packing of pseudo-continuously closed triangles in the SVV subdomain IIa short oligonucleotide construct.....	213
Figure 5.14:	Side view of pseudo-continuously closed large triangle in crystals of the SVV subdomain IIa long oligonucleotide construct.....	214
Figure 5.15:	Face-to-face dimer formation of pseudo-continuously closed large triangles in crystals of the subdomain IIa long oligonucleotide construct.....	215
Figure 5.16:	End-to-end and side-to-side packing of pseudo-continuously closed large triangles in crystals of the long subdomain IIa oligonucleotide construct.....	216
Figure 5.17:	Self-assembling small RNA triangle constructs of identical size but varying oligonucleotide termini and number of overhanging nucleotides.....	220
Figure 5.18:	Native PAGE gel in 5mM MgCl ₂ of self-assembling small RNA triangle constructs ST5-1 (5 base pair overlapping sequence) and ST4-2 (4 base pair overlapping sequence).....	224
Figure 5.19:	Native PAGE gel in 5mM MgCl ₂ of self-assembling small triangle constructs.....	225
Figure 5.20:	Self-assembling small triangle constructs of various size and varying numbers of overhanging or underhanging nucleotides.....	228
Figure 5.21:	Native PAGE gel in 5mM MgCl ₂ of self-assembling small triangle and small triangle (+1bp) constructs.....	231
Figure 5.22:	Native PAGE gel in 5mM MgCl ₂ of self-assembling small RNA triangle constructs of various sizes described left to right.....	232
Figure 5.23:	Native PAGE gels in 1, 2, 5 or 10mM MgCl ₂ (left to right) run for various lengths of time for the small triangle (+2bp) construct	233

Figure 5.24:	Native PAGE gel in 10mM MgCl ₂ for the small triangle (+2bp) construct.....	234
Figure 5.25:	Self-assembling large RNA triangle constructs of identical size but varying oligonucleotide termini and number of overhanging nucleotides.....	236
Figure 5.26:	Native PAGE gel in 5mM MgCl ₂ of self-assembling large RNA triangle construct LT5	239
Figure 5.27:	Native PAGE gel in 10mM MgCl ₂ of self-assembling large RNA triangle constructs LT4 (4 residue overlapping sequence) and LT5 (5 residue overlapping sequence).....	240
Figure 5.28:	Programmable large triangle (PLT) assembled in 5mM MgCl ₂ for 10min at 25° or 37°	242
Figure 5.29:	PLT assembled in 2mM MgCl ₂ for 20min at 25° or 37°.....	243
Figure 5.30:	PLT assembled in 10mM MgCl ₂ for 30min at 25° or 37°.....	244
Figure 5.31:	PLT assembled in 20mM MgCl ₂ for 30min at 25° or 37°.....	245
Figure 5.32:	PLT assembled in 2mM MgCl ₂ for 30min at 25°.....	246
Figure 5.33:	Comparison of the small triangle's resistance to RNase A (endoribonuclease) degradation compared to tRNA and the Iia-short construct.....	247
Figure 5.34:	Small triangle and Iia-short construct stability when treated with RNase A (endoribonuclease).....	248
Figure 5.35:	Small triangle and Iia-short construct stability when treated with RNase A.....	249
Figure 5.36:	UV thermal melting experiments of small and large triangles as well as subdomain Iia motifs representing single corners of triangles as seen in crystal structure packing.....	250
Figure 5.37:	Native PAGE analysis in 5mM MgCl ₂ of SLT assembled in the presence of 0, 0.5, or 2 mM guanine.....	252
Figure 5.38:	Native PAGE analysis in 5mM MgCl ₂ of RNA square.....	254

Figure 5.39:	PLT assembled in the presence of 0 or 500 μM ligand in ideal assembly conditions of 2mM MgCl_2 for 20min at 25°.....	255
Figure 5.40:	Native PAGE analysis in 0.5, 2, or 5 mM MgCl_2 (left to right) of the small triangle assembly in the presence of 0, 0.1, or 0.5 mM ligand.....	256
Figure 5.41:	Native PAGE in 2mM Mg of the small triangle (+1bp) (Figure 5.20a) assembled in the presence of 0, 0.1, or 0.5 mM ligand.....	257
Figure 5.42:	Photos of nanostructure crystals.....	259
Figure 5.43:	AFM images of small RNA triangle.....	263

LIST OF TABLES

Table 2.1:	Crystallographic data collection and refinement statistics for the SVV IRES.....	44
Table 2.2:	Crystallographic data collection and refinement statistics for the SVV IRES subdomain IIa extended RNA structure.....	46
Table 2.3:	Summary of activity of “guanine-like” small molecules.....	81
Table 2.4:	Summary of activity of purine derivatives P-1 to P-6 in the FRET assay experiments shown in Figures 2.29-2.34.....	83
Table 2.5:	Summary of activity of 2-amino-quinazoline derivatives Q-1 to Q-9 in the FRET assay experiments shown in Figures 2.35-2.43...	90
Table 2.6:	Summary of activity of 2-amino-quinazoline derivatives Q-10 to Q-11 in the FRET assay experiments shown in Figures 2.44-2.45.	100
Table 2.7:	Summary of activity of pteridine derivatives Pt-1 to Pt-5 in the FRET assay experiments shown in Figures 2.46-2.51.....	102
Table 2.8:	Summary of activity of tetrahydropteridine derivatives T-1 to T-2 in the FRET assay experiments shown in Figures 2.52-2.53.....	109
Table 2.9:	Summary of activity of dihydropteridine derivatives D-1 to D-2 in the FRET assay experiments shown in Figures 2.54-2.55.....	112
Table 2.10:	Oligonucleotide sequences for the cloning of HCV Δ II mutants in bicistronic reporter plasmids.....	121
Table 2.11:	Oligonucleotide sequences for site-directed mutagenesis of bicistronic reporter constructs.....	123
Table 3.1:	Taxonomy of viruses studied here.....	135
Table 4.1:	Crystallographic data collection and refinement statistics for ^{Se} U-modified A-site RNA.....	176
Table 5.1:	Crystallographic data collection and refinement statistics for the self-assembling RNA nanotriangle.....	211
Table 5.2:	RNA triangle crystal constructs.....	258

Table 5.3:	DNA and DNA/RNA hybrid triangle crystal constructs.....	261
-------------------	---	-----

ACKNOWLEDGEMENTS

I would like to thank Prof. Thomas Hermann for his guidance and mentorship throughout my time at UCSD. He is always available with a listening ear to answer questions and to offer advice. He has taught me to be a well-rounded scientific researcher, writer, educator, and person. Most admirably, he cares a great deal about the human beneath the lab coat and goggles. It has been a privilege to work in his laboratory.

I would also like to thank all past and present undergraduate, graduate, and postdoc researchers in the Hermann lab who have helped me in a variety of ways. Specifically, I would like to thank Sergey, Kevin, Shu, Nick, Keija, and Brian who helped me from the very beginning of my time in the lab. I would also like to specifically thank several talented undergraduates, Jing, Andrew, and Lulu who collaborated in my work for extended periods of time.

I would also like to thank my other classmates, colleagues, and friends at UCSD who made this time so rewarding.

The material in Chapter 2, in part, has been published in Proceedings of the National Academy of Sciences USA (**2014**, *111*, 15952-15957) titled *Functional conservation despite structural divergence in ligand-responsive RNA switches* with Sergey Dibrov, Jing Gu, David Wyles, and Thomas Hermann as co-authors. The dissertation author was the primary investigator and author of this material.

The material in Chapter 3, in part, has been published in RNA Biology (**2015**, *12*, 780-786) titled *Ligand-responsive RNA mechanical switches* with Thomas Hermann as co-author. The dissertation author was the primary investigator and author of this material.

The material in Chapter 4, in part, has been published in *Methods* (**2015**, *91*, 38-39) titled *Conformational flexibility of viral RNA switches studied by FRET* with Thomas Hermann as co-author. The dissertation author was the primary investigator and author of this material.

The material in Chapter 4, in part, is currently being prepared for submission for publication titled *Fluorescence detection of RNA-ligand binding and crystal structure determination of ribosomal decoding site RNA using a heavy atom containing fluorescent ribonucleoside* with Ashok Nuthanakanti, Seergazhi Srivatsan, and Thomas Hermann as co-authors. The dissertation author was the co-primary investigator and author of this material.

The material in Chapter 5, in part, has been published in *Angewandte Chemie International Edition* (**2016**, *55*, 4097-4100) titled *Crystal structure-guided design of self-assembling RNA nanotriangles* with Sergey Dibrov and Thomas Hermann as co-authors. The dissertation author was the primary investigator and author of this material.

The material in Chapter 5, in part, is currently being prepared for submission for publication titled *Design and Crystallography of Self-Assembling RNA Nanostructures* with Thomas Hermann as co-author. The dissertation author was the primary investigator and author of this material.

VITA

- 2011 Bachelor of Science, Point Loma Nazarene University
Chemistry
- 2011 Bachelor of Arts, Point Loma Nazarene University
International Development Studies
- 2013 Master of Science, University of California, San Diego
Chemistry
- 2016 Doctor of Philosophy, University of California, San Diego
Chemistry

PUBLICATIONS

- M. A. Boerneke**, S. M. Dibrov, T. Hermann, “Crystal structure-guided design of self-assembling RNA nanotriangles,” *Angewandte Chemie International Edition*, 55, 4097–4100 (2016).
- M. A. Boerneke**, T. Hermann, “Conformational flexibility of viral RNA switches studied by FRET,” *Methods*, 91, 35-39 (2015).
- M. A. Boerneke**, T. Hermann, “Ligand-responsive RNA mechanical switches,” *RNA Biology*, 12, 780–786 (2015).
- M. A. Boerneke**, S. M. Dibrov, J. Gu, D. L. Wyles, T. Hermann, “Functional conservation despite structural divergence in ligand responsive RNA switches,” *Proceedings of the National Academy of Sciences*, 111, 15952–15957 (2014).
- K. D. Rynearson, B. Charrette, C. Gabriel, J. Moreno, **M. A. Boerneke**, S. M. Dibrov, and T. Hermann, “2-Aminobenzoxazole ligands of the hepatitis C virus internal ribosome entry site,” *Bioorganic & Medicinal Chemistry Letters*, 24, 3521–3525 (2014).
- K. Ding, A. Wang, **M. A. Boerneke**, S. M. Dibrov, T. Hermann, “Aryl-substituted aminobenzimidazoles targeting the hepatitis C virus internal ribosome entry site,” *Bioorganic & Medicinal Chemistry Letters*, 24, 3113–3117 (2014).
- S. M. Dibrov, J. Parsons, M. Carnevali, S. Zhou, K. D. Rynearson, K. Ding, E. Garcia Sega, N. D. Brunn, **M. A. Boerneke**, M. P. Castaldi, and T. Hermann, “Hepatitis C virus translation inhibitors targeting the internal ribosomal entry site,” *Journal of Medicinal Chemistry*, 57, 1694–1707 (2014).

ABSTRACT OF THE DISSERTATION

Ligand-responsive RNA switches: viral translation regulators, therapeutic targets, and tunable building blocks for nanotechnology

by

Mark A. Boerneke

Doctor of Philosophy in Chemistry

University of California, San Diego, 2016

Professor Thomas Hermann, Chair

Ligand-responsive RNA mechanical switches represent a new class of simple and small switching modules which regulate viral translation initiation by adopting well-defined ligand-free and bound conformational states without undergoing large secondary structure rearrangements, distinguishing them from metabolite-sensing riboswitches. Initially discovered in the internal ribosome entry site (IRES) of hepatitis C virus (HCV), RNA switch motifs have now been discovered in the genomes of diverse other viruses.

Although large variations are seen in sequence and local secondary structure of these switches, their function in viral translation initiation requiring recognition of identical cognate ligands is conserved. These simple ligand-responsive viral RNA switches may be the target for therapeutic intervention and may be incorporated as tunable building blocks for nanotechnology.

Chapter 1:

Introduction to RNA switches involved in the regulation of viral translation and their therapeutic potential

Introduction

Functional non-coding RNA and its potential as a drug target

Over the course of the last decades, structural and functional studies have expanded our understanding of the numerous roles that RNA plays in addition to transferring genetic information from DNA into proteins (protein-coding RNA), including its ability to regulate diverse cellular processes such as transcription, RNA processing, and translation (non-coding RNA or ncRNA).^{1,2} These ncRNA elements adopt complex three-dimensional architectures, undergo unique conformational changes, and interact with metabolite, protein, or nucleic acid partners in accomplishing their function, all of which may be the target of therapeutic intervention. The potential of ncRNA as a therapeutic target has been validated by a growing number of RNA targets which have successfully been targeted by small molecule inhibitors. The most widely studied of these targets is ribosomal RNA of the bacterial ribosome which is the target for antibiotics such as aminoglycosides, macrolides, tetracyclins, and oxazolidinones.^{3,4} Other well studied RNA targets include the human immunodeficiency virus (HIV) trans-activation response (TAR) element and Rev response element (RRE),⁵ the HCV internal ribosome entry site (IRES, studied here),⁶ and bacterial riboswitches.^{7,8}

Highly conserved ncRNA elements found in viral genomes play essential roles in the life cycle of many viruses, providing a high barrier to the rise of viral mutants resistant to therapeutics targeting these RNA elements.⁶ New antiviral targets and agents are critically needed to fight viruses for which there are no current therapies, and to address clinical concerns of overcoming viral drug-resistance with agents that broadly act

against multiple genotypes of a virus and have unique mechanisms of action that can be used in combination with existing therapies.

Hepatitis C Virus (HCV)

HCV is a serious viral pathogen which affects 2-3% of the world's population (130-150 million people globally including 3-4 million people in the USA)^{9,10} with 3-4 million new patient diagnoses every year.¹¹ HCV is a leading cause of liver disease including cirrhosis and hepatocellular carcinoma, and is now responsible for a higher mortality rate than HIV in the USA.¹² Currently there is no effective vaccine for HCV. The virus is bloodborne and infection most commonly occurs through unsafe percutaneous injection drug use, inadequate sterilization of medical equipment or needlestick injuries, and unscreened contaminated blood transfusions. There are six major HCV genotypes and over 50 reported subtypes. Genotype 1 is most prevalent in the USA,^{13,14} and the majority of studies here were performed on targets derived from genotype 1b sequences. Prior to the last several years, standard HCV therapy involved innate immune response stimulating pegylated interferon-alpha and the broad-spectrum antiviral ribavirin, but this regimen suffered from low efficacy and serious side effects.¹⁵ In the last several years, a therapeutic revolution has ushered in a wave of new direct-acting antivirals for the treatment of HCV infection with increased efficacy and greater tolerability, which have achieved virus cure rates greater than 90%.¹⁶ In 2011, two protease inhibitors were the first direct-acting antivirals to be approved.¹⁷ This was followed by approval of additional antivirals inhibiting the viral protease as well as new antivirals inhibiting the viral RNA polymerase and the nonstructural protein NS5A

replication complex.^{18,19} HCV-infected patients have reaped enormous benefits from these new therapies, and yet the prevalence of HCV quasispecies with preexisting drug-resistant mutations calls for the need to develop distinct antivirals with unique mechanisms of action to be used in resistance-suppressing combination therapy.^{20,21}

HCV is a positive-strand, enveloped RNA virus in the *Hepacivirus* genus in the *Flaviviridae* family whose single stranded RNA genome serves directly as messenger RNA (mRNA) coding for its viral proteins. This viral genome is 9.6 kilobases long and contains highly conserved and structured 5' and 3' untranslated regions (UTR) flanking a single open reading frame (ORF) encoding 3 structural (S) and 7 nonstructural (NS) viral proteins (Figure 1.1).

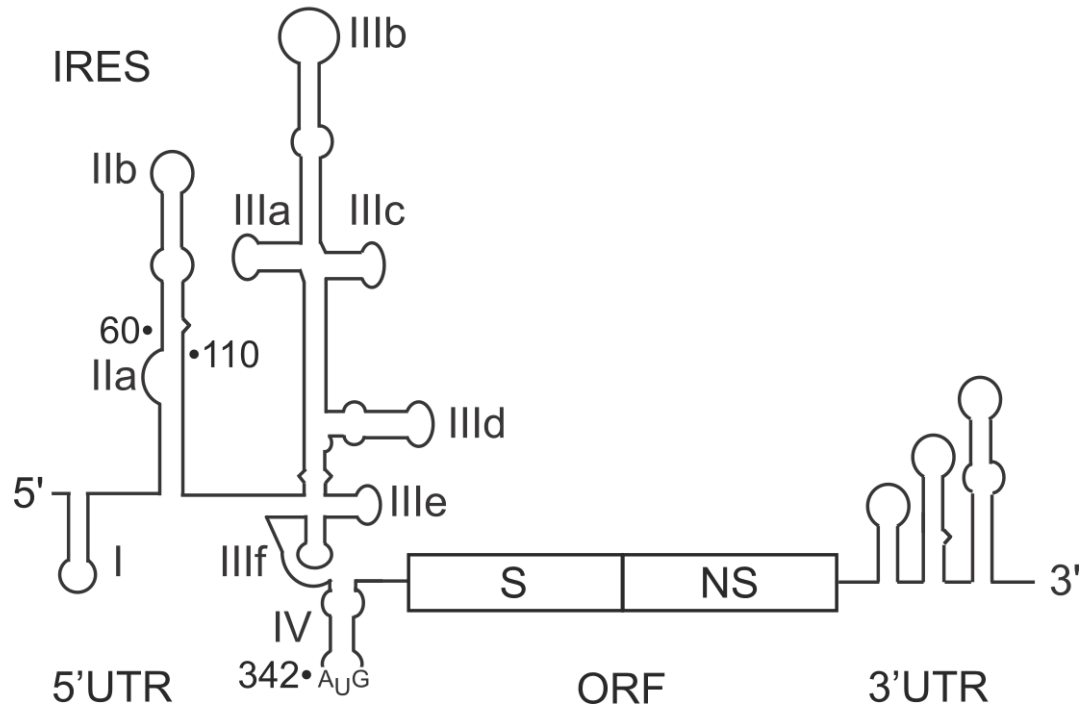


Figure 1.1: Hepatitis C virus (HCV) genome. The highly conserved and structured internal ribosome entry site (IRES) sits in the 5' untranslated region (UTR). The open reading frame (ORF) encodes 3 structural (S) and 6 nonstructural (NS) proteins. The AUG start codon is sequestered in the IRES domain IV hairpin which unwinds during translation initiation. The genome also contains a structured 3' UTR.

HCV Internal Ribosome Entry Site (IRES)

The 5' UTR of HCV contains a structured internal ribosome entry site (IRES) which functions by hijacking host cell's ribosomes for the synthesis of its viral proteins (Figure 1.1).²² This IRES initiates translation by assembling functional ribosomes directly at the start codon, bypassing the need for 5' cap recognition, ribosome scanning, and the requirement for most eukaryotic host initiation factors (eIF) (Figure 1.2a).²³⁻²⁶ This is in contrast to canonical eukaryotic cap-dependent translation initiation which involves the ordered assembly of eukaryotic initiation factors at a modified nucleoside cap structure [7-methyl guanosine (m^7G)] on the 5' end of a messenger RNA (mRNA)

leading to the recruitment of the 40S ribosome and scanning to the start codon where the 60S ribosomal subunit can associate, forming a translationally competent 80S ribosome complex (Figure 1.2b).^{23,27}

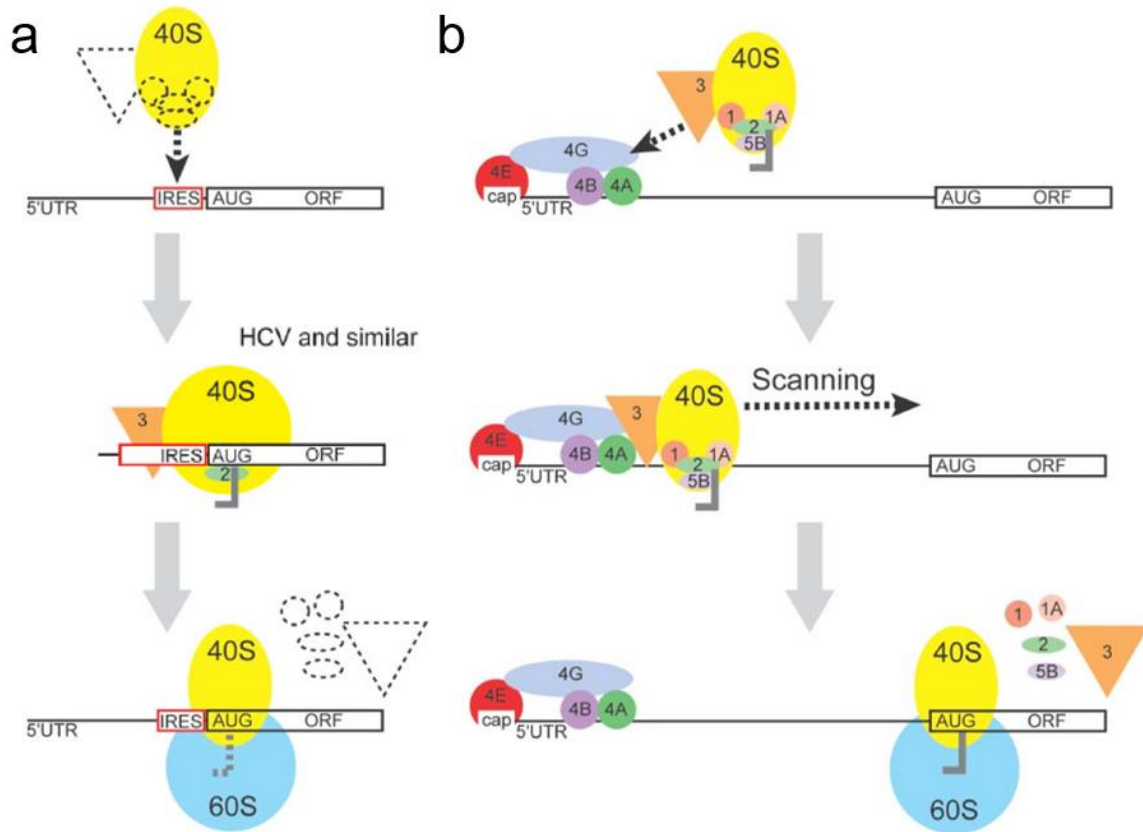
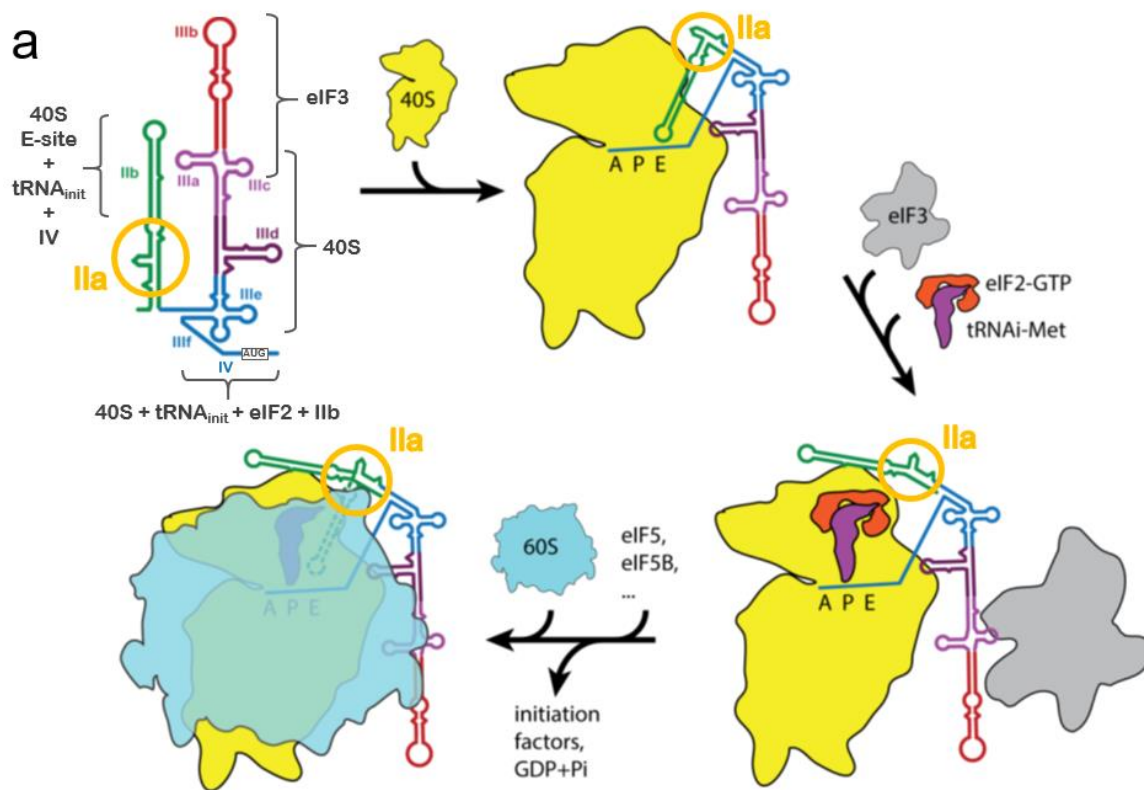


Figure 1.2: General translation initiation mechanisms. a) Cap-independent IRES-mediated translation initiation which involves the direct binding of the 40S ribosomal subunit by an IRES without requiring a cap structure, ribosome scanning for the AUG start codon, or many initiation factors (dotted shapes indicate general initiation factors not required for IRES-mediated translation initiation). The center panel shows the requirement of HCV and HCV-like IRES elements for eIF3 and eIF2. b) Cap-dependent eukaryotic translation initiation requires a cap structure (cap), many initiation factors (colored shapes), and ribosome scanning for the start codon. Figure modified from ref²³, 2012, John Wiley & Sons Ltd.

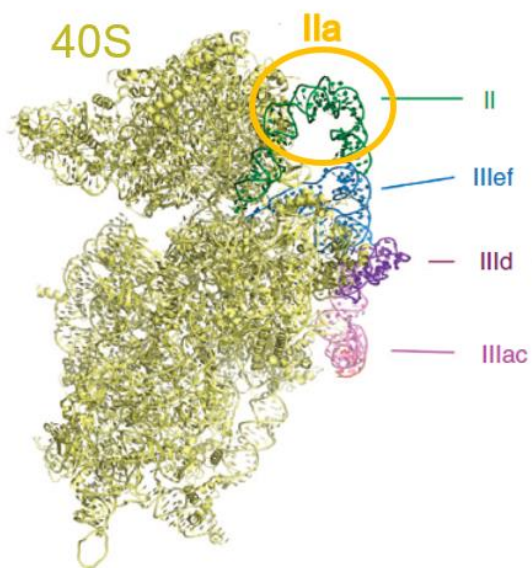
The HCV IRES spans the viral genome from position 40 in the 5' UTR to position 372 in the protein coding frame, 30 nucleotides beyond the start codon.^{6,22} The

IRES is composed of three independently folding domains (II-IV) connected by flexible single stranded RNA (ssRNA) linkers and which are required for functional translation initiation (Figure 1.1).²⁸ The domain I hairpin structure of the 5' UTR is not required for IRES function but participates in virus replication through interactions with the 3' UTR.²⁹ IRES domain III directly recruits the 40S ribosomal subunit by forming intermolecular interactions with the ribosome via subdomains IIIc, IIIId, and IIIe (Figure 1.3).^{30,31} Additionally subdomain IIIb binds the eIF3 complex³⁰ and stem 2 (S2) of the subdomain IIIf pseudoknot interacts nonspecifically with the ribosome at the exit of the mRNA channel and aids in the correct positioning of the start codon.^{30,32,33} The domain IV hairpin melts during 40S binding to the IRES to allow precise placement of the AUG start codon in the ribosomal decoding groove.^{31,34,35} Additionally, domain IV binds the ternary complex including eIF2, GTP, and initiator methionyl-tRNA (tRNAⁱ_{Met}),^{25,26,36} though in times of cellular stress the HCV IRES can undergo eIF2-independent translation initiation.^{37,38}

Figure 1.3: Model of HCV IRES translation initiation with the role of subdomain IIa highlighted. a) The structured HCV IRES cartoon is color coded by domains and subdomains. Domain IV containing the AUG start codon is shown in its unwound form as it exists when bound to the 40S subunit. The subdomains of the free IRES are labeled with their known interactions with the ribosome and other IRES domains as seen in various structures and as demonstrated throughout the translation initiation pathway cartoon. Subdomain IIa is highlighted with a gold circle throughout. Subdomains IIIa,c,d,e,f of the IRES recruit the 40S subunit while subdomain IIa's bent structure directs IIb to the ribosomal E-site. Subdomains IIIa,b,c then recruit eIF3 while the ternary complex including eIF2, GTP, and tRNA^{iMet} assemble at the start codon, likely paired with domain II's displacement from the E-site by capture of subdomain IIa in an elongated conformation,³⁹ as the position of domain II on the E-site is incompatible with the ternary complex's binding to the 40S.^{33,40,41} Domain II and eIF5 then activate GTP hydrolysis by eIF2, which is followed by eIF2 and eIF3 removal, 60S joining, and adjustment of tRNA^{iMet} to form the 80S complex, with subdomain IIa adopting both bent and elongated conformations.^{41,42} b) Atomic model of HCV IRES bound to the human 40S. Figure modified from ref³³, 2015, Macmillan Publishers Limited.



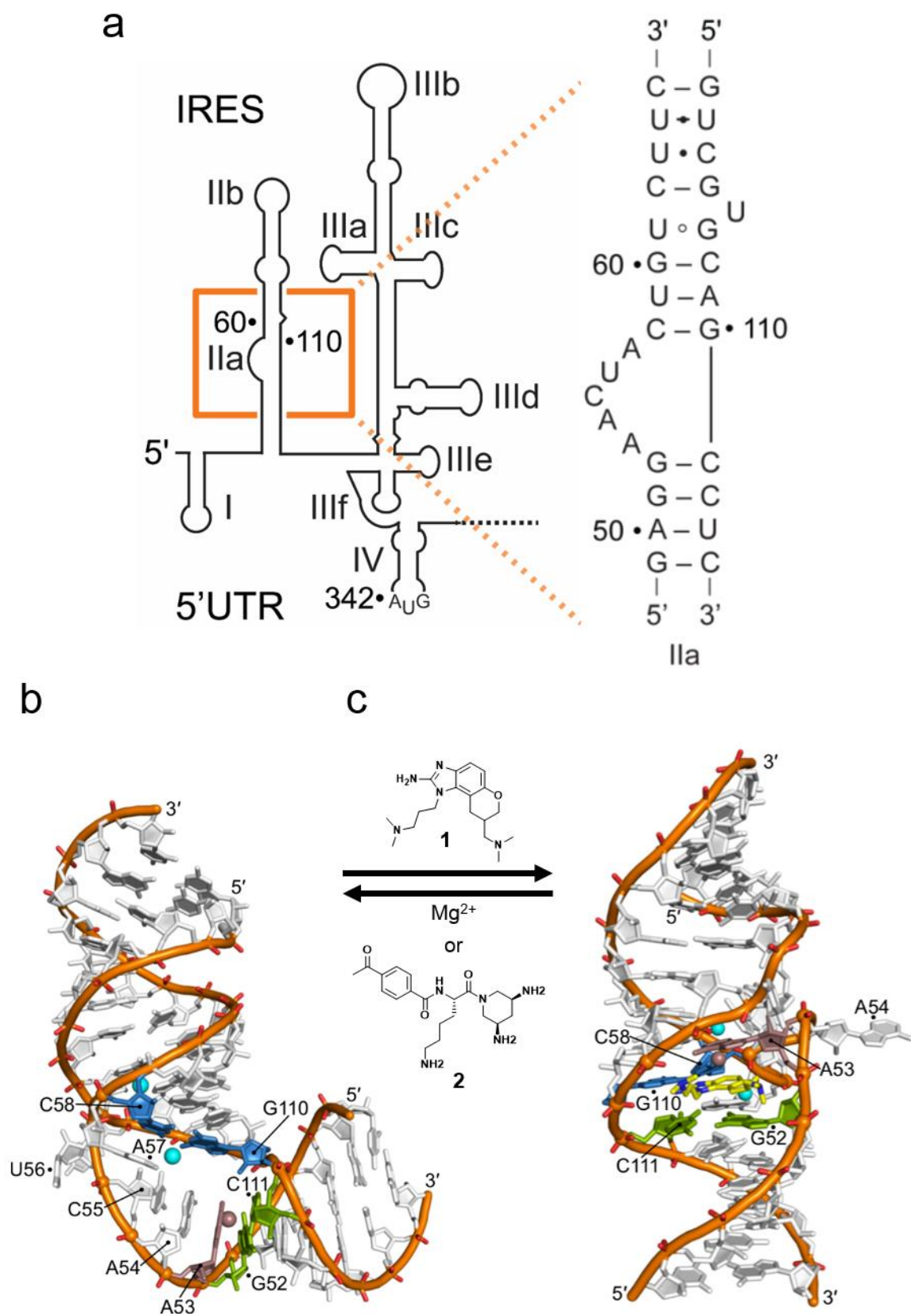
b



HCV IRES Domain II

Domain II is not required for IRES binding of the 40S,³⁰ but it promotes stable entry of HCV mRNA at the decoding groove of the 40S subunit (Figure 1.3).^{35,36,43} It is important for the activation of GTP hydrolysis by eIF2 and required for eIF2 removal prior to 60S ribosomal subunit joining^{26,44,45} as well as adjustment of tRNAⁱ_{Met} orientation.⁴⁶ HCV domain II directly interacts with tRNAⁱ_{Met}⁴² and forms tertiary interactions with unwound domain IV.⁴¹ Subdomain IIa folds into an L-shaped motif⁴⁷ (Figure 1.4) that introduces a 90° bend in domain II⁴⁸ and directs the IIb hairpin towards the E-site at the ribosomal subunit interface (Figure 1.3).^{34,49} Here, domain II induces head-body tilt conformational changes in the 40S.^{33,34,41,42,50,51} Domain II can adopt a number of conformations on the ribosome,⁵² although all three dimensional structures of the HCV IRES in both the presence and absence of ribosomal subunits reveal a bent or L-shaped domain II^{33,34,41,42,47,48,51} (Figures 1.3, 1.4b) except in IRES-80S complexes with bound initiator tRNA^{41,42} or the isolated IRES subdomain IIa bound by a synthetic inhibitor²¹ (Figures 1.3, 1.4c) where domain II is seen in elongated conformations.

Figure 1.4: Structure of HCV IRES subdomain IIa RNA in absence and presence of a binding ligand. The upper closing base pair of the internal loop is highlighted in blue, the lower closing base pair in green, and A53 stabilized by a magnesium ion and forming the roof of the ligand-binding pocket highlighted in mauve. a) Secondary structure of the HCV IRES with subdomain IIa highlighted. b) Crystal structure of the ligand-free IIa RNA shown with stabilizing magnesium ions and flanking helices adopting a bent conformation. This ligand-free bent conformation may also be locked in place by diaminopiperidine inhibitor **2**. Image was prepared from PDB coordinate file 2NOK. c) Crystal structure of the ligand-captured IIa RNA adopting an elongated conformation when bound to 2-amino-benzimidazole inhibitor **1**. Image was prepared from PDB coordinate file 3TZR.



HCV IRES Subdomain IIa

Domain II is highly conserved in thousands of clinical isolates^{6,21,53} and contains an internal loop in its lower stem, subdomain IIa (Figure 1.4a), which is the target for viral translation inhibitors⁵⁴ that bind to the internal loop and block translation by capturing distinct conformational states of the RNA.⁶ The three dimensional structure of subdomain IIa was determined by x-ray crystallography⁴⁷ (Figure 1.4b), which revealed the RNA adopting a 90° bent architecture, in agreement with cryo-EM and NMR studies.^{33,34,41,42,48,51} FRET experiments demonstrated that subdomain IIa undergoes a conformational switch in the presence of benzimidazole viral translation inhibitors.^{53,54} The co-crystal structure of subdomain IIa in complex with benzimidazole inhibitor **1** showed capture of the RNA in an extended conformation in which the stems flanking the internal loop are coaxially stacked on both sides of the ligand-binding pocket (Figure 1.4c).²¹ In contrast, diaminopiperidine compounds such as **2** bind and lock the IIa RNA in a bent conformation that corresponds to the ligand-free state (Figure 1.4b,c).⁵⁵ The ligand pocket in subdomain IIa was recently demonstrated to selectively recognize guanine and serves as a fortuitous binding site for structurally similar 2-aminobenzimidazoles.³⁹ Based on this observation in conjunction with other circumstantial evidence,³¹ we hypothesized that a guanosine within an RNA sequence of the viral genome or ribosomal RNA acts as a trigger of the IIa conformational switch.³⁹ This capture of the IIa conformational switch will be discussed in great detail in Chapter 2. This IIa conformational switch as well as other viral IIa conformational switches discussed in Chapters 2-4 will be referred to as ligand-responsive RNA switches (or simply RNA switches) and their distinction from metabolite sensing riboswitches with regard to their

small size as well as intrinsic stability and structural definition of each constitutive conformations state will be discussed in Chapter 3.

RNA Switches in Other Viral IRES Elements

A number of other viruses in the *Flaviviridae* and *Picornaviridae* families also contain HCV-like IRES elements (classified as type IV IRES elements) in their viral genomes with varying degrees of structural and functional similarity but only limited sequence conservation (Figure 1.5).^{39,56,57} Despite the differences, many of these type IV IRES elements are organized into similar domains as HCV, contain analogous domain II elements, and facilitate translation initiation in a similar fashion. A number of these viral domain II elements have been found to contain RNA switches functionally homologous to HCV.^{39,58} The three-dimensional (3D) structure of the switch from Seneca Valley Virus (SVV, now named *Senecavirus A*) was determined by X-ray crystallography and revealed a bent structure with a nearly identical 3D structure to the HCV switch, despite very different sequences and predicted secondary structures.³⁹ These viruses with analogous domain II elements containing Ila RNA switches will be discussed in great detail in Chapters 2-3.

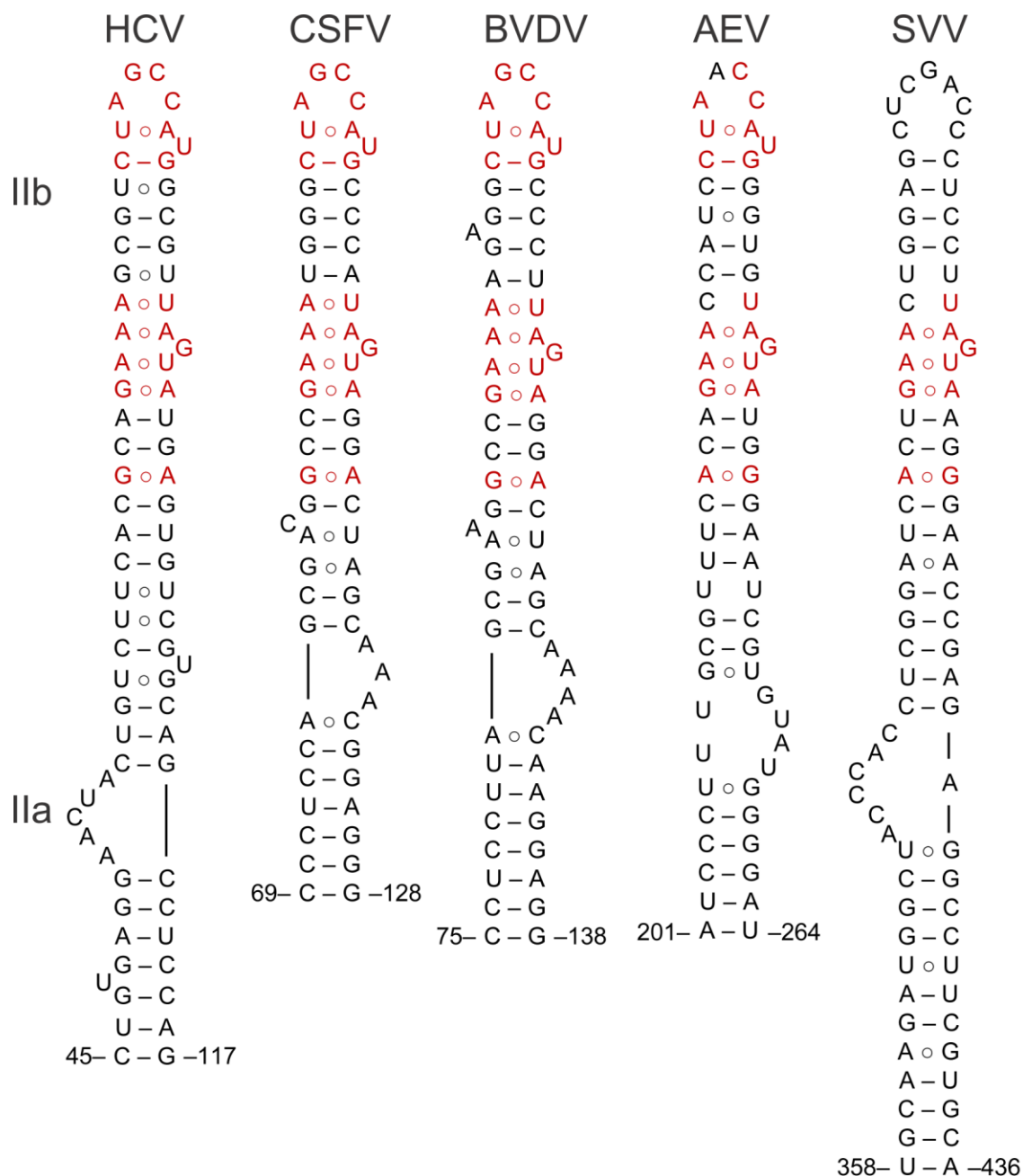


Figure 1.5: Secondary structure predictions of domain II motifs in viral IRES elements from HCV and other flaviviruses including classic swine fever virus (CSFV) and bovine viral diarrhea virus (BVDV) as well as picornaviruses such as avian encephalomyelitis virus (AEV) and Seneca Valley virus (SVV). Non-Watson-Crick base pairs are indicated by the \circ symbol. Conservation of the apical loop and the internal loop E motif in subdomain IIb is indicated in red font. NCBI reference sequences: HCV (NC_004102), CSFV (NC_002657.1), BVDV (NC_001461.1), AEV (NC_003990.1), SVV (NC_011349.1).

X-ray crystallography for studying RNA structure and ligand binding

A growing knowledge and understanding of the structures of biomacromolecules has played a pivotal role in understanding their diverse functions. Among the methods which provide structural information about biomolecules, X-ray crystallography provides the highest resolution unambiguous molecular snapshot of a biomolecule's structure. While not every biomacromolecule is amenable to crystallography, this method has been widely used to study molecular structure, function, and ligand binding. X-ray crystallography has been enormously successful in determining protein structures, and there has been increasing success in determining nucleic acid structures.

As mentioned previously, the Hermann lab has utilized X-ray crystallography to determine the ligand-free and bound structures of the HCV IRES subdomain Iia RNA switch.^{21,47} In these studies, the crystal structure of the Iia RNA switch from the SVV IRES was determined and found to be a structural and functional replacement of the HCV Iia switch. This crystal structure will be discussed in detail in Chapters 2-4.^{39,58} This SVV RNA switch was incorporated as a building block in self-assembling nanoarchitectures, and the crystal structure determination of a self-assembling RNA nanotriangle will be discussed in Chapter 5.⁵⁹ Attempts were made to characterize the structure of the model HCV Iia RNA switch with an incorporated heavy atom containing fluorescent nucleoside analog (selenophene-conjugated uridine, ^{Se}U) useful for detecting RNA folding and ligand binding and providing phase information from its scattering properties for crystal structure determination. While this modified RNA successfully detected ligand binding, no diffracting crystals could be grown. A model RNA construct from the ribosomal decoding site (A-site) which also contained this ^{Se}U, did crystallize, and its determined

structure revealed a minimally structure perturbing heavy atom containing fluorescent nucleoside useful for the utilities described above. This structure will be discussed in Chapter 4.^{60,61}

Fluorescence tools for studying RNA conformational changes, and ligand binding

Many tools utilizing fluorescence properties have been developed or adapted to study RNA conformational changes and ligand binding. Förster resonance energy transfer (FRET) experiments involve a distance-dependent energy transfer between fluorophores, and can be used to monitor conformational changes in functional RNA molecules. In this study, RNA FRET constructs were constructed with 5' terminal Cy3 or Cy5 modification and used to characterize switch folding and conformational capture by ligands of new viral IRES IIa RNA switches (Figure 1.6a).³⁹ A modular system was also developed to screen a potential larger number of RNA switches utilizing universal FRET-labelled DNA probes (Figure 1.6b).⁶² These FRET tools will be discussed in detail in Chapter 4 and their applications will be described in Chapters 2-4.

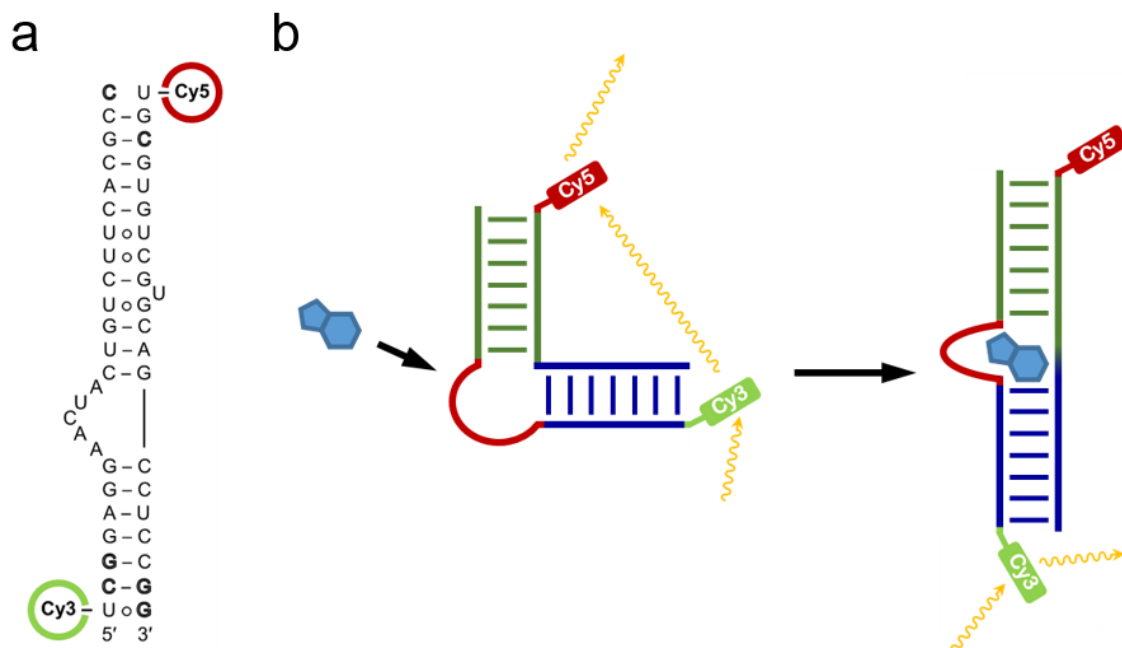


Figure 1.6: RNA switch FRET experiments. a) Model FRET construct of the HCV IRES IIa RNA switch representative of other switch FRET constructs. b) Cartoon demonstrating maximum FRET efficiency for an RNA switch in a bent conformation and maximum Cy3 fluorescence in a ligand-captured elongated conformation (discussed in Chapter 4).

Fluorescent nucleoside analogs which retain base pairing abilities and can be incorporated chemically or enzymatically into nucleic acid oligonucleotides have been developed as probes to study local dynamics, conformational changes, and ligand binding of functional RNA molecules. In another vein, heavy atom containing nucleoside analogs which also retain base pairing abilities and can be incorporated chemically or enzymatically into nucleic acid oligonucleotides have been developed to provide initial phase estimates in solving novel nucleic acid X-ray crystal structures. In this study, a selenophene-conjugated uridine analog ($^{\text{Se}}\text{U}$) was developed as a dual function probe with fluorescence properties able to detect RNA conformational changes and heavy atom

anomalous scattering properties useful for phase determination in crystal structure determination, and was incorporated into model RNA constructs of the HCV Ila RNA switch and the bacterial ribosomal decoding site (A-site) (Figure 1.7).^{60,61} These ^{Se}U-modified constructs both successfully monitor ligand binding by fluorescence changes and the 3D structure of the ^{Se}U-modified A-site RNA construct was determined by X-ray crystallography^{60,61} and will be discussed in Chapter 4.

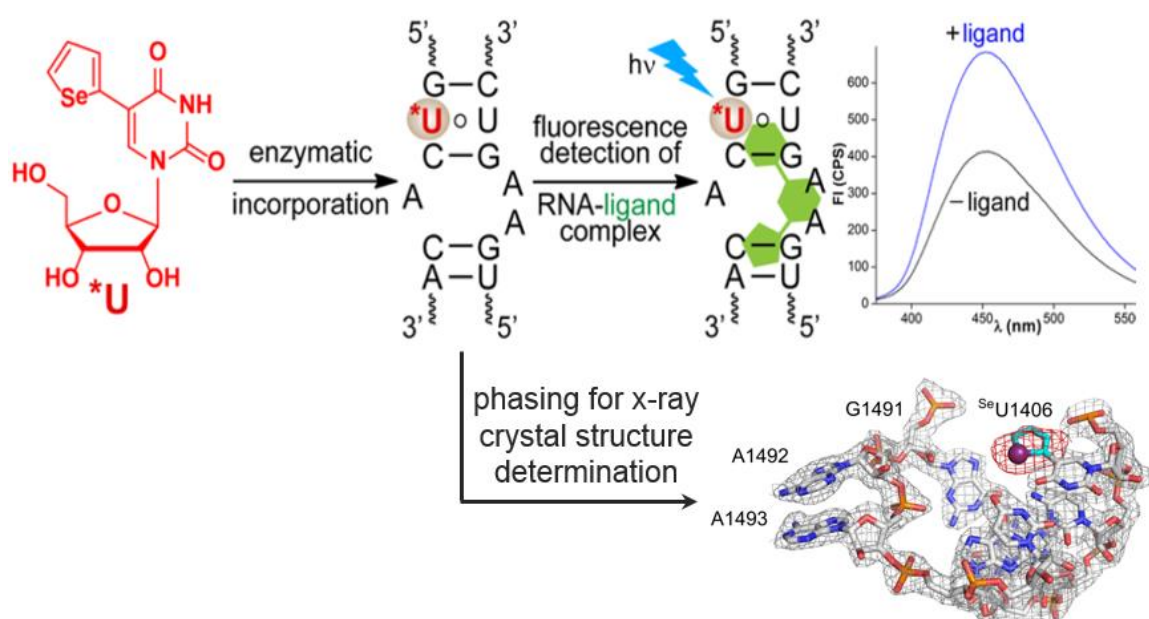


Figure 1.7: Utility of dual functional selenophene-conjugated uridine analog. This ^{Se}U was enzymatically incorporated into a model ribosomal A-site RNA construct whose fluorescence increased upon ligand binding. The crystal structure determination of this ^{Se}U-modified A-site RNA revealed only minimal ^{Se}U-perturbation of the wild type (WT) RNA structure and provides an anomalous scattering selenium atom for phase determination in X-ray crystallography. The upper half of this figure was adapted from ref.⁶¹ 2013, American Chemical Society.

Use of RNA Switches as Tunable Building Blocks for Nanotechnology

Nucleic acid (RNA, DNA, and nucleic acid mimetics) has been used extensively to build nano-sized architectures by controlling assembly through designed base pair interactions. RNA, specifically, has been used in nanotechnology for its structural complexity and diversity as well as its array of biological functions and catalytic properties in addition to its programmability via base pair interactions. The RNA switches introduced previously have unique advantages as building blocks for nanotechnology as they are responsive to ligands and adopt distinct and stable 90° or 180° conformations. A more detailed introduction to this topic as well as a detailed discussion on the incorporation of RNA switches into self-assembling and ligand-responsive nanotriangles characterized by X-ray crystallography will be discussed in Chapter 5 (Figure 1.8).⁵⁹

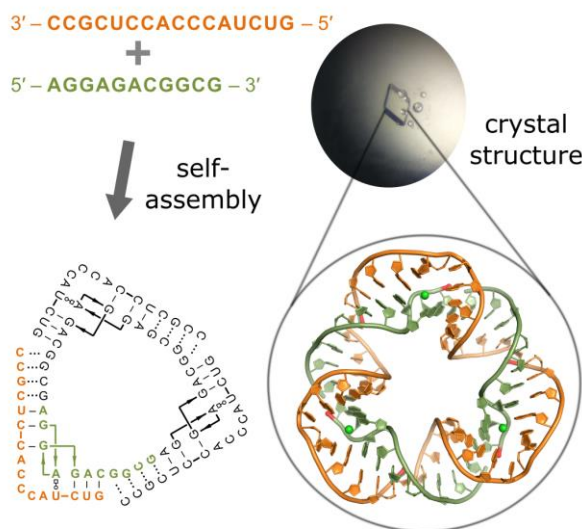


Figure 1.8: Design and characterization of RNA nanostructures. Two short oligonucleotides self-assemble cooperatively in solution to form the so far smallest circularly closed nanotriangle made entirely of double-stranded RNA. This nanoobject formed crystals, and its structure was determined by X-ray diffraction at 2.6 Å resolution.

Chapter 2:

Functional conservation and capture by a common ligand despite structural differences in diverse viral RNA switches

Parts of this work are published in:

“Functional conservation despite structural divergence in ligand-responsive RNA switches” *Proceedings of the National Academy of Sciences*, 111, 15952–15957 (2014).

Abstract

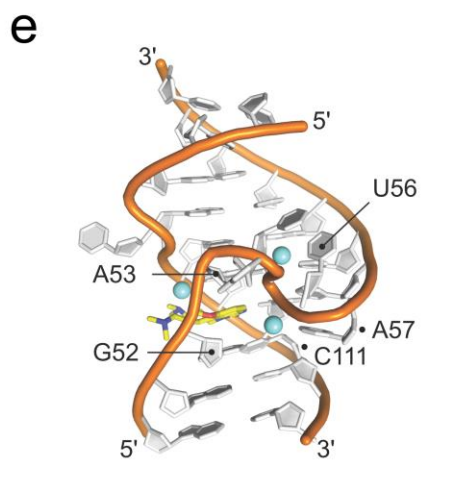
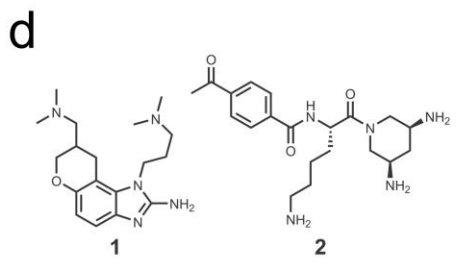
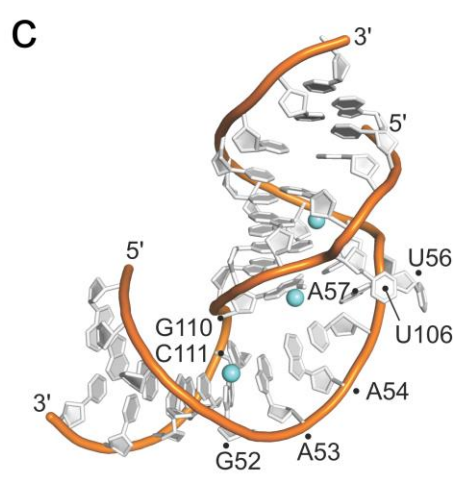
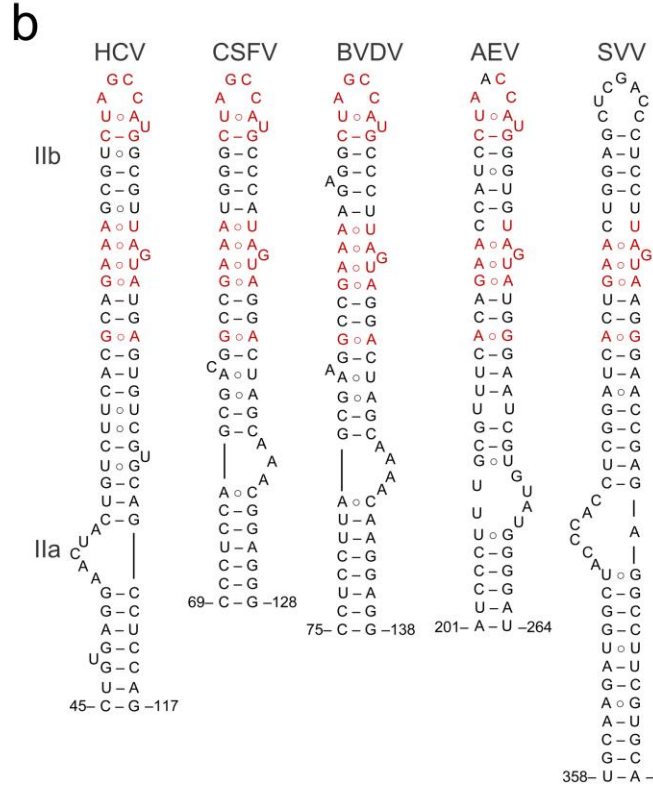
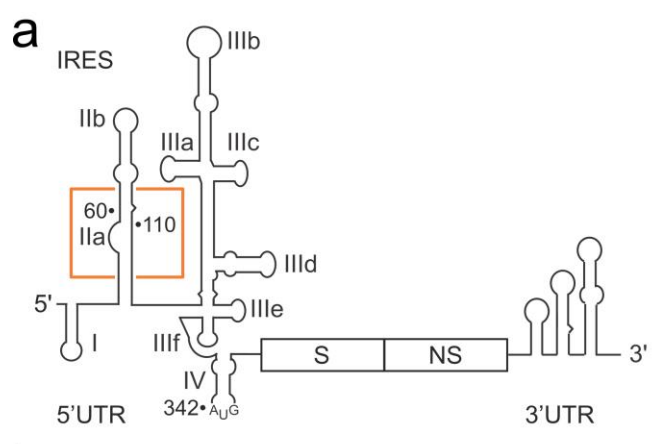
An internal ribosome entry site (IRES) initiates protein synthesis in RNA viruses including the hepatitis C virus (HCV). Ligand-responsive conformational switches have been discovered in viral IRES elements. Modular RNA motifs of greatly distinct sequence and local secondary structure have been found to serve as functionally conserved switches that are involved in viral IRES-driven translation and may be captured by identical cognate ligands. The RNA motifs described here constitute a new paradigm for ligand-captured switches that differ from metabolite-sensing riboswitches with regard to their small size as well as intrinsic stability and structural definition of the constitutive conformational states. These viral RNA modules represent the simplest form of ligand-responsive mechanical switches in nucleic acids.

Introduction

Internal ribosome entry site (IRES) elements provide an alternative mechanism for translation initiation by directing the assembly of functional ribosomes directly at the start codon in a process that does not require 5' cap recognition or ribosomal scanning and which is independent of many host initiation factors.²³⁻²⁶ The genomes of *Flavi-* and *Picornaviridae* contain elements that share similarity with the archetypical hepatitis C virus (HCV) IRES in overall domain organization but not sequence or details of secondary structure.⁵⁶ The HCV IRES adopts a complex architecture of four independently folding domains (Figure 1.1a).²⁸ Domain II is near 100% conserved in clinical isolates⁶ and has analogous counterparts in other viral IRES elements, all of which display some secondary structure similarity but significant sequence variation in

their subdomain IIa internal loops (Figure 1.1b). Domain II has been shown to promote stable entry of HCV and CSFV mRNA at the decoding groove of the 40S subunit^{35,36,43} and is required for initiation factor removal prior to ribosomal subunit joining⁴⁴ as well as adjustment of initiator tRNA orientation.⁴⁶ The transition from initiation to elongation stages of translation depends critically on domain II.⁶³ Recently, direct interaction of HCV domain II with initiator tRNA has been demonstrated.⁴² In HCV, subdomain IIa folds into an L-shaped motif⁴⁷ (Figure 1.1c) that introduces a 90° bend in domain II⁴⁸ and directs the IIb hairpin towards the E-site at the ribosomal subunit interface.^{34,49}

Figure 2.1: Structures and ligands of viral internal ribosome entry sites (IRES). (A) The IRES in the 5' untranslated region (UTR) of the hepatitis C virus (HCV) genome. The location of subdomain IIa is highlighted by an orange box. The viral genome encodes structural (S) and nonstructural (NS) proteins and contains a structured 3' UTR. (B) Secondary structure predictions of domain II motifs in viral IRES elements from HCV and other flaviviruses including classic swine fever virus (CSFV) and bovine viral diarrhea virus (BVDV) as well as picornaviruses such as avian encephalomyelitis virus (AEV) and Seneca Valley virus (SVV). Non-Watson-Crick base pairs are indicated by the \circ symbol. Conservation of the apical loop and the internal loop E motif is indicated in red font. (C) Crystal structure of the subdomain IIa RNA from HCV. (D) Benzimidazole (**1**) and diaminopiperidine (**2**) inhibitors of IRES-driven translation that target the HCV subdomain IIa. (E) Crystal structure of the HCV subdomain IIa RNA in complex with inhibitor **1**.



The HCV IRES subdomain IIa is the target for viral translation inhibitors (Figure 2.1d) that bind to the internal loop and block translation by capturing distinct conformational states of the RNA.⁶ Structure analysis revealed that benzimidazole inhibitors such as compound **1**^{54,64} interact with an extended architecture of IIa in which the stems flanking the internal loop are coaxially stacked on both sides of the ligand binding pocket (Figure 2.1e).²¹ In addition, we have found aryl-substituted 2-aminobenzimidazoles⁶⁵ and 2-aminobenzoxazole⁶⁶ compounds which also bind and capture the extended IIa architecture and for which structure-binding activity relationships have been established. In contrast, diaminopiperidine compounds such as **2** bind and lock the IIa RNA in a bent conformation that corresponds to the ligand-free state.⁵⁵ Conformational capture of the subdomain IIa switch by ligands in solution was demonstrated by FRET experiments and established as a mechanism of IRES inhibition.⁵³ Based on these findings it was proposed that subdomain IIa may be the target for a cognate biological ligand whose adaptive recognition by the RNA motif may facilitate ribosome release from the IRES-bound complex.⁶

Here, we have explored potential candidates for a cognate ligand of the subdomain IIa switch and investigated the structural and functional conservation of similar ligand responsive switch motifs in other IRES RNAs.

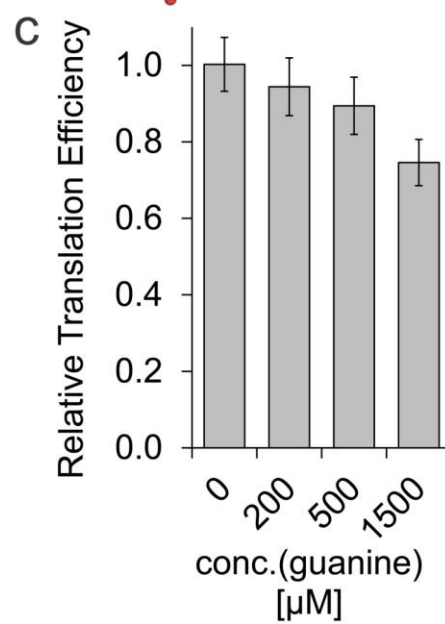
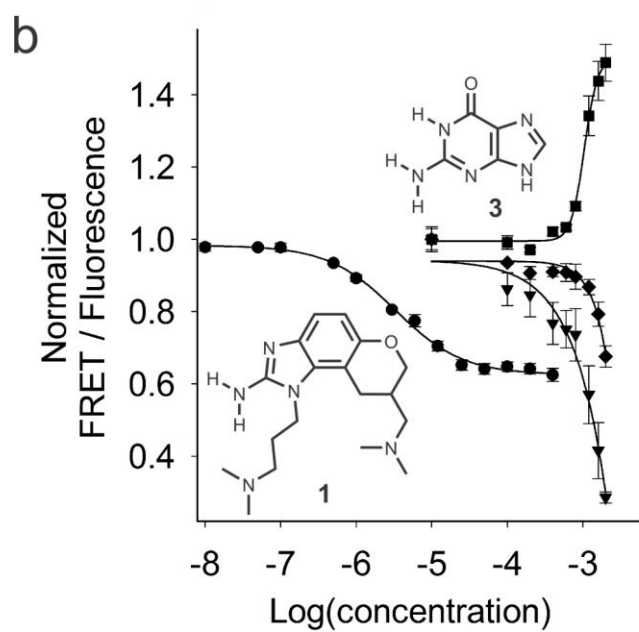
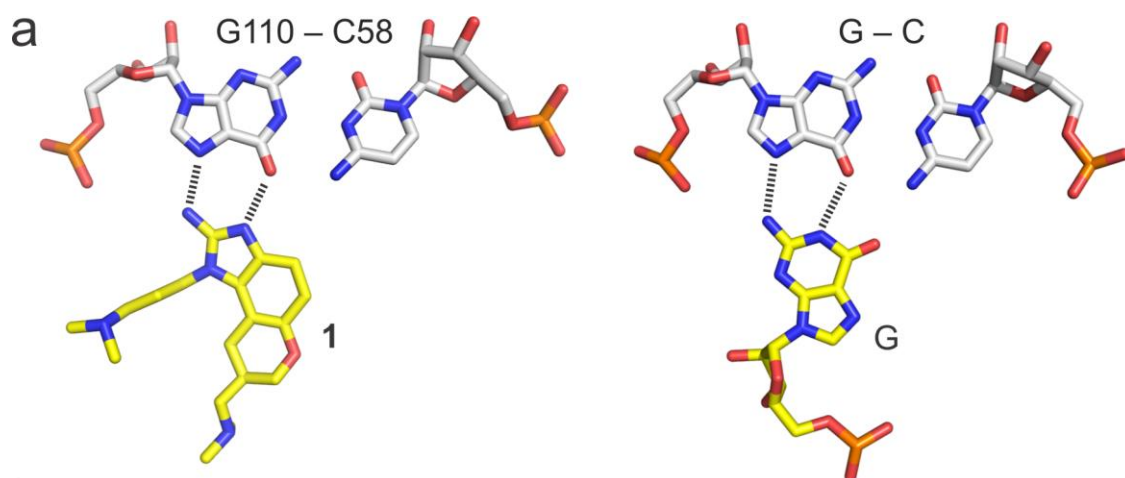
Results and Discussion

Guanine Captures an Extended Conformation of the Subdomain IIa Switch

Cryo-electron microscopy studies of the HCV IRES bound to the ribosome reveal the bent domain II interacting with the 40S subunit in a curved topology that would

prevent the progression of the ribosome from initiation to elongation.^{34,49} Conformational dynamics in the subdomain IIa RNA switch along with ligand capture of the extended state may facilitate removal of domain II from the ribosomal E site.²¹ Support for this hypothesis comes from recent cryo-EM studies that reveal major differences in the conformation of domain II in 40S-bound binary and 80S-bound initiation complexes of the HCV IRES.⁴² Synthetic benzimidazole translation inhibitors (such as **1**)⁵⁴ which capture an extended conformation of subdomain IIa^{21,53} appear to be fortuitous ligands of this RNA motif.⁶ The near perfect conservation across clinical isolates of residues in subdomain IIa⁶ along with the adaptive formation of a deep binding pocket that encapsulates the small molecule ligand reminiscent of riboswitches led us to speculate about a cognate biological ligand. Binding of compounds such as **1** to subdomain IIa depends critically on two hydrogen bonds to the Hoogsteen edge of a G-C base pair, closely resembling isosteric hydrogen bonding patterns observed for G-C pairs interacting with arginine in protein-RNA complexes^{67,68} and with guanosine in C-G⊙G base triples (Figure 2.2a).^{69,70}

Figure 2.2: Ligand binding to the HCV IRES. a) Docking of the benzimidazole inhibitor **1** at the C58-G110 base pair in the crystal structure of the HCV subdomain IIa target complex in comparison to the geometry of a C-G \circ G triple which is a frequently occurring motif in RNA architectures. b) Titrations of Cy3/Cy5-labelled HCV subdomain IIa RNA with benzimidazole **1** and guanine **3**. Curves show normalized FRET signal for **1** (●) and guanine (▼) as well as the normalized fluorescence signals of the donor Cy3 (■) and acceptor Cy5 (◆) for guanine. Fitting of a single-site binding curve to the benzimidazole **1** FRET response gave an EC₅₀ value for ligand binding of 3.4±0.3μM. Because the FRET signal did not reach saturation in the guanine titration, affinity was estimated by fitting single-site binding to the Cy3 emission which resulted in an EC₅₀ value for ligand binding of 1050±69μM (guanine). Error bars represent ± 1s.d. calculated from triplicate experiments. c) Impact of guanine on IRES-driven translation as measured in an *in vitro* translation assay. Error bars represent ± 1s.d. calculated from triplicate experiments.



Inspired by Yarus's classical study of arginine and G binding to self-splicing group I introns,⁷¹ we used a previously established FRET assay^{53,72} to test arginine, guanine and derivatives thereof for binding to the subdomain Iia RNA (see also the *Extended discussion of guanine-like molecules* later in this chapter). Conformational capture of the dye-labelled RNA switch (Figure 2.3) in the extended state leads to dose-dependent reduction of the FRET signal and concurrent increase of Cy3 donor dye fluorescence due to diminished resonance energy transfer. The FRET assay has previously been used to demonstrate conformational capture of the Iia switch by benzimidazole ligands.⁵³ Neither arginine nor its derivatives showed binding to the Iia RNA construct. In contrast, titration of guanine (**3**) elicited a dose-dependent FRET reduction indicative of ligand binding to the extended conformation of subdomain Iia (Figure 2.2b). While the binding activity of guanine was around 1 mM (EC₅₀ value), 300-times weaker than that of the benzimidazole inhibitor **1**, the concurrent increase of fluorescence from the Cy3 donor dye substantiated a selective ligand interaction leading to conformational capture. Binding was specific for guanine, either as the base itself or in the nucleoside guanosine, while other nucleobases were inactive in the FRET assay (Figure 2.4a, b). The 2-aminopyrimidinone ring of guanine was essential for the interaction with the subdomain Iia RNA as indicated by the binding activity of 2-aminoquinazolinone (compound **4**, Figure 2.4c) and its derivatives (*Extended discussion of guanine-like molecules*; Tables 2.3, 2.5) as well as pterin derivatives (*Extended discussion of guanine-like molecules*; Tables 2.3, 2.6), which share the pyrimidinone but not the imidazole heterocycle. The quinazolinone **4** bound the target about twofold tighter than guanine (EC₅₀ value of 483 μM), perhaps due to more favorable stacking interactions

with the larger benzene ring in **4**. Structure-activity relationships for guanine and the 2-aminopyrimidinone-containing quinazoline and pteridine scaffolds and their derivatives will be discussed in greater detail in the *Extended discussion of guanine-like molecules* later in this chapter.

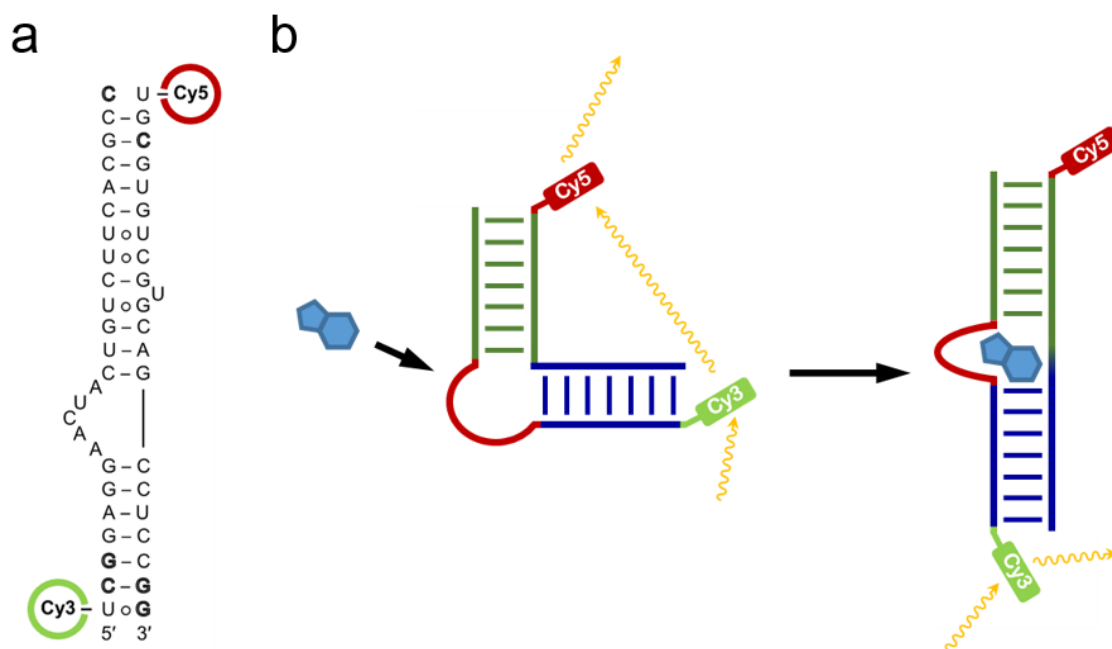


Figure 2.3: RNA switch FRET experiments. a) Structure and sequence of dye-labeled HCV IRES subdomain IIa construct for FRET experiments.^{53,62} b) Cartoon demonstrating maximum FRET efficiency for an RNA switch in a bent conformation and maximum Cy3 fluorescence in a ligand-captured elongated conformation.

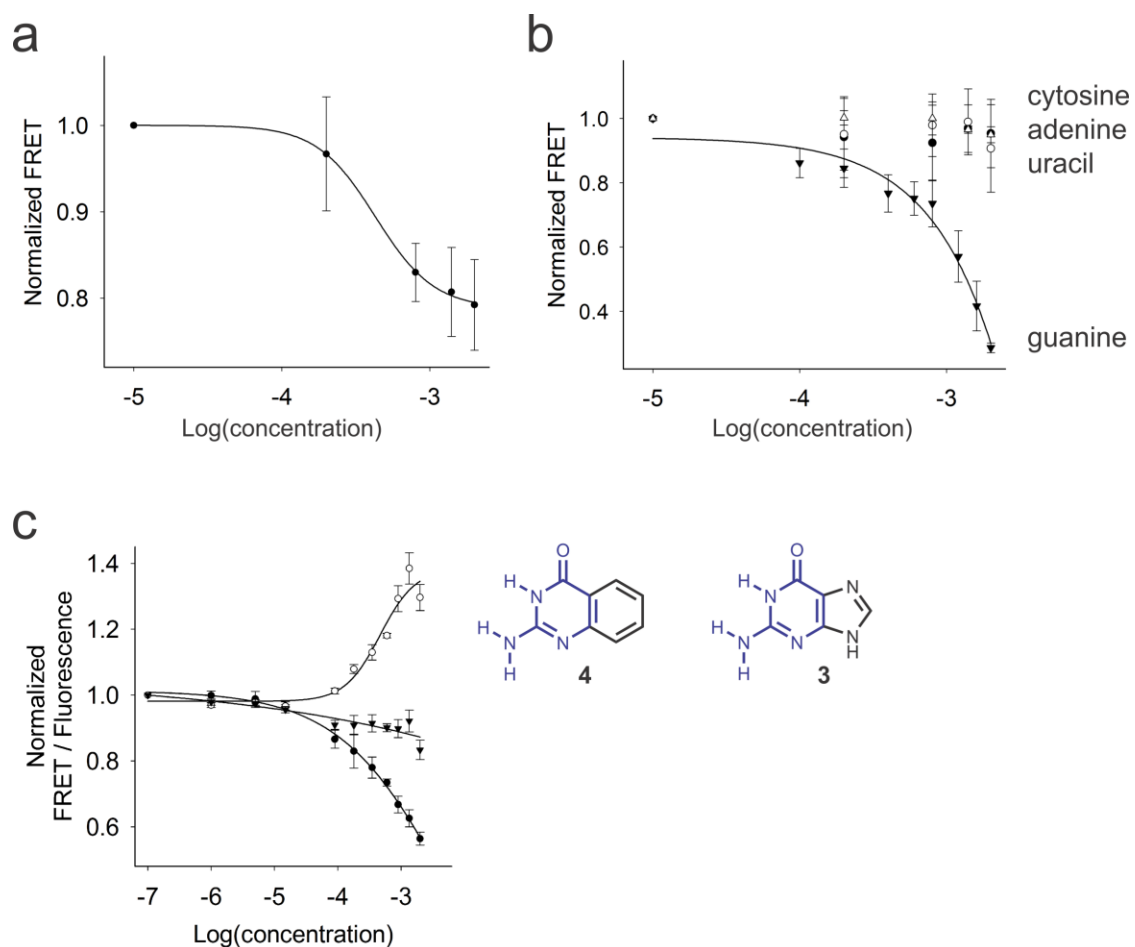


Figure 2.4: HCV FRET experiments with guanine control compounds including guanosine, the other three nucleobases (cytosine, adenine, uracil), and a 2-aminoquinazolinone. a) Titration of Cy3/Cy5-labelled HCV subdomain IIa with guanosine (●). b) Titrations of Cy3/Cy5-labelled HCV subdomain IIa with guanine (▼), cytosine (●), adenine (○) and uracil (Δ). c) Titration of Cy3/Cy5-labelled HCV subdomain IIa with 2-aminoquinazolin-4(3H)-one **4** which contains the 2-aminopyrimidinone heterocycle (blue) of guanine **3** but has the imidazole ring replaced by a benzene. Curves show normalized FRET signal for **4** (●) as well as the normalized fluorescence signals of the donor Cy3 (○) and acceptor Cy5 (▼). Because the FRET signal did not reach saturation, affinity was estimated by fitting single-site binding to the Cy3 emission which resulted in an EC₅₀ value for ligand binding of 483±120μM. In all panels, error bars represent ± 1s.d. calculated from triplicate experiments.

To test for the impact of the guanine interaction on IRES function, we measured luciferase reporter expression levels from a bicistronic construct in an *in vitro* translation assay (Figure 2.2c).⁷² The bicistronic reporter construct allows measuring effects on IRES-driven translation of Renilla luciferase while also providing an internal control of a cap-initiated firefly luciferase (Figure 2.5f). In agreement with the observed binding of guanine to subdomain IIa, the nucleobase selectively inhibited expression of the IRES-driven reporter (Figure 2.2c). Translation inhibition required a higher concentration of guanine compared to benzimidazole **1** (Figure 2.5a, e), again reflecting the weaker binding of guanine relative to the synthetic inhibitor. Inhibition of reporter expression was selective for guanine as none of the other three nucleobases showed this effect (Figure 2.5b-d).

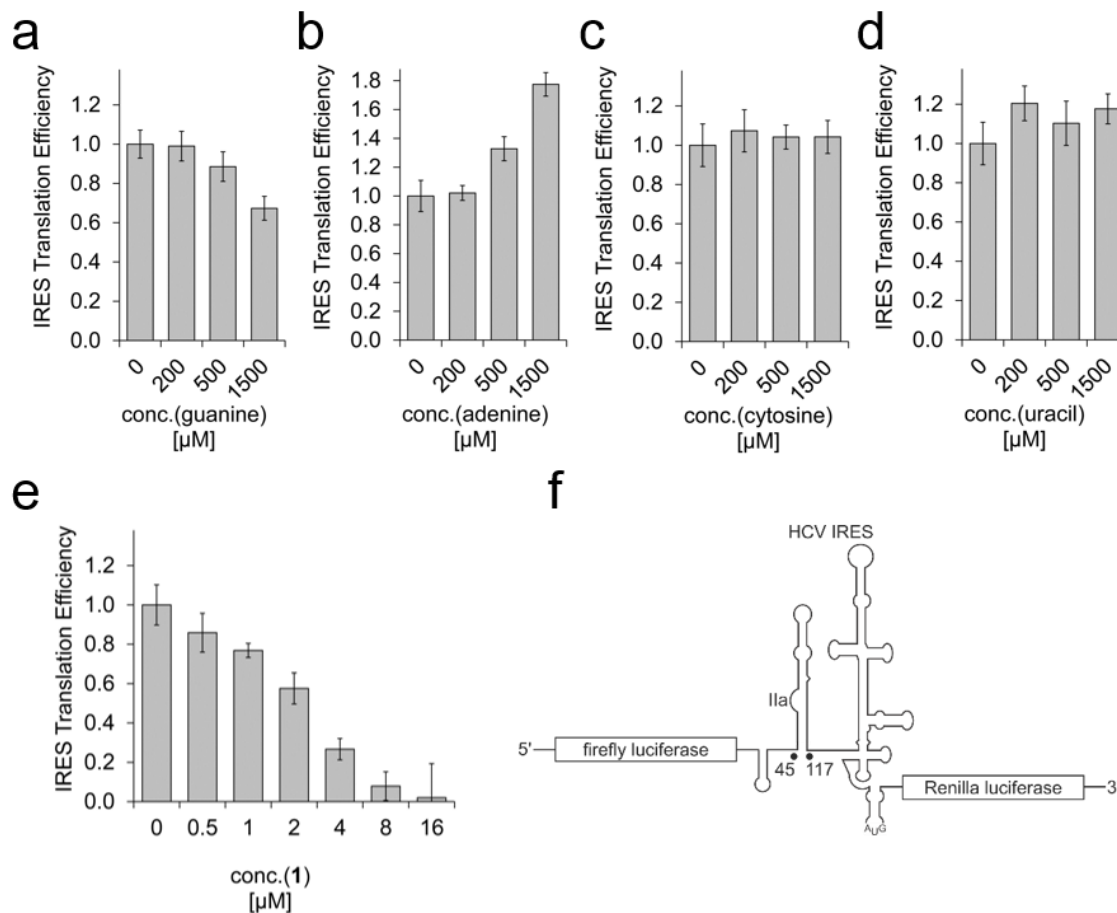


Figure 2.5: Impact of a) guanine b) adenine c) cytosine d) uracil and e) the benzimidazole **1** on IRES-driven translation as measured in an *in vitro* translation assay. Error bars represent \pm 1s.d. calculated from triplicate experiments. f) Bicistronic (dual) luciferase reporter construct used for the IVT experiments.

Guanine also selectively binds a dye-labelled HCV Ila FRET construct containing an A57U mutation seen in more than 10% of HCV clinical isolates (Figure 2.6).^{21,53} This result is consistent with the ability of benzimidazole **1** to bind this A57U mutant.²¹ This mutation decreases viral translation efficiency in an *in vitro* translation assay⁵⁸ (Chapter 3: Figure 3.3c; This and other HCV Ila mutations are discussed in greater detail in

Chapter 3) and in human cells transfected with HCV replicon⁵³ but does not appear to decrease ligand binding in the FRET assay to the same extent.

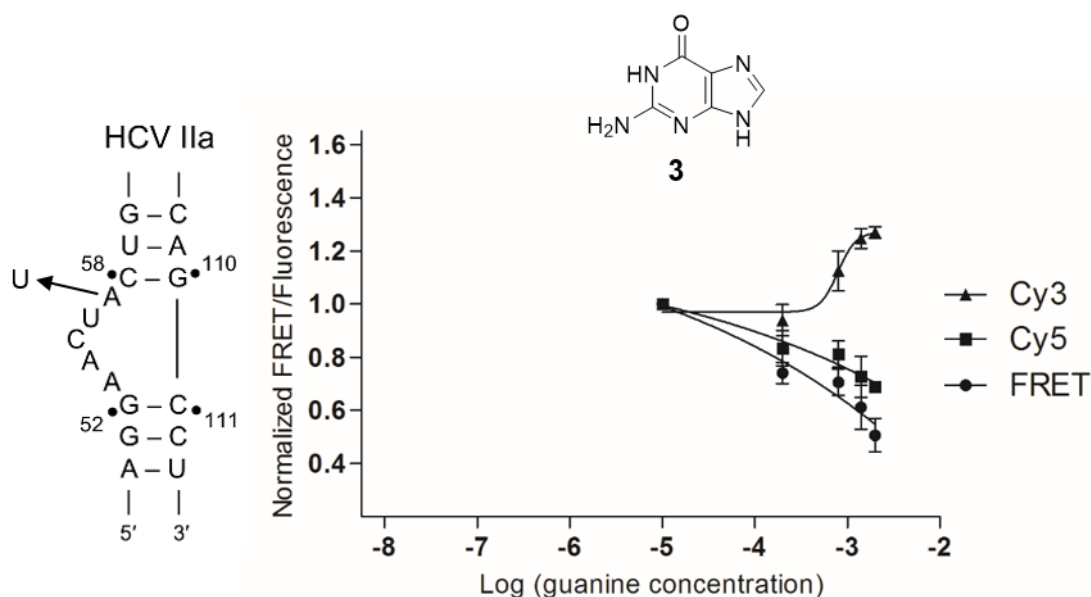


Figure 2.6: Titration of **3** (guanine) with FRET-labelled HCV IRES IIa RNA carrying an A57U mutation. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 s.d. calculated from triplicate experiments. Fitting of a single-site binding curve to Cy3 emission (FRET signal did not reach saturation) gave an estimated EC_{50} value for ligand binding of $800 \pm 140 \mu\text{M}$.

Domain II-like motifs from other IRES RNAs are unrelated in sequence but fold like the HCV archetype

IRES elements and their function in translation initiation differ fundamentally in distinct families of positive-sense RNA viruses including *Flaviviridae* and *Picornaviridae*. The IRES domains of flaviviruses such as HCV, classic swine fever virus (CSFV) and bovine viral diarrhoea virus (BVDV) recruit 40S subunits independent of eukaryotic initiation factors (eIF) while IRES elements of most picornaviruses require

involvement of eIF4 components.⁷³ Despite the differences, RNA domains have been identified in the IRES elements of picornaviruses, including avian encephalomyelitis virus (AEV) and Seneca Valley virus (SVV),⁷⁴ which resemble analogous motifs in flaviviruses in overall secondary structure.^{56,57} RNA structures corresponding to the HCV domain II occur in the IRES elements of CSFV, BVDV, AEV, SVV (Figure 2.1b) as well as several other RNA viruses.⁵⁶ Domain II-like RNAs share common motifs including a hairpin loop in subdomain IIb and a loop E motif⁷⁵ which are conserved in sequence and separation by base pairs (Figure 2.7a). The internal loop of subdomain IIa shows no sequence conservation and differs both in size and location in the 5' or 3' proximal strand of the lower stem. NMR analysis of domain II from CSFV, which revealed a bent subdomain IIa,⁴⁴ hinted at the structural similarity of subdomain IIa motifs despite their distinctions in sequence, size and strand location.

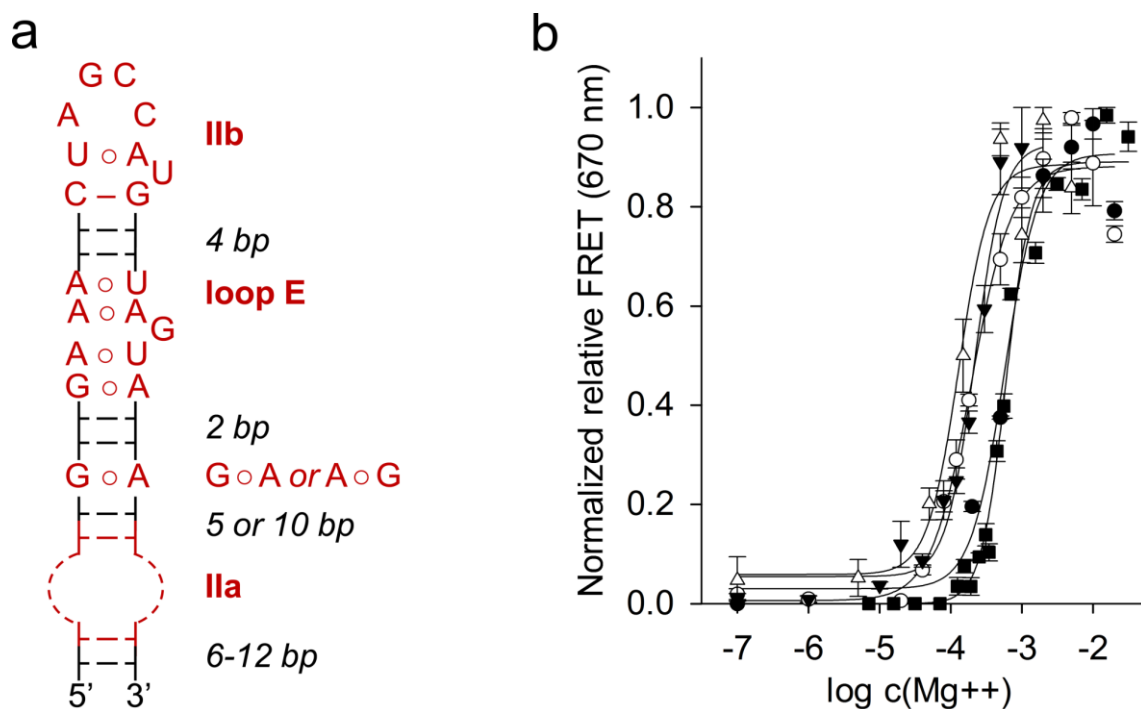


Figure 2.7: Domain II-like RNAs in viral IRES elements. (A) Consensus secondary structure of domain II analogs from viral IRES elements. The internal loop of subdomain IIa consists of an unpaired stretch of nucleotides in either the 5' or 3' proximal strand of the lower stem. (B) Mg²⁺-induced folding of viral subdomain IIa motifs. FRET signal from terminally Cy3/Cy5 labeled RNA constructs was monitored while increasing Mg²⁺ concentration. See Figure 2.8 for RNA construct structures. Fitting of dose-response curves resulted in EC50 values for Mg²⁺-induced folding at 598 ± 25 μM (HCV, ■) (23), 196 ± 21 μM (CSFV, ○), 219 ± 19 μM (BVDV, ▼), 131 ± 21 μM (AEV, Δ), and 579 ± 86 μM (SVV, ●). Error bars represent ±1 s.d. calculated from triplicate experiments.

To investigate if subdomain IIa motifs from different viral IRES elements share the ability to adopt structurally similar bent architectures, we studied magnesium ion-induced folding of the RNAs by monitoring FRET. We had previously used the FRET assay to establish magnesium-dependent folding of the HCV subdomain IIa.⁵³ We titrated dye labeled constructs containing viral subdomain IIa motifs (Figure 2.8), which were initially free of divalent metal ions, with increasing amounts of Mg²⁺ (Figure 2.7b).

FRET was not observed for any of the constructs in the absence of Mg^{2+} when the internal loop of subdomain IIa does not stably fold and the RNA adopts an extended or flexible conformation which places the dyes beyond the Förster radius. Upon addition of Mg^{2+} the FRET signal appeared in a dose-dependent fashion, indicating folding into a bent architecture for constructs of all viruses. Control RNA in which the internal loop was deleted did not show an increase of FRET (Figure 2.9a). Effective Mg^{2+} concentrations for folding were of similar magnitude for all subdomain IIa constructs, ranging from ~130–600 μ M (Figure 2.7b). Constructs that were more closely related with respect to overall secondary structure required similar amounts of magnesium for folding, delineating two distinct subtypes of subdomain IIa architectures (CSFV, BVDV, AEV: 131–219 μ M Mg^{2+} ; SVV, HCV: 579–598 μ M Mg^{2+}). Separation between these subtypes did not coincide with family divisions between *Flaviviridae* and *Picornaviridae*, but rather coincided with location of the subdomain IIa internal loop on the 5' or 3' proximal strand of the lower domain II stem.

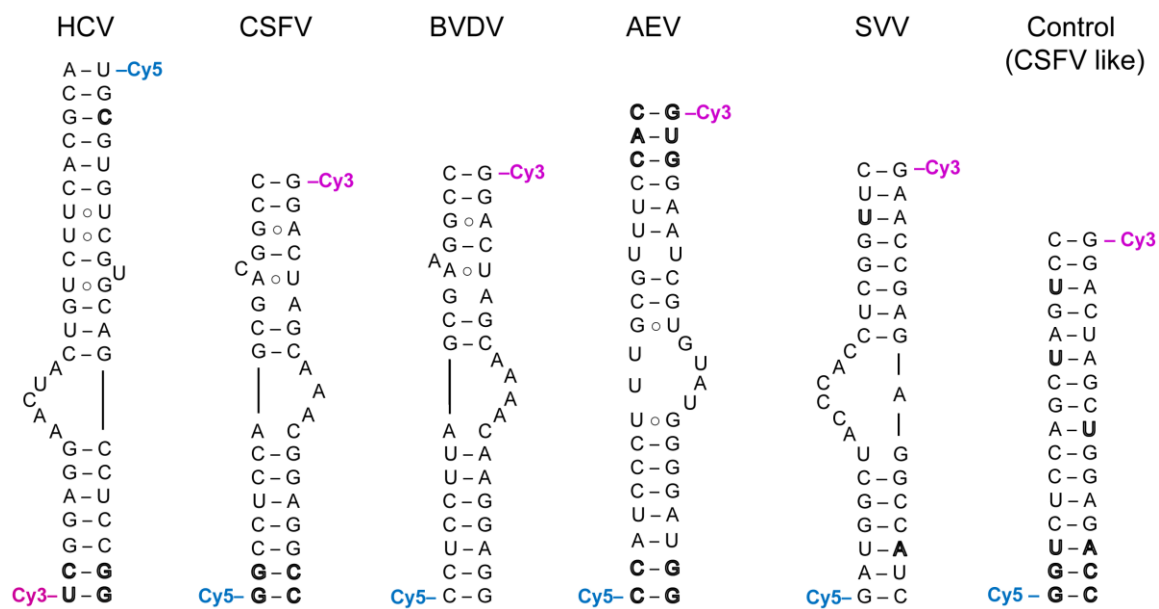


Figure 2.8: Structures and sequences of dye-labeled viral IRES subdomain IIa constructs for FRET experiments.

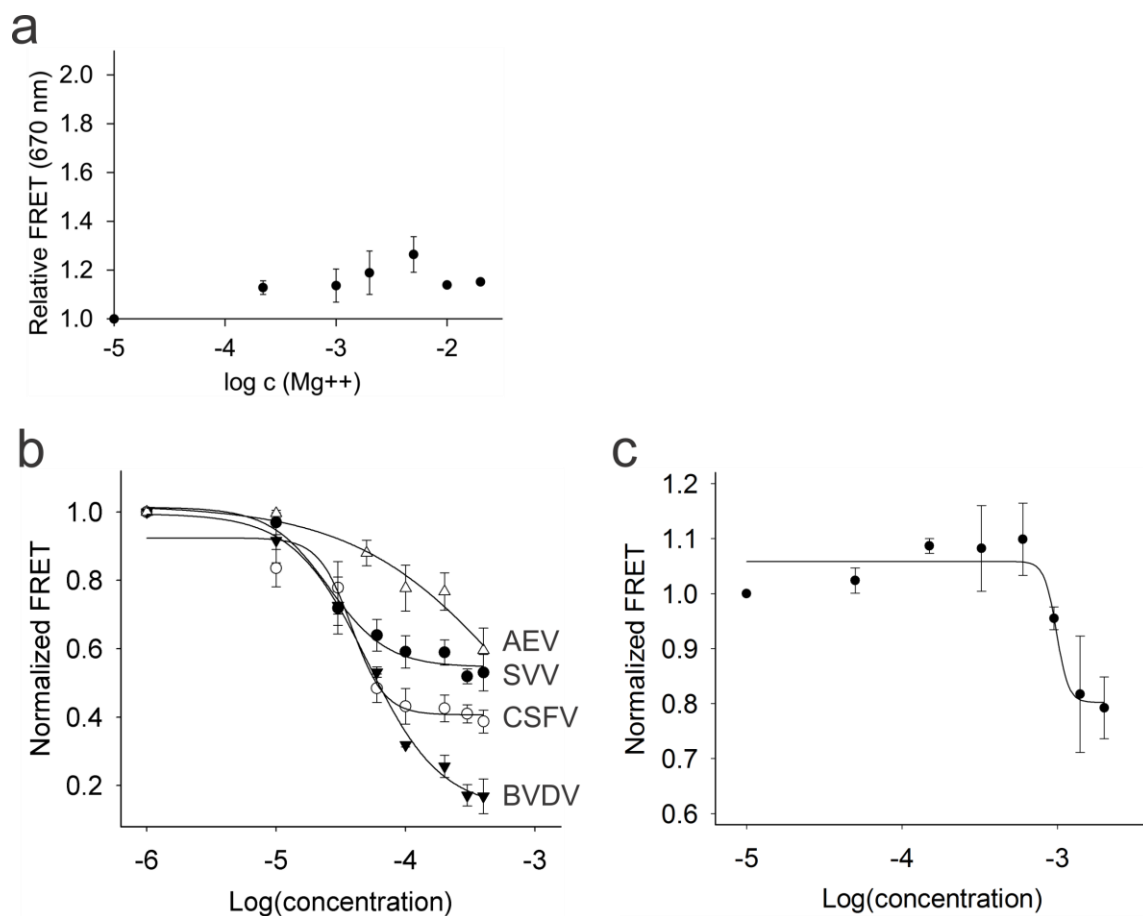


Figure 2.9: FRET experiments with constructs shown in Figure 2.8. a) Titration of terminally Cy3/Cy5 labeled RNA Control construct with Mg²⁺. b) Titrations of Cy3/Cy5-labelled viral subdomain IIa RNA constructs with benzimidazole **1**. Fitting of single-site binding curves resulted in EC₅₀ values for ligand binding of 39±5μM (CSFV ○), 51±7μM (BVDV ▼) and 26±5μM (SVV ●). Titration of the AEV RNA did not result in a saturating binding curve (Δ). c) Titration of the SVV RNA with guanine. Fitting of a single-site binding curve resulted in an EC₅₀ value for ligand binding of 990±170μM. In all panels, error bars represent ± 1s.d. calculated from triplicate experiments.

To further explore the ability for conformational switching that subdomain IIa analogs may share in common with the HCV archetype motif, we used the FRET assay to test other viral motifs for interaction with the benzimidazole translation inhibitor **1**. Titration of the compound showed that ligand binding captured the subdomain IIa RNAs

of CSFV, BVDV and SVV in an extended state (Figure 2.9b), similar as previously demonstrated for the HCV target (Figure 2.2b), albeit requiring an 8 to 15-fold higher concentration of **1**. The motif from AEV showed weaker interaction with the compound and saturation of the ligand bound state was not achieved. Because subdomain IIa of SVV bound the benzimidazole tightest among the viral analogs, this RNA was also tested for its interaction with guanine. Similar to the subdomain IIa motif of HCV, the SVV analog was captured in an extended state by guanine, at a comparable EC₅₀ value of 990 μ M (Figure 2.9c).

Subdomain IIa RNA from the SVV IRES Element Adopts a 90° Bent Fold.

Since experiments to study metal-induced folding indicated that diverse viral subdomain IIa analogs adopt bent architectures, we attempted to elucidate the three-dimensional structure of the RNAs. Crystal structures of two different RNA constructs of the SVV subdomain IIa were determined. A high resolution structure (1.86 \AA) was obtained for the internal loop (Figure 2.10, 2.11; Table 2.1) and a lower resolution structure (3.2 \AA) for an extended construct that included the region connecting to the loop E motif (Figure 2.12, Table 2.2). Both structures showed identical architecture for the internal loop and neighboring base pairs.

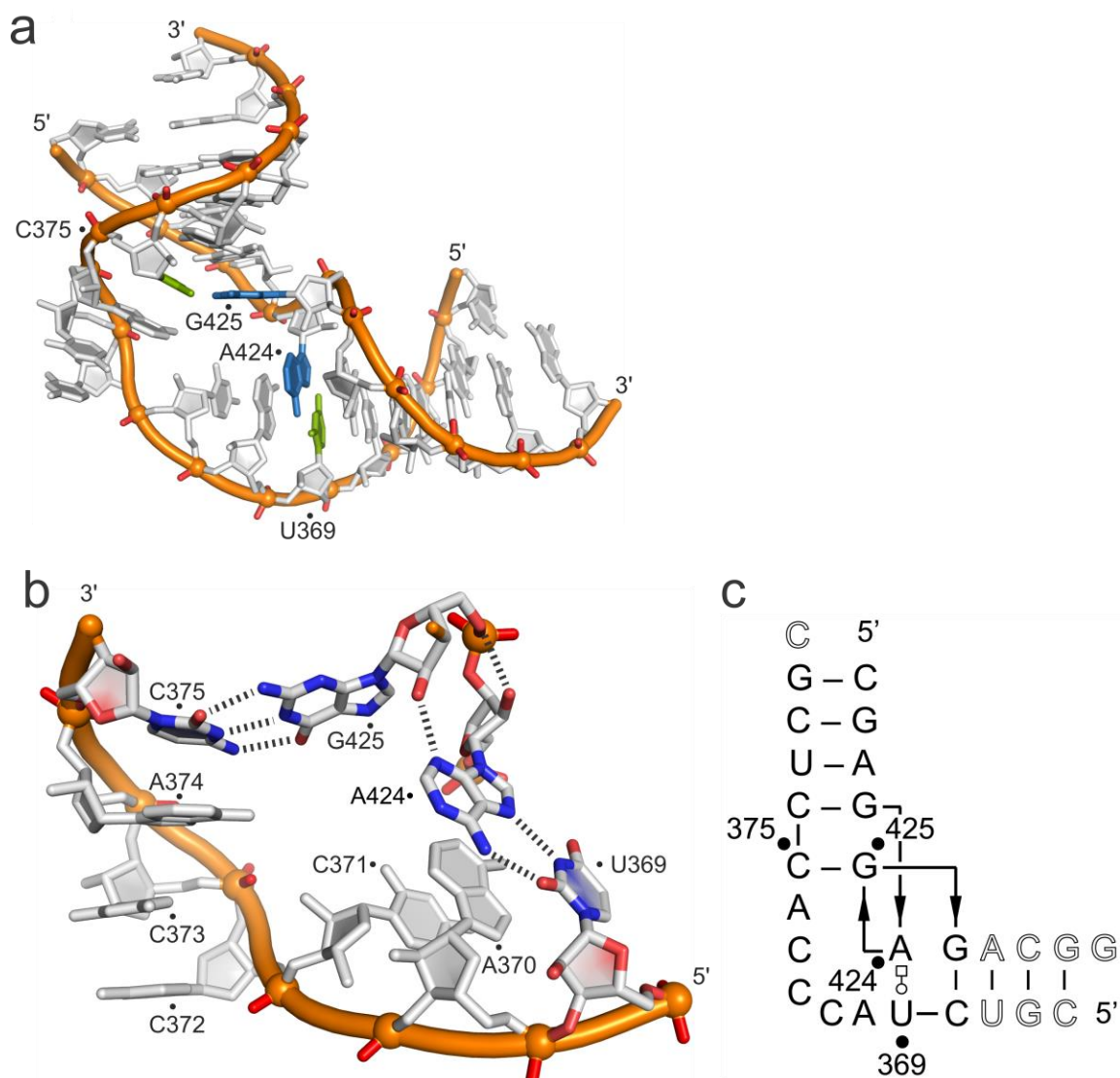


Figure 2.10: Structure of the SVV IRES subdomain IIa RNA. a) Crystal structure of the bent SVV subdomain IIa RNA fold which is stabilized by the perpendicular U369○A424 trans-Watson Crick-Hoogsteen base pair and the Watson Crick C375-G425 base pair (highlighted in color). b) Detailed view of the internal loop RNA fold showing additional hydrogen bonds between A424 and G425 in the perpendicular base pairs (N3_{A424}...O2'_{G425} and O2'_{A424}...O5'_{G425}). c) Secondary structure schematic illustrating base pairing and consecutive stacking interactions. All base pairs adopt canonical Watson-Crick geometry except U369 and A424 which form a reverse Hoogsteen base pair. Nucleotides deviating from the SVV wild type sequence are shown in outlined font.

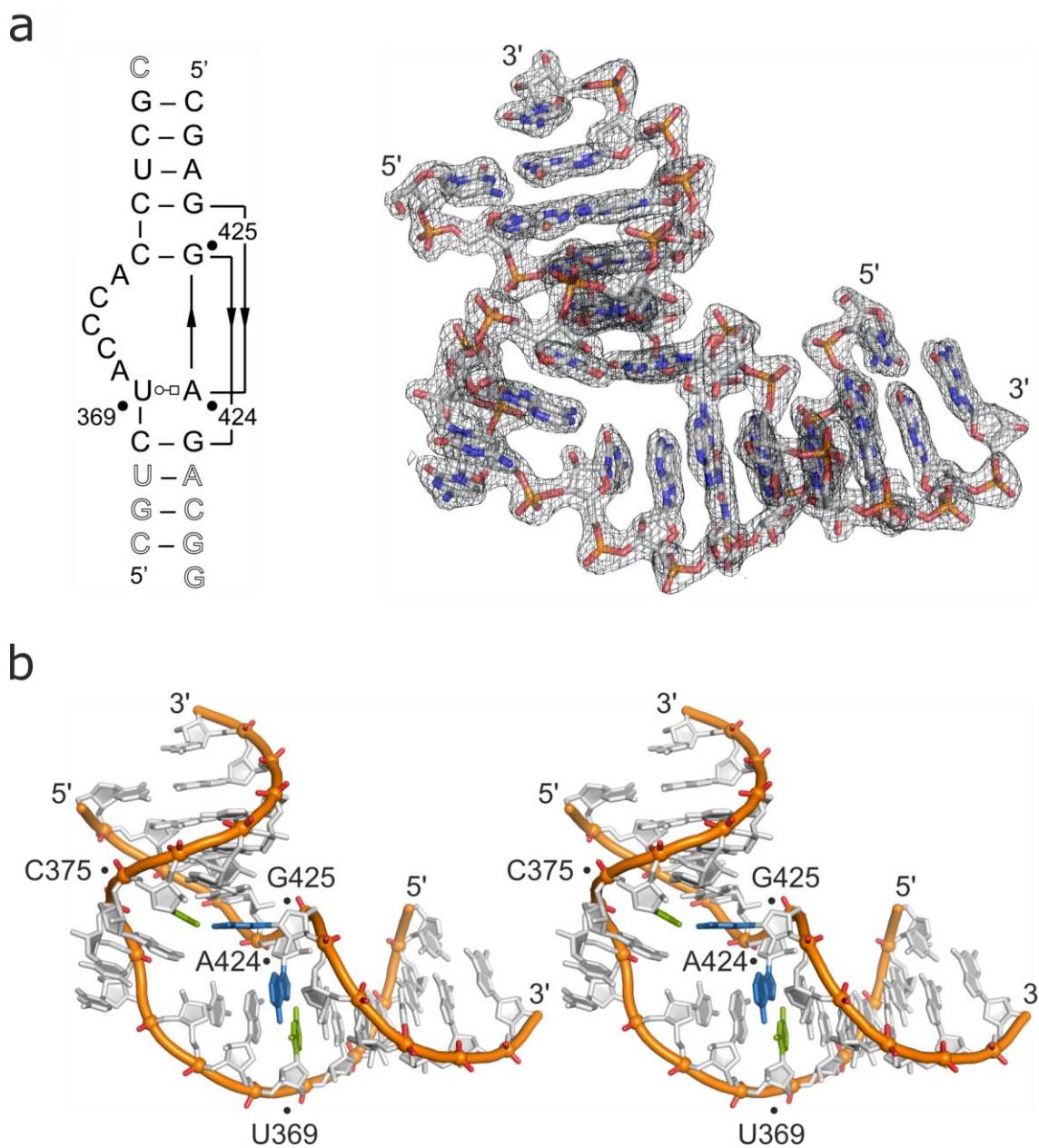


Figure 2.11: Crystal structure of the SVV subdomain IIa RNA, determined at 1.86Å resolution (Table 2.1; PDB entry 4P97). a) The RNA construct that was used for crystallization is shown on the left, with secondary structure indicated as observed in the crystal. A $2F_o - F_c$ electron density map is shown contoured at 1σ . b) Stereo view of the crystal structure.

Table 2.1: Crystallographic data collection and refinement statistics for the SVV IRES subdomain IIa RNA structure (PDB entry 4P97).

Data Collection	
Wavelength (Å)	1.54
High-resolution limit (Å)	1.86
Low-resolution limit (Å)	19.56
Redundancy ^a	8.2 (3.2)
Completeness (%) ^a	99.12 (90.3)
$I/\sigma(I)$ ^a	10.78 (3.67)
Total reflections	13932
Unique reflections	1699
Refinement	
Space group	H3
Cell dimensions (Å)	
<i>a</i>	60.53
<i>b</i>	60.53
<i>c</i>	122.68
α	90
β	90
γ	120
R_{work}/R_{free}	0.19 / 0.23
No. atoms	
RNA atoms	1136
Solvent atoms	111
Metal ions	2 Ca ²⁺
Mean <i>B</i> factors (Å ²)	
RNA	35.3
Solvent	40.1
Metal	46.1
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.163
Dihedral angles (°)	9.519

^aNumbers in parentheses are for the highest-resolution shell.

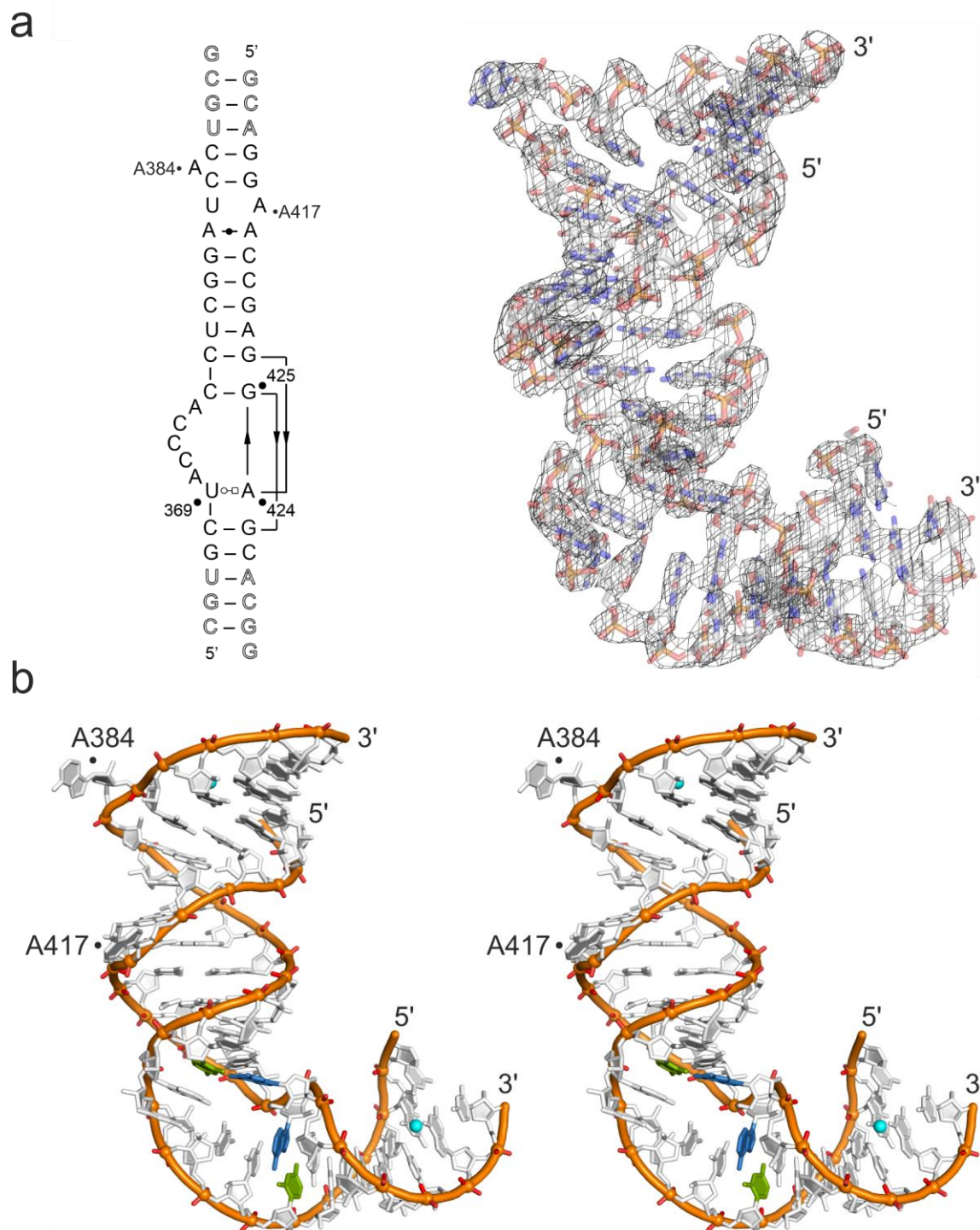


Figure 2.12: Crystal structure of an extended SVV subdomain IIa RNA construct, determined at 3.2Å resolution (Table 2.2; PDB entry 4PHY). a) The RNA construct that was used for crystallization is shown on the left, with secondary structure indicated as observed in the crystal. A $2F_o - F_c$ electron density map is shown contoured at 1.5σ . (B) Stereo view of the crystal structure.

Table 2.2: Crystallographic data collection and refinement statistics for the SVV IRES subdomain IIa extended RNA structure (PDB entry 4PHY).

Data Collection	
Wavelength (Å)	1.54
High-resolution limit (Å)	3.10
Low-resolution limit (Å)	19.90
Redundancy ^a	28.0 (15.5)
Completeness (%) ^a	88.4 (56.4)
$I/\sigma(I)$ ^a	22.28 (2.48)
Total reflections	92932
Unique reflections	3319
Refinement	
Space group	P6322
Cell dimensions (Å)	
<i>a</i>	79.61
<i>b</i>	79.61
<i>c</i>	100.94
α	90
β	90
γ	120
R_{work}/R_{free}	0.21 / 0.27
No. atoms	
RNA atoms	1136
Ligand	2 CH ₃ COO ⁻
Metal ions	2 Mg ²⁺
Mean <i>B</i> factors (Å ²)	
RNA	84.1
Ligand	81.8
Metal	63.42
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.070
Dihedral angles (°)	17.489

^aNumbers in parentheses are for the highest-resolution shell.

The SVV IRES subdomain IIa adopts a bent fold that arranges the flanking RNA helices at an angle of 90° (Figures 2.10a, 2.11, 2.12). The overall architecture of subdomain II from SVV was nearly identical to that of the motif from HCV (Figure 2.13), despite differences in sequence and secondary structure. The folding of the bend in the SVV motif has no precedent in other RNA architectures. Except for the reverse Hoogsteen pair U369○A424, all base pairs adopt canonical Watson-Crick geometry and all unpaired bases are continuously stacked on one of the flanking helices. The U369○A424 and C375-G425 pairs are arranged orthogonally and provide the closing ends that interface the flanking helices to the internal loop (Figure 2.10a). The perpendicular alignment of consecutive bases A424 and G425 is stabilized by two hydrogen bonds involving the sugar backbone of these residues (Figure 2.10b). Resulting from the orthogonal orientation of A424 and G425, the U369○A424 and C375-G425 pairs form a cross-over motif, and strand directionality is locally inverted in the secondary structure (Figure 2.10c).

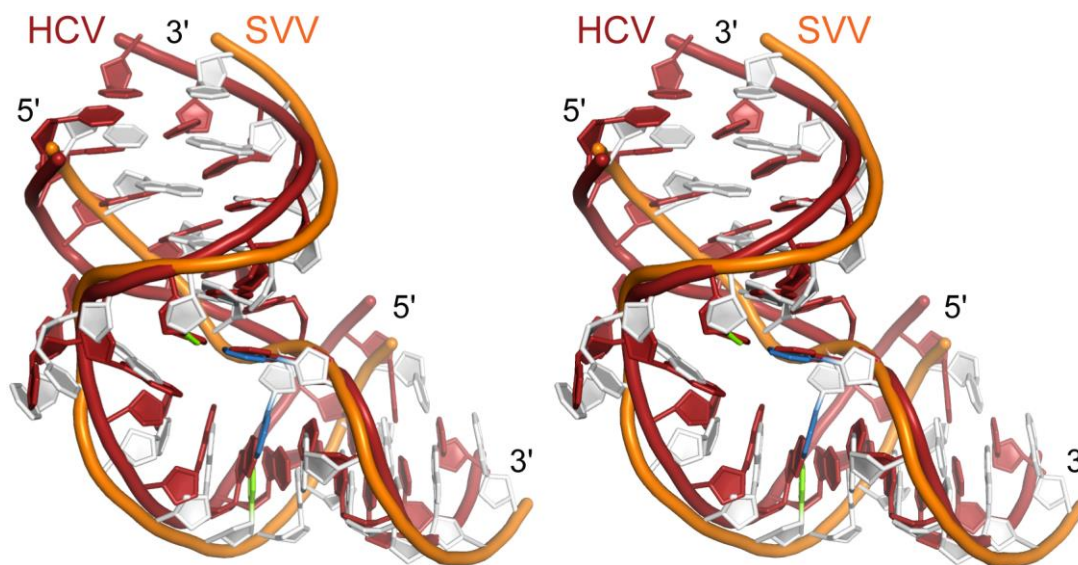


Figure 2.13: Stereo view of a superposition of the subdomain IIa crystal structures from the HCV (red) and SVV (orange/white) IRES IIa elements. The cross-stacked U369-A424 reverse Hoogsteen base pair and the Watson-Crick C375-G425 base pair in the SVV RNA structure are highlighted in green and blue (see Figure 2.10a).

Domain II-like RNAs are interchangeable functional modules despite their sequence dissimilarity.

To determine if domain II analogs are autonomous RNA motifs that are functional as interchangeable modules, we investigated reporter expression from chimera IRES constructs in which the cognate HCV domain II was substituted by analogs from other viruses (Figure 2.14). Replacement of the HCV domain II by any of the other viral analogs resulted in an IRES element that was proficient for translation initiation, except for the SVV domain which does not carry a homologous apical hairpin sequence (IIb, Figure 2.15a). This IIb hairpin sequence along with the loop E motif are necessary for translational activity, and the replacement of the IIb motif with the corresponding motif from SVV which shows no sequence conservation but does contain the same number of

residues rescues translational activity to roughly 40% of wt (Figure 2.16). Domain II replacement observations are in agreement with a previous report on translational activity for a similar HCV IRES chimera that had the domain II inserted from CSFV.⁷⁶ Deletion of the subdomain IIa-like region in the chimera IRES elements, which ablates the bend in domain II, led to significant loss of function in all constructs (Figure 2.15a, Δ IIa). Removal of two base pairs between the subdomain IIb hairpin loop and the conserved loop E motif, which affects both the distance and rotational orientation of the two motifs, resulted in a reduction of activity that was also similar across the different chimeras (Figure 2.15a, Δ 2bp).

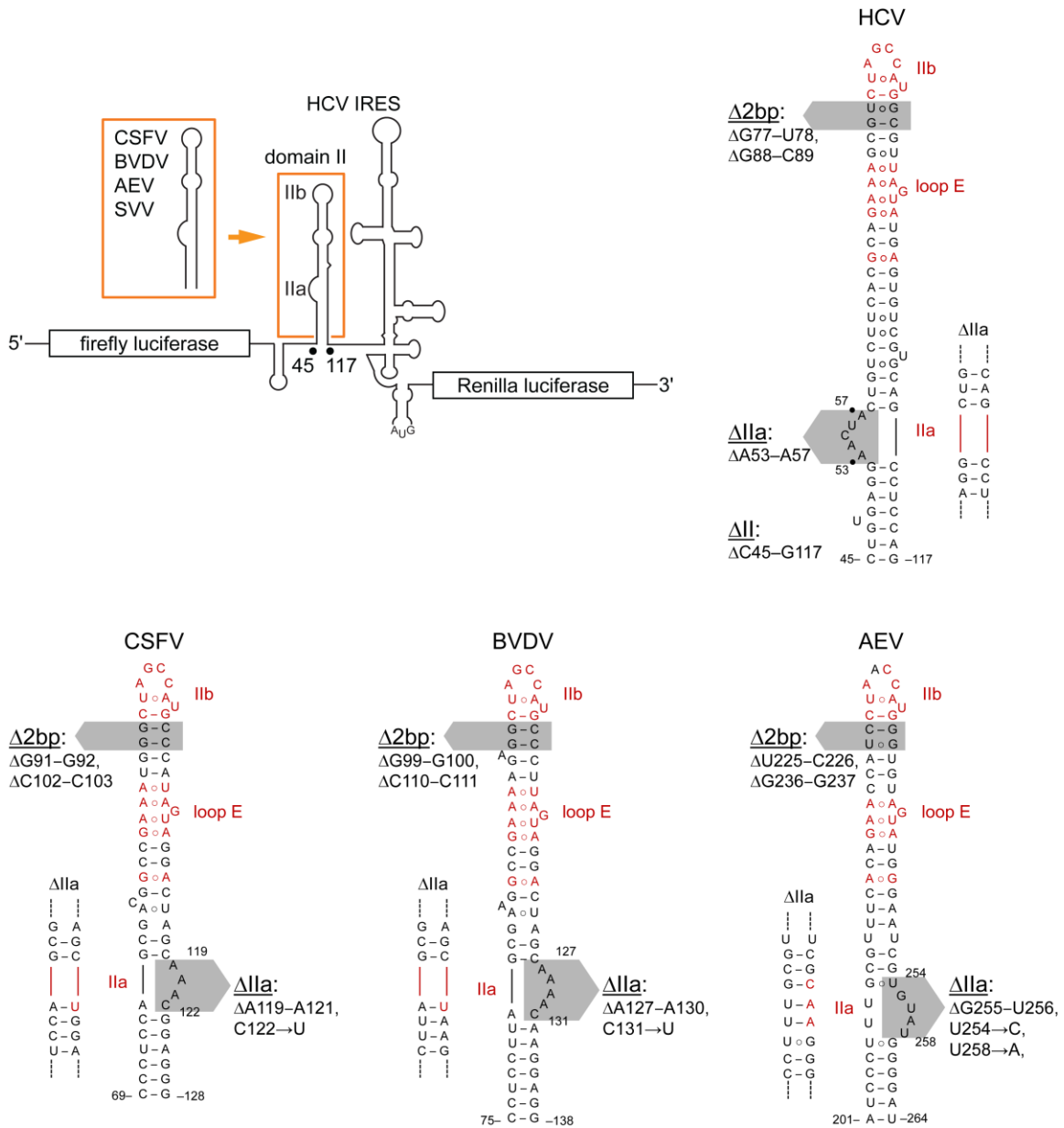
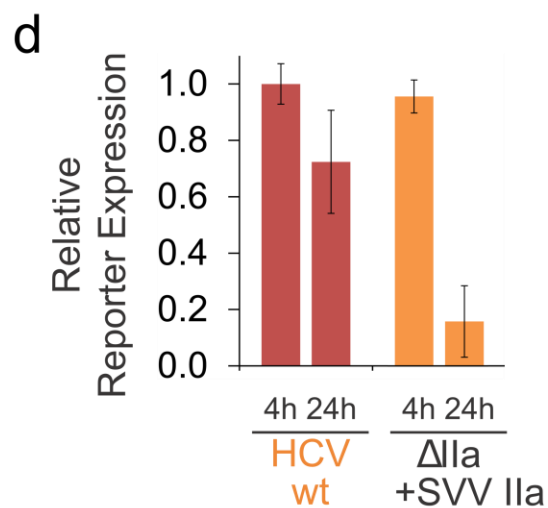
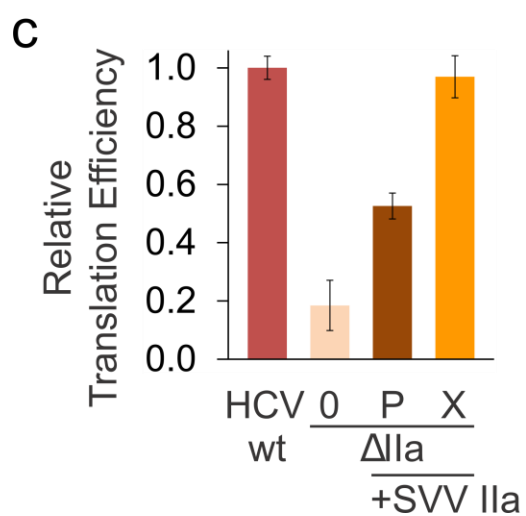
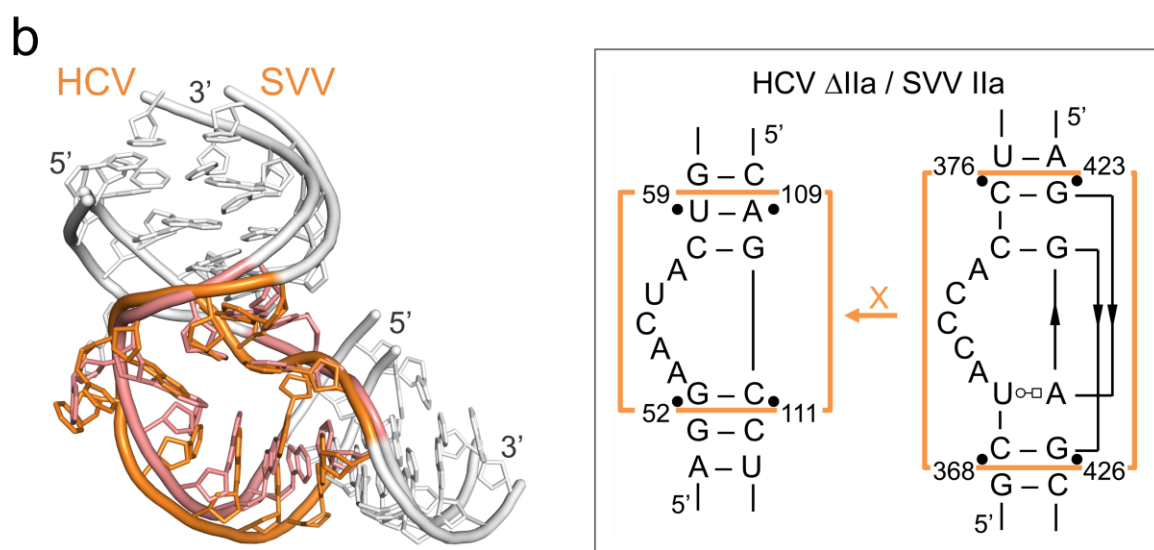
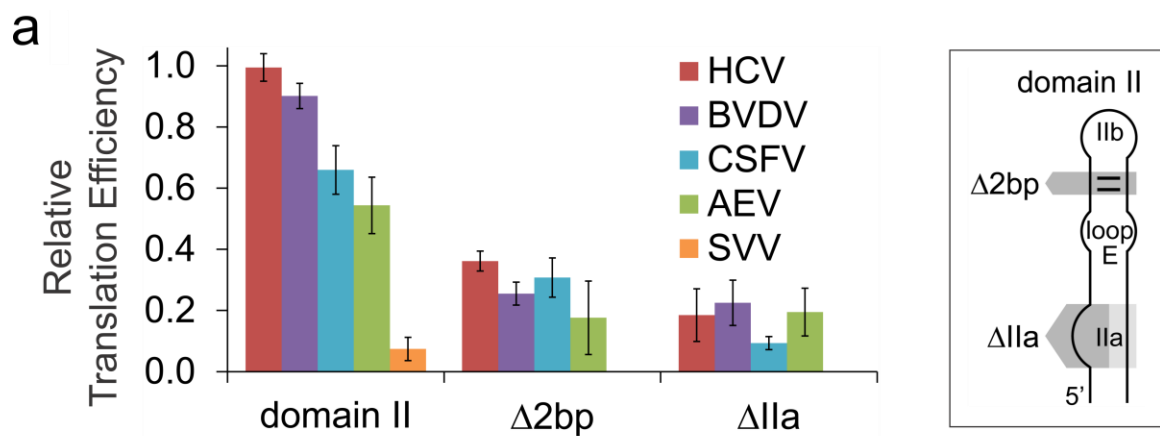


Figure 2.14: Structures and sequences of HCV IRES domain II chimera constructs. Bicistronic (dual) luciferase reporter constructs were used for the IVT experiments. Domain II was mutated or replaced by corresponding RNA motifs from other viruses as indicated. NCBI reference sequences: HCV (NC_004102), CSFV (NC_002657.1), BVDV (NC_001461.1), AEV (NC_003990.1), SVV (NC_011349.1).

Figure 2.15: Functional competence of chimera HCV IRES elements. a) Effect on *in vitro* translation of domain II replacement in the HCV IRES by corresponding motifs from the IRES elements of other viruses (see Figure 2.1b) and additional deletion of two base pairs between the hairpin loop IIb and the loop E motif ($\Delta 2\text{bp}$), or deletion of the internal loop IIa (ΔIIa). See Figure 2.14 for IRES domain II chimera structure and sequences. b) Construction of an IRES chimera X (orange) by an exact structural motif swap based on superimposition of the subdomain IIa crystal structures from HCV (red/white) and SVV (orange/white). c) Effect on *in vitro* translation of subdomain IIa replacement in the HCV IRES by the corresponding motif from SVV (chimera X, outlined in b). The chimera P (brown) is a control construct that contains a subdomain IIa swap with a one base pair offset relative to the X construct (Figure 2.17). The HCV IRES wt and subdomain IIa deletion (ΔIIa , 0) are shown as controls. Translation efficiencies were normalized to the cap driven expression in bicistronic dual reporter constructs. d) Function of the HCV/SVV IRES chimera X in replicon-transfected human cells. Reporter expression was measured 4h and 24h after transfection with the subgenomic replicon RNA carrying the IRES subdomain IIa replaced by the corresponding motif from SVV. Error bars represent $\pm 1\text{s.d.}$ calculated from triplicate experiments, except for panel d where triplicates of three biological replicates (9 values) were used.



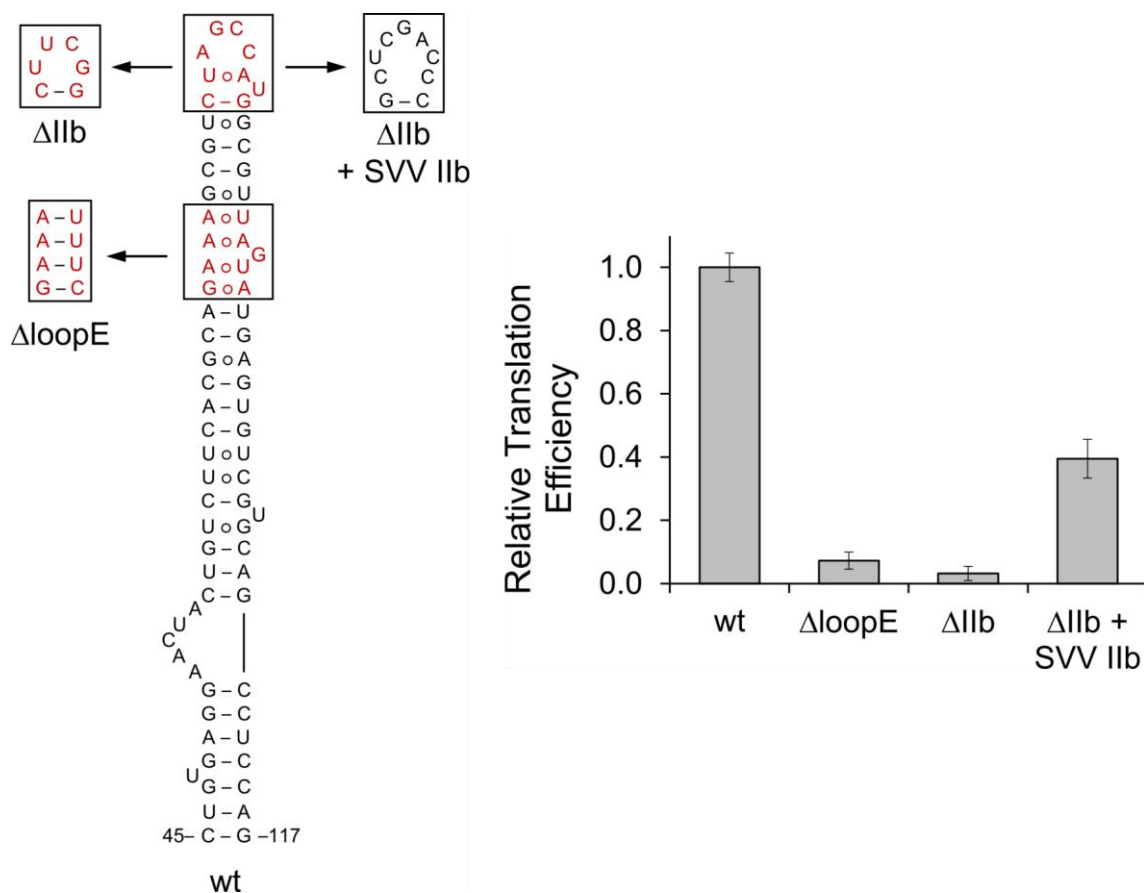


Figure 2.16: HCV IRES domain II mutations' effect on translation efficiency in an *in vitro* translation assay. The wt HCV IRES domain II is shown with mutation sequences. Δ loopE introduced mutations to the loop E motif to Watson-Crick (WC) base pair all wt residues on the 5' proximal strand of this motif. Δ Iib replaced the Iib hairpin with a stable tetraloop. Δ Iib + SVV Iib replaced the 9 residues of the wt Iib motif with the corresponding 9 residues from the wt SVV Iib motif. Translation efficiencies were normalized to the cap driven expression in bicistronic dual reporter constructs. Error bars represent \pm 1 s.d. calculated from triplicate experiments.

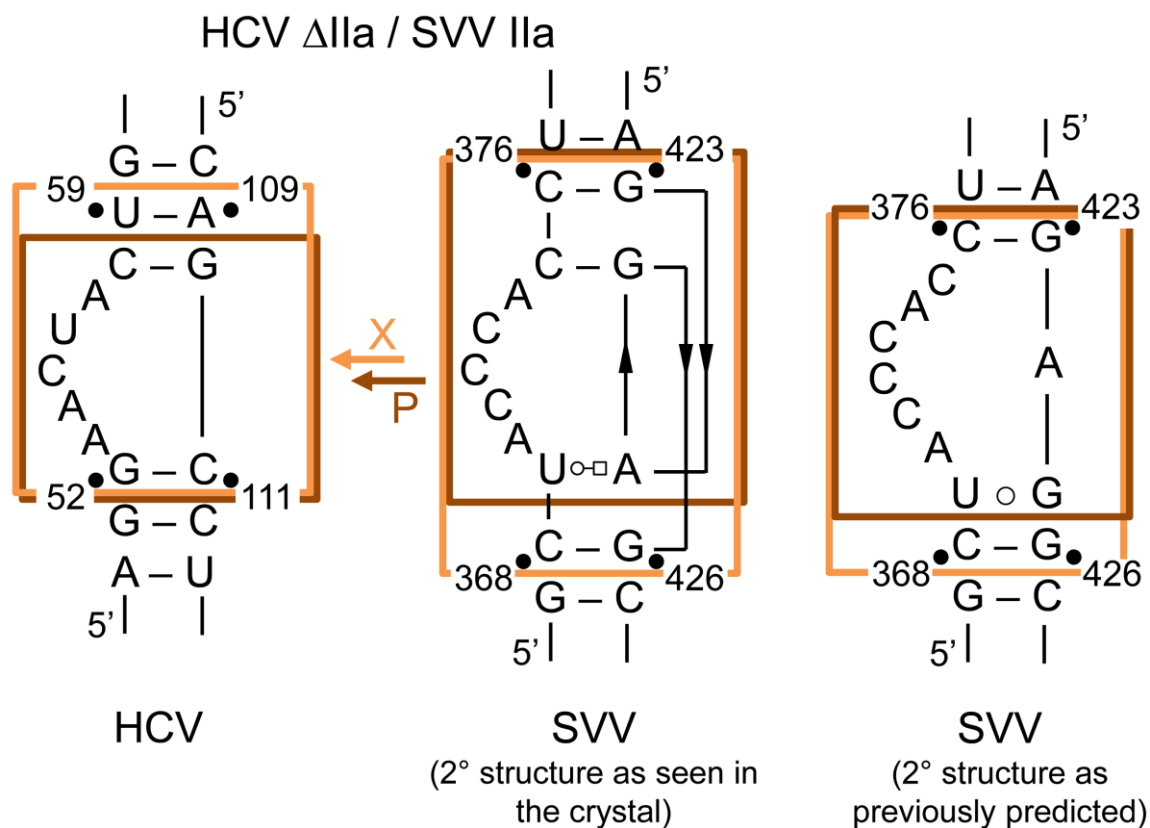


Figure 2.17: Construction of HCV IRES chimeras. The chimera construct X (orange) is an exact structural motif swap based on superimposition (Figure 2.13) of the subdomain Ila crystal structures from HCV (see Figure 2.1c) and SVV (see Figure 2.10a). The chimera P (brown) is a control construct that contains a subdomain Ila swap with a one base pair offset that is guided by the secondary structure of the RNAs without regard of the 3D structure information from crystallography. Two representations of the subdomain Ila secondary structure in SVV are shown.

The availability of a crystal structure for the SVV subdomain Ila allowed us to determine a minimal architectural module in this RNA that corresponded to the bend in the HCV motif. Guided by the superposition of RNA structures from both viruses, we substituted in the HCV IRES reporter construct the internal loop comprised of residues G52...U59/A109...C111 by the corresponding module of SVV consisting of C368...C376/G423...G426 (Figure 2.15b). While the swapped RNA modules shared no

sequence similarity, the resulting chimera IRES was fully proficient in translation initiation (Figure 2.15c, construct X). Even a small deviation from the structure-guided module swap, by introduction of a single base pair offset (construct P, Figure 2.17), resulted in a 50% loss in IRES activity (Figure 2.15c). The chimera construct X also conferred full translation activity of HCV replicon in human cells 4h after transfection when viral translation occurs but not yet replication (Figures 2.15d, 2.18).⁷⁷ However, expression activity from the chimera replicon was greatly diminished after 24h when the cumulative reporter signal is expected to be dominated by replication. Apparently, the motif swap of subdomain IIa modules negatively affected communication between the 5' and 3' UTR which has been implicated in viral replication.^{78,79}

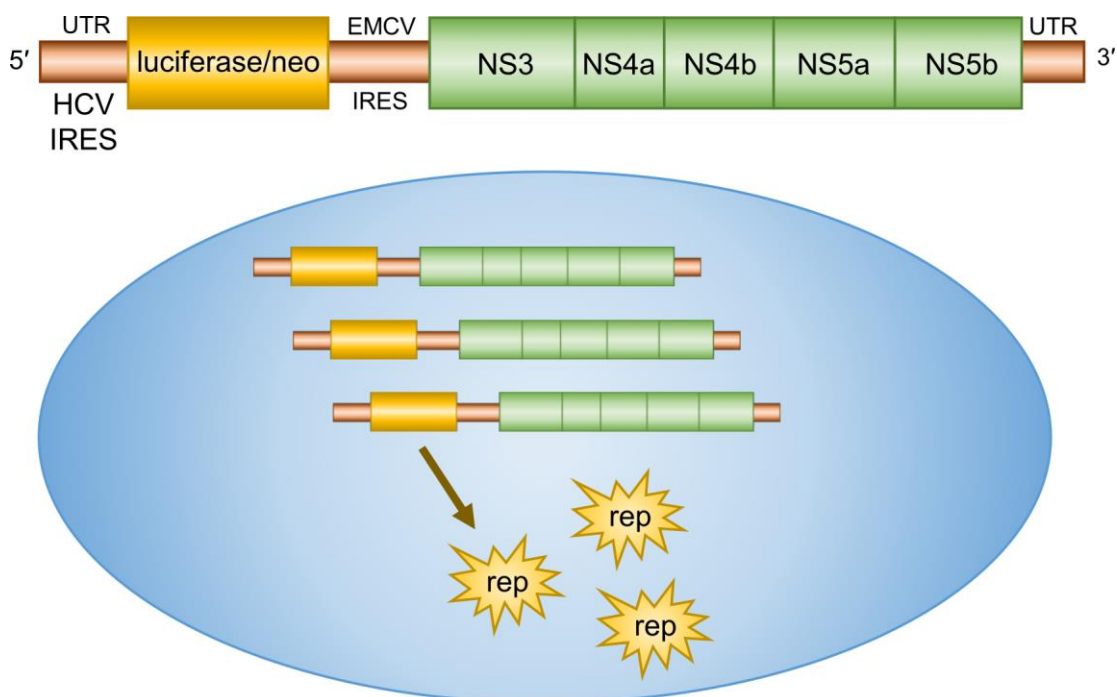


Figure 2.18: HCV replicon experiments in Huh-7.5.1 cells. SGR-JFH1 FEO⁸⁰ or Δ SVV-IRES-IIa³⁹ replicon RNA transcripts carrying an HCV IRES or HCV- Δ SVV-IIa IRES driven luciferase/neo fusion protein gene (firefly luciferase, neomycin phosphotransferase) replacing HCV structural protein genes and the EMCV (encephalomyocarditis virus) IRES driven HCV replication protein genes (NS3, 4a, 4b, 5a, and 5b) followed by the structured HCV 3' UTR. Replicon transcripts are transfected into cells and luciferase reporter (rep) expression is read at 4 hours post-transfection when viral translation occurs, but not yet replication, and 24 hours post-transfection when cumulative reporter expression is expected to be dominated by replication (Figure 2.15d). Competent replication requires HCV replication proteins and communication between the HCV 5' and 3' UTR. The SVV IIa swap in the HCV IRES possibly affects this 5'-3' UTR communication (Figure 2.15d).

Replacement of the subdomain IIa motif in HCV by the corresponding analog from the CSFV IRES was performed to investigate the effect of relocating the internal loop from the 5' to the 3' proximal strand of the domain II lower stem. Secondary structure-guided swapping of the internal loop modules did not result in a translation-proficient IRES. An active chimera was obtained only after shortening the distance by 5

base pairs between the subdomain IIa internal loop and the conserved loop E in combination with relocating a bulged-out pyrimidine from the 3' to the 5' proximal strand (Figure 2.19). These changes were motivated by the observation that in domain II analogs which carry the internal loop in the 3' proximal strand, the distance to the loop E motif is shorter by 5 base pairs, corresponding to a half turn of an A-form RNA helix. A bulged-out pyrimidine, separated by 4 base pairs above the subdomain IIa internal loop, is found consistently in the strand opposite to the internal loop (Figure 2.1b). Similarly, a translation-competent IRES containing a subdomain IIa motif replacement from AEV, was obtained only after replacing a 9 base pair segment between the subdomain IIa internal loop and the conserved loop E of HCV with a 4 base pair segment from AEV which also contained a bulged-out pyrimidine (Figure 2.20). The AEV subdomain IIa replacement was guided by a revised domain II secondary structure prediction (pred 2, Figure 2.20) from the secondary structure shown in Figure 2.1b based on a previous prediction.⁵⁶ This previous prediction shows a subdomain IIa which differs from both the HCV/SVV (5' IIa internal loop) and BVDV/CSFV (3' IIa internal loop) subdomain IIa classes. This AEV subdomain IIa contains an internal loop with unpaired bases in both the 5' and 3' proximal strands, a 7 base pair segment between IIa and loop E (not the typical 5 or 10 base pairs), and no bulged out pyrimidine 4 base pairs above and opposite the internal loop. The revised domain II secondary structure prediction reveals an AEV subdomain IIa internal loop more similar to BVDV/CSFV, which still contains internal loop unpaired bases in both the 5' and 3' proximal strands, but which is located 5 base pairs away from the loop E motif and contains a bulged out pyrimidine on the opposite strand 4 base pairs above and opposite the internal loop (Figure 2.20). This revised

secondary structure prediction for AEV IRES IIa was further tested by making replacements of the subdomain IIa motif in the HCV Δ dII + BVDV dII chimera by corresponding AEV IIa predictions (Figure 2.21; see also Chapter 3, Figure 3.5). This revised AEV IIa prediction (pred 2) had similar activity in the HCV/BVDV chimera construct as it did in the HCV construct, which was significantly higher than two additional AEV IIa predictions arranged to be more like CSFV (pred 3) or BVDV (pred 4) (Figure 2.21). Changing the length of the lower domain II helix had various effects on domain II function in the five viral IRES elements, and demonstrated a specific size range which is required for this domain II base stem (Figure 2.23). These combined observations from AEV and BVDV as well as HCV, SVV, and CSFV provided the basis for a topological model of the modular domain II architecture in viral IRES elements (Figure 2.22).

Figure 2.19: Effect on translation efficiency of subdomain IIa replacement in the HCV IRES by the corresponding motif from CSFV, and mutants thereof, as measured in an *in vitro* translation assay. The structure of CSFV chimera constructs is outlined on the top. The HCV wt and CSFV wt chimeras are shown as controls (see Figure 2.15a). Translation efficiencies were normalized to the cap driven expression in bicistronic dual reporter constructs. Error bars represent \pm 1s.d. calculated from triplicate experiments.

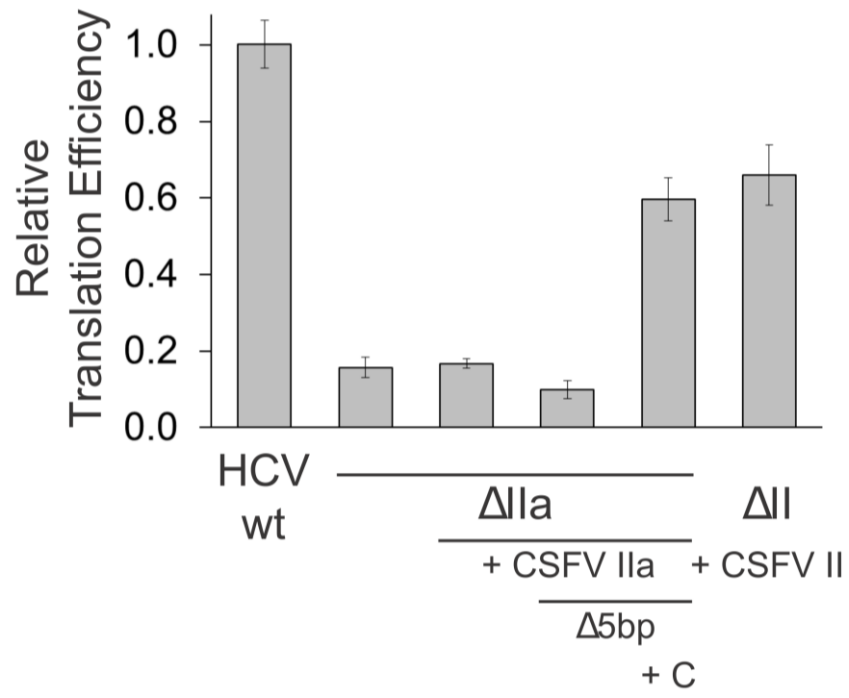
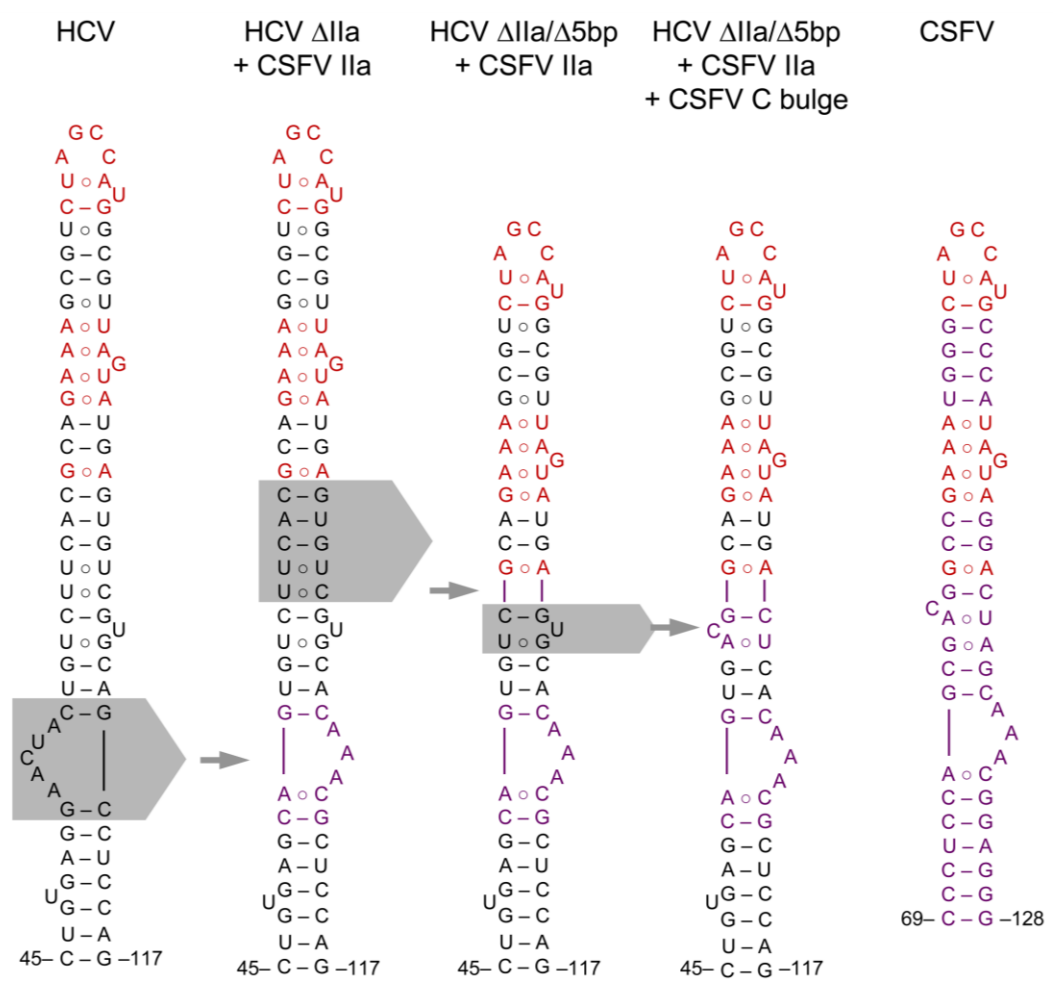


Figure 2.20: Effect on translation efficiency of subdomain IIa replacement in the HCV IRES by the corresponding motif from AEV, and mutants thereof, as measured in an *in vitro* translation assay. The structure of AEV chimera constructs is outlined on the top. Conserved sequences in both HCV and AEV are shown in red above the horizontal grey line. The HCV wt and AEV wt chimeras are shown as controls (see Figure 2.15a). AEV domain II secondary structure prediction 1 (pred 1) is based on a previous prediction⁵⁶ and secondary structure prediction 2 (pred 2) is revised based on the developed model in Figure 2.7a and the CSFV experiments in Figure 2.19. Translation efficiencies were normalized to the cap driven expression in bicistronic dual reporter constructs. Error bars represent \pm 1s.d. calculated from triplicate experiments.

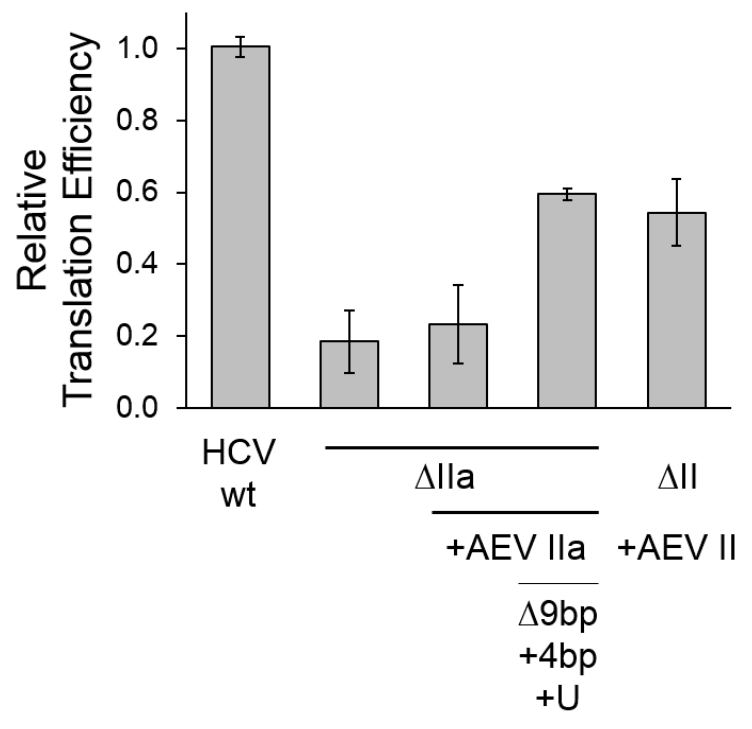
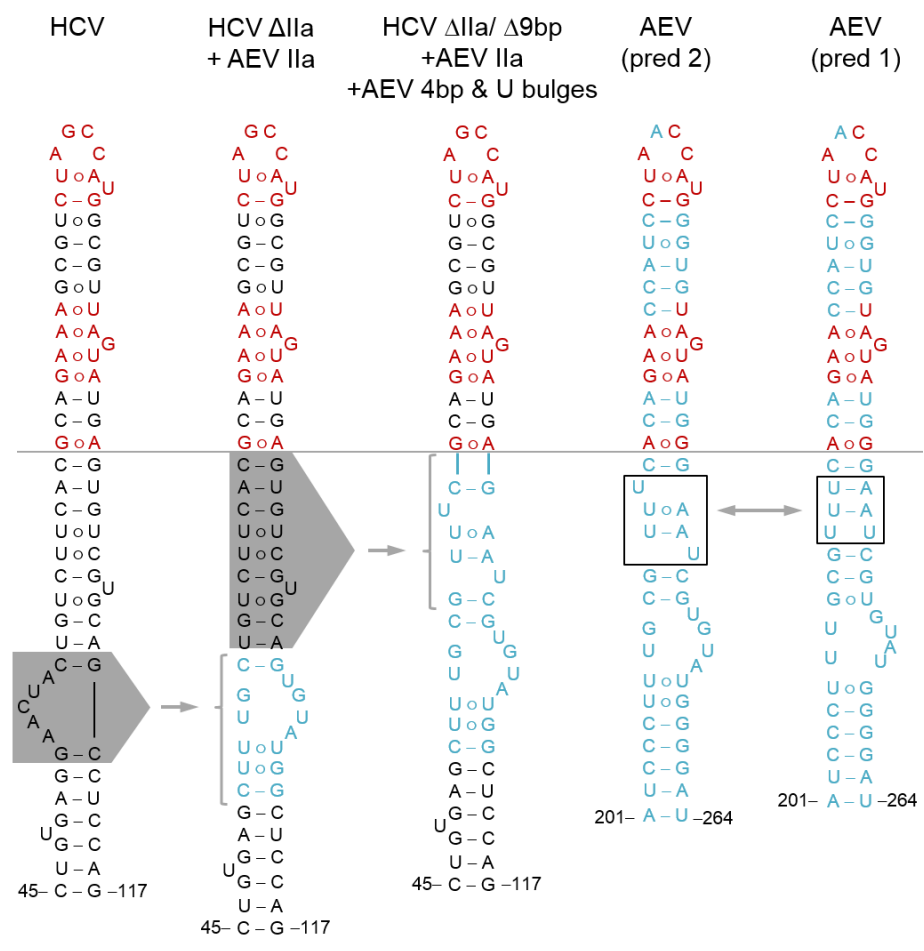
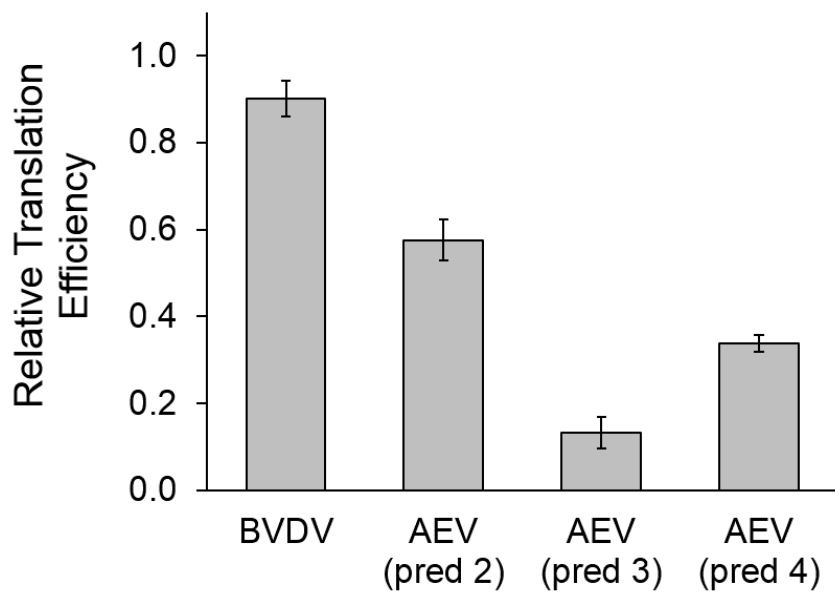
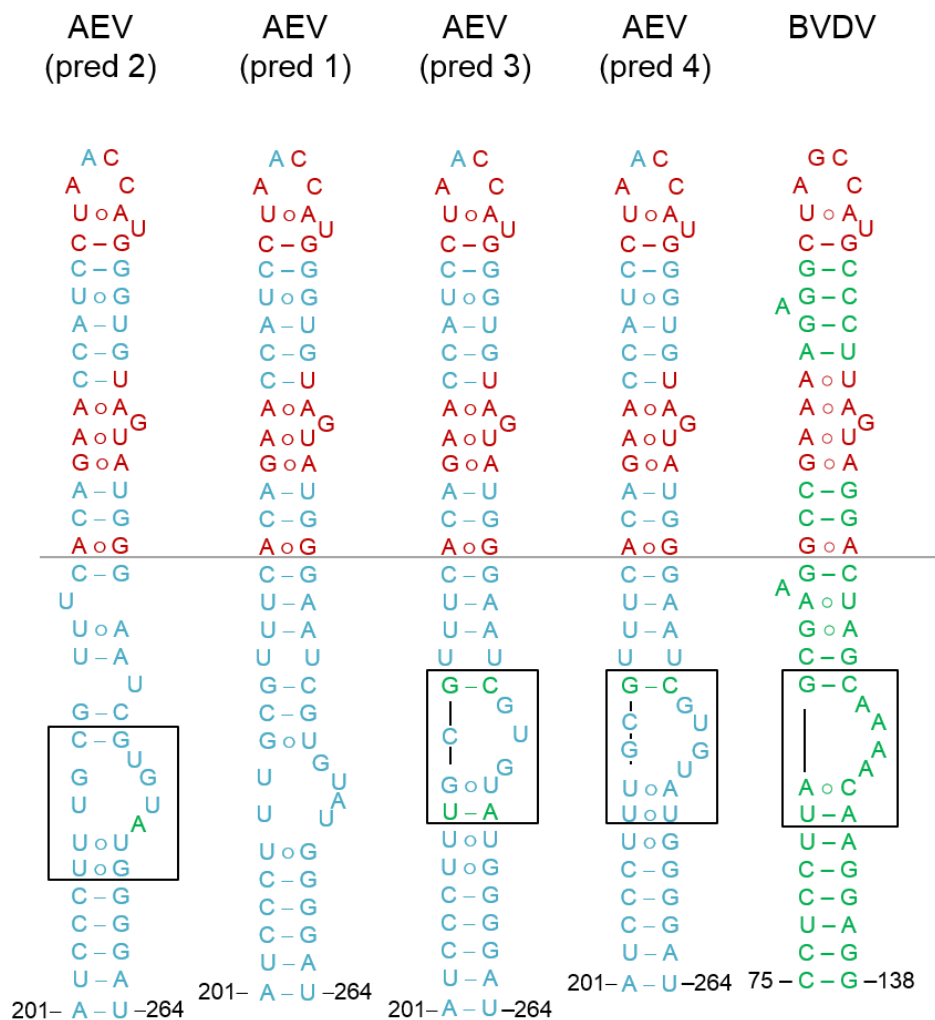


Figure 2.21: Effect on translation efficiency of subdomain IIa replacement in the HCV Δ dII + BVDV dII chimera IRES by the predicted corresponding motif from AEV as measured in an *in vitro* translation assay. The structure of AEV domain predictions is outlined on the top. Conserved sequences in HCV, BVDV, and AEV are highlighted in red above the horizontal grey line. AEV domain II prediction 1 (pred 1) is shown as a reference. AEV IIa predictions 2, 3, and 4 are boxed and these sequences were used to replace the corresponding boxed BVDV IIa. Conserved BVDV sequences in AEV are shown in green. The Δ dII + BVDV dII chimera (BVDV) is shown as a control (see also Figure 2.15a). Prediction 2 has the highest activity in both the context of the modified HCV IRES and the HCV Δ dII + BVDV dII chimera IRES. Translation efficiencies were normalized to the cap driven expression in bicistronic dual reporter constructs and the wt HCV signal was normalized to 1 (not shown on graph). Error bars represent \pm 1s.d. calculated from triplicate experiments.



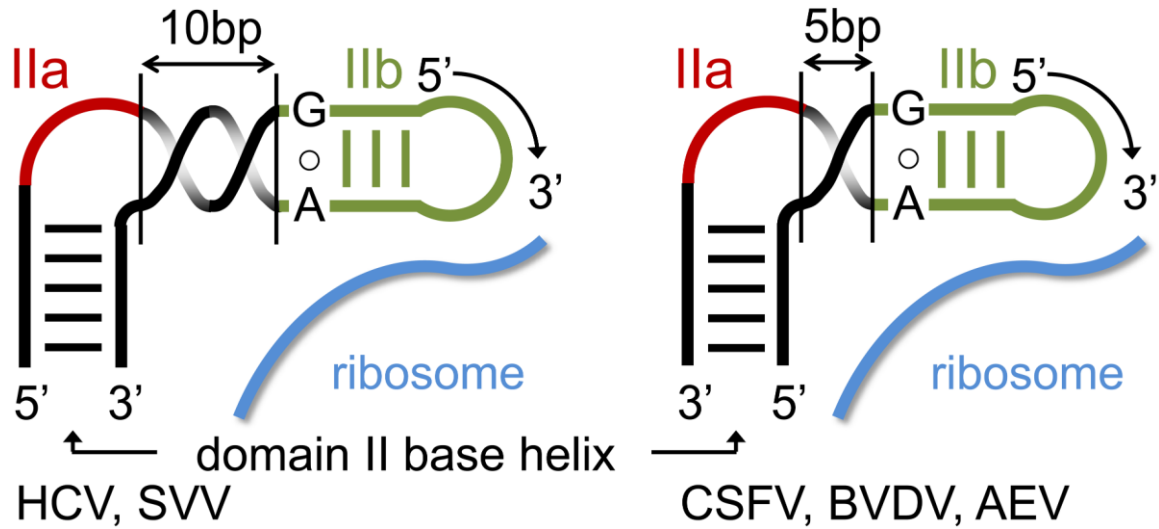


Figure 2.22: Model for the modular domain II architecture in the IRES elements of RNA viruses. For the correct positioning of the conserved hairpin loop IIb at the 40S ribosomal subunit, the bent internal loop IIa module has to be located in the 5' or 3' strand of the base helix, depending on the number of base pairs in the spacer segment between the subdomains IIa and IIb. See also Figure 2.1 where the conserved G∘A pair is highlighted in red. A spacer of 10 base pairs, as in the IRES elements of HCV and SVV, corresponds to a full turn of an A-form RNA helix while a spacer of 5 base pairs, as in the IRES elements of CSFV, BVDV, and AEV, provides a half turn.

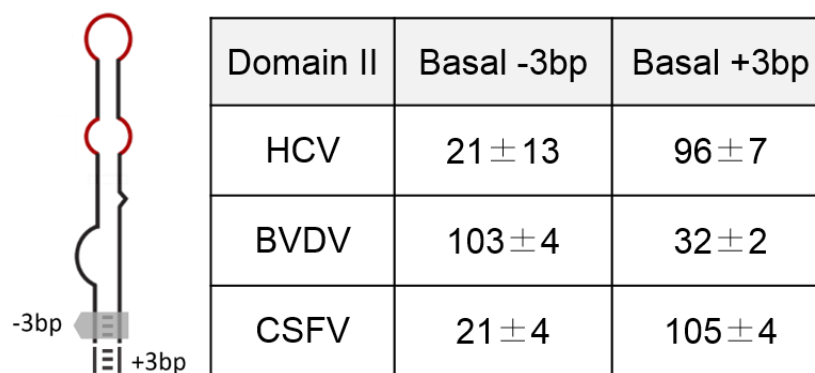


Figure 2.23: Functional competence of chimera HCV IRES with domain II elements shortened or lengthened by 3 base pairs (-3bp, +3bp) in the basal helix. Numerical values represent the average percentage of wt translational activity. Error represents \pm 1s.d. calculated from triplicate experiments.

Discussion of viral switches captured by a guanosine

The highly conserved subdomain IIa in the HCV IRES had previously been established as an RNA conformational switch that is able to convert from a ligand-free bent state to a ligand-bound extended conformation.⁶ The bent RNA fold is stabilized by magnesium ions and ensures the correct positioning of the IRES on the 40S ribosomal subunit and docking of domain II at the E site.^{34,51} Removal of domain II from the ribosome during translation initiation may be topologically achieved by a conformational change in subdomain IIa transitioning from the bent to the extended state.²¹ Recent cryo-EM studies confirm the conformational flexibility of domain II through comparison of 40S-bound binary and 80S-bound initiation complexes of the HCV IRES.⁴² Motivated by the discovery of a riboswitch-like binding site for benzimidazole translation inhibitors⁵⁴ in the extended form of subdomain IIa,²¹ we hypothesized that a cognate ligand may serve as the biological actuator of the RNA switch that ultimately facilitates ribosomal release from the IRES. Inspection of the crystal structure previously determined for an

inhibitor complex of subdomain IIa suggested that benzimidazole derivatives might be fortuitous ligands that exploit a recognition site for arginine or a guanine base. Here, we used a FRET assay to demonstrate binding and capture of the subdomain IIa RNA in an extended conformation by guanine and structurally related derivatives. Guanine binding to the subdomain IIa target led to inhibition of IRES-driven translation albeit requiring a ~300-fold higher concentration compared to potent benzimidazole derivatives. Structural comparison shows that the guanine base may engage in similar interactions with subdomain IIa as those observed for benzimidazole inhibitors which engage in hydrogen bonding at the Hoogsteen edge of the G110-C58 base pair (Figures 2.1e, 2.2a). The resulting arrangement corresponds to the interaction of guanosine with a Watson-Crick G-C pair in a geometry that accounts for one of the most frequently occurring base triples in RNA folds.⁷⁰

Modeling of guanine in the ligand binding pocket of the benzimidazole-RNA crystal structure revealed suboptimal stacking by the nucleobase on neighboring residues (Figure 2.24). Better stacking is achieved with the larger quinazolinone **4** which has a ~2-fold higher affinity for the RNA target than guanine (Figure 2.2b). The smaller size in combination with the neutral character of the ligand may explain the weaker binding affinity of guanine compared to the benzimidazole inhibitors. The nucleobase as a putative biological ligand of the subdomain IIa RNA switch may function as a transient actuator that readily dissociates from the IRES, unlike the benzimidazole inhibitors, which arrest the IRES topology and thereby inhibit translation initiation. It is conceivable that a guanosine in an RNA sequence such as ribosomal RNA or the viral genome emerging from the ribosome may serve as a trigger facilitating IRES translation initiation

and release from the ribosome. Docking of IRES fragment crystal structures to a cryo-EM map of an IRES-40S ribosomal complex revealed a cluster of guanines from the central domain pseudoknot (III_f, Figure 2.1a) and domain IV (Figure 2.25) residing in a single-stranded stretch upstream of the viral initiation codon and located in close proximity to subdomain IIa.³¹ Involvement of a guanine from the viral genome as a trigger for the subdomain IIa switch would capitalize on a locally high concentration of the ligand as a part of the same RNA strand. In contrast, participation of cellular guanine or guanosine is unlikely due to their low physiological concentration of around 97 μM and 0.9 μM, respectively.⁸¹ Guanosine triphosphate (GTP), while present in human cells at a higher level (~300 μM), is a less suitable ligand for the IIa target due to electrostatic repulsion by the phosphate groups. GTP along with GMP (monophosphate) were found to have no activity in FRET binding experiments.

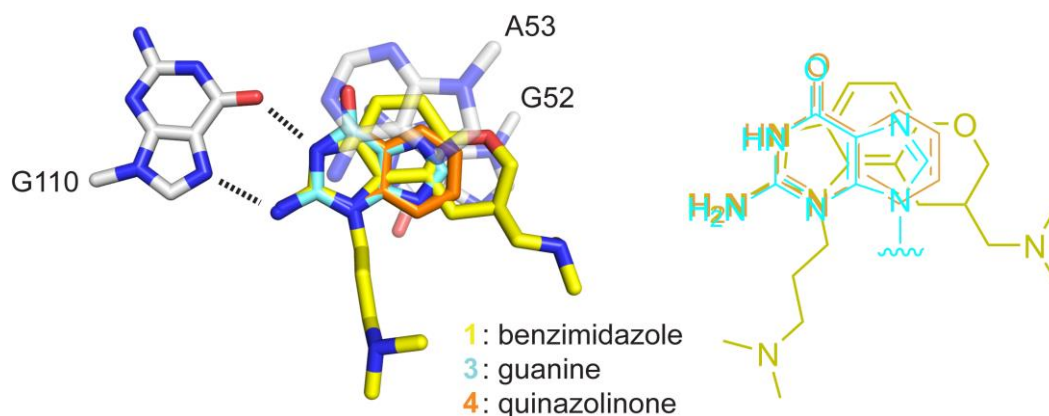


Figure 2.24: Modeling of guanine **3** and quinazolinone **4** by superimposition on the benzimidazole **1** in the ligand binding site of the subdomain IIa complex crystal structure. Neighboring stacking bases are shown along with the docking site at G110. A two-dimensional overlay shown on the right.

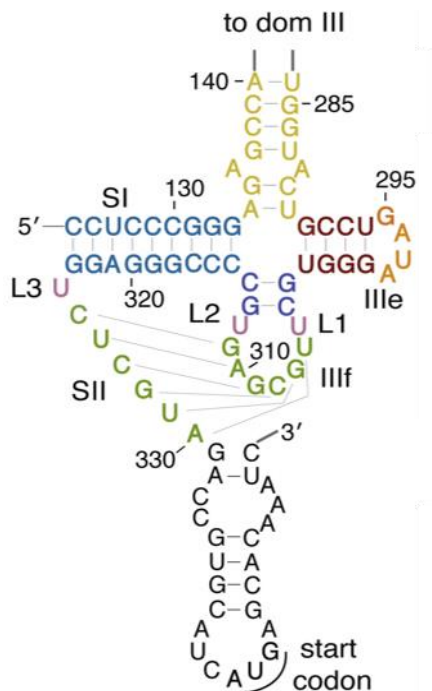


Figure 2.25: HCV IRES central domain secondary structure showing a pseudoknot in domain III and the domain IV hairpin holding the AUG start codon in black. Figure is adapted from ref³¹, 2011, Elsevier Ltd.

Regulation of the HCV IRES activity by the subdomain IIa switch through interaction in *cis* with a guanosine from domain IV would provide an efficient autoregulatory mechanism in which the trigger G residue is sequestered in a hairpin loop of the unbound IRES to allow ribosome assembly supported by the bent IIa RNA motif. Placement of the viral start codon at the decoding site requires melting of the domain IV hairpin which in turn unmasks the trigger G residue that facilitates IRES release during initiation through capture of the IIa extended state (Figure 2.26).

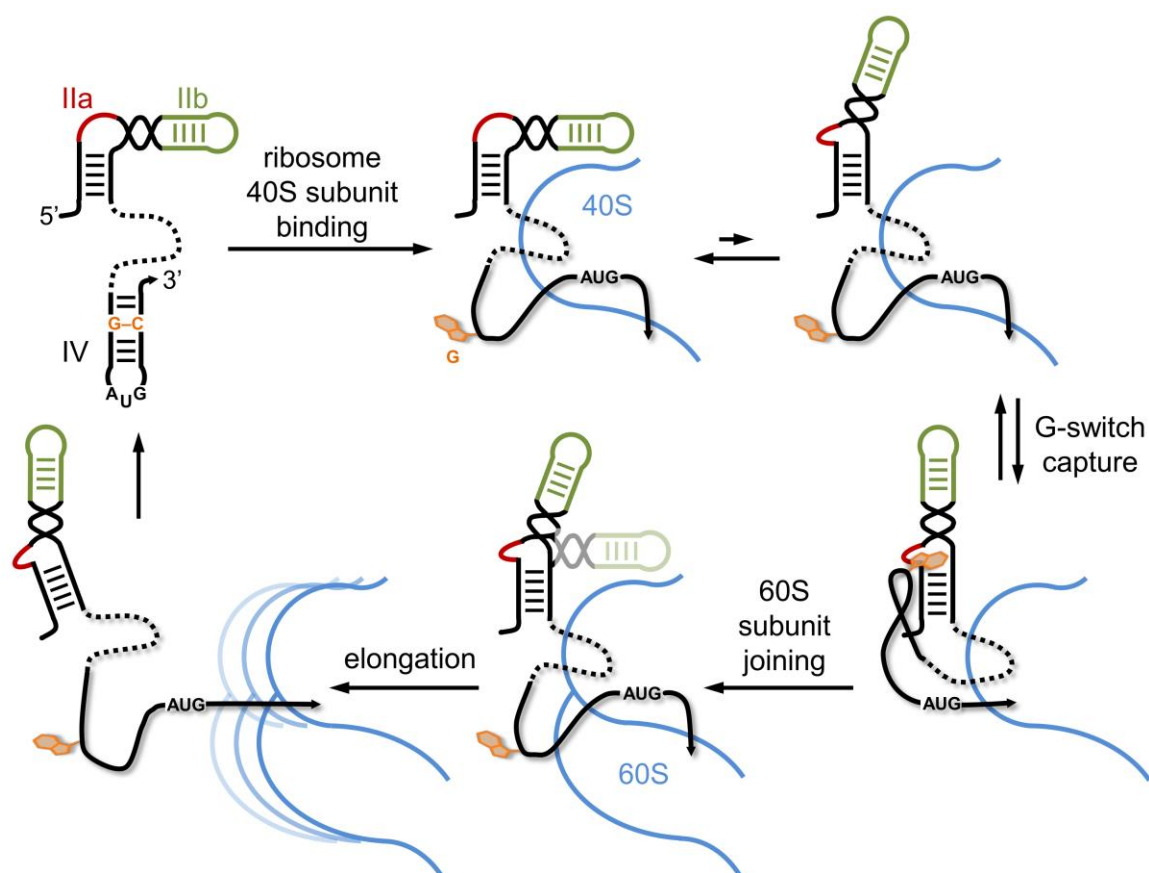


Figure 2.26: Proposed mechanism for the regulation of IRES activity by the subdomain Iia switch through an interaction with a G residue from domain IV. Ribosome binding of the IRES entails melting of the domain IV hairpin to provide access of the viral start codon to the decoding site, which in turn unmasks the trigger G residue that facilitates functional IRES translation initiation by capture of the Iia extended state. After the ribosome transitions to elongation, the unbound IRES may refold to the original state.

Attempts were made to identify the location of this guanosine trigger. A sixteen nucleotide long oligonucleotide sequence from the G-containing domain IV of the HCV IRES (Figure 2.25), which comes in close proximity to domain II when bound to the 40S ribosomal subunit, beginning at the end of the domain III pseudoknot and ending 1 residue after the start codon (5'-AGA CCG UGC AUC AUG A-3') was synthesized and tested for activity in the FRET assay but did not show increased binding over control

sequences. RNA and DNA oligonucleotides 5 to 18 nucleotides in length representing various domain IV sequences along with control sequences lacking G residues were also tested but gave inconclusive results. Attempts to characterize interactions between domain IIa and domain IV sequences (including a G residue) were also attempted by electrophoretic mobility shift assay, but no RNA-RNA interactions were observed. Mutagenesis studies using the *in vitro* translation assay were also inconclusive. G to C mutations to domain IV residues G331, G335, and G337 with the corresponding paired C to G mutations to maintain domain IV hairpin base pairs all resulted in IRES elements with less than 50% of wt translation activity (G331C/C354G: ~10%; G335C/C349G: ~40%; G337C/C347G: ~40%). These mutations in combination with a mutation to the guanosine recognizing base pair of the subdomain IIa switch (C58G/G110C) and which could result in a G-C \circ C base triple interaction (versus the wt C-G \circ G) did not result in translation rescue. Attempts were made to co-crystallize several domain IV RNA oligonucleotide sequences (6, 10, and 16 nucleotides long) with two different previously crystallized HCV IIa crystal constructs,^{21,47} but no diffracting crystals were obtained.

The discovery of the IRES subdomain IIa as a ligand-responsive RNA switch in HCV raised the question if corresponding motifs in other viruses adopt analogous folds whose conformational state may be captured by ligand binding. Here, we demonstrated that RNA motifs from several flavi- and picornaviruses fold into magnesium-stabilized bent architectures similar to the HCV archetype despite their sequence dissimilarity (Figure 2.7). Earlier NMR studies of domain II from CSFV performed at lower salt concentration already showed a bent structure for this RNA.⁴⁴ The crystal structure of the internal loop motif from SVV (Figure 2.10) revealed a fold that is overall identical to the

HCV subdomain IIa despite that the two motifs share little sequence or local secondary structure similarity (Figure 2.17). In addition to the resemblance of static structure, analogous subdomain IIa RNAs from other viral IRES elements interact with the benzimidazole **1** as well as guanine to be captured in an extended conformation (Figure 2.9b, c). Binding of **1** to the subdomain IIa RNAs from other viruses was ~10-fold weaker than the affinity measured for the HCV target, perhaps reflecting the extensive optimization of the benzimidazole derivative for inhibition of the HCV IRES.^{54,82}

Discussion of functional similar switches in five viruses despite sequence dissimilarity

The analogies between subdomain IIa RNA motifs from various viruses extend beyond the similarity of static structure and the ability to adopt two distinct conformational states. Domain swap experiments in which crystal structure information was used to precisely replace subdomain IIa in the HCV IRES by the corresponding motif from SVV demonstrated that biological function is completely conserved between these intrinsically distinct RNA building blocks (Figure 2.15b, c). HCV-SVV chimera IRES elements were fully functional both *in vitro* and in replicon-infected cells. Even the structurally less accurate replacement of the whole domain II in the HCV RNA with analogous motifs from BVDV, CSFV and AEV yielded functional chimera IRES elements (Figure 2.15a). By inference from these findings, we conclude that subdomain IIa motifs of greatly distinct sequence and local secondary structure may serve as functionally conserved RNA conformational switches that are involved in viral IRES-driven translation and may be captured by highly similar or even identical ligands. We

propose that the biological capture ligand shared between the different viral RNA motifs is likely a guanosine.

Conclusions

The functionally conserved subdomain IIa motifs constitute a new paradigm for ligand-captured RNA switches that differ from metabolite-sensing riboswitches with regard to their small size as well as intrinsic stability and structural definition of the constitutive conformational states. They require no large secondary structure rearrangements in order to facilitate functional translation initiation. These viral RNA modules represent the simplest form of ligand-responsive mechanical switches in nucleic acids. They do not inherently serve as “on” or “off” switches for viral translation, though they may be exploited as such by synthetic inhibitors which lock them in one of two states required for function. This requirement for a dynamic and conformationally changing RNA is no new phenomena, but rather the aspect that this small and simple RNA conformational switch of four to six unpaired nucleotides may be stabilized by a ligand in two mechanically different states (same secondary structure but different tertiary structure) and is not just moved around by protein and nucleic acid factors due to its inherent flexibility. This type of RNA switching motif which is able to be transiently captured by a ligand (nucleic acid residue, protein residue, or other metabolite) may also be utilized in other places in biology. It is conceivable that other viral or cellular functional RNA elements use similar ligand-responsive switches which may be temporarily displaced for certain structural or catalytic functions. While identifying these switches is currently difficult due to their small size, their lack of sequence conservation,

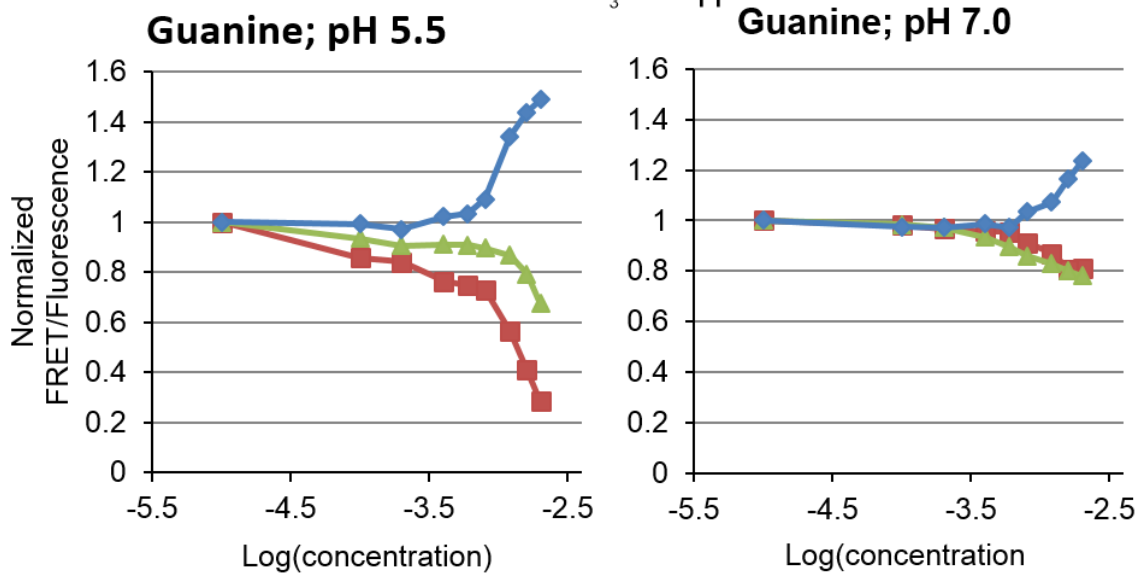
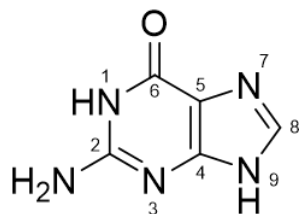
and a limited number of available 3D RNA structures, future methods may better enable their discovery and characterization leading to additional therapeutic opportunities.

Extended discussion of guanine-like molecules

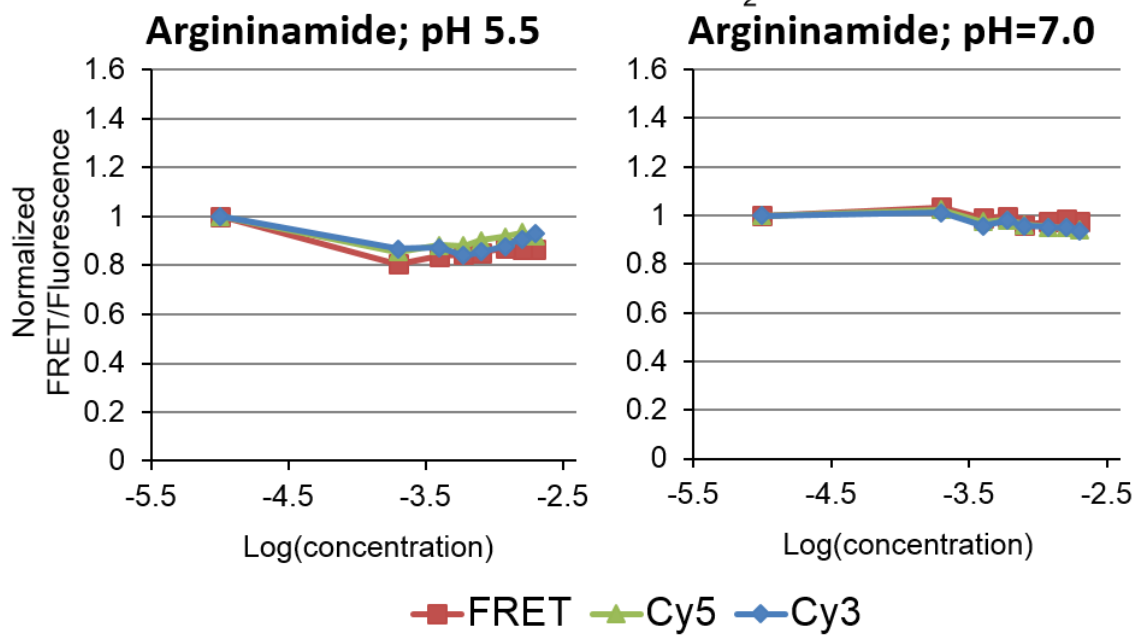
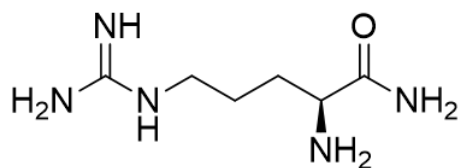
As mentioned previously in this chapter, neither arginine nor its derivatives showed binding to the IIa RNA FRET construct (example shown in Figure 2.27), and a list of tested derivatives is shown in Figure 2.28. In contrast, guanine does bind and capture the extended conformation of the HCV subdomain IIa switch. After the discovery of guanine binding in the HCV IIa RNA FRET assay, it was found that pH has an effect on the binding of “guanine-like” molecules (see Figure 2.27). A slightly acidic pH of 5.5 enhanced binding of these molecules, and this can be seen for guanine in Figure 2.27. This is in contrast to 2-amino-benzimidazole compounds which bind with similar affinities at pH 5.5 or 7.0 (data not shown). This may be due to a requirement for the N1 of guanine (Figure 2.27) to be protonated to bind the IIa RNA. A more acidic pH is required as this nitrogen is less basic than the corresponding nitrogen of the 2-amino-benzimidazole scaffold (Figure 2.2a).

Figure 2.27: pH dependence of guanine-like scaffolds binding affinity for the HCV Ila target. a) FRET binding of guanine at pH 5.5 or 7.0. At pH 5.5, a dose-dependent decrease in FRET and dose-dependent increase in Cy3 emission correspond to a higher affinity (lower EC_{50}) than at pH 7.0. Guanine's purine core is numbered for reference. b) FRET binding of argininamide at pH 5.5 or 7.0 reveals no affinity for the HCV Ila target. All data points shown are the average of triplicate experiments.

a



b



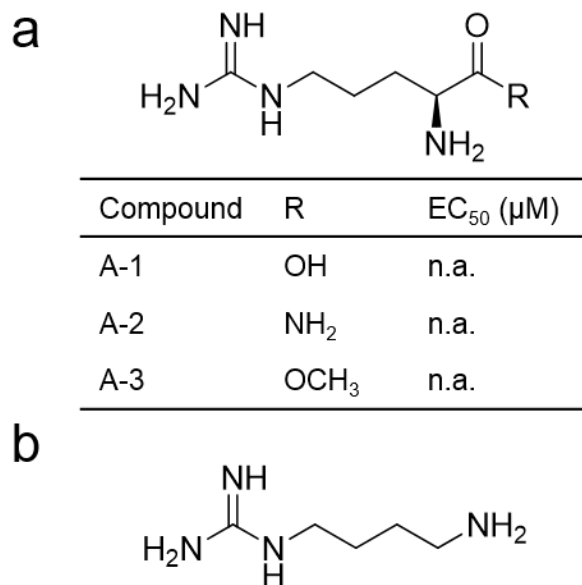


Figure 2.28: Arginine and derivative testing in the FRET assay. a) Activity of L-arginine (**A-1**) and derivatives L-argininamide (**A-2**) and L-arginine methyl ester (**A-3**) in the FRET assay. None of these compounds had any effect in the FRET assay and are listed as having no activity (n.a.) in the EC₅₀ column. b) Agmatine, an arginine derivative replacing arginine’s formic acid (CH₂OOH) tail with a hydrogen, and more closely resembling RNA-binding polyamines, was also tested and found to have no activity for the HCV Ila target. Several other guanidynylated (guanidine-containing) compounds were tested (not shown) and found to have no activity for the HCV Ila target.

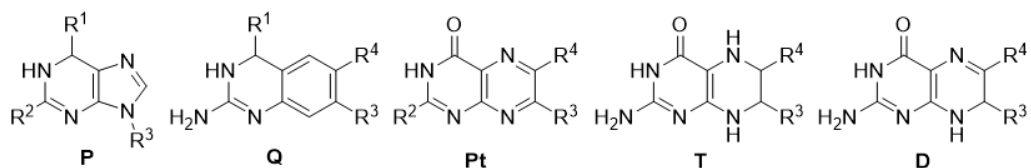
Derivatives of guanine and other 2-aminopyrimidinone-containing molecules were tested for their affinity for the HCV Ila target in the FRET assay. Six purine derivatives (**P-1** to **P-6**, Tables 2.3, 2.4; Figures 2.29-2.34), eleven 2-amino-quinazoline derivatives (**Q-1** to **Q-11**, Tables 2.3, 2.5, 2.6; Figures 2.35-2.45), six pteridine derivatives (**Pt-1** to **Pt-6**, Tables 2.3, 2.7; Figures 2.46-2.51), two tetrahydropteridine derivatives (**T-1** to **T-2**, Tables 2.3, 2.8; Figures 2.52-2.53), and two dihydropteridine derivatives (**D-1** to **D-2**, Tables 2.3, 2.9; Figures 2.54-2.55) comprising five different scaffolds were tested in the FRET assay. Many of these “guanine-like” molecules are

important intermediates, byproducts, cofactors, and therapeutics in biological processes and are readily commercially available. Xanthine (**P-4**) is involved in the purine degradation pathway.⁸³ Aciclovir (**P-6**) is an acyclic guanosine analog and an important antiviral therapeutic used to treat a number of viral infections including herpesvirus.⁸⁴ Folic acid (**Pt-3**) is involved in uridine monophosphate (UMP) and methionine metabolism.⁸⁵ L-sepiapterin (**D-1**) is involved in folic acid biosynthesis.⁸⁶ Lumazine (**Pt-4**) is involved in riboflavin biosynthesis.⁸⁷ Tetrahydrobiopterin (THB; **T-1**) is an essential cofactor of phenylalanine, tyrosine, and tryptophan hydroxylase enzymes.⁸⁶ Neopterin (**Pt-2**) is involved in THB metabolism and isoxanthopterin (**D-2**) is involved in the THB degradation pathway.⁸⁵ These biologically important molecules, several other commercially available compounds, along with several other compounds synthesized in the Hermann lab (**Q-5** to **Q-11**)^{88,89} were tested for binding to the HCV IRES IIa target in the FRET assay (Figures 2.29-2.55). Many of these derivatives had very low affinities for the target and many dose response experiments did not result in saturating concentrations allowing EC₅₀ values for ligand binding to be calculated. In many of these cases ligand binding specific to the target was still observed as FRET widening. FRET widening involves the capture of the target RNA in an elongated conformation as evidenced by a dose-dependent decrease in FRET and increase in Cy3 emission. Several significant conclusions as well as important structure-activity relationships were drawn from this study. First, the guanidine moiety is essential for binding, and any changes to the 2-amino (**R**², Table 2.3) substituent result in non-active compounds. A carbonyl group at the 6 position (**R**¹, Table 2.3) was important for all binding molecules, though primary or tertiary amines (dimethyl-amino) were tolerated at this position on purine (**P**) or 2-amino-

quinaoline (**Q**) scaffolds, albeit twofold weaker affinity. The most active of these 2-amino, 6-carbonyl-substituted scaffolds is the 2-amino-quinazolinone (**Q-1**, Figures 2.4c, 2.35). Various substitutions at R³ and R⁴ positions on all 5 scaffolds were tolerated but only several resulted in higher activity compounds. Notably, a ribose substitution at R³ of the purine scaffold (**P-5**, guanosine, Figure 2.33) resulted in a twofold increase in activity over guanine, but an acyclic guanosine derivative (**P-6**, acyclovir, Figure 2.34) had a fourfold weaker activity. Also notable, was THB (**T-1**, Figure 2.52), a tetrahydropteridine derivative which had an activity only slightly lower than the active 2-amino-quinazolinone. Ultimately, the 5-membered imidazole ring of guanine may be replaced by larger 6-membered benzene (as in the quinazolines) or pyrazine (as in the pteridines) rings which may also participate in favorable stacking interactions in the HCV Ila binding pocket. Initial efforts to substitute at the R³ position of the 2-amino-quinazolinone core with phenyl or aniline substituents resulted in compounds which all precipitated the FRET RNA before reaching saturating concentrations allowing EC₅₀ values for ligand binding to be calculated. Dimethyl-amino-propyl linked substituents at the N1 or N6 of the 2-amino-quinazolinone core (**Q-10**, **11**; Table 2.6, Figures 2.44-2.45), resembling a helpful substituent of the model 2-aminobenzimidazole compound, did not result in compounds with greater activity.⁸⁹ While these structure-activity relationships emerging from the guanosine cognate ligand provide a useful starting point towards viral inhibitor design, further synthetic efforts will be needed to develop higher affinity compounds with these or similar scaffolds with better drug-like qualities. Attempts were made to co-crystallize **P-1**, **Q-1**, and **T-1** with the HCV Ila RNA, as these

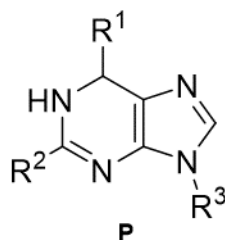
were the highest binding molecules from each respective scaffold, but these attempts yielded no highly diffracting crystals.

Table 2.3: Summary of activity of “guanine-like” small molecules, including purine derivatives **P-1** to **P-6**, 2-amino-quinazoline derivatives **Q-1** to **Q-9**, pteridine derivatives **Pt-1** to **Pt-5**, tetrahydropteridine derivatives **T-1** to **T-2**, and dihydropteridine derivatives **D-1** to **D-2** in the FRET assay experiments shown in Figures 2.29-2.43, 2.46-2.55. EC_{50} values are calculated from the ligand binding curve of the experiment listed in parenthesis (FRET, Cy3) and are listed with ± 1 SD calculated from triplicate experiments. No activity (n.a.) is listed for compounds which had no effect in the FRET assay. Compounds which precipitated RNA at any concentration or before reaching concentrations sufficient to calculate EC_{50} values of ligand binding are labelled (precip.).



Compound	R ¹	R ²	R ³	R ⁴	EC ₅₀
P-1	=O	NH ₂	H		1100 ± 100 (Cy3)
P-2	NH ₂	NH ₂	H		> 2000 (FRET)
P-3	N(CH ₃) ₂	NH ₂	H		> 2000 (FRET)
P-4	=O	=O	H		n.a.
P-5	=O	NH ₂			430 ± 340 (FRET)
P-6	=O	NH ₂			1600 ± 4100 (FRET)
Q-1	=O		H	H	480 ± 120 (Cy3)
Q-2	NH ₂		H	H	810 ± 2000 (Cy3)
Q-3	H		H	H	precip.
Q-4	NH ₂		H	NH ₂	precip.
Q-5	=O		Cl	H	> 2000 (FRET)
Q-6	=O			H	precip.
Q-7	=O		C ₆ H ₅	H	precip.
Q-8	=O			H	precip.
Q-9	=O			H	precip.
Pt-1		NH ₂	H	H	> 2000 (Cy3)
Pt-2		NH ₂	H		n.a.
Pt-3		NH ₂	H		> 2000 (FRET)
Pt-4		OH	H	H	n.a.
Pt-5		OH		see R ³	precip.
Pt-6	NH ₂	NH ₂	H	CH ₂ OH	precip.
T-1			H		640 ± 50 (FRET)
T-2			CH ₃	CH ₃	1800 ± 2000 (FRET)
D-1			H		> 2000 (FRET)
D-2			=O	H	n.a.

Table 2.4: Summary of activity of purine derivatives **P-1** to **P-6** in the FRET assay experiments shown in Figures 2.29-2.34. EC_{50} values are calculated from the ligand binding curve of the experiment listed in parenthesis (FRET, Cy3) and are listed with ± 1 SD calculated from triplicate experiments. No activity (n.a.) is listed for P-4 which had no effect in the FRET assay.



Compound	R ¹	R ²	R ³	EC_{50} (μ M)
P-1	=O	NH ₂	H	1100 \pm 100 (Cy3)
P-2	NH ₂	NH ₂	H	> 2000 (FRET)
P-3	N(CH ₃) ₂	NH ₂	H	> 2000 (FRET)
P-4	=O	=O	H	n.a.
P-5	=O	NH ₂		430 \pm 340 (FRET)
P-6	=O	NH ₂		1600 \pm 4100 (FRET)

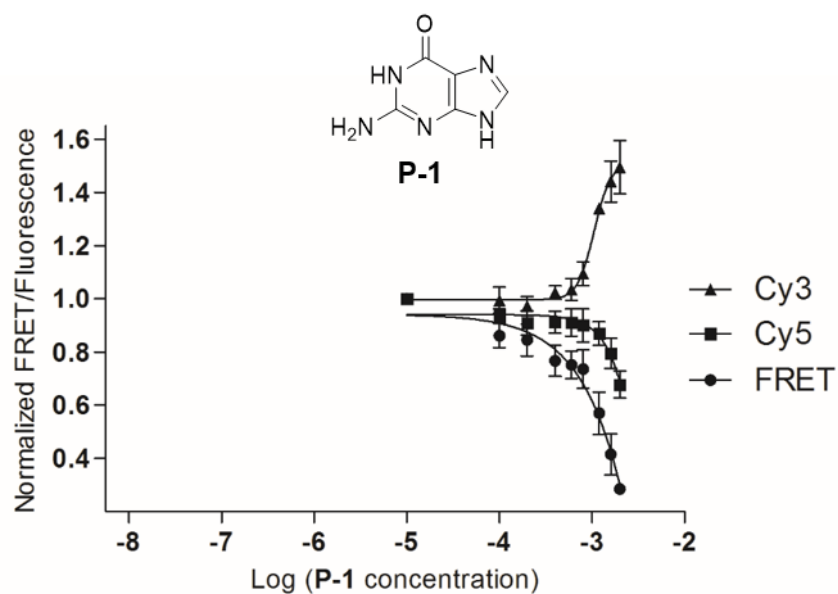


Figure 2.29: P-1 (guanine) titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. Fitting of a single-site binding curve to Cy3 emission (FRET signal did not reach saturation) gave an EC_{50} value for ligand binding of $1100 \pm 100 \mu\text{M}$.

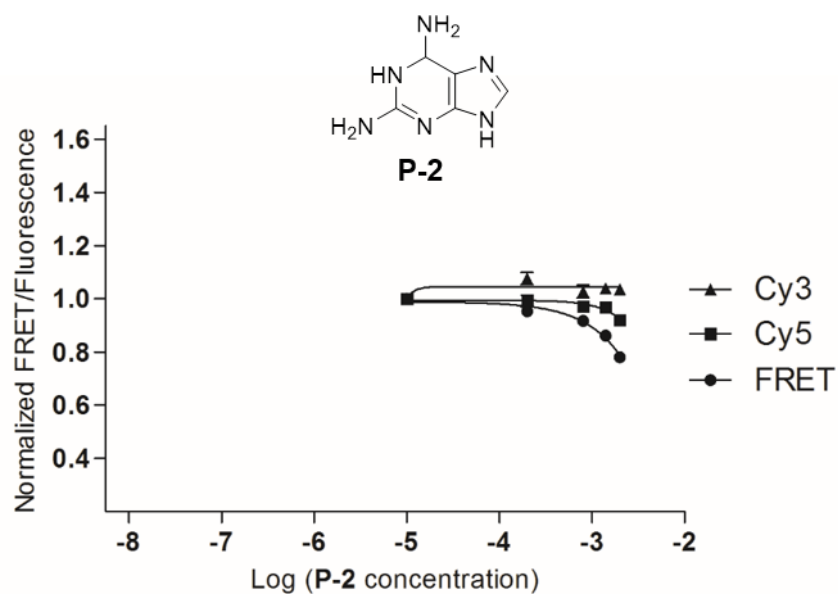


Figure 2.30: P-2 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. Although FRET or Cy3 signals did not reach saturation, FRET widening is observed and an EC_{50} value for ligand binding $> 2000 \mu\text{M}$ is estimated.

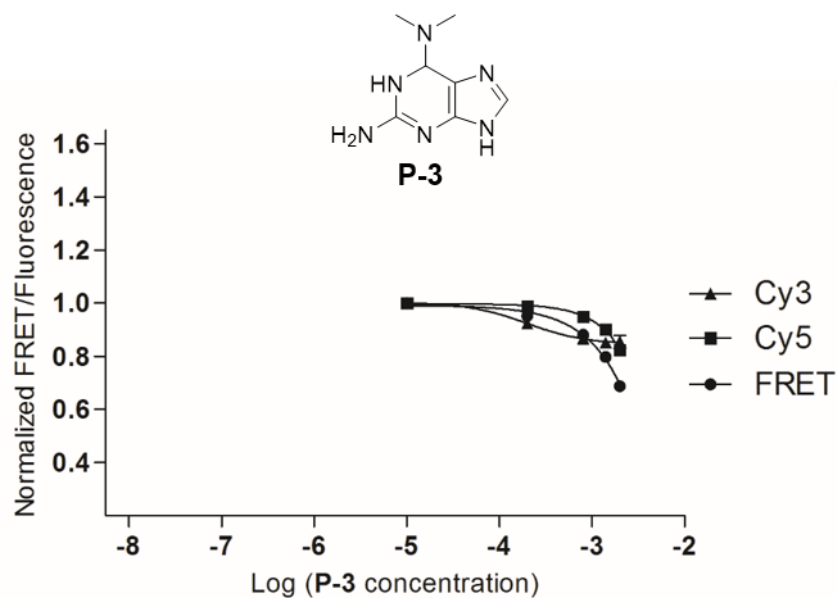


Figure 2.31: P-3 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. Although FRET or Cy3 signals did not reach saturation, FRET widening is observed and an EC_{50} value for ligand binding $> 2000 \mu\text{M}$ is estimated.

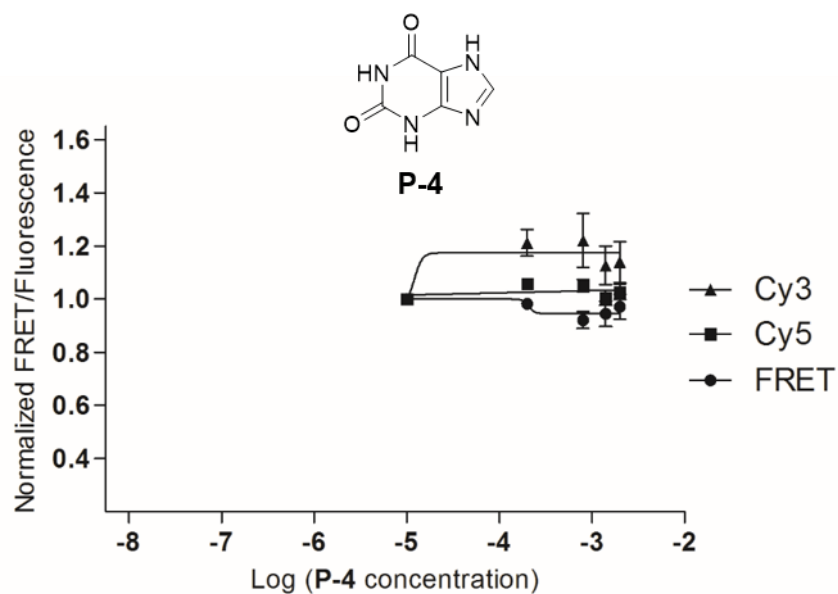


Figure 2.32: P-4 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. No compound activity is observed.

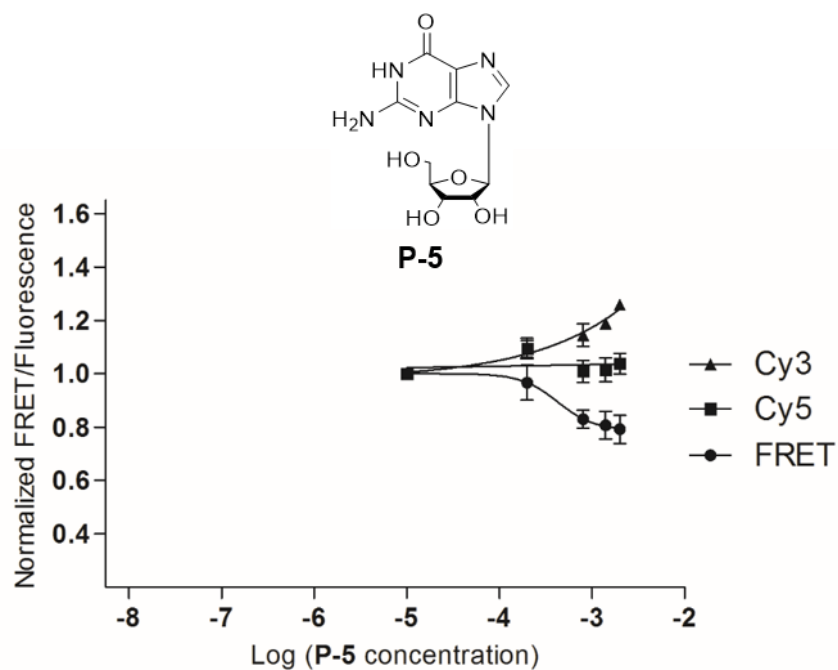


Figure 2.33: P-5 (guanosine) titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. Fitting of a single-site binding curve to FRET response gave an EC_{50} value for ligand binding of $430 \pm 340 \mu\text{M}$.

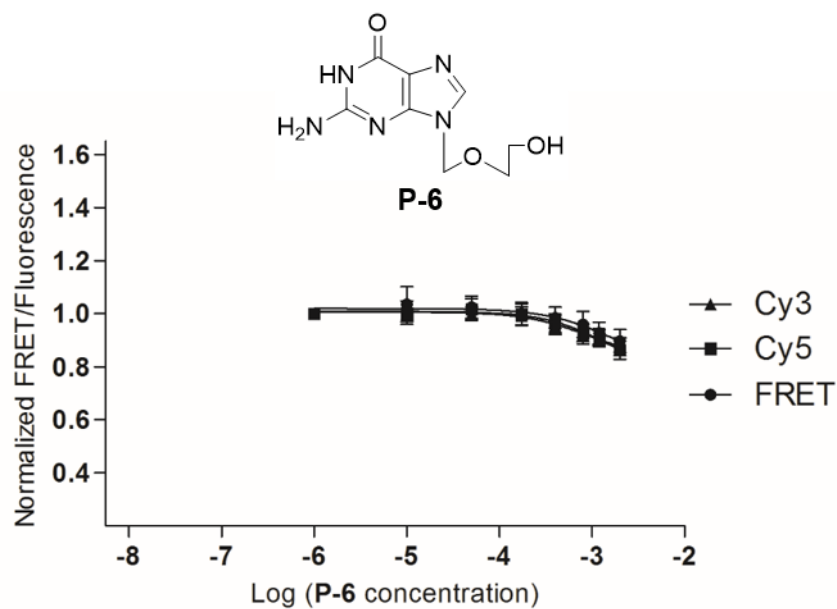
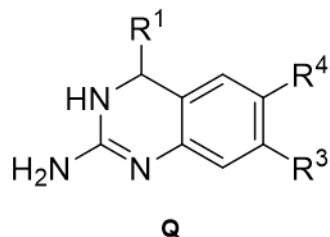
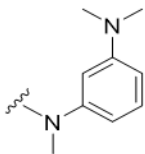
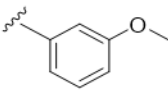
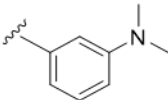


Figure 2.34: P-6 (acyclovir) titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. Although FRET or Cy3 signals did not reach saturation, possible slight FRET widening is observed and an EC_{50} value for ligand binding of $1600 \pm 4100 \mu\text{M}$ is estimated using the FRET response curve.

Table 2.5: Summary of activity of 2-amino-quinazoline derivatives **Q-1** to **Q-9** in the FRET assay experiments shown in Figures 2.35-2.43. EC_{50} values are calculated from the ligand binding curve of the experiment listed in parenthesis (FRET, Cy3) and are listed with ± 1 SD calculated from triplicate experiments. Compounds which precipitated RNA at any concentration are labelled (precip.).



Compound	R ¹	R ³	R ⁴	EC_{50} (μ M)
Q-1	=O	H	H	480 \pm 120 (Cy3)
Q-2	NH ₂	H	H	810 \pm 2000 (Cy3)
Q-3	H	H	H	precip.
Q-4	NH ₂	H	NH ₂	precip.
Q-5	=O	Cl	H	> 2000 (FRET)
Q-6	=O		H	precip.
Q-7	=O	C ₆ H ₅	H	precip.
Q-8	=O		H	precip.
Q-9	=O		H	precip.

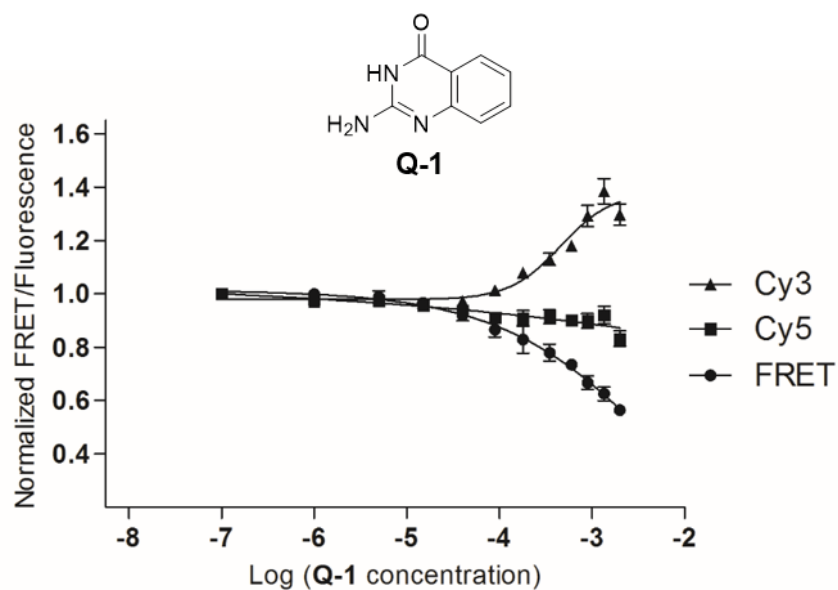


Figure 2.35: Q-1 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. Fitting of a single-site binding curve to Cy3 emission (FRET signal did not reach saturation) gave an EC_{50} value for ligand binding of $480 \pm 120 \mu\text{M}$.

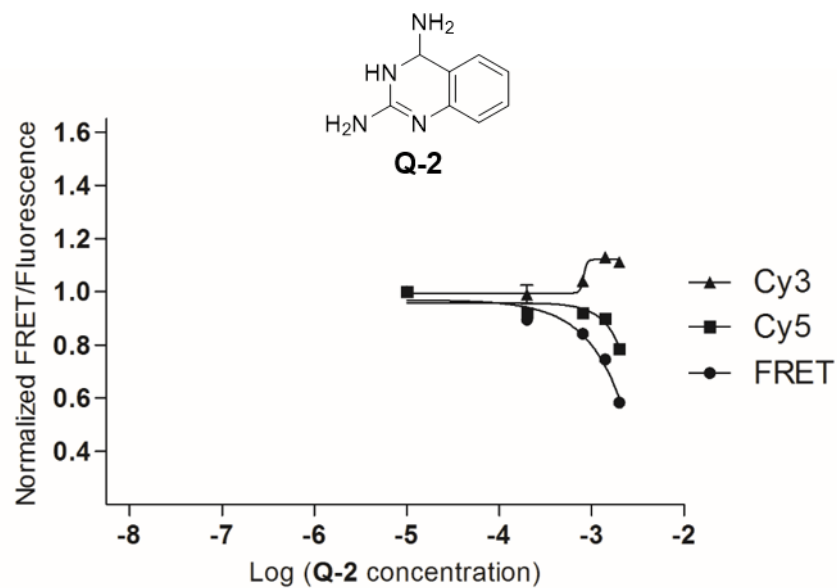


Figure 2.36: Q-2 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. Fitting of a single-site binding curve to Cy3 emission (FRET signal did not reach saturation) gave an estimated EC_{50} value for ligand binding of $810 \pm 2000 \mu\text{M}$.

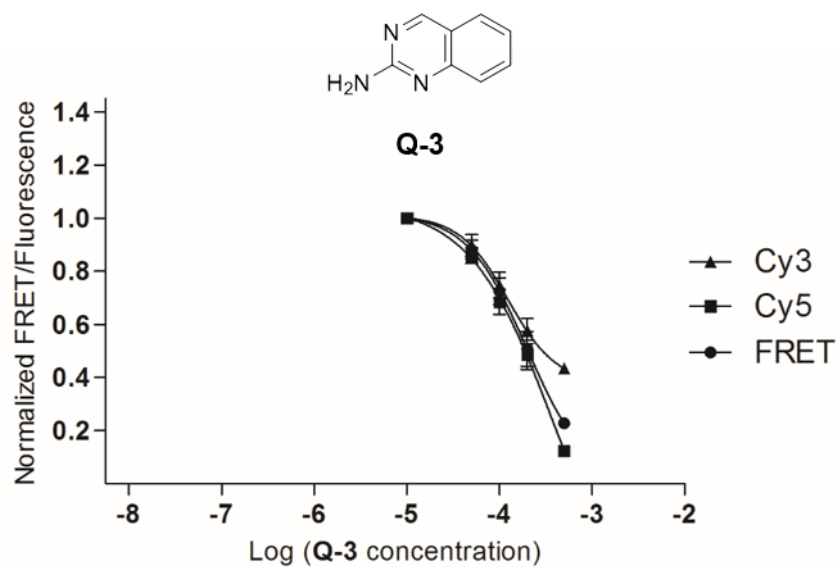


Figure 2.37: Q-3 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. An observed decrease in Cy3, Cy5, and FRET signals may be due to non-specific interactions between Q-3 and the RNA resulting in precipitation at all Q-3 concentrations.

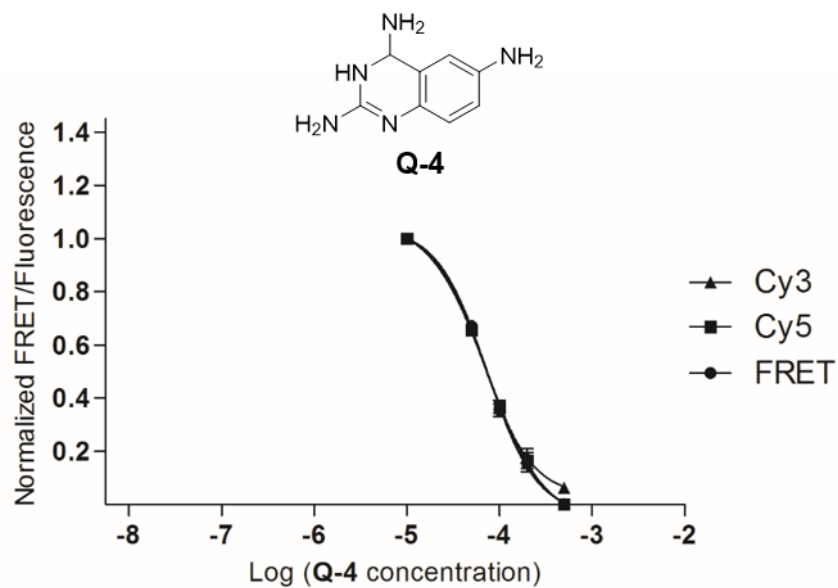


Figure 2.38: Q-4 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. An observed decrease in Cy3, Cy5, and FRET signals may be due to non-specific interactions between Q-4 and the RNA resulting in precipitation at all Q-4 concentrations.

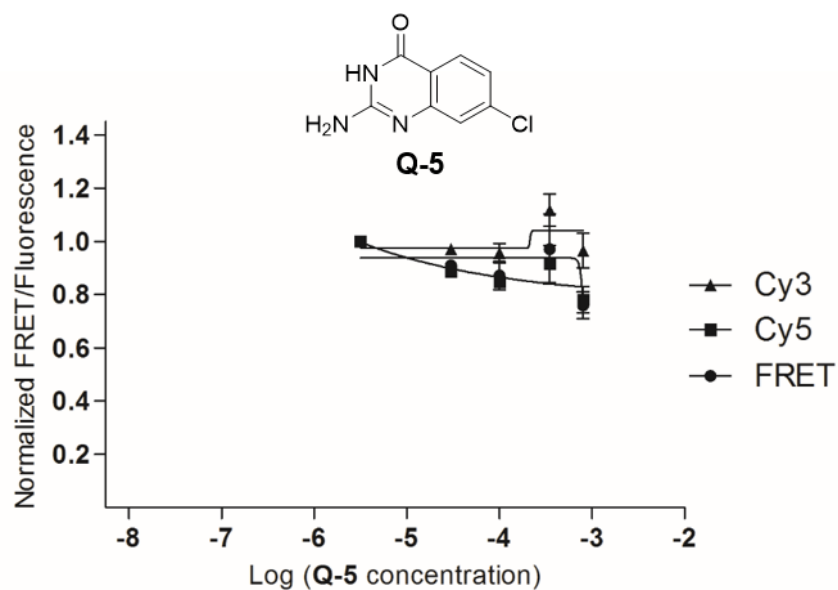


Figure 2.39: Q-5 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. Although FRET or Cy3 signals did not reach saturation, FRET widening is observed and an EC_{50} value for ligand binding $> 2000 \mu\text{M}$ is estimated.

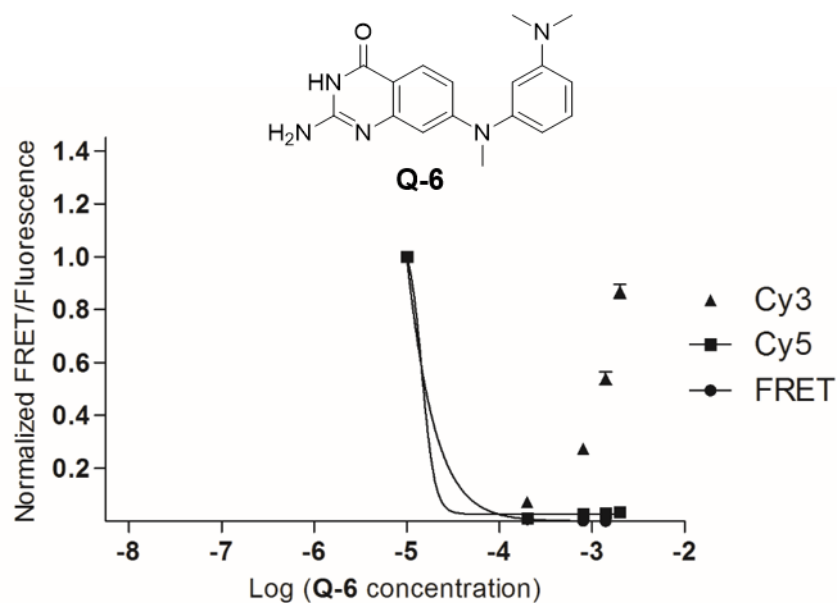


Figure 2.40: Q-6 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. An observed decrease in Cy3, Cy5, and FRET signals may be due to non-specific interactions between Q-6 and the RNA resulting in precipitation at all Q-6 concentrations. An increase in Cy3 emission at concentrations greater than 200 μ M was later discovered to be the result of Q-6 fluorescence at the same excitation and emission wavelengths as Cy3.

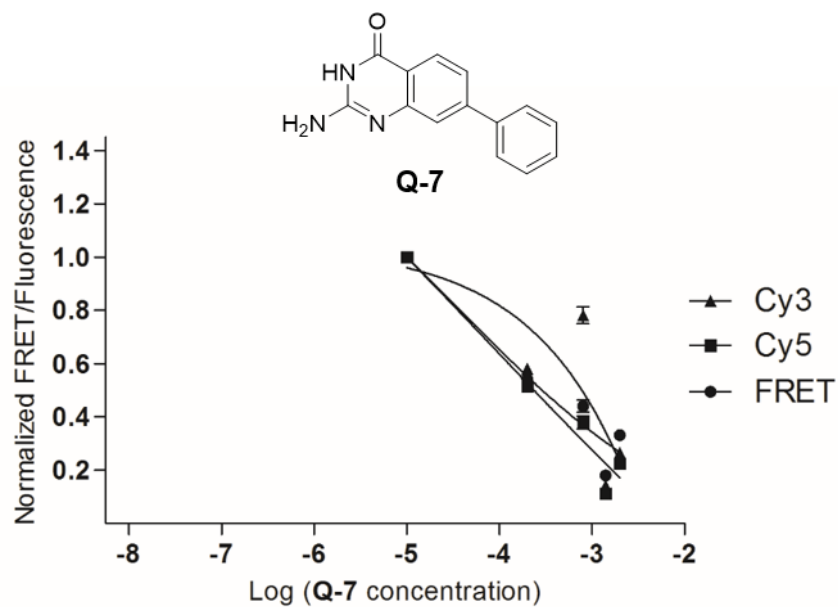


Figure 2.41: Q-7 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. An observed decrease in Cy3, Cy5, and FRET signals may be due to non-specific interactions between Q-7 and the RNA resulting in precipitation at all Q-7 concentrations.

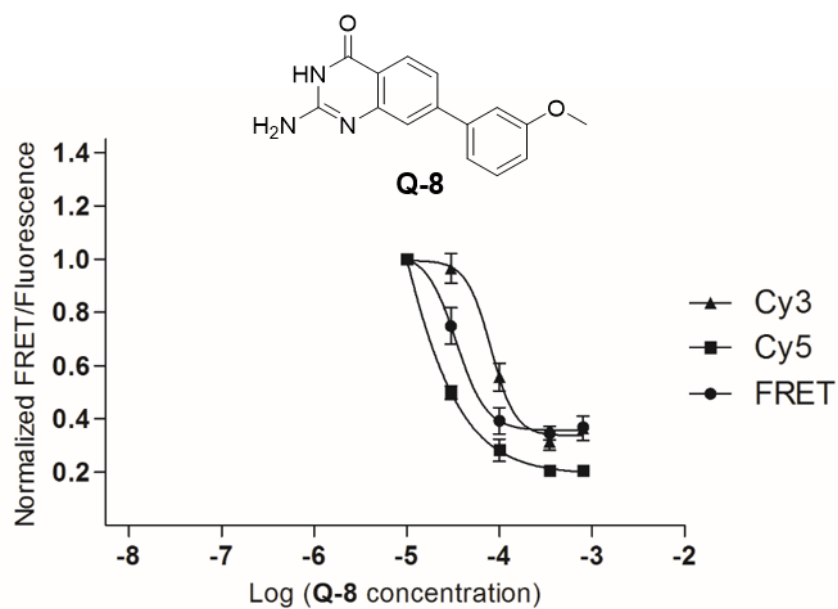


Figure 2.42: Q-8 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. An observed decrease in Cy3, Cy5, and FRET signals may be due to non-specific interactions between Q-8 and the RNA resulting in precipitation at all Q-8 concentrations.

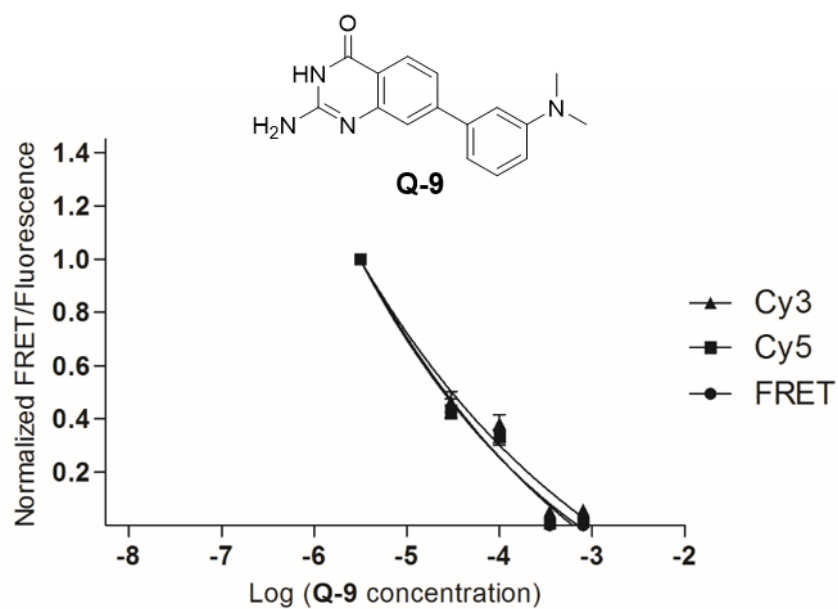
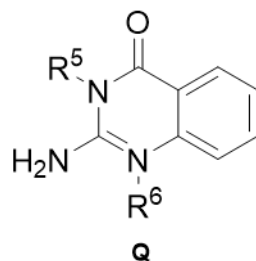


Figure 2.43: Q-9 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. An observed decrease in Cy3, Cy5, and FRET signals may be due to non-specific interactions between Q-9 and the RNA resulting in precipitation at all Q-9 concentrations.

Table 2.6: Summary of activity of 2-amino-quinazoline derivatives **Q-10** to **Q-11** in the FRET assay experiments shown in Figures 2.44-2.45. These two derivatives are shown in a separate table from Q1-9 as these are the only two compounds of this quinalazine scaffold series (or purine and pteridine scaffold series) with functionalization at the N1 and N3 nitrogens of the 2-amino-pyrimidinone ring. EC_{50} values are calculated from the ligand binding curve of the experiment listed in parenthesis (FRET, Cy3) and are listed with ± 1 SD calculated from triplicate experiments.



Compound	R ⁵	R ⁶	EC ₅₀ (μM)
Q-10	H		770 ± 190 (Cy3)
Q-11		H	> 400 μM (ambiguous)

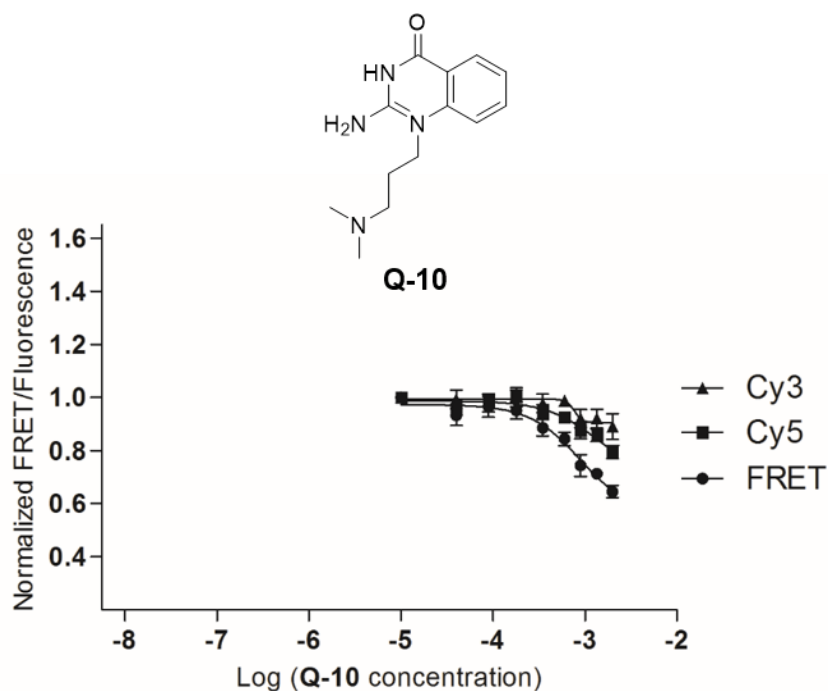


Figure 2.44: Q-10 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments.

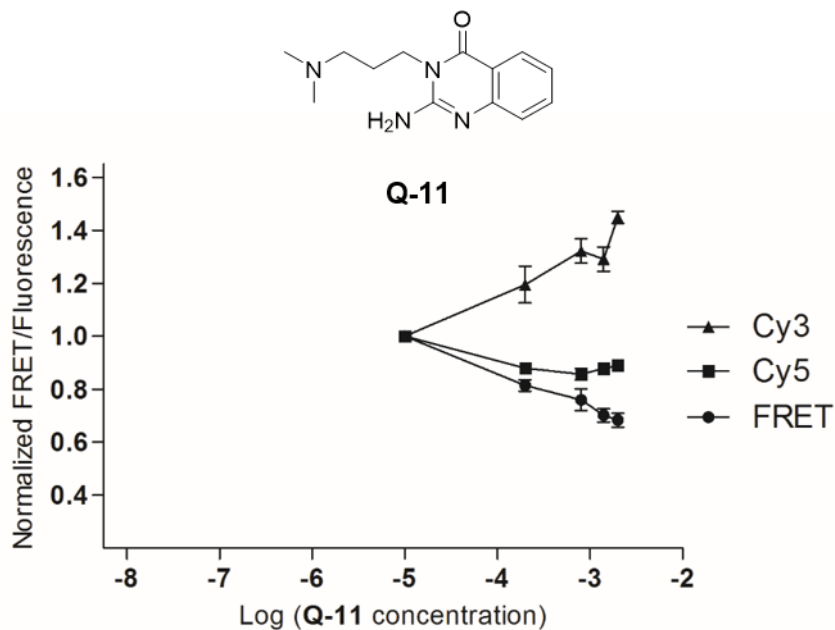
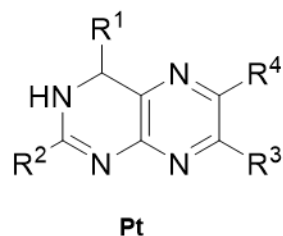


Figure 2.45: Q-11 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments.

Table 2.7: Summary of activity of pteridine derivatives **Pt-1** to **Pt-5** in the FRET assay experiments shown in Figures 2.46-2.51. EC₅₀ values are calculated from the ligand binding curve of the experiment listed in parenthesis (FRET, Cy3) and are listed with ± 1 SD calculated from triplicate experiments. No activity (n.a.) is listed for compounds which had no effect in the FRET assay. Pt-5 precipitated RNA before reaching concentrations sufficient to calculate an EC₅₀ value of ligand binding and is labelled (precip.).



Compound	R ¹	R ²	R ³	R ⁴	EC ₅₀ (μM)
Pt-1	=O	NH ₂	H	H	> 2000 (Cy3)
Pt-2	=O	NH ₂	H		n.a.
Pt-3	=O	NH ₂	H		> 2000 (FRET)
Pt-4	=O	OH	H	H	n.a.
Pt-5	=O	OH		see R ³	precip.
Pt-6	NH ₂	NH ₂	H	CH ₂ OH	precip.

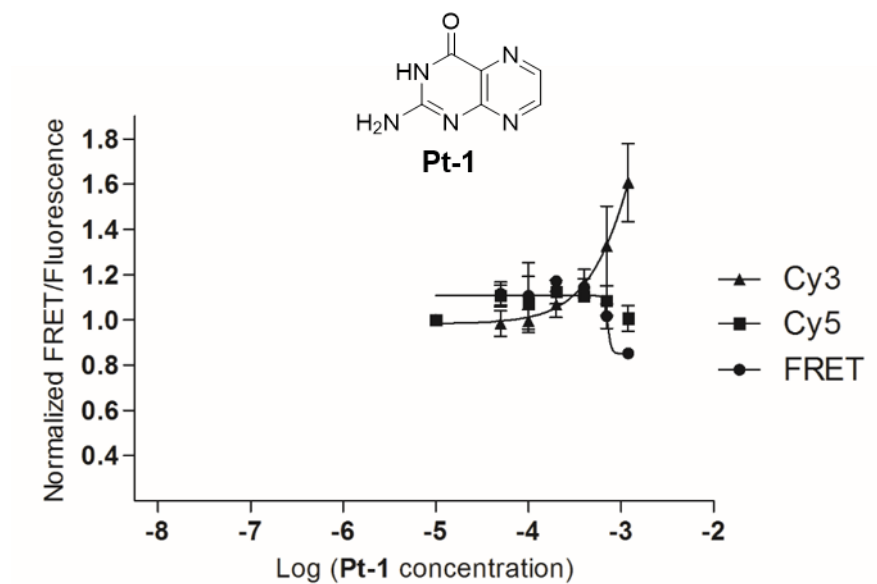


Figure 2.46: Pt-1 (pterin) titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. Although FRET or Cy3 signals did not reach saturation, FRET widening is observed and an EC_{50} value for ligand binding $> 2000 \mu\text{M}$ is estimated from the Cy3 curve.

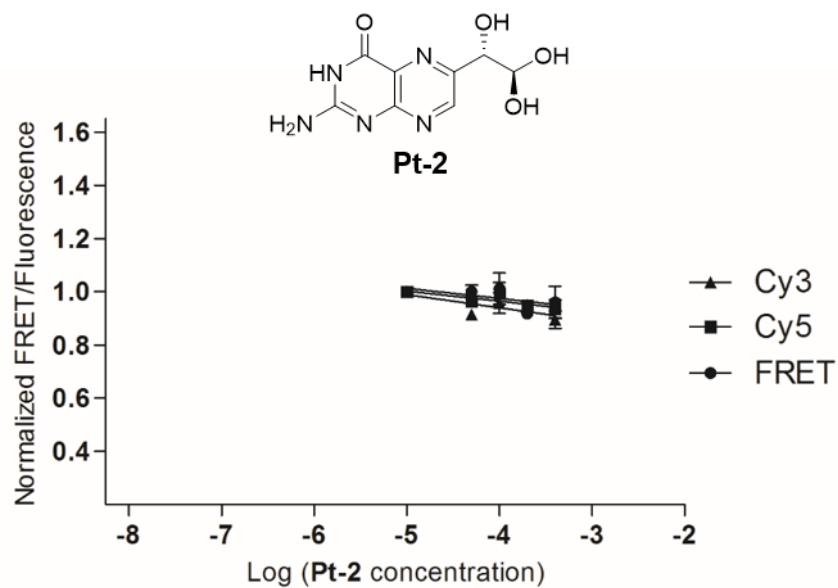


Figure 2.47: Pt-2 (neopterin) titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. No compound activity is observed up to 500 μ M (Pt-2 was insoluble in DMSO or H₂O and concentrations higher than 500 μ M).

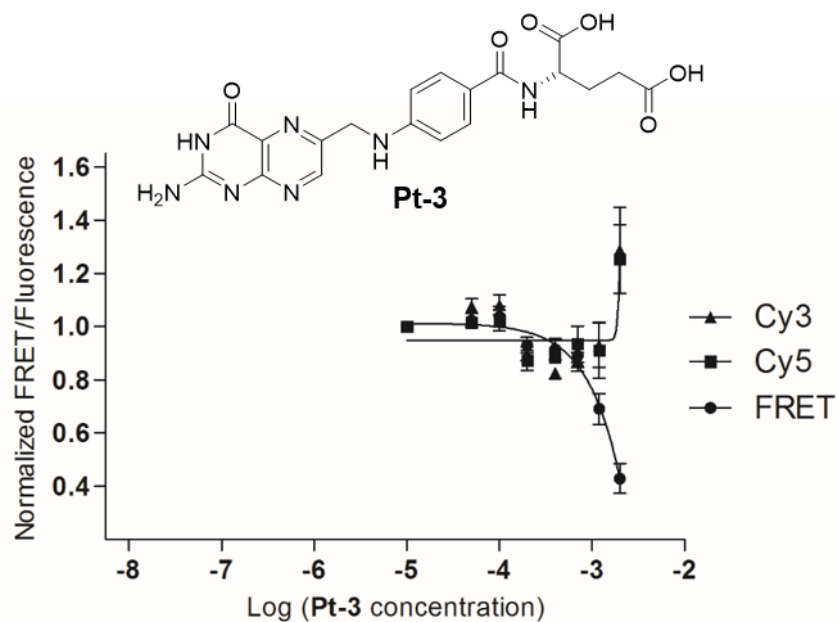


Figure 2.48: Pt-3 (folic acid) titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. Although FRET or Cy3 signals did not reach saturation, FRET widening is observed and an EC_{50} value for ligand binding $> 2000 \mu\text{M}$ is estimated from the FRET curve.

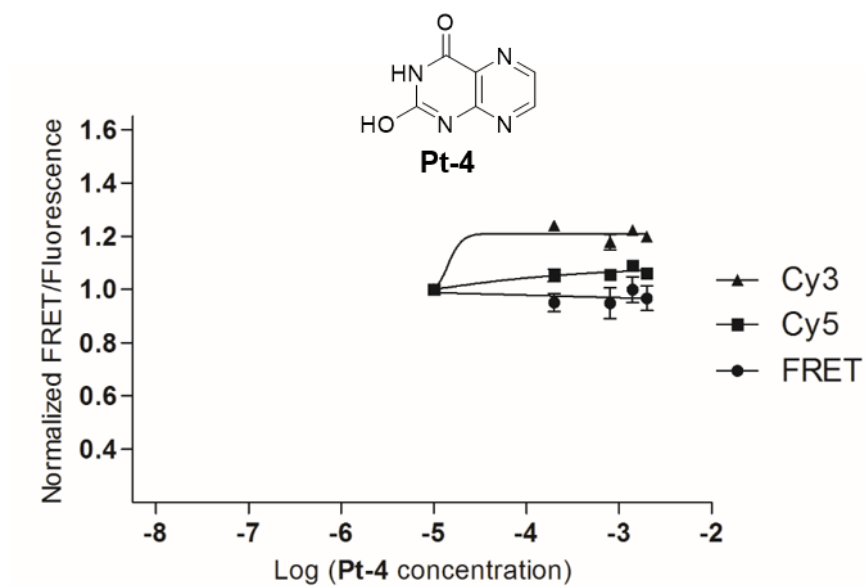


Figure 2.49: Pt-4 (lumazine) titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. No compound activity is observed.

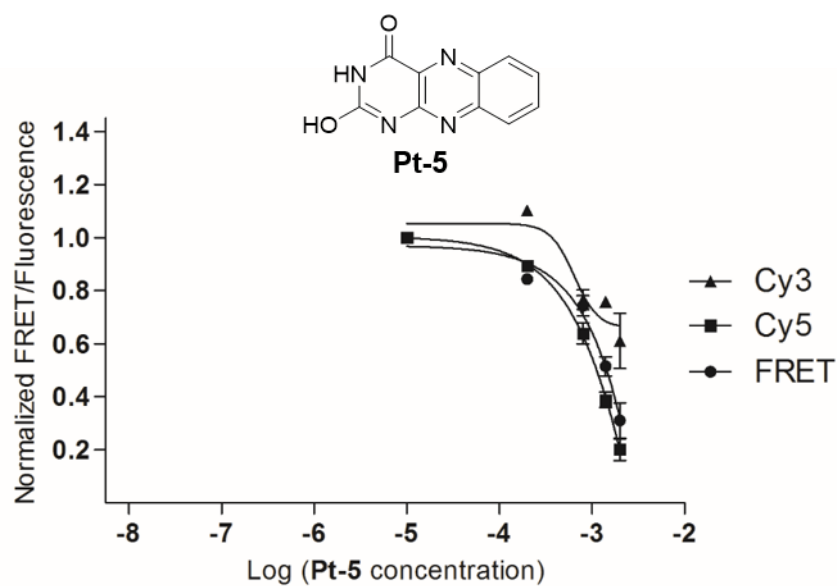


Figure 2.50: Pt-5 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. An observed decrease in Cy3, Cy5, and FRET signals may be due to non-specific interactions between Pt-5 and the RNA resulting in precipitation before reaching Pt-5 concentrations sufficient to calculate an EC_{50} value for ligand binding.

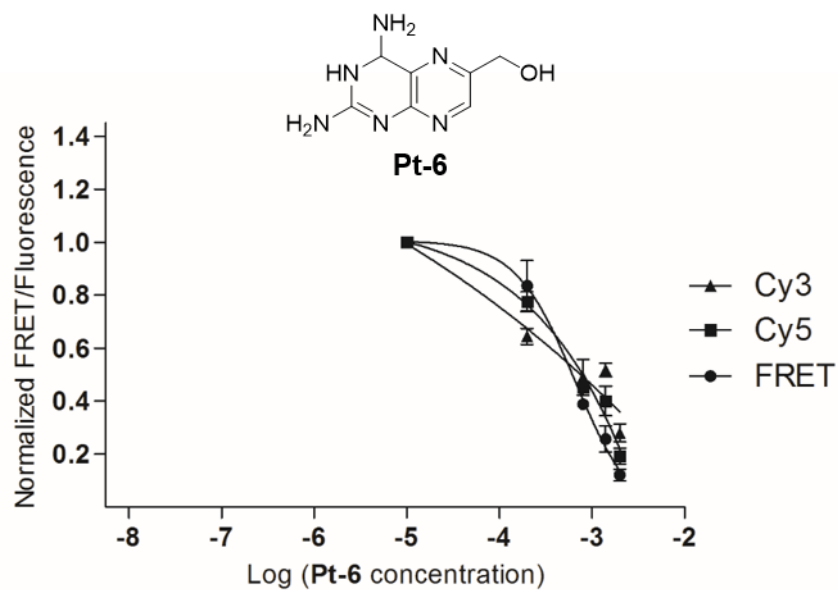
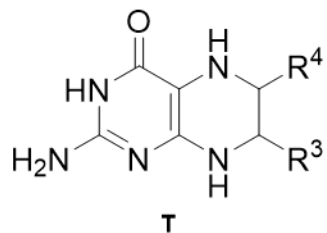
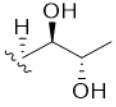


Figure 2.51: Pt-6 titration with FRET-labelled HCV IRES Ila RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. An observed decrease in Cy3, Cy5, and FRET signals may be due to non-specific interactions between Pt-5 and the RNA resulting in precipitation before reaching Pt-6 concentrations sufficient to calculate an EC_{50} value for ligand binding.

Table 2.8: Summary of activity of tetrahydropteridine derivatives **T-1** to **T-2** in the FRET assay experiments shown in Figures 2.52-2.53. EC_{50} values are calculated from FRET ligand binding curves and are listed with ± 1 SD calculated from triplicate experiments.



Compound	R ³	R ⁴	EC ₅₀ (μM)
T-1	H		640 ± 40 (FRET)
T-2	CH ₃	CH ₃	1800 ± 2000 (FRET)

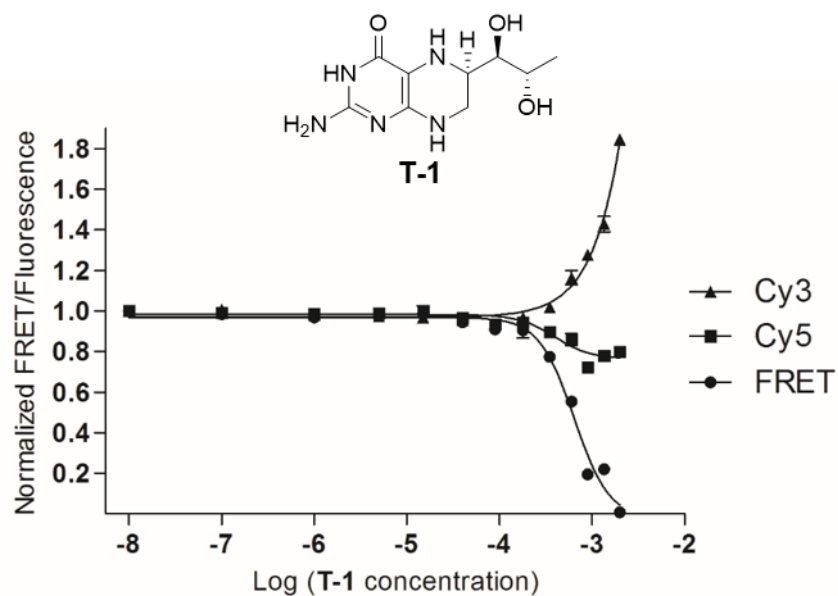


Figure 2.52: T-1 (tetrahydrobiopterin, THB) titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. Fitting of a single-site binding curve to FRET response gave an EC_{50} value for ligand binding of $640 \pm 40 \mu\text{M}$.

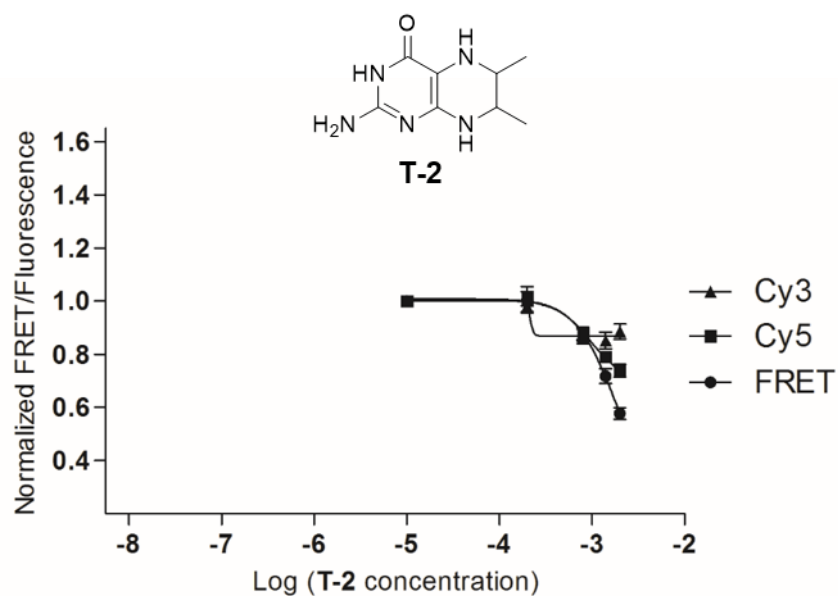
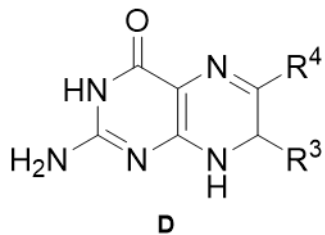


Figure 2.53: T-2 titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. Fitting of a single-site binding curve to FRET response gave an estimated EC_{50} value for ligand binding of $1800 \pm 2000 \mu\text{M}$.

Table 2.9: Summary of activity of dihydropteridine derivatives **D-1** to **D-2** in the FRET assay experiments shown in Figures 2.54-2.55. D-1's EC_{50} value is estimated from the FRET ligand binding curve calculated from triplicate experiments. No activity (n.a.) is listed for D-2 which had no effect in the FRET assay.



Compound	R ³	R ⁴	EC ₅₀ (μM)
D-1	H		> 2000 (FRET)
D-2	=O	H	n.a.

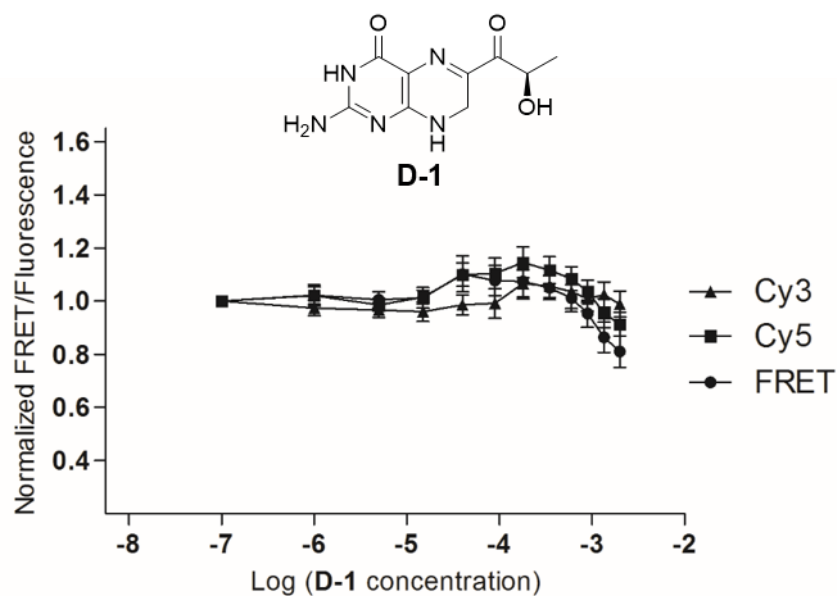


Figure 2.54: D-1 (L-sepiapterin) titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. Although FRET or Cy3 signals did not reach saturation, a slight FRET widening is observed and an EC_{50} value for ligand binding $> 2000 \mu\text{M}$ is estimated from the FRET curve. The curves shown connect individual Cy3, Cy5, or FRET data points as single site binding curves could not be calculated.

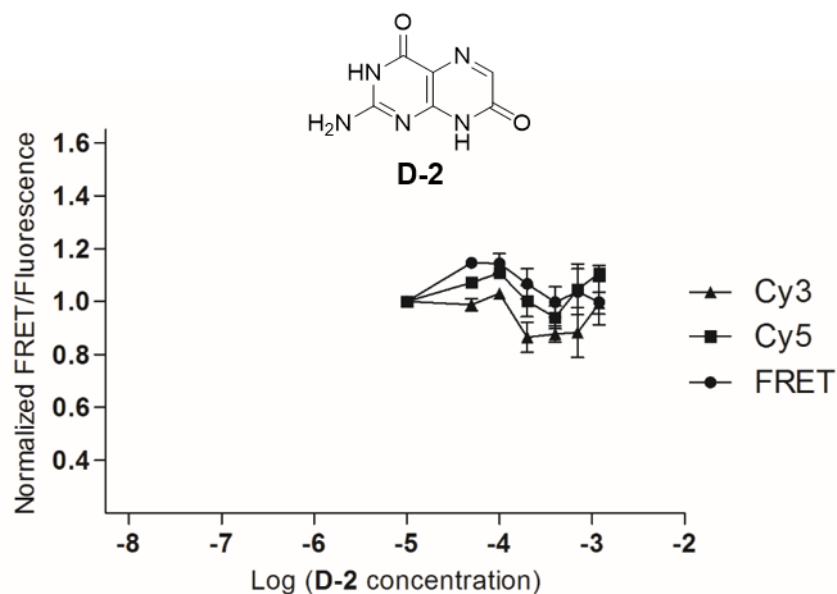


Figure 2.55: D-2 (isoxanthopterin) titration with FRET-labelled HCV IRES IIa RNA. Normalized Cy3 and Cy5 fluorescence and normalized FRET is plotted against the log compound concentration. Error bars represent ± 1 SD calculated from triplicate experiments. No compound activity is observed. The curves shown connect individual Cy3, Cy5, or FRET data points as single site binding curves could not be calculated.

Materials and Methods

RNA Preparation

Cyanine dye labeled and unlabeled RNA oligonucleotides were obtained from chemical synthesis and purified by HPLC (Integrated DNA Technologies, Coralville, IA). Stock solutions were prepared by dissolving lyophilized oligonucleotides in 10mM sodium cacodylate buffer, pH 6.5.

FRET Folding Experiments

Terminally Cy3/Cy5-labeled IIa RNA constructs were annealed from single strands by heating to 65°C for 5 min followed by snap cooling in 10mM HEPES (4- (2-

hydroxyethyl)-1-piperazineethanesulfonic acid) buffer, pH 7.0. FRET folding experiments were performed as described previously⁷² on a Spectra Max Gemini monochromator plate reader (Molecular Devices, Sunnyvale, CA) at 25°C. RNA was brought to a final concentration of 100nM in 10mM HEPES buffer, pH 7.0. Emission filters were set at 550 and 665 nm. The Cy3 label was excited at 520 nm and transferred fluorescence was read as Cy5 emission at 670 nm. FRET folding of labeled Iia RNA constructs was monitored while increasing Mg^{2+} concentration. Data sets were analyzed and FRET calculated as described previously.⁷²

FRET Compound Screening Experiments

FRET compound screening experiments were performed as described in the FRET folding experiments above, except RNA was brought to a final concentration of 100nM in 10mM sodium cacodylate buffer, pH 5.5, containing 2mM $MgCl_2$ (see Figure 2.27). Ligand induced FRET changes in terminally Cy3/Cy5-labelled Iia RNA constructs was monitored while increasing ligand concentration. To record fluorescence of the Cy3 dye, excitation was done at 520nm and emission read at 570nm with an emission filter at 550nm. To record fluorescence of the Cy5 dye, excitation was done at 620nm and emission read at 670nm with an emission filter at 665nm.

In vitro translation experiments

The *in vitro* translation assay (IVT) was performed using the TNT Quick coupled reticulocyte lysate system (Promega, Madison, WI) and an HCV bicistronic luciferase reporter as previously described.⁹⁰ The bicistronic luciferase reporter contains the

sequence coding for the HCV IRES-Renilla luciferase preceded by a cap-initiated firefly luciferase internal control. Briefly, reactions were carried out according to the manufacturer's instruction at a volume of 7.5 μ L, containing 1.5 μ L reporter DNA plasmid (100ng/ μ L), 1.5 μ L H₂O or compound solution, and 4.5 μ L reaction buffer containing reticulocyte lysate, SP6 polymerase, RNase inhibitor, and amino acids. Detection of firefly and Renilla luciferase levels was done using the Dual-Glo Luciferase Assay System (Promega, Madison, WI) as previously described.⁷² Relative translation efficiencies were calculated as a ratio of IRES-driven Renilla luciferase levels to the internal control firefly luciferase levels.

Compound testing for in vitro translation inhibition

Compounds were dissolved in DMSO and added to the assay solution at the desired compound concentrations of 0, 200, 500, and 1500 μ M at a final DMSO concentration of 1.5vol%. Compound testing for *in vitro* translation inhibition was conducted using the IVT assay as described above. Relative translation efficiencies were normalized to 0 μ M levels.

HCV Bicistronic Reporter Mutagenesis

Mutations were introduced into the HCV bicistronic luciferase reporter DNA plasmid via standard molecular cloning techniques and site directed mutagenesis (Q5 Site-Directed Mutagenesis Kit, New England Biolabs, Ipswich, MA). Oligonucleotides for the cloning of HCV Δ II mutant bicistronic reporter plasmids are listed in Table 2.10.

Briefly, chemically synthesized oligonucleotides were phosphorylated and then annealed pairwise (xxx-nT with xxx-nB). Segment 2 was first ligated to segment 3 followed by the ligation of segment 1 to segment 2-3. Segment 4 was generated from a polymerase chain reaction (PCR) using the WT HCV bicistronic reporter plasmid as the template and HCV Segment 4 forward and reverse primers. The full length insert was then generated from the ligation product (segment 1-2-3) and PCR product (segment 4) using PCR by overlap extension. The resulting full length PCR product was purified with a QIAquick PCR Purification Kit and digested with EcoRI and AccI. Likewise, the WT HCV bicistronic reporter plasmid was digested with EcoRI and AccI and ligated with the EcoRI-AccI digested full length PCR product. The product of this ligation was then transformed into NEB 5-alpha competent E. coli cells, plated on LB-Amp medium and grown overnight at 37°C. Individual colonies with the correct insertion were confirmed by automated DNA sequencing.

Oligonucleotides for site-directed mutagenesis of HCV Δ II mutant bicistronic reporter plasmids are outlined in Table 2.11. Site-directed mutagenesis experiments were carried out according to the manufacturer's instructions. The sequences of all mutant bicistronic reporter plasmids were verified by DNA sequencing. Mutational studies were conducted using the IVT assay as described above. Relative translation efficiencies were normalized to HCV WT levels.

Crystallization and Data Collection

SVV subdomain IIa RNA (Figure 2.11a) was annealed from stoichiometric amounts of the single strands by heating to 65°C for 4min followed by slow cooling to

room temperature. After cooling, the RNA was crystallized at 16 °C by hanging drop vapor diffusion. For crystallization, 1 μ L of 0.2mM RNA was mixed with an equal volume of precipitating solution containing 10mM calcium chloride, 200mM ammonium chloride, 50mM Tris hydrochloride buffer, pH 8.5, and 25% w/v polyethylene glycol 4,000. Cube-shaped crystals appeared and grew to full size over 2-4 days of equilibration against 700 μ L of well solution containing precipitating solution.

Extended SVV Subdomain IIa RNA (Figure 2.12a) was crystallized using the same methods as above after mixing 1 μ L of 0.2mM RNA with an equal volume of precipitating solution containing 100mM magnesium acetate, 200mM potassium chloride, 50mM sodium cacodylate buffer, pH 6.5, and 10% w/v polyethylene glycol 8,000. Plate-shaped crystals grew over 2 months of equilibration against 700 μ L of well solution containing precipitating solution.

Crystals were flash-cooled in liquid nitrogen. X-ray diffraction data were collected at 110K on a Rigaku rotating anode X-ray generator ($\lambda = 1.54 \text{ \AA}$) equipped with a MAR345 imaging plate detector system. Datasets were processed, integrated, and scaled with the HKL2000 package.⁹¹

Structure Solution and Refinement

The three-dimensional structure of the SVV subdomain IIa RNA was solved by molecular replacement with the program Phaser⁹² using A-form RNA duplexes as search models and refined by the program Refmac⁹³ both within the CCP4 package.⁹⁴ Subsequent iterative rounds of manual building and refinement, alternating between Refmac and manual rebuilding in Coot,⁹⁵ were based on the obtained 2Fo-Fc and Fo-Fc

maps. Final refinement was carried out in PHENIX⁹⁶ with individual isotropic atomic displacement parameters and water picking (Tables 2.1, 2.2). Coordinates and structure factors for both SVV subdomain IIa structures have been deposited in the RCSB Protein Data Bank under accession codes 4P97 and 4PHY.

HCV Replicon Mutagenesis

SVV IRES IIa mutations were introduced into the SGR-JFH1 FEO⁸⁰ DNA plasmid via site directed mutagenesis (Q5 Site-Directed Mutagenesis Kit, New England Biolabs, Ipswich, MA). The sequence of the Δ SVV-IRES-IIa SGR-JFH1 FEO plasmid was verified by automated DNA sequencing. Sequences of the mutagenic oligonucleotides for SGR-JFH1 FEO were as follows, with lowercase letters indicating mutated residues:

5'-GCCATGGCGTTAGTATGAGTGTCGTAC_{gagg}CTCCAGGCCCCCCCTCC-3'

(sense)

5'-

TAGGCGCTTTCTGCGTGAAGAC_{ggtggtag}CTCACAGGGGAGTGATTCATGGCG

-3' (anti-sense)

HCV Replicon Assay

SVV IRES IIa mutations were introduced into the SGR-JFH1 FEO⁸⁰ DNA plasmid as outlined above. SGR-JFH1 FEO and Δ SVV-IRES-IIa RNAs were generated from the corresponding DNA plasmid using T7 RNA polymerase as previously described.^{80,97} Briefly, plasmids were linearized with XbaI and then digested with mung

bean nuclease to generate an authentic 3' end. *In vitro* transcription (T7 RiboMAX Express Large Scale RNA Production System, Promega, Madison, WI) was carried out according to the manufacturer's instructions to yield SGR-JFH1 FEO and Δ SVV-IRES-IIa RNAs. Transfection was performed as previously described.^{53,97} Briefly, 400 μ l of a Huh-7.5.1 cell suspension (107 cells/ml) was placed in a 2mm cuvette with 10 μ g SGR-JFH1 FEO or Δ SVV-IRES-IIa RNA. The mixture was electroporated (Bio-Rad Gene Pulser, Hercules, CA) with an exponential protocol (140V, 950 μ F). Cells were then seeded into 96-well plates at a density of 20,000 cells/well and resuspended in 100 μ L complete media. All conditions were run in triplicate. Luciferase activity was determined at 4h and 24h post-transfection using a combined lysis buffer and luciferin reagent (OneGlo, Promega, Madison, WI) according to the manufacturer's instructions. Luciferase activity was determined using a microplate luminometer (Veritas microplate luminometer; Turner Biosystems, Sunnyvale, CA). 4h and 24h luciferase levels were normalized to WT luciferase level at 4h.

Table 2.10. Oligonucleotide sequences for the cloning of HCV Δ II mutants in bicistronic reporter plasmids.

Oligonucleotide Name	Sequence (5' → 3')*
HCV Segment 1T (sense)	CTAGAGaattCCCAGCCCCCGATTGGGGGCGACACTCC ACCATAGATC
HCV Segment 1B (antisense)	GGAGTGATCTATGGTGGAGTGTGCGCCCCAATCGGGG CTGGGaatctCTCTAG
HCV Segment 4 Forward Primer (sense)	CCCTCCCGGGAGAGCCATA
HCV Segment 4 Reverse Primer (antisense)	CGAAGGATTCGTGCTCATGG
HCV Δ II + BVDV II-2T (sense)	ACTCCCCCTCCTTAGCGAAGGCCGAAAAGAGGCTAGCCA TGCCCT
HCV Δ II + BVDV II-2B (antisense)	TACTAAGGGCATGGCTAGCCTCTTTTCGGCCTTCGCTA AGGAGG
HCV Δ II + BVDV II-3T (sense)	TAGTAGGACTAGCAAACAAGGAGGACCCCCCTCCCG GGAGAGC
HCV Δ II + BVDV II-3B (antisense)	GCTCTCCCGGGAGGGGGGTCTCCTTGTTTTGCTAGT CC
HCV Δ II + CSFV II-2T (sense)	ACTCCCCCTCAGCGACGGCCGAAATGGGCTAGCCATGC CCA
HCV Δ II + CSFV II-2B (antisense)	TACTATGGGCATGGCTAGCCATTTTCGGCCGTCGCTGG AGG
HCV Δ II + CSFV II-3T (sense)	TAGTAGGACTAGCAAACGGAGGACCCCCCTCCCGGGA GAGC
HCV Δ II + CSFV II-3B (antisense)	GCTCTCCCGGGAGGGGGGTCTCCTCGTTTTGCTAGTCC
HCV Δ II + SVV II-2T (sense)	ACTCCGATGGCTACCCACCTCGGATCACTGAACTGGAG CTCGACCCT
HCV Δ II + SVV II-2B (antisense)	TAAGGAGGGTCGAGCTCCAGTTCAGTGATCCGAGGTGG GTAGCCATC
HCV Δ II + SVV II-3T (sense)	CCTTAGTAAGGGAACCGAGAGGCCTTCACCCCCCTCC CGGGAGAGC
HCV Δ II + SVV II-3B (antisense)	GCTCTCCCGGGAGGGGGGTGAAGGCCTCTCGGTTCCT TTAC
HCV Δ II + BVDV II/ Δ IIa-2T (sense)	ACTCCCCCTCCTTAGCGAAGGCCGAAAAGAGGCTAGCCA TGCCCT
HCV Δ II + BVDV II/ Δ IIa-2B (antisense)	TACTAAGGGCATGGCTAGCCTCTTTTCGGCCTTCGCTA AGGAGG

Table 2.10. Oligonucleotide sequences for the cloning of HCV Δ II mutants in bicistronic reporter plasmids. *Continued.*

HCV Δ II + BVDV II/ Δ IIa-3T (sense)	TAGTAGGACTAGCTAAGGAGGACCCCCCTCCCGGGAGAGCCATA
HCV Δ II + BVDV II/ Δ IIa-3B (antisense)	TATGGCTCTCCCGGGAGGGGGGGTCTCCTTAGCTAGTCC
HCV Δ II + CSFV II/ Δ IIa-2T (sense)	ACTCCCCTCCAGCGACGGCCGAAATGGGCTAGCCATGCCA
HCV Δ II + CSFV II/ Δ IIa-2B (antisense)	TACTATGGGCATGGCTAGCCATTTTCGGCCGTCGCTGGAGG
HCV Δ II + CSFV II/ Δ IIa-3T (sense)	TAGTAGGACTAGCTGGAGGACCCCCCTCCCGGGAGAGCCATA
HCV Δ II + CSFV II/ Δ IIa-3B (antisense)	TATGGCTCTCCCGGGAGGGGGGGTCTCCAGCTAGTCC
HCV Δ 2bp-2T (sense)	ACTCCCCTGTGAGGAACTACTGTCTTCACGCAGAAAGCTAGCCATGGT
HCV Δ 2bp-2B (antisense)	TACTAACCATGGCTAGGCTTTCTGCGTGAAGACAGTAGTTCTCACAGG
HCV Δ 2bp-3T (sense)	TAGTATGAGTGTCGTGCAGCCTCCAGGACCCCCCTCCCGGGAGAGCCATA
HCV Δ 2bp-3B (antisense)	TATGGCTCTCCCGGGAGGGGGGGTCTGGAGGCTGCACGACTCA
HCV Δ II + BVDV II/ Δ 2bp-2T (sense)	ACTCCCCTCCTTAGCGAAGGCCGAAAAGACTAGCCATGCT
HCV Δ II + BVDV II/ Δ 2bp-2B (antisense)	TACTAAGCATGGCTAGTCTTTTCGGCCTTCGCTAAGGAGG
HCV Δ II + BVDV II/ Δ 2bp-3T (sense)	TAGTAGGACTAGCAAAACAAGGAGGACCCCCCTCCCGGGAGAGC
HCV Δ II + BVDV II/ Δ 2bp-3B (antisense)	GCTCTCCCGGGAGGGGGGGTCTCCTTGTTTTGCTAGTCC
HCV Δ II + CSFV II/ Δ 2bp-2T (sense)	ACTCCCCTCCAGCGACGGCCGAAATGCTAGCCATGCA
HCV Δ II + CSFV II/ Δ 2bp-2B (antisense)	TACTATGCATGGCTAGCATTTCGGCCGTCGCTGGAGG
HCV Δ II + CSFV II/ Δ 2bp-3T (sense)	TAGTAGGACTAGCAAACGGAGGACCCCCCTCCCGGGAGAGC
HCV Δ II + CSFV II/ Δ 2bp-3B (antisense)	GCTCTCCCGGGAGGGGGGGTCTCCGTTTGCTAGTCC

*Lowercase letters indicate EcoRI restriction site.

Table 2.11. Oligonucleotide sequences for site-directed mutagenesis of bicistronic reporter constructs.

Parent Construct	Oligonucleotide Name	Sequence (5' → 3')*
HCV WT	HCV Δ II sense (sense)	GACCCCCCTCCCGGGAG
HCV WT	HCV Δ II antisense (antisense)	GGGAGTGATCTATGGTGGAGTGTCG
HCV WT	HCV Δ II + AEV II (sense)	atgggtgtagtatgggaatcgtgtatggggatACCCC CCCTCCCGGGAGA
HCV WT	HCV Δ II + AEV II (antisense)	ggttaggatggttctgtgaaacgcaaagggatGGAGT GATCTATGGTGGAGTGTCGC
HCV WT	HCV Δ Ia (sense)	CTGTCTTCACGCAGAAAG
HCV WT	HCV Δ Ia (antisense)	CCTCACAGGGGAGTGATC
HCV Δ II + AEV II	HCV Δ II + AEV II/ Δ Ia (sense)	ATGGGAATCGcaaGGGGATACCCCC
HCV Δ II + AEV II	HCV Δ II + AEV II/ Δ Ia (antisense)	ACTACACCCATGGTTAGG
HCV Δ II + AEV II	HCV Δ II + AEV II/ Δ 2bp (sense)	catgTGTAGTATGGGAATCGTG
HCV Δ II + AEV II	HCV Δ II + AEV II/ Δ 2bp (antisense)	gttagTGGTTCTGTGAAACGCAAAG
HCV WT	HCV Δ Ia + SVV Ila Predicted (sense)	atggcgtagtatgagtgctcgtgcAGAGGCCTTCACC CCCCC
HCV WT	HCV Δ Ia + SVV Ila Predicted (antisense)	ggctagacgctttctgcgtgaagacAGGTGGGTAGCC ATCGGAG
HCV WT	HCV Δ Ia + SVV Ila X-Ray (sense)	gccatggcgtagtatgagtgctcgtgcgaggCTCCAG GACCCCCCTCC
HCV WT	HCV Δ Ia + SVV Ila X-Ray (antisense)	tagacgctttctgcgtgaagacgggtgggtagCTCACA GGGAGTGATCTATGGTG
HCV WT	HCV Δ Ia + CSFV Ila (sense)	tggcgtagtatgagtgctcgtgcacaaacgCTCCAGG ACCCCCCTCC
HCV WT	HCV Δ Ia + CSFV Ila (antisense)	tggctagacgctttctgcgtgaagacactgCTCACAG GGGAGTGATCTATGGTG
HCV WT	HCV Δ Ia + CSFV Ila/ Δ 5bp (sense)	tggcgtagtatgagtgacacaaacgCTCCAGGACCCC CCCTCC
HCV WT	HCV Δ Ia + CSFV Ila/ Δ 5bp (antisense)	tggctagacgctttctgcgacactgCTCACAGGGGAG TGATCTATGGTG

Table 2.11. Oligonucleotide sequences for site-directed mutagenesis of bicistronic reporter constructs. *Continued.*

HCV Δ Ila + CSFV Ila	HCV Δ Ila + CSFV Ila/ Δ 5bp/+C (sense)	catggcgtagtatgactagCAAACGCTCCAGGACCC C
HCV Δ Ila + CSFV Ila	HCV Δ Ila + CSFV Ila/ Δ 5bp/+C (antisense)	gctagacgctttctgccgctcgCTGCTCACAGGGGAGT GATC

*Lowercase letters indicate a substitution from the parent construct. The absence of lowercase letters indicates a simple deletion.

Acknowledgements

This material, in part, has also been published in Proceedings of the National Academy of Sciences USA (2014, 111, 15952-15957) titled *Functional conservation despite structural divergence in ligand-responsive RNA switches* with Sergey Dibrov, Jing Gu, David Wyles, and Thomas Hermann as co-authors. The dissertation author was the primary investigator and author of this material.

Chapter 3:

A closer look at ligand-responsive RNA mechanical switches and their widespread discovery in diverse viral IRES elements

Parts of this work are published in:

“Ligand-responsive RNA mechanical switches” *RNA Biology*, 12, 780–786 (2015).

Abstract

Ligand-responsive RNA mechanical switches represent a new class of simple switching modules that adopt well-defined ligand-free and bound conformational states, distinguishing them from metabolite-sensing riboswitches. Initially discovered in the internal ribosome entry site (IRES) of hepatitis C virus (HCV), we have now discovered RNA switch motifs in the genomes of diverse other viruses. Although large variations are seen in sequence and local secondary structure of the switches, their function in viral translation initiation that requires selective ligand recognition is conserved. In Chapter 2, the crystal structure of an RNA switch from Seneca Valley virus (SVV) which is able to functionally replace the switch of HCV was determined. The switches from both viruses recognize identical cognate ligands despite their sequence dissimilarity. Here, we describe the discovery of seven new switches in addition to the previously established five examples. We highlight structural and functional features unique to this class of ligand-responsive RNA mechanical switches and discuss implications for therapeutic development and the construction of RNA nanostructures.

Introduction

Ligand-responsive RNA mechanical switches versus metabolite-sensing riboswitches

Over the course of the last decades, structural and functional studies have expanded our understanding of the numerous roles that RNA plays in addition to transferring genetic information from DNA into proteins, including its ability to regulate diverse cellular processes.¹ Many of these roles involve a switching transformation

between two different structural states of an RNA, each of which exerts a distinct function. The most common of these switches are known as riboswitches which act as genetic regulatory on-off switches in bacterial mRNA by directly binding small molecule metabolites. The binding of a ligand to a riboswitch sensor domain induces secondary structure changes within a second domain, the expression platform, which regulates downstream transcription and translation of the mRNA (Figure 3.1a). Exhaustive rearrangement of base pairs in the switching sequence results in different secondary structures of riboswitches in the presence or absence of a binding ligand.⁹⁸⁻¹⁰⁰

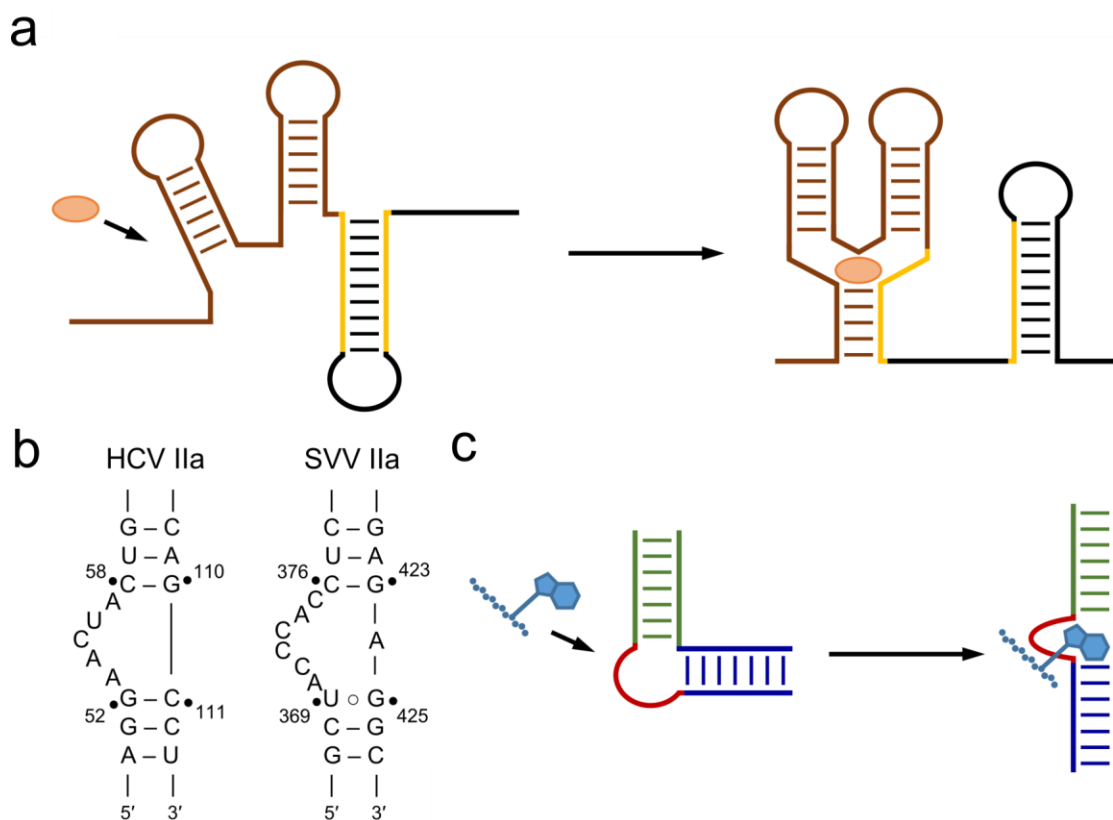


Figure 3.1: Comparison of ligand-responsive RNA mechanical switches in contrast to metabolite-sensing riboswitches. **a)** Metabolite-sensing riboswitches consist of a sensor domain (brown) linked to an expression platform (yellow and black) and exist in structurally diverse metabolite-free or bound conformations. The binding of a metabolite (orange) induces structural changes to the RNA switching sequence (yellow), resulting in folding and base-pairing patterns that are distinct from the metabolite-free conformation. **b)** Predicted secondary structure for HCV and SVV IIa RNA switches. **(C)** Ligand-responsive RNA mechanical switches are composed of a small internal loop (red) of 3-6 unpaired bases and two flanking helices (green and dark blue) of double stranded RNA and adopt distinct ligand-free (bent) and ligand-captured (elongated) conformations. The ligand-free bent conformation is stabilized by magnesium ions (not shown) and continuous stacking interactions of unpaired bases in the internal loop. Ligand (light blue) binding captures the elongated conformation of the RNA in a process that involves rearrangement of magnesium ions and unpaired bases but does not affect the integrity of base pairs of the flanking helices.

Results and Discussion

Unlike riboswitches, the RNA switches described here are unique in their small size (Figure 3.1b) and mechanical responsiveness to ligand binding which gives rise to two distinct and stable conformational states implicated with a switching function during viral translation initiation (Figure 3.1c).³⁹ The viral RNA switches adopt an L-shaped conformation stabilized by magnesium ions (Figures 3.2a, 3.3a) or cross-over base pairing (Figure 3.3b), and are captured in a straightened conformation by ligands binding in a deep pocket (Figures 3.2b, 3.4).^{6,21,39,47,53,55,65,66} The individual conformational states interconvert as a result of the rearrangement of magnesium ions and unpaired bases in the internal loop RNA, and without breaking or new formation of any base pairs in the secondary structure as occurs in riboswitches. The requirement for two stable switch conformations where each internal loop base has specific functional interactions in both the ligand-bound and unbound structural state leads to the high level of sequence conservation seen in virus clinical isolates.^{6,21} Conformational dynamics of the switches is required for biological function, as capture in one of the individual states by synthetic inhibitor binding inactivates the switch.⁶ These switches do not undergo secondary structure changes but rather a simple mechanical switching between two distinct and stable conformations that are required as functional states during translation initiation, similar to a light switch whose moving parts translocate and rotate but do not undergo reassembly during actuation.

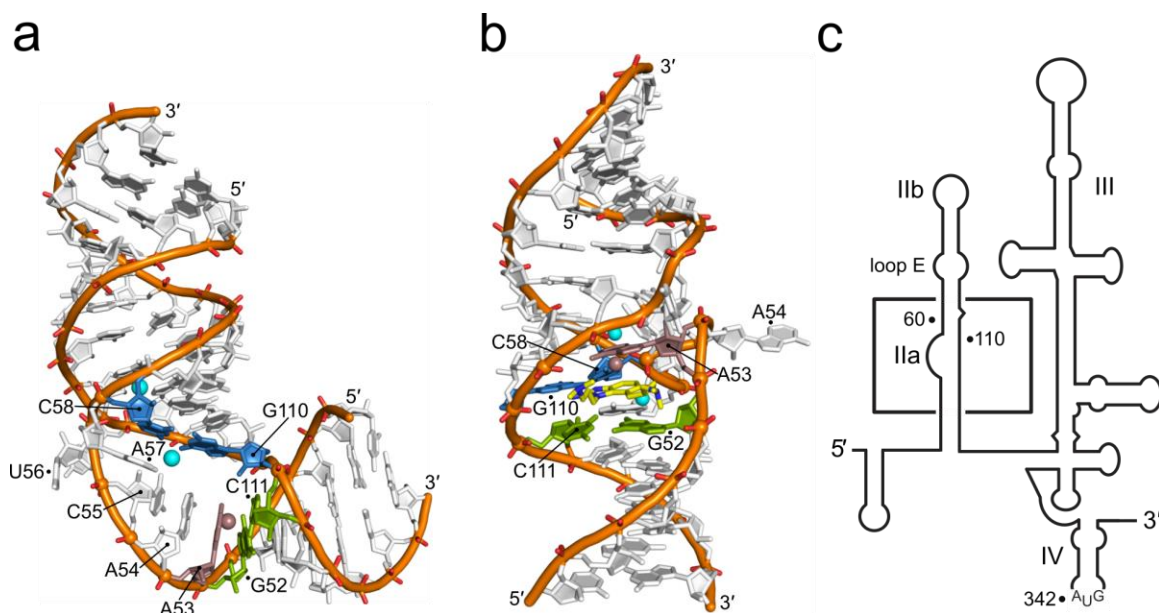


Figure 3.2: Structure of HCV IRES subdomain IIa RNA switch in absence and presence of binding ligand. Colors of highlighted bases are described in Figure 3.3. a) Crystal structure of ligand-free IIa RNA switch from HCV shown with stabilizing magnesium ions and flanking helices adopting a bent conformation. b) Crystal structure of ligand-captured IIa RNA switch from HCV adopting an elongated conformation. c) Secondary structure of the HCV IRES element in the 5' untranslated region of the viral genome. Subdomain IIa is highlighted inside a box.

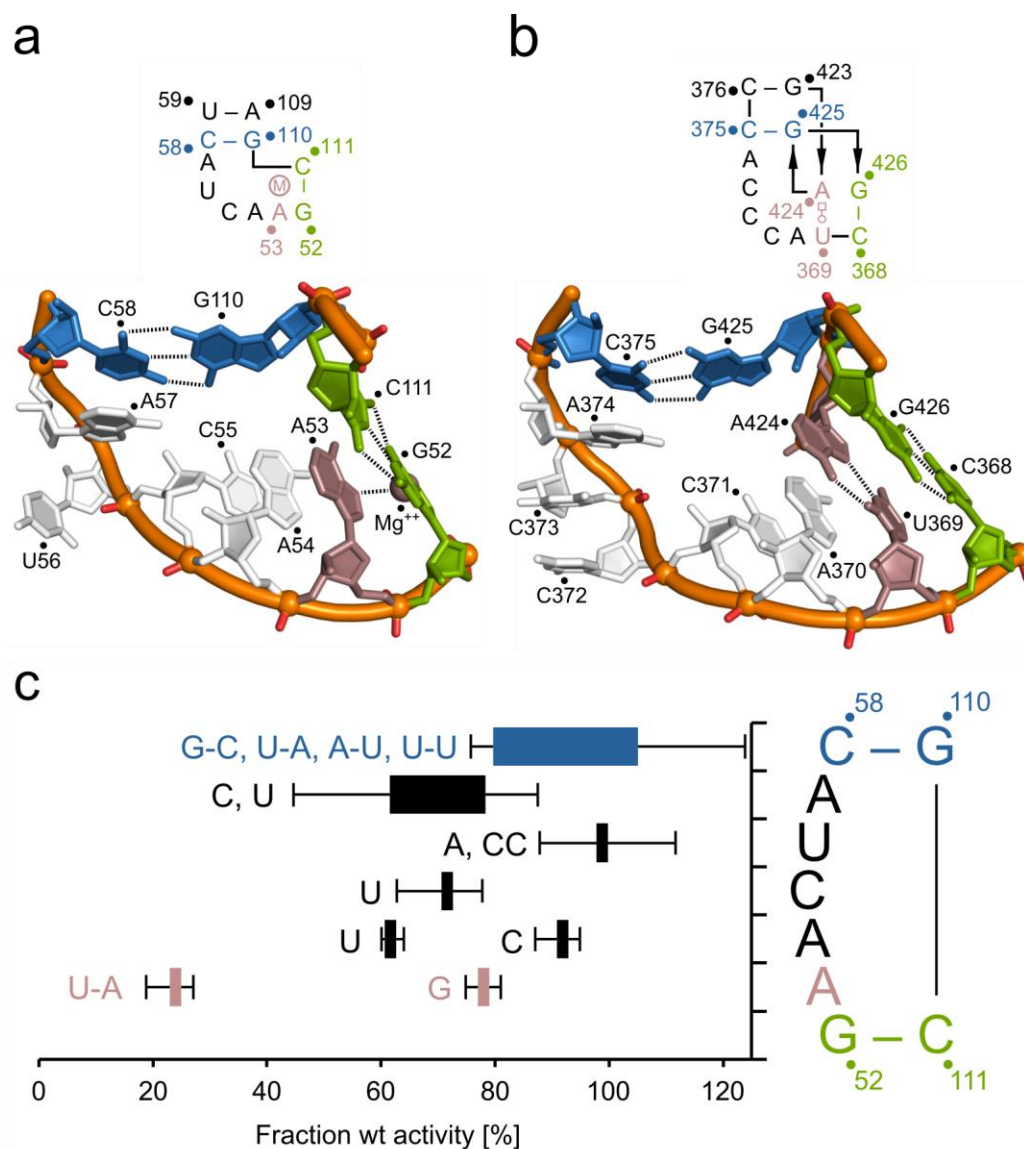


Figure 3.3: Structures of IRES subdomain IIa RNA switches from SVV and HCV. Colors highlight corresponding elements in each switch. a) Secondary structure schematic and crystal structure of HCV IIa highlighting closing base pairs (top-blue; bottom-green) and A53 stabilized by a magnesium ion (mauve, Mg^{++}) that occupies a similar space as the U \circ A Hoogsteen pair in the SVV motif. b) Secondary structure schematic and crystal structure of SVV IIa highlighting closing base pairs (top-blue; bottom-green) and unusual U \circ A Hoogsteen pair (mauve). Local inversion of strand directionality is indicated by arrows. c) Mutational analysis of HCV IIa switch activity studied in the context of an *in vitro* translation (IVT) assay. Box width indicates the range of average activity for each mutation as a percent of wild type activity. Error bars indicate minimum and maximum observed values for each set of mutations.

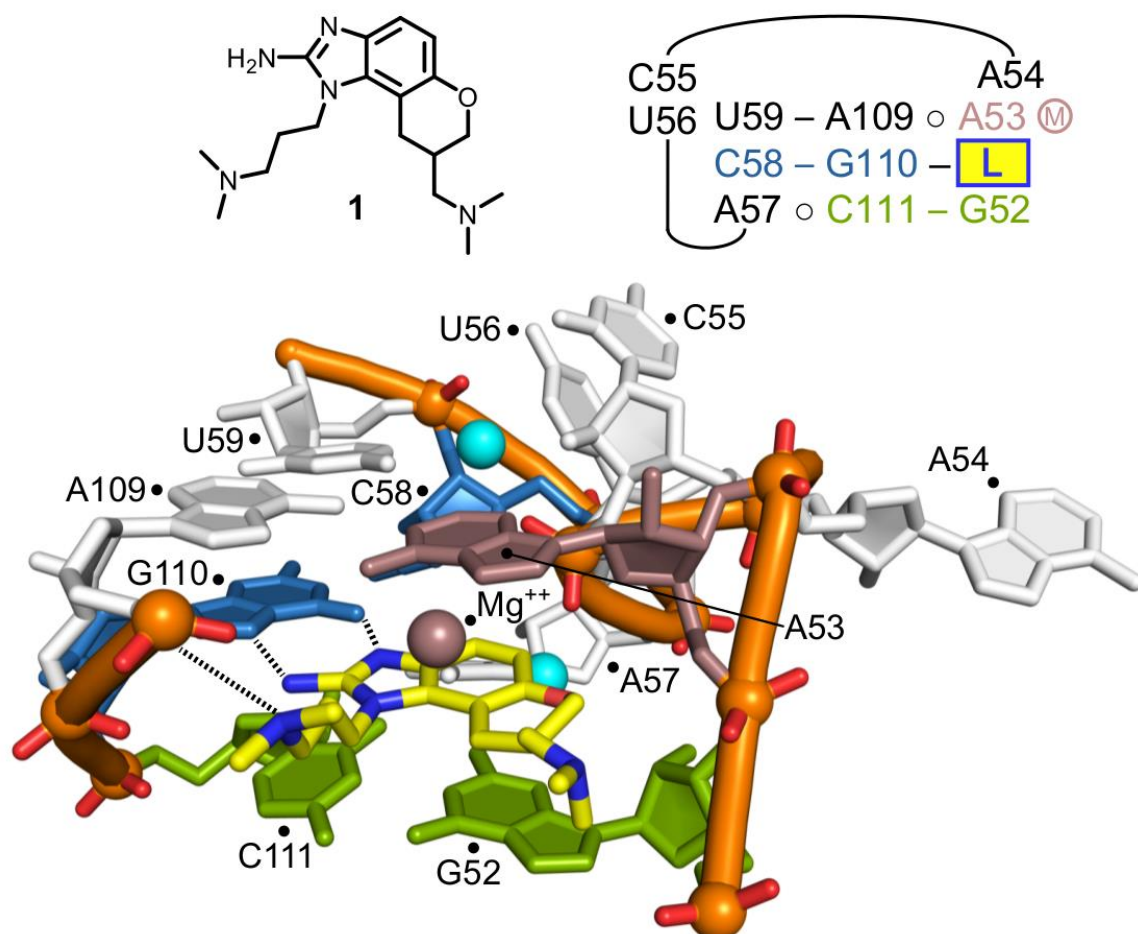


Figure 3.4: Structure of the HCV IRES subdomain IIa RNA switch in complex with ligand **1**. The color scheme is identical to Figure 3.3 with the addition of ligand **1** in yellow and two additional magnesium ions (cyan). When the HCV IIa switch is bound by ligand **1**, closing base pairs (blue and green) and adjacent helices maintain Watson-Crick base pairing. Rearrangement of unpaired internal loop bases and magnesium ions occurs to form a deep ligand-binding pocket. The roof of the binding pocket involves the U59-A109 \circ A53 base triple, while the floor is formed by the A57 \circ C111-G52 triple. The guanidinium ion-like moiety of ligand **1** interacts with the Hoogsteen edge of G110 forming a C58-G110 \circ Ligand pseudo-base triple-like complex. The phosphate of U56 is shown coordinated to two magnesium ions (cyan). Secondary structure schematic indicates the stacking of these three base triples, and the flipped out bases U56, C55, and A54.

Discovery of the RNA switches in viral IRES elements

The archetype of these switches was discovered in the internal ribosome entry site (IRES) of hepatitis C virus (HCV). The 5' untranslated region of some eukaryotic and viral mRNAs contains a structured IRES element that initiates translation by assembling functional ribosomes directly at the start codon and bypassing the need for 5' cap recognition, ribosome scanning, as well as the requirement for most initiation factors.^{23–26} Within the HCV IRES, which is composed of several domains (Figure 3.2c), domain II is highly conserved in thousands of clinical isolates^{6,21,53} and contains an internal loop in its lower stem, subdomain IIa (Figure 3.2c), which is the target for viral translation inhibitors.⁵⁴ The three dimensional structure of subdomain IIa was determined by x-ray crystallography⁴⁷ (Figures 3.2a, 3.3a), which revealed the RNA adopting a 90° bent architecture, in agreement with cryo-EM and NMR studies.^{34,48,51} FRET experiments demonstrated that the subdomain IIa undergoes a conformational switch in the presence of benzimidazole viral translation inhibitors.^{53,54} The co-crystal structure of subdomain IIa in complex with benzimidazole inhibitor **1** showed capture of the RNA in an extended conformation (Figures 3.2b, 3.4).²¹ It was demonstrated in Chapter 2 that the ligand pocket in subdomain IIa selectively recognizes guanine and serves as a fortuitous binding site for structurally similar benzimidazoles. Based on this observation in conjunction with other circumstantial evidence,³¹ it was hypothesized that a guanosine within an RNA sequence of the viral genome or ribosomal RNA acts as a trigger of the IIa conformational switch.³⁹

A number of other viruses in the flavivirus and picornavirus families also contain HCV-like IRES elements with varying degrees of structural and functional similarity but

only limited sequence conservation.^{39,56} Despite the differences, most of these IRES elements are organized into similar domains and contain analogous domain II elements, twelve of which will be discussed here. The taxonomy of these viruses is shown in Table 3.1. The subdomain IIb hairpin loop and internal loop E motifs (Figure 3.2c) are conserved in hepatitis C virus (HCV), classical swine fever virus (CSFV), bovine viral diarrhea virus (BVDV), avian encephalomyelitis virus (AEV), border disease virus (BDV), non-primate hepacivirus (NPHV), and giraffe pestivirus (GPV) while Seneca Valley virus (SVV, senecavirus), simian picornavirus (SPV), duck picornavirus (DPV), GB Virus (GBV), and duck hepatitis virus (DHV) maintain sequence conservation of their loop E motif but not the IIb hairpin loop. All twelve of these viruses contain a subdomain IIa internal loop switch structure in the 5'- or 3'-proximal strand of the lower domain II stem. The switch motifs range in size from 9-12 bases, 3 to 6 of which are unpaired, and display little to no sequence conservation (Figure 3.5).³⁹

Table 3.1: Taxonomy of viruses studied here. These viruses fall into two families and seven genera.

Species	Proposed Species	Family	Genus	Proposed Genus
HCV		Flaviviridae	Hepacivirus	
GBV-B		Flaviviridae	Hepacivirus	
NPHV		Flaviviridae	Pegivirus	
BVDV		Flaviviridae	Pestivirus	
CSFV		Flaviviridae	Pestivirus	
BDV		Flaviviridae	Pestivirus	
GPV (Giraffe)		Flaviviridae	Pestivirus	
SVV	Senecavirus A	Picornaviridae	Senecavirus	
AEV (Tremovirus)	Tremovirus A	Picornaviridae	Tremovirus	
DHV		Picornaviridae	Enterovirus	Avihepatovirus
DPV		Picornaviridae	Sapelovirus	Anatovirus
SPV	Simian Sapelovirus 3	Picornaviridae	Sapelovirus	

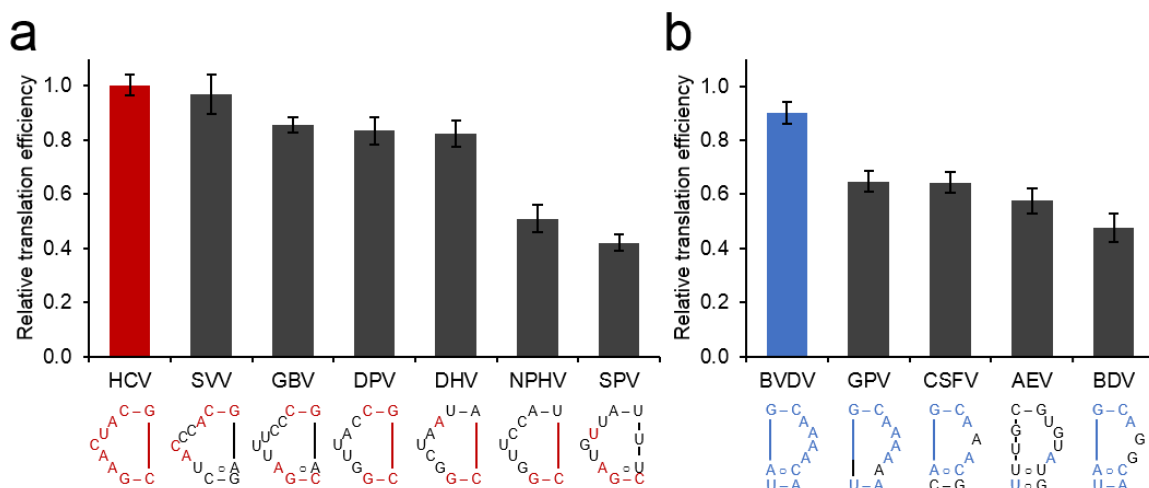


Figure 3.5: Functional activity and secondary structures of IRES IIA switches. a) In vitro translation activity of HCV IRES chimeras with subdomain IIA replacement by corresponding HCV-like viral IRES IIA elements. Sequences conserved with HCV are shown in red. Secondary structure for SVV is drawn to indicate how bases overlay with HCV in three-dimensional space as seen in the crystal structures. Other IIA element secondary structures are drawn as predicted to overlay with HCV IIA. b) In vitro translation activity of HCV IRES chimeras with domain II replacement with BVDV and subdomain IIA replacement by corresponding BVDV-like viral IRES IIA elements. Values were normalized to HCV wt. Sequences conserved with BVDV are shown in blue. IIA element secondary structures are drawn as predicted to overlay with BVDV IIA. NCBI reference sequences: HCV (NC_004102), SVV (NC_011349.1), GBV (NC_001655.1), DPV (AY563023.1), DHV (DQ249299.1), NPHV (JQ434001.1), SPV (AY064717.2), BVDV (NC_001461.1), GPV (NC_003678.1), CSFV (NC_002657.1), AEV (NC_003990.1), BDV (NC_003679.1).

Structure of the ligand-free RNA switches in HCV and SVV

Sequence and structure analyses of HCV-like IRES elements have allowed us to develop consensus models of domain II secondary structure and subdomain IIA switch motifs. X-ray crystallography has been used to determine the three-dimensional structure of the HCV and SVV switches in the ligand-free state^{39,47} and the HCV switch in complex with a viral translation inhibitor ligand. The CSFV switch has been characterized by NMR spectroscopy.⁴⁴

While HCV and SVV Ila switches have quite different sequences and predicted secondary structures (Figure 3.1a), their overall three dimensional structures, as determined by x-ray crystallography, are almost identical (Figure 3.3a, b).^{39,53} Both switches have a C-G pair (blue, Figure 3.3) closing the internal loop at the upper flanking helix. For SVV, this C-G pair is formed by G425 from the 3'-proximal side of the Ila stem, crossing over and forming a Watson-Crick base pair with C375, which diverges from the predicted secondary structure (Figure 3.1b) and inverts local strand directionality between G423 and G426 (Figure 3.3b). Mutational analysis revealed that this crossover Watson-Crick C375-G425 base pair must be a C-G or G-C pair, as changes to a U-A or any non-pairing bases resulted in a ~40% decrease in translation activity (Figure 3.6). The Ila switches from both HCV and SVV share an A base (A57 in HCV, A374 in SVV) stacking beneath the C (C58 in HCV, C375 in SVV) of the closing base pair. In SVV, two C residues (C373, C372) stack beneath A374, while in HCV, a single U base (U56) fills this same space and packs against the ribose of A57. This upper part of the switch, including half of the Ila internal loop and the flanking upper helix, is oriented at a 90° angle relative to the lower helix which connects through perpendicularly arranged bases of the lower Ila loop. The G52-C111 Watson-Crick pair closes the bottom of the internal loop in HCV, and corresponds to the C368-G426 pair in SVV (green, Figure 3.3). Stacking above the C368-G426 pair in SVV, a reverse Hoogsteen pair U369◦A424 is formed in which the Watson-Crick edge of U369 pairs with the Hoogsteen edge of A424 (mauve, Figure 3.3b). In HCV, this space is occupied by A53 which interacts with a magnesium ion (mauve, Figure 3.3a) and stacks above the G52-C111 closing pair. Two bases (A54, C55) stack above A53 in HCV, which correspond to

A370 and C371 stacking above U369 in SVV. All unpaired bases in the IIa motifs from both viruses stack continuously on either of the flanking helices with the exception of U56 in HCV whose stacking interaction may be interrupted by crystal packing, and is seen stacking in some solution conformations observed by NMR.⁴⁸ The folds of both the SVV and HCV switch motifs are unprecedented in other RNA architectures. Part of the SVV motif (C368-G426, G425, and U369○A424) resembles a UA-handle which is found as a recurring minimal building block in numerous other RNA architectures.^{101,102} Systematic mutational analyses of residues in the HCV subdomain IIa reveal that some structural sites tolerate change while others do not, generally in agreement with interactions observed for these residues in the crystal structure (Figure 3.3c). Consequently, certain combinations of multiple mutations lead to functionally fit switches while the constituting individual base changes may cause reduction in translation activity (Figure 3.5).

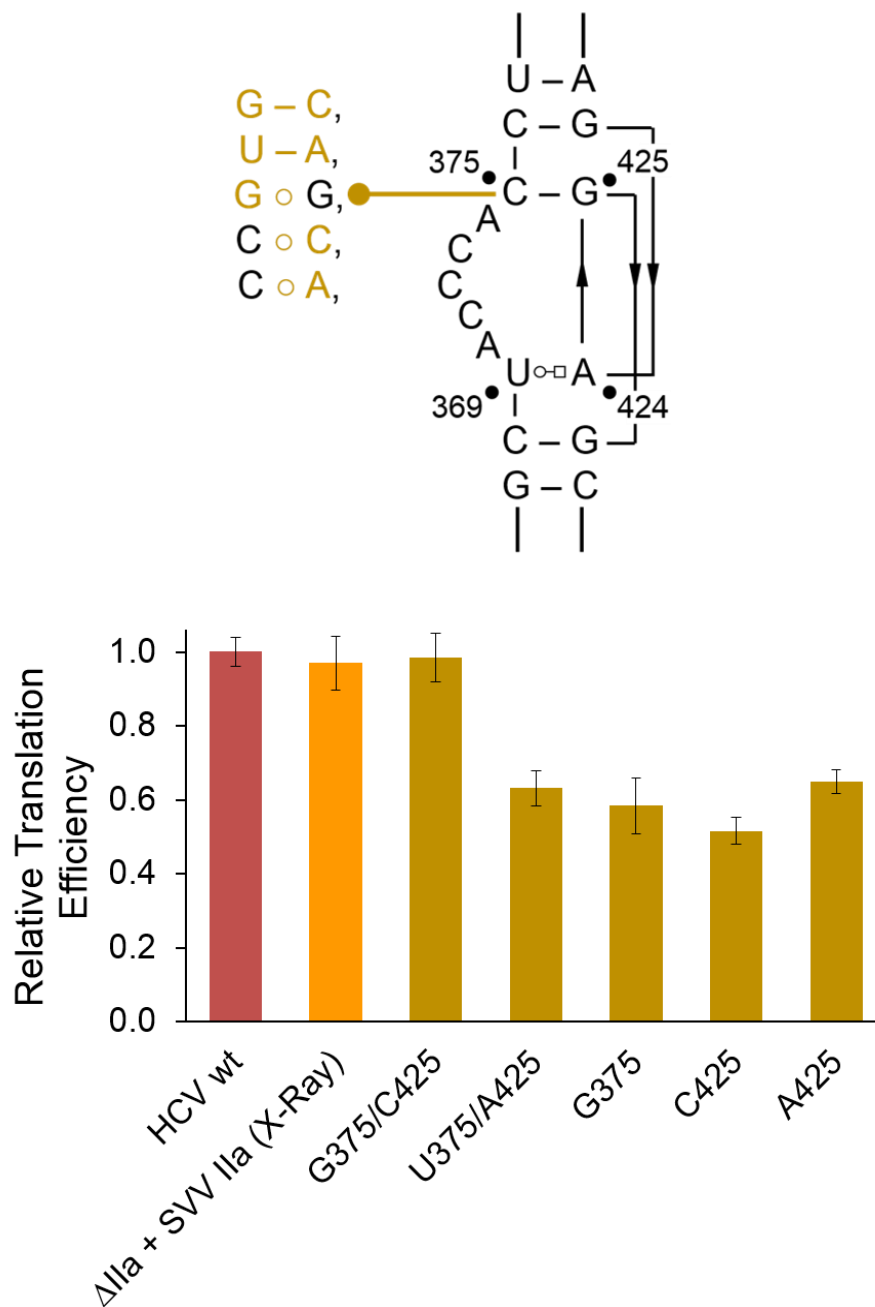


Figure 3.6: Mutational analysis of SVV Ila switch. The secondary structure of SVV Ila is shown with five mutations to the C375-G425 pair shown in yellow. In vitro translation activity of SVV Ila mutants in the context of a HCV Δ Ila + SVV Ila (X-ray) chimera IRES which has the same activity as the wild type HCV IRES. Translation efficiencies were normalized to the cap driven expression in bicistronic dual reporter constructs and the wt HCV signal was normalized to 1. Error bars represent \pm 1 s.d. calculated from triplicate experiments.

Structure of a ligand-bound switch

X-ray crystal structure analysis of the HCV IIa switch in complex with a viral translation inhibitor ligand revealed an elongated conformation of the RNA which maintains all base pairs seen in the bent conformation.²¹ A deep ligand binding pocket is formed by the rearrangement of magnesium ions and unpaired internal loop bases to form additional base triple interactions. The roof of the binding pocket is formed by base triple U59-A109∘A53 with the Watson-Crick edge of A53 pairing to the Hoogsteen edge of A109. The N7 atom of A53 is coordinated to a magnesium ion, unchanged from the ligand-free state of the RNA (mauve, Figure 3.4). The U59-A109 pair remains stacked above the C58-G110 pair as it does in the ligand-free state (not shown in Figure 3.3b). A53 stacks above the binding ligand (yellow, Figure 3.4) which engages in a pseudo-base triple interaction with C58-G110 (blue, Figure 3.4). The Hoogsteen edge of G110 recognizes the guanidinium cation-like moiety of the amino-benzimidazole ligand. The floor of the binding pocket is formed by base triple A57∘C111-G52, with G52 stacking beneath the ligand, and the sugar edge of C111 (green, Figure 3.4) pairing with the Watson-Crick edge of A57. U56, C55, and A54 are all flipped out of the helix with U56 and C55 stacking onto each other. The phosphate group of U56 is rotated into the back of the ligand-binding pocket, where it coordinates two magnesium ions (cyan, Figure 3.4). One of the metal ions interacts with U59 in the roof of the binding pocket while the other is coordinated to A57 in the floor. An additional hydrogen bonding interaction occurs between the protonated dimethylamino propyl side chain of the ligand and the phosphate group of A109.

Ila Switch function

Ila RNA switch motifs have been shown to facilitate viral translation initiation by assuming bent and elongated conformations, and directing the Iib hairpin to the E-site at the ribosomal subunit interface.^{34,39,51} We recently demonstrated that switching between bent and elongated conformations occurs in Ila switches from diverse viruses including HCV, BVDV, CSFV, SVV, and AEV by FRET.^{39,53} Additionally, Ila switch functionality in initiating HCV IRES translation has been demonstrated using an *in vitro* translation (IVT) assay and replicon-transfected human cells.³⁹ HCV chimeric IVT reporter constructs obtained by replacing HCV whole domain II with analogous domains from BVDV, CSFV, or AEV yield IRESs proficient in initiating translation, with BVDV domain II displaying near wild-type activity (blue, Figure 3.5b).³⁹ Replacement of the HCV subdomain Ila switch only with the analogous motif from SVV, guided by x-ray crystal structures, furnished a fully functional IRES both in an *in vitro* translation and HCV replicon assay, despite only marginal sequence similarity shared between the viral Ila motifs (Figures 3.1b, 3.5a).³⁹ These experiments established the distinct viral RNA Ila motifs as minimal architectural switch modules participating in a conserved biological function. Inspired by these findings, we set out to search for similar switch motifs in other viruses. Guided by a consensus secondary structure model for HCV domain II-like fragments (Chapter 2: Figure 2.7), seven new distinct switches were identified in addition to those from HCV, SVV, BVDV, CSFV, and AEV. The new switches were found in the viral IRES elements of GBV, DPV, DHV, NPHV, SPV, GPV, and BDV. All switches, in addition to those from HCV and SVV, demonstrated proficiency in translation initiation when tested in the context of HCV IRES chimeras in the IVT assay (Figure 3.5). Ila

switches with an internal loop occurring in the 5' proximal strand of the lower domain II stem were deemed to be HCV-like and were used to replace the HCV subdomain IIa in the chimeric IVT reporter construct as described above for SVV (Figure 3.5a). Ila switches with an internal loop in the 3' proximal strand of the lower domain II stem were deemed to be BVDV-like and were used to replace the BVDV subdomain IIa in the HCV/BVDV domain II chimeric IVT reporter construct as described above (Figure 3.5b). Ila switch sequences used to replace respective HCV or BVDV switches are shown in Figure 3.5. Attempts to crystallize these various viral RNA switches will be discussed in Chapter 4. A revised consensus secondary structure model for HCV domain II-like fragments was developed after this study (Figure 3.7), and includes all 12 viruses studied here as well as an additional DHV genotype which contains a different switch sequence. A number of additional potential switches have been identified using this model combined with recently published alignments and secondary structure predictions for type IV IRES elements in *Picornaviridae*.⁵⁷ Identified potential switches are found in (but not limited to):

Feline sakobuvirus, Rodent hepacivirus, Guereza hepacivirus, Swine pasivirus, Bungowannah pestivirus, Pronghorn pestivirus, Antelope pestivirus, Hobi-like pestivirus, Tunisian sheep virus, Guereza hepacivirus, Norway Rat pestivirus, Aalivirus A, Turkey hepatitis, Feline picornavirus, Porcine kobuvirus, Ferret kobuvirus, Pigeon picornavirus B, mesivirus 1 and 2, Tortoise rafivirus, Manhattan parechovirus, and Ferret parechovirus.

Further studies are needed to determine if these RNA switches function in a similar fashion. The modular nature of these distinct switches suggests that they have adapted in their different viral contexts to perform corresponding biological functions. It has been suggested that IRES architectures may have been exchanged between viral genomes by

horizontal transfer and multiple recombination events.⁵⁶ Transfer of IRES elements has been observed within viral families and may also occur between families and through capture of cellular RNA elements.⁵⁶ Widely divergent switches, such as those involving variation of the internal loop motif occurring in the 3' versus 5' proximal strand of the lower domain II stem, in distinct viral families, may also have originated from repeated adaptation of independently evolving motifs that eventually converged to provide similar functions.

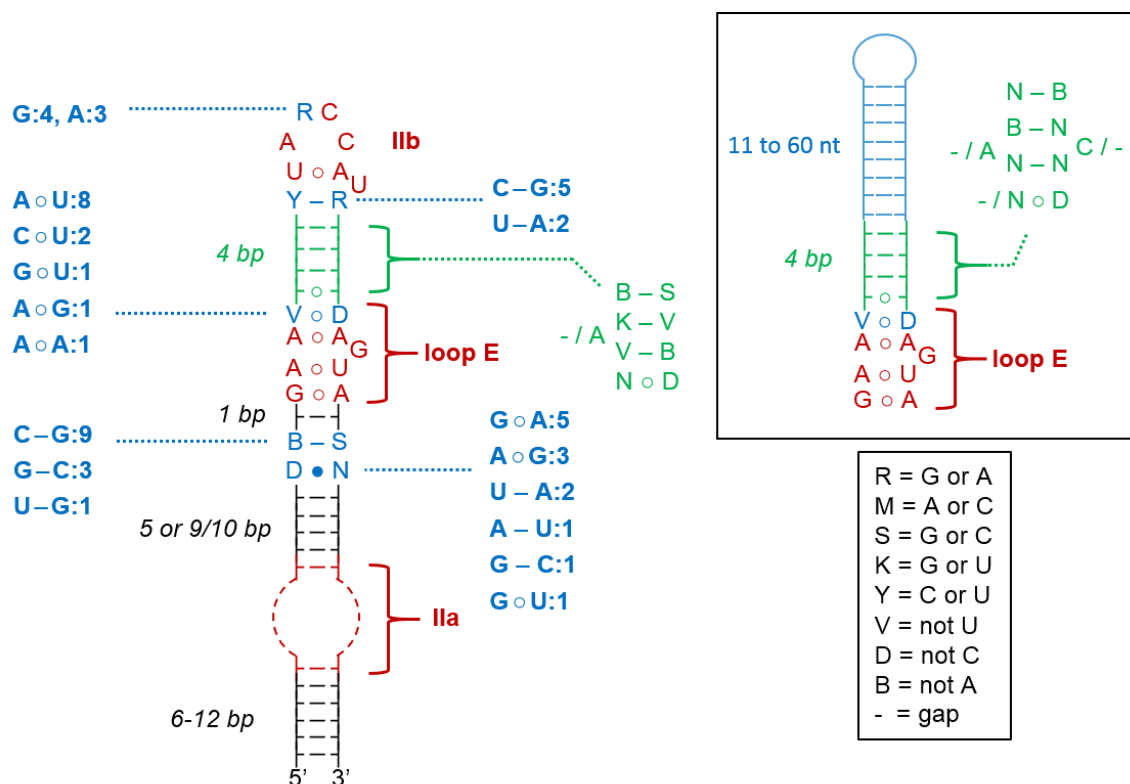


Figure 3.7: Consensus secondary structure of domain II analogs from viral IRES elements studied here. The domain II region containing loop E and below provides secondary structure and sequence information from 13 viral IRES elements which all contain the red loop E consensus sequence. The region above loop E, including the 4bp green stem and Iib hairpin, provides secondary structure and sequence information from 7 of the 13 viruses which contain loop E and Iib consensus sequences in red. Sequence and length requirements for the region above loop E for the remaining 6 viruses (with no Iib conservation) is shown in the upper right box. A legend for nucleotide letter codes for incompletely specified bases is shown in the lower right box and is based on recommendations by the Nomenclature Committee of the International Union of Biochemistry (NC-IUB).¹⁰³ The internal loop of subdomain Ila consists of an unpaired stretch of nucleotides in either the 5' or 3' proximal strand of the lower stem and examples can be seen in Figure 3.5.

Significance and applications of the Ila switches

The IRES subdomain Ila switch has been exploited as a therapeutic target to selectively suppress viral translation with synthetic inhibitors that capture specific

conformations of the RNA.^{6,39,53,55,65,66} High levels of conservation in HCV clinical isolates and the presence of a deep solvent-excluded ligand binding pocket render the Ila RNA switch a target for the development of antiviral drugs. In the context of viral IRES function, the Ila switches are involved in the dynamic interplay of ribosomal subunit recruitment and the transitioning from translation initiation to elongation,^{34–36,42–44,46,51,63} which are facilitated by transient interaction with a G residue that captures a conformational change in the Ila RNA.³⁹

It is conceivable that similar ligand-responsive switches occur in other noncoding RNAs to regulate interactions with nucleic acid or protein partners. Mechanical switches that do not undergo global changes in secondary structure, involving the kinetically slow disruption and formation of base pairs as is common in riboswitches, may be preferred for transient regulatory or pausing events that are quickly reversible. While the dynamics of IRES binding and ribosome release during initiation is not yet fully understood, it has been observed that ribosomes stall on the IRES in the presence of Ila-targeting inhibitor ligands or when the Ila switch is deleted.^{6,44,46} Ligand-responsive mechanical switches provide a simple mechanism to regulate binding and release events of RNA-RNA and RNA-protein interactions contingent on the presence of a trigger ligand.

The Ila RNA switch motifs have also been exploited as rectangular building blocks to rationally design and construct nanostructures with a variety of potential applications. We used the HCV Ila switch to build a self-assembling RNA square that adopts a well-defined structure as revealed by x-ray structure analysis.¹⁰⁴ Recently, the same RNA switch has been incorporated into a self-assembling nanoprism.¹⁰⁵ Other RNA switches, including the previously discovered SVV subdomain Ila³⁹ and the Ila-like viral

motifs described here, provide a growing tool box of bent RNA motifs to use as building blocks for complex RNA nanostructures that self-assemble from short oligonucleotides and have the potential to change conformation in response to selective ligand binding. These applications will be discussed in greater detail in Chapter 5.

Conclusions

In conclusion, we have described what are currently the simplest known ligand-dependent mechanical switching motifs found in noncoding RNA architectures. The HCV Ila switch has been characterized by x-ray crystallography for both the ligand-free and ligand-bound conformation. The crystal structure of a functionally analogous Ila switch from SVV in its ligand-free conformation was described, revealing structural and functional interchangeability with the archetypical HCV switch despite a lack of sequence conservation. In addition to HCV and SVV, Ila-like RNA switches were discovered in ten other viruses and were able to functionally replace the HCV Ila switch to create translationally competent IRES elements. These switches contain a recognition site for a guanosine residue and are fortuitous targets for synthetic viral translation inhibitors. Unlike metabolite-sensing riboswitches, these ligand-responsive mechanical switches represent a new class of simple RNA modules that are structurally well-defined in both ligand-free and bound states despite their small size. The viral subdomain Ila-like modules may represent the simplest form of ligand-responsive mechanical switches in nucleic acids. While functioning in a dynamic fashion during viral translation initiation and elongation, these RNA motifs represent a unique therapeutic target as well as well-defined ligand-responsive building blocks for the construction of RNA nanostructures.

Materials and Methods

In Vitro Translation Experiments

The in vitro translation assay (IVT) was performed using the TNT Quick coupled reticulocyte lysate system (Promega, Madison, WI) and an HCV bicistronic luciferase reporter as previously described.⁹⁰ The bicistronic luciferase reporter contains the sequence coding for the HCV IRES-Renilla luciferase preceded by a cap-initiated firefly luciferase internal control. Briefly, reactions were carried out according to the manufacturer's instruction at a volume of 7.5 μ L, containing 1.5 μ L reporter DNA plasmid (100ng/ μ L), 1.5 μ L H₂O, and 4.5 μ L reaction buffer containing reticulocyte lysate, SP6 polymerase, RNase inhibitor, and amino acids. Detection of firefly and Renilla luciferase levels was done using the Dual-Glo Luciferase Assay System (Promega, Madison, WI) as previously described.⁷² Relative translation efficiencies were calculated as a ratio of IRES-driven Renilla luciferase levels to the internal control firefly luciferase levels.

HCV Bicistronic Reporter Mutagenesis

Mutations were introduced into the HCV bicistronic luciferase reporter DNA plasmid via site directed mutagenesis (Q5 Site-Directed Mutagenesis Kit, New England Biolabs, Ipswich, MA). Construction of the HCV Δ Ila + SVV Ila (X-ray) and HCV Δ II + BVDV Ila constructs are described in Chapter 2. Site-directed mutagenesis experiments were carried out according to the manufacturer's instructions. The sequences of all mutant bicistronic reporter plasmids were verified by automated DNA sequencing. Mutational studies were conducted using the IVT assay as described above. Relative translation efficiencies were normalized to HCV WT levels.

Acknowledgements

The material in Chapter 2, in part, has been published in RNA Biology (2015, 12, 780-786) titled *Ligand-responsive RNA mechanical switches* with Thomas Hermann as co-author. The dissertation author was the primary investigator and author of this material.

Thanks to Andrew Bergdorf for help with several mutagenesis experiments and for helping to compile sequence data for Figure 3.7 model.

Chapter 4:

Tools and Techniques for Studying RNA Structure, Folding, and Ligand Binding

Parts of this work are published in:

“Conformational flexibility of viral RNA switches studied by FRET” *Methods*, 91, 35–39 (2015).

Parts of this work are in preparation for submission to be published in:

“Fluorescence detection of RNA-ligand binding and crystal structure determination of ribosomal decoding site RNA using a heavy atom containing fluorescent ribonucleoside” (2016).

Abstract

RNA molecules accomplish a diverse array of functions by adopting unique structures, undergoing large or small conformational changes, and binding small molecule, nucleic acid, or protein ligands. Many tools and techniques have been developed or adapted to study RNA structure, conformational changes, and ligand binding. Here, the use of X-ray crystallography has been used to determine the structure of four RNA architectures. Förster resonance energy transfer (FRET) experiments involving the conjugation of fluorophores to RNA or DNA have been used to study RNA folding and conformational ligand capture. A new heavy atom containing fluorescent nucleoside analog has been used to fluorescently measure ligand binding to model RNA systems and to provide phase information in determining a crystal structure without disrupting native RNA folding

Introduction

RNA crystallography

A growing knowledge and understanding of the structures of biomacromolecules has played a pivotal role in understanding their diverse functions. Among the methods which provide structural information about biomolecules, X-ray crystallography provides the highest resolution unambiguous molecular snapshot of a biomolecule's structure. While not every biomacromolecule is amenable to crystallography, this method has been widely used to study molecular structure, function, and ligand binding. X-ray crystallography has been enormously successful in determining protein structures, and there are currently 99,437 protein-containing X-ray crystal structure entries in the Protein

Data Bank (redundancy not considered).¹⁰⁶ While there are currently only 1729 nucleic acid-containing X-ray crystal structure entries in the Protein Data Bank,¹⁰⁷ this number has grown rapidly in the past two decades, and there has been increasing success in determining nucleic acid structures by X-ray crystallography.

Förster resonance energy transfer (FRET)

Many tools utilizing fluorescence properties have been developed or adapted to study RNA folding and ligand binding. Förster resonance energy transfer (FRET) experiments involve a distance-dependent energy transfer between fluorophores, and can be used to monitor conformational changes in functional RNA molecules. In this study, RNA FRET constructs for newly identified viral IRES IIa switches were constructed with 5' terminal Cy3 or Cy5 modification and used to characterize switch folding and conformational capture by ligands (Figure 4.1).³⁹ A modular system was also developed to screen a potential larger number of RNA switches utilizing universal FRET labelled DNA probes hybridizing to a core RNA switch module.⁶²

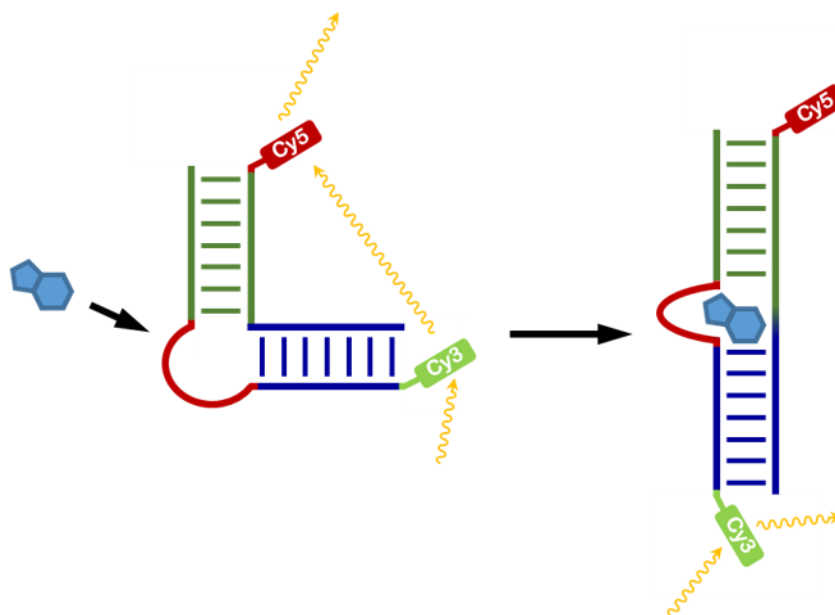


Figure 4.1: RNA switch FRET experiments. Cartoon showing dye-labeled RNA FRET construct demonstrating maximum FRET efficiency for an RNA switch in a bent conformation and maximum Cy3 fluorescence in a ligand-captured elongated conformation.

Heavy atom containing fluorescent ribonucleoside analog

Fluorescent nucleoside analogs which retain base pairing abilities and can be incorporated chemically or enzymatically into nucleic acid oligonucleotides have been developed as probes to study local dynamics, conformational changes, and ligand binding of functional RNA molecules. In another vein, heavy atom containing nucleoside analogs which also retain base pairing abilities and can be incorporated chemically or enzymatically into nucleic acid oligonucleotides have been developed to provide initial phase estimates in solving novel nucleic acid X-ray crystal structures. In this study, a selenophene-conjugated uridine analog ($^{\text{Se}}\text{U}$) was developed as a dual function probe with fluorescence properties able to detect RNA conformational changes and heavy atom anomalous scattering properties useful for phase determination in crystal structure

determination. This probe was incorporated into well studied model RNA constructs of the HCV IRES IIa RNA switch (Chapters 1-3) and the bacterial ribosomal decoding site (A-site) RNA.^{60,61} These ^{Se}U-modified model constructs both successfully monitor ligand binding by fluorescence changes. In addition, the 3D structure of the ^{Se}U-modified A-site RNA construct was determined by X-ray crystallography and revealed this ^{Se}U to be nondisruptive of native RNA structure.^{60,61}

Results and Discussion

RNA crystallography

As mentioned previously, the Hermann lab has utilized X-ray crystallography to determine the ligand-free and bound structures of the HCV IIa switch.^{21,47} Here, attempts were made to crystallize analogous IIa switches in other viral IRES elements. Constructs for identified switches in bovine viral diarrhea virus (BVDV), classical swine fever virus (CSFV), Seneca Valley virus (SVV, senecavirus), avian encephalomyelitis virus (AEV), simian picornavirus (SPV), duck hepatitis virus (DHV), GB Virus (GBV), and giraffe pestivirus (GPV) were designed and tested for their amenability to crystallize (Figures 4.2-4.6). Switches from BVDV, CSFV, AEV, and GPV are all located in the 3' proximal strand of the lower stem of domain II. Switches from SVV, SPV, DHV, and GBV are all located in the 5' proximal strand of the lower stem of domain II, similar to the HCV switch.

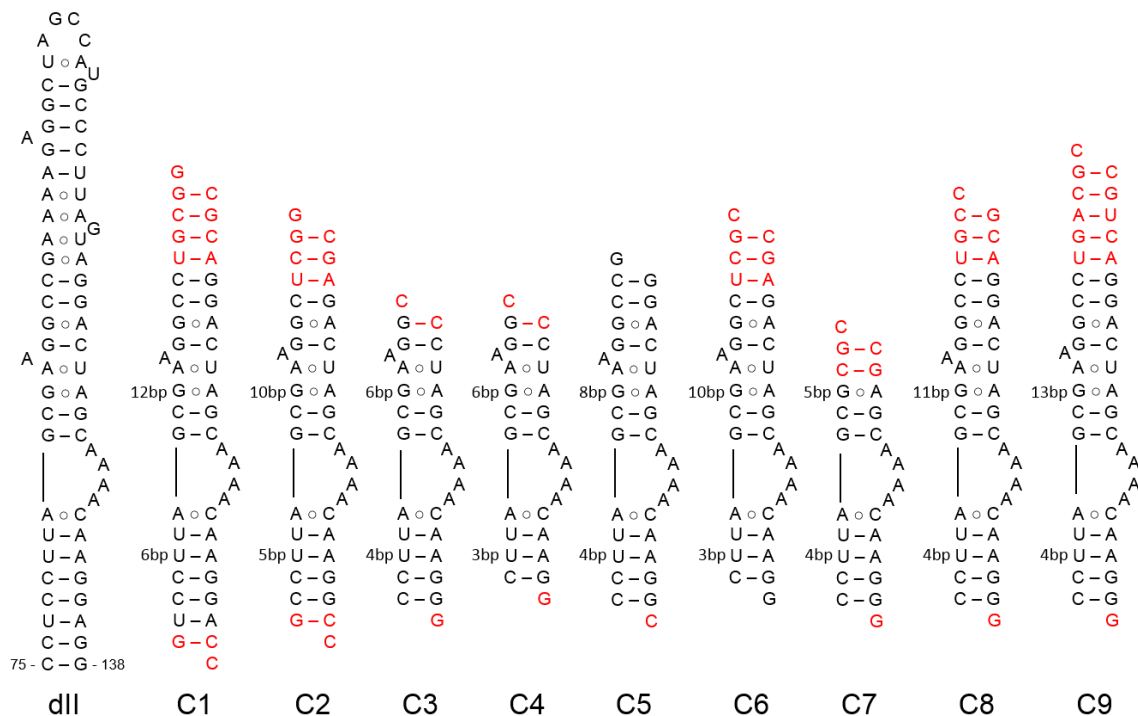


Figure 4.2: BVDV IRES subdomain IIa switch RNA crystal constructs. The full domain II (dII) secondary structure is shown for reference. Sequences that deviate from wild type BVDV sequence in the crystal constructs are highlighted in red. BVDV-C1 is a similar size to the HCV Ila construct which crystallized, diffracted, and had its structure determined.⁴⁷ BVDV-C7 is a similar size to the short SVV Ila construct (SVV-C6, Figure 4.4) which crystallized, diffracted, and had its structure determined.³⁹ NCBI reference sequence: BVDV (NC_001461.1).

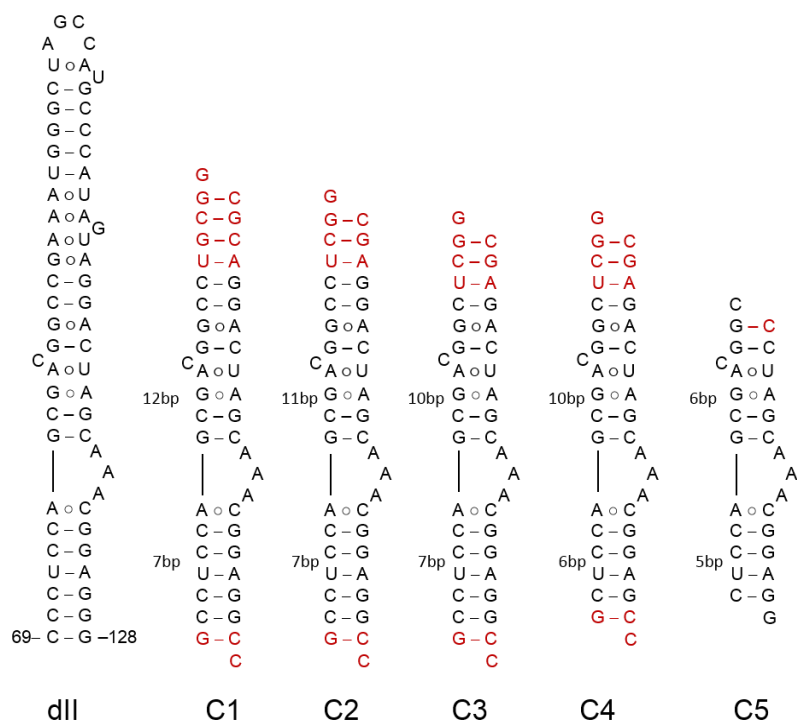


Figure 4.3: CSFV IRES subdomain IIa switch RNA crystal constructs. The full domain II (dII) secondary structure is shown for reference. Sequences that deviate from wild type CSFV sequence in the crystal constructs are highlighted in red. NCBI reference sequence: CSFV (NC_002657.1).

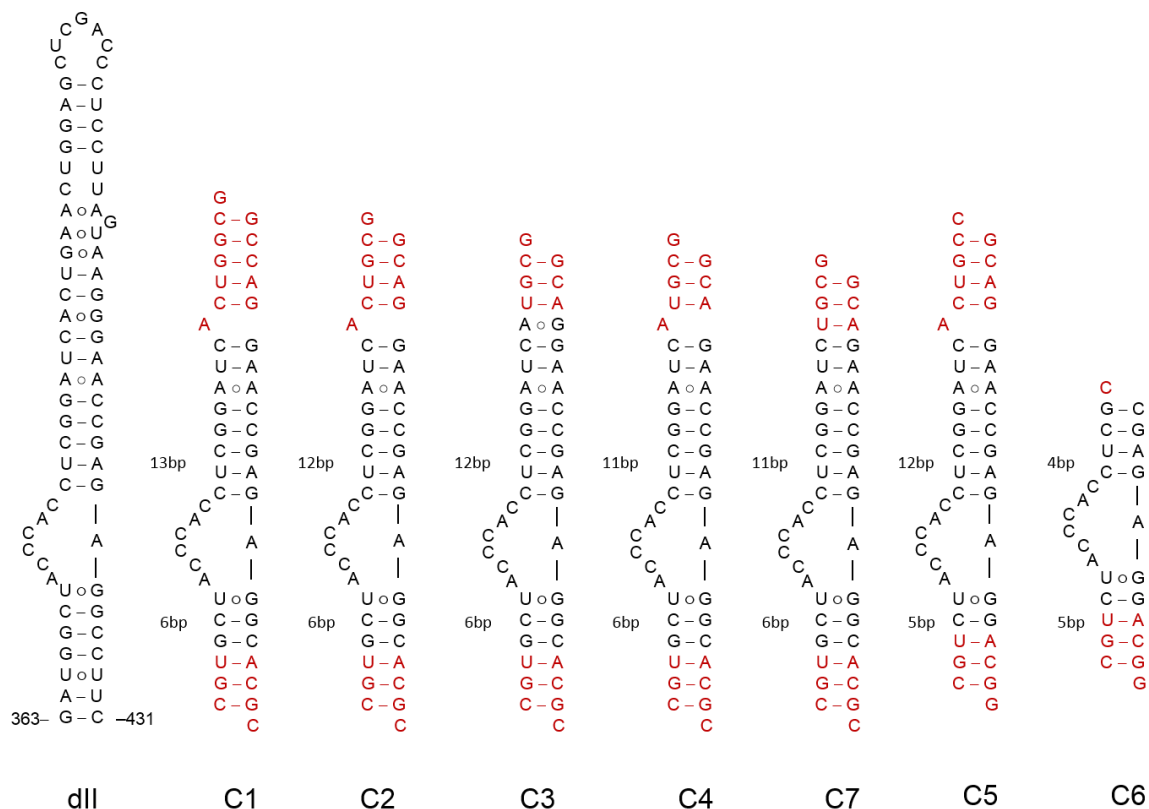


Figure 4.4: SVV IRES subdomain IIa switch RNA crystal constructs. The full domain II (dII) secondary structure is shown for reference. Sequences that deviate from wild type SVV sequence in the crystal constructs are highlighted in red. SVV-C4 and C7 are a similar size to the HCV Ila construct which crystallized, diffracted, and had its structure determined.⁴⁷ SVV-C2 crystallized, diffracted, and its structure was determined at 3.2Å (PDB entry 4PHY; Chapter 2: Figure 2.12, Table 2.2).³⁹ SVV-C6 crystallized, diffracted, and its structure was determined at 1.86Å (PDB entry 4P97; Chapter 2: Figures 2.10-2.11, Table 2.1).³⁹ NCBI reference sequence: SVV (NC_011349.1).

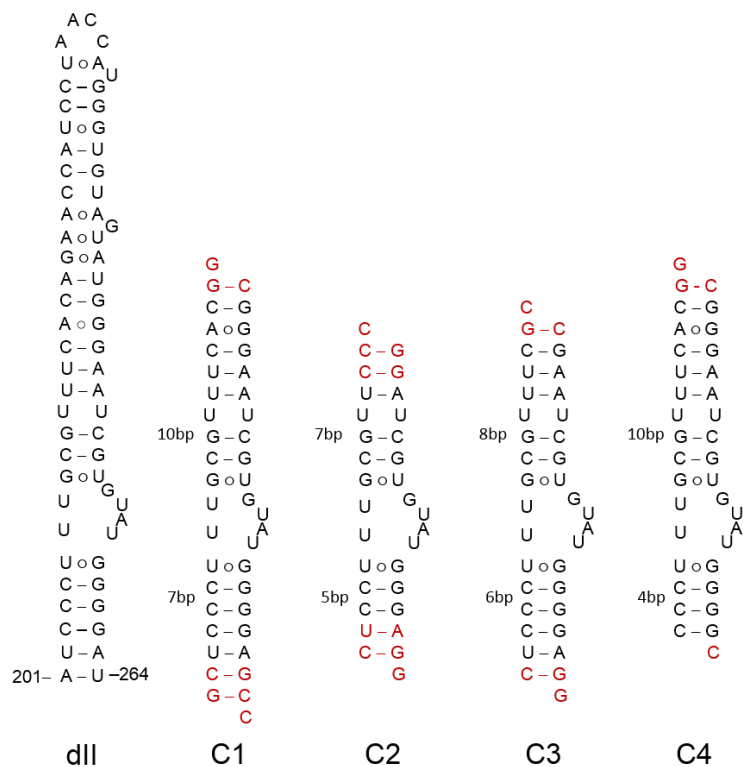


Figure 4.5: AEV IRES subdomain IIa switch RNA crystal constructs. The full domain II (dII) secondary structure is shown for reference. Sequences that deviate from wild type AEV sequence in the crystal constructs are highlighted in red. NCBI reference sequence: AEV (NC_003990.1).

Figure 4.6: Viral IRES subdomain IIa switch RNA crystal constructs. Domain II (dII) secondary structure predictions are shown for reference. Hairpin IIb and loop E sequences that are conserved with HCV are outlined. Sequences that deviate from wild type sequences in the crystal constructs are highlighted in red. a) SPV IRES domain II and crystal constructs. b) DHV IRES domain II and crystal constructs. c) GBV IRES domain II and crystal constructs. GBV domain II (dII*) prediction shown here is only one stem and hairpin from a much larger predicted domain II and the prediction here deviates greatly from previous predictions.¹⁰⁸ d) GPV IRES domain II and crystal constructs. NCBI reference sequences: SPV (AY064717.2), DHV (DQ249299.1), GBV (NC_001655.1), GPV (NC_003678.1).

Short bipartite RNA model constructs were designed for viral IRES IIa switches. For each viral switch, constructs were designed one or several at a time, and individual chemically synthesized RNA oligonucleotide constructs were screened one or two at a time in crystallization trials. Original switch constructs were designed to be similar to successfully crystallized constructs from the model HCV switch.^{21,47} These constructs contain the IIa switch internal loop flanked by two double stranded helices. They also include one or two overhanging nucleotides on the 5' or 3' termini to aid intermolecular interactions in crystal packing. In subsequent constructs for each viral switch, base pairs were added or removed from one or both stems flanking the internal loop. After success crystallizing and determining the structures of two SVV constructs (C2 and C6, Figure 4.4),³⁹ other viral switch constructs were designed to be similar to these. As the SVV-C6 construct was the shortest (single strands of 11 and 16 nucleotides) and cheapest to chemically synthesize of these two SVV constructs as well as the two successfully crystallized HCV switch constructs, efforts were focused on designing SVV-C6-similar constructs for other viral switches (BVDV-C7, Figure 4.2; CSFV-C5, Figure 4.3; SPV-C1, DHV-C1, GBV-C1, and GPV-C1, Figure 4.6) Weighted exploration of “construct space” (constructs of different size, shape, and containing various crystallization facilitation motifs) versus “condition space” (different crystallization conditions)¹⁰⁹ led to greater success in these studies.^{39,59,61}

All RNA switch constructs for BVDV, CSFV, AEV, GBV, and GPV (Figures 4.2, 4.3, 4.5, 4.6) yielded no or poorly diffracting crystals. The second construct from SVV (SVV-C2, Figure 4.4) yielded crystals which diffracted to 3.2Å. This medium resolution structure was partially determined, and later further refined after a high resolution

structure of SVV-C6 was determined at 1.86Å.³⁹ Highly diffracting crystals were also grown for DHV-C1 ($\geq 1.86\text{\AA}$, Figure 4.6). Datasets were collected for several DHV-C1 crystals, but none of these full datasets were able to be indexed and integrated into a single space group. Diffracting crystals were also grown for SPV-C1 ($\geq 3.1\text{\AA}$, Figure 4.6), and datasets were collected for several crystals. Several of these were successfully indexed and processed, but molecular replacement attempts have yet yielded no structure solutions. Efforts are ongoing to grow better diffracting SPV-C1 crystals.

The crystal structure of selenophene-modified uridine incorporated into a model RNA construct of the ribosomal A-site will be discussed later in this chapter (see *Ribosomal A-site RNA with an incorporated heavy atom containing fluorescent ribonucleoside analog*). The crystal structure of a self-assembling nanotriangle will be discussed in Chapter 5.

Förster resonance energy transfer (FRET)

Model oligonucleotides representing viral IRES switches enable the investigation of conformational changes in isolation from the background of the complex IRES-ribosome interaction. In previous studies, the crystal structure of the HCV IRES subdomain IIa switch was used to design RNA constructs carrying 5' terminal cyanine dye labels for FRET experiments (Figure 4.7a).⁵³ The length of the stems flanking the internal loop was chosen such that the distance of the cyanine-modified 5' termini in the bent state of the switch is close to the Förster radius of the dye pair. Therefore, maximum FRET signal is achieved for the folded switch in the ligand-free state. RNA modular FRET constructs were designed from optimized conventional model FRET

constructs^{53,110} with several modifications. The lower stem of this construct was lengthened from seven to eight base pairs to accommodate a Cy3-conjugated DNA oligonucleotide of five bases while maintaining three Watson-Crick base pairs on the lower side of the internal loop for stability of the central RNA core. The upper stem was shortened from thirteen to twelve base pairs to partially offset the change in distance between dyes due to the lengthening of the lower stem. Additionally, the U○U closing pair of the RNA core's upper stem was changed to a Watson-Crick G–C pair for added construct stability. We chose a length of five bases for the dye-modified DNA oligonucleotides as we thought this might be the minimum length required for stable hybridization to the RNA core.

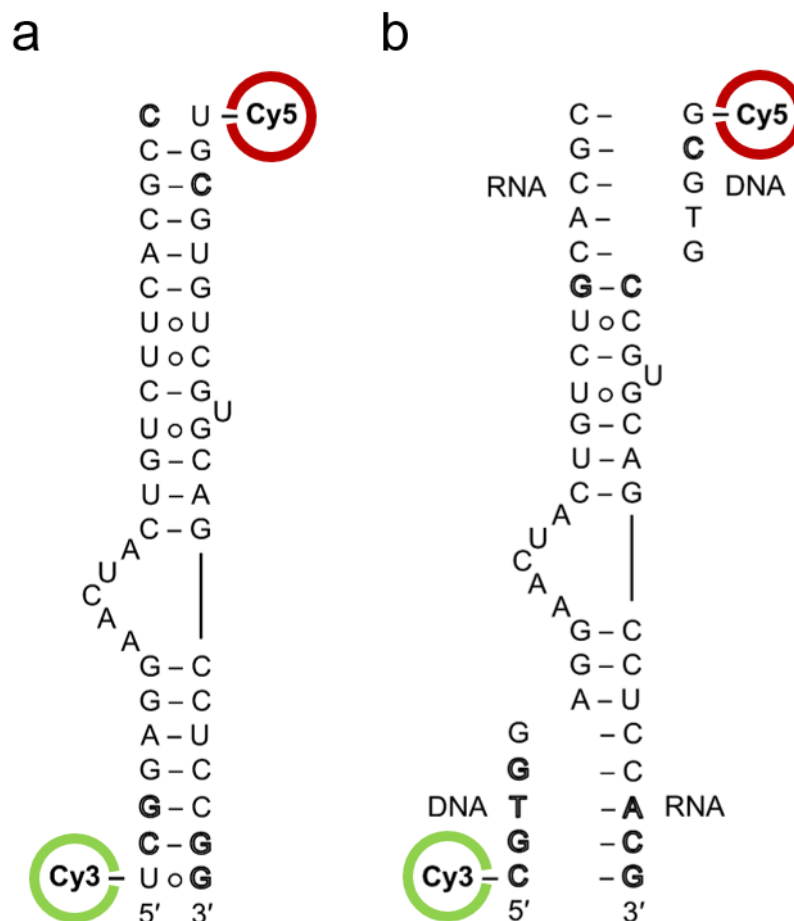


Figure 4.7: Secondary structure and folding of viral RNA switches monitored by FRET experiments with cyanine dye-labeled oligonucleotides. a) Secondary structure of the previously developed, conventional model construct of the HCV IRES subdomain IIA switch 5' terminally conjugated with Cy3 and Cy5 dyes. b) Modular FRET construct consisting of unmodified oligonucleotides that contain the HCV RNA switch and carry overhanging single strands which hybridize with cyanine dye-conjugated DNA oligonucleotides. In both panels A and B, outlined letters represent nucleotides deviating from HCV genotype 1b sequence.

To validate the FRET approach, folding of the RNA switch from HCV induced by metal ion binding was investigated. The crystal structure of the HCV IRES subdomain IIA revealed three tightly bound Mg^{2+} ions which are integral to the RNA switch architecture. We anticipated that the internal loop of the subdomain IIA RNA may not be stably folded

in the absence of Mg^{2+} , leaving the flanking helices arranged in an overall extended conformation and separating the terminally attached dyes beyond the Förster radius. Folding of the switch RNA was induced by addition of an increasing amount of Mg^{2+} and monitored by FRET measurement of the Cy5 acceptor dye emission while exciting the Cy3 donor. A dose-dependent increase of the FRET signal was observed which indicated localization of the dye labels within the Förster radius due to formation of the bent RNA switch architecture (Figure 4.8). The apparent EC50 value of the Mg^{2+} dose response for folding of the HCV subdomain IIa RNA was determined at $730 \pm 40 \mu M$, comparable to previously measured values with similar constructs.^{39,53} Corresponding dye-labeled model oligonucleotides for RNA switches from four other viruses, including classic swine fever virus (CSFV), bovine viral diarrhea virus (BVDV), avian encephalomyelitis virus (AEV) and Seneca Valley virus (SVV) were designed with guidance from secondary structure comparisons with HCV subdomain IIa RNA as no high-resolution structural data has been determined for BVDV and AEV. CSFV and SVV RNA FRET construct design was also supported by NMR⁴⁴ and X-ray crystal structures.³⁹ These RNA constructs were tested in analogous FRET experiments, resulting in EC50 values for folding and Mg^{2+} binding of $130\text{-}220 \pm 20 \mu M$ (CSFV, BVDV, AEV) and $580 \pm 90 \mu M$ (SVV) (Figure 4.9b; see also discussion in Chapter 2).³⁹

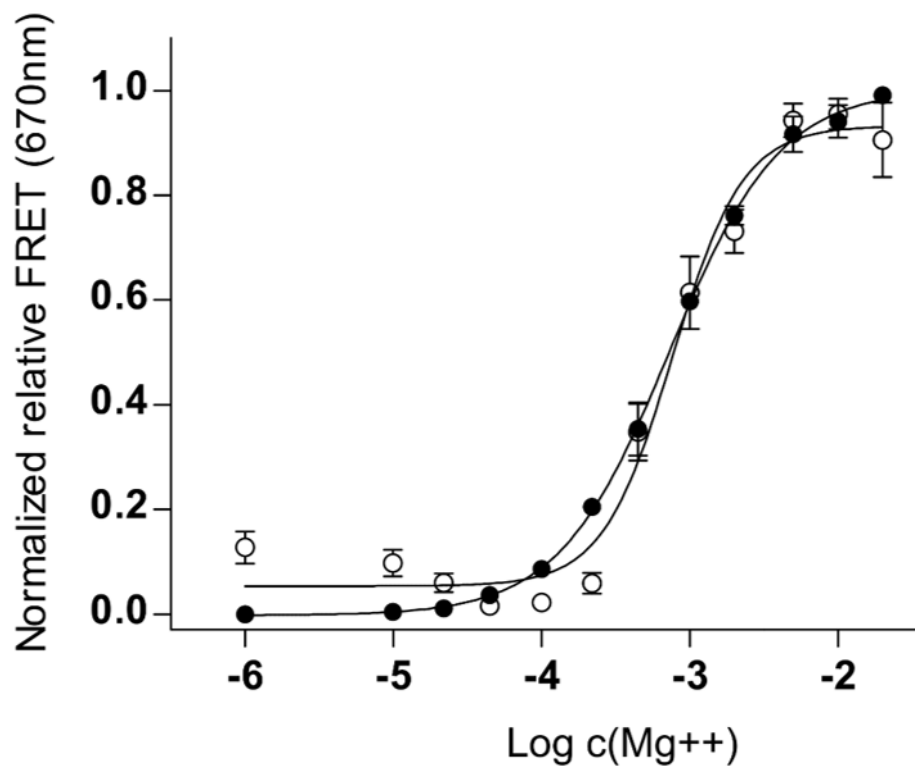


Figure 4.8: Normalized relative FRET signal for the Mg^{2+} titration of the HCV conventional (●) and modular (○) FRET constructs. Dose-response fitting curves gave EC_{50} values for Mg^{2+} dependent folding of $730 \pm 40 \mu\text{M}$ (conventional) and $770 \pm 80 \mu\text{M}$ (modular), respectively, for the two different constructs. Error bars represent ± 1 s.d. calculated from triplicate experiments.

Figure 4.9: a) Structures and sequences of dye-labeled viral IRES subdomain IIa constructs for FRET experiments. b) Mg^{2+} -induced folding of viral subdomain IIa motifs. FRET signal from terminally Cy3/Cy5 labeled RNA constructs was monitored while increasing Mg^{2+} concentration. Fitting of dose–response curves resulted in EC50 values for Mg^{2+} -induced folding at $598 \pm 25 \mu M$ (HCV, ■) (23), $196 \pm 21 \mu M$ (CSFV, ○), $219 \pm 19 \mu M$ (BVDV, ▼), $131 \pm 21 \mu M$ (AEV, Δ), and $579 \pm 86 \mu M$ (SVV, ●). Error bars represent ± 1 SD calculated from triplicate experiments.

While these previous investigations required the preparation of covalently dye-modified oligonucleotides for each viral IRES switch, we have now developed a modular FRET approach in which unmodified model RNA constructs are hybridized with dye-labeled DNA oligonucleotides (Figure 4.7b). The modular design for the HCV RNA switch was tested in the salt-induced folding experiment. Titration with Mg^{2+} resulted in a dose-dependent increase of the FRET signal with an EC_{50} value of $770 \pm 80 \mu M$ which was in excellent agreement with the affinity determined with the conventional construct (Figure 4.8).

The modular constructs will be advantageous to investigate conformational dynamics of other candidate RNA switches, especially those identified in subdomain IIa of various IRES-containing viruses (Figure 4.10; see also Chapter 3),⁵⁸ since the same pair of universal Cy3- and Cy5-conjugated DNA sequences can be used to prepare various dye-labeled FRET constructs. The cyanine dye-conjugated DNA strands are only 5 nucleotides in length and therefore readily prepared by robust chemical synthesis from which they are obtained in high purity by HPLC and significantly more cost effective than longer modified RNA oligonucleotides. Labeling of RNA by hybridization with dye-modified DNA oligonucleotides has been used extensively in fluorescence studies of complex RNA architectures, such as the ribosome, for which site-specific covalent modification of large nucleic acid components is difficult to achieve.^{110,111} A prime concern associated with the fluorescent probe hybridization approach is the integrity of the target RNA structure which ideally is only minimally disturbed by binding of the dye-conjugated DNA oligonucleotide. In the modular FRET constructs of viral RNA switches described here, structure and conformational flexibility of the switch motif are not

affected by the DNA probes which hybridize distantly from the internal loop (Figure 4.7b), as attested by the consistent EC50 values for Mg^{2+} -induced folding of the conventional and modular RNAs (Figure 4.8). The implementation of this method for studying other RNA switches may require core RNA length modification to achieve optimal FRET distances as the required number of base pairs flanking an internal loop of interest for stability may vary. Changes in stem length may also place a dye closer or further from its FRET pair based on its helical position. Alternatively, a suspected RNA switch may not be amenable to this method of study if it does not adopt well-defined stable conformations but remains flexible.

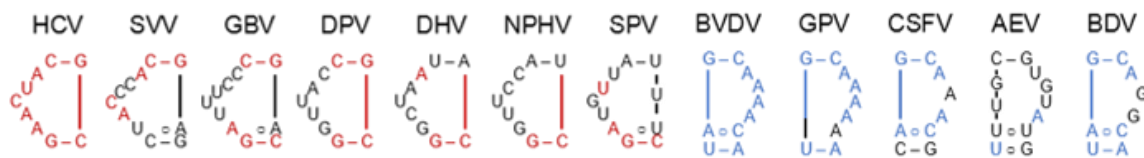


Figure 4.10: Secondary structures of functional viral IRES IIa switches.

Monitoring ligand binding at viral RNA switches

While a bent conformation is preferred by RNA switches from viral IRES elements in the ligand-free state, binding of benzimidazole translation inhibitors or guanosine captures an extended structure containing a deep pocket that encapsulates the ligand. The transition between conformational states can be monitored by FRET measurements similar to the salt-dependent folding experiments outlined in the previous section. The same model constructs have been used to investigate conformational changes during ligand binding, including the conventional dye-conjugated RNA and a modular oligonucleotide system (Figure 4.7a, b). Previously, cyanine dye-labeled RNA,

similar to the construct shown in Figure 4.7a, was instrumental in establishing target binding and elucidating the molecular mechanism of action of benzimidazole inhibitors of HCV IRES-driven translation.^{6,53} A related RNA construct was used to develop a high-throughput screen for translation inhibitor ligands targeting the HCV subdomain IIa RNA switch.⁷²

To monitor compound binding to viral RNA switches, dye-labeled oligonucleotide constructs were incubated in the presence of Mg^{2+} at physiological concentration (2 mM) to obtain the stably folded bent conformation, followed by addition of ligand and incubation. Dose-dependent decrease of the FRET signal coinciding with increased Cy3 donor emission is indicative of ligand capture of the extended switch conformation (Figures 4.1, 4.11). When the dyes move beyond the Förster radius, reduced efficiency of energy transfer by FRET causes growing Cy3 donor emission which, at the same time, provides a selectivity control for compounds that stabilize the extended switch state rather than nonspecifically quenching fluorescence. Since ligands capture extended states from an equilibrium of switch conformations, and bent structures are preferred by the unbound switch, complex formation is a relatively slow event and requires incubation of the RNA target with compound for 10-30 minutes to achieve maximum signal change.⁷² Binding of benzimidazole translation inhibitor **1** (Figure 4.11a), which was previously used to determine the structure of the RNA target complex,²¹ captured the extended state of the HCV switch at an EC50 value of 22 ± 12 μ M (Figure 4.11a, b).

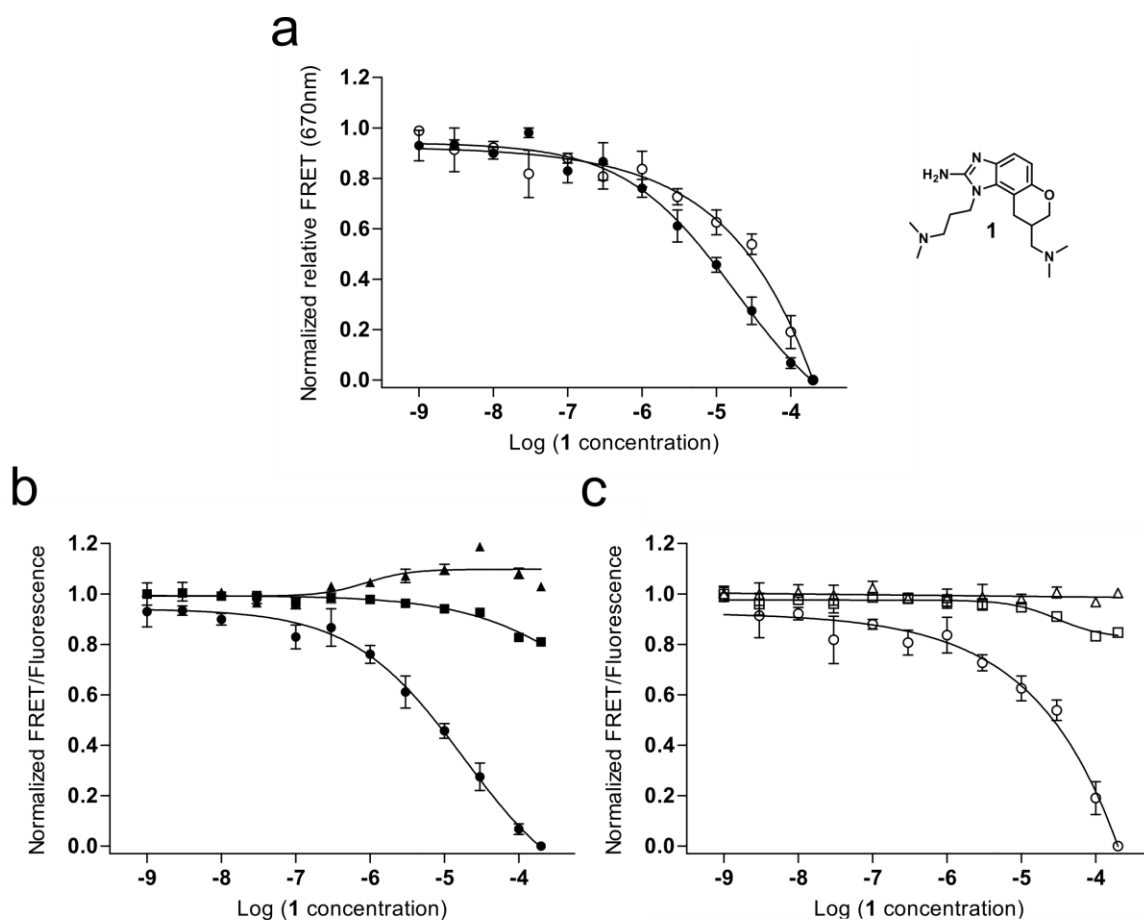


Figure 4.11: Binding of a translation inhibitor ligand **1**^{39,64} to the HCV RNA switch as monitored by FRET experiments with cyanine dye-labeled oligonucleotides. a) Normalized relative FRET signal for the titration of **1** to a conventional model construct of the HCV IRES subdomain IIa switch 5' terminally conjugated with Cy3 and Cy5 dyes (●) and a modular FRET construct consisting of unmodified oligonucleotides that contain the HCV RNA switch and carry overhanging single strands which hybridize with cyanine dye-conjugated DNA oligonucleotides (○). Dose-response fitting curves gave an EC_{50} value for capture of the extended switch state of $22 \pm 12 \mu\text{M}$ for the conventional construct and $37 \pm 21 \mu\text{M}$ for the modular construct. As the FRET signal for the modular system did not reach saturation, affinity of compound **1** was calculated by dose-response fitting to the Cy5 emission signal. b) Individual dye and FRET signals for the conventional model construct. (●)=FRET; (■)=Cy5; (▲)=Cy3. c) Individual dye and FRET signals for the modular construct. (○)=FRET; (□)=Cy5; (Δ)=Cy3. Error bars represent ± 1 s.d. calculated from triplicate experiments.

Similar to the folding studies outlined previously, we have extended the scope of the FRET assay for monitoring ligand binding to viral RNA switches by using modular dye-labeled constructs. For the HCV switch, the modular oligonucleotide system (Figure 4.7b) reported an EC50 value for capture of the extended RNA by compound 1 at 37 ± 21 μM which was in good agreement with the affinity determined with the conventional non-modular construct (Figure 4.11a, c). As was observed in the folding experiments, hybridization of the RNA switch with dye-labeled DNA probes did not interfere with the FRET signal change induced by ligand binding.

Ribosomal A-site RNA with an incorporated heavy atom containing fluorescent ribonucleoside analog

Fluorescent nucleoside analogs are useful tools to study local dynamics, conformational changes, and ligand binding of functional RNA molecules. In this study, a selenophene-conjugated uridine analog ($^{\text{Se}}\text{U}$) was developed as a dual function probe with fluorescence properties able to detect RNA conformational changes and heavy atom anomalous scattering properties useful for phase determination in crystal structure determination, and was incorporated into well-studied model RNA constructs of the HCV IIa RNA switch near the internal loop (Figure 4.7, discussed in Chapters 1-4) and the bacterial ribosomal decoding site (A-site) (Figure 4.12),^{60,61} which is involved in protein synthesis fidelity by monitoring mRNA codon-tRNA anticodon base pairing and which is the binding site for aminoglycoside antibiotics (Figure 4.13).¹¹²⁻¹¹⁵ These $^{\text{Se}}\text{U}$ -modified constructs both successfully detect ligand binding with an increase in fluorescence, and an example is shown in Figure 4.12 for the modified A-site RNA.⁶⁰

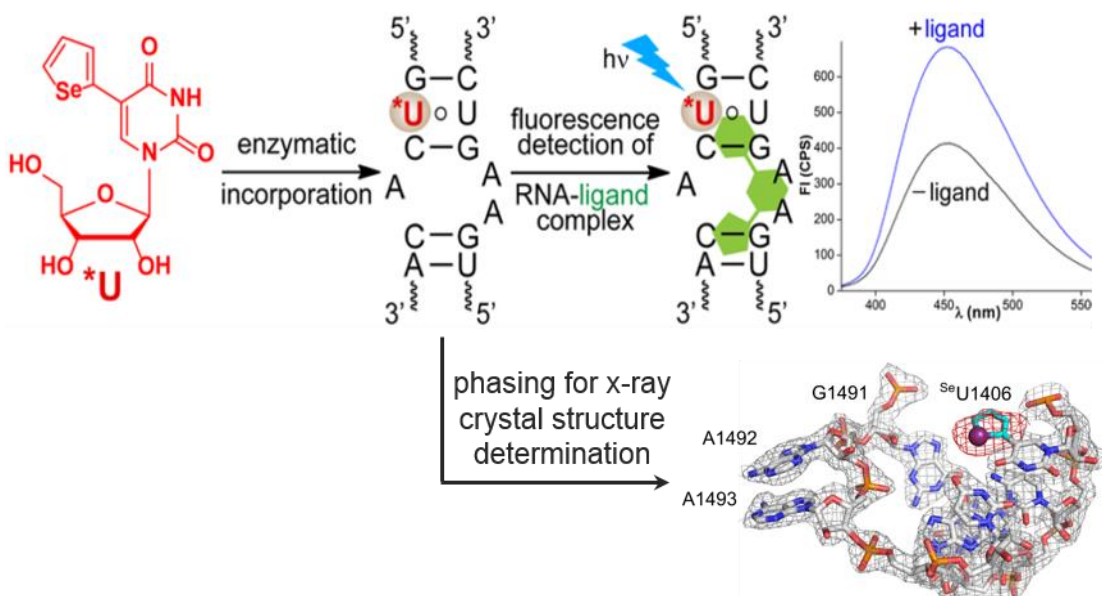


Figure 4.12: Utility of dual functional selenophene-conjugated uridine analog. This ^{76}SeU was enzymatically incorporated into a model ribosomal A-site RNA construct whose fluorescence increased upon ligand binding. The crystal structure determination of this ^{76}SeU -modified A-site RNA revealed only minimal ^{76}SeU -perturbation of the wild type (WT) RNA structure and provides an anomalous scattering selenium atom for phase determination in X-ray crystallography. The upper half of this figure was adapted from ref⁶¹, 2013, American Chemical Society.

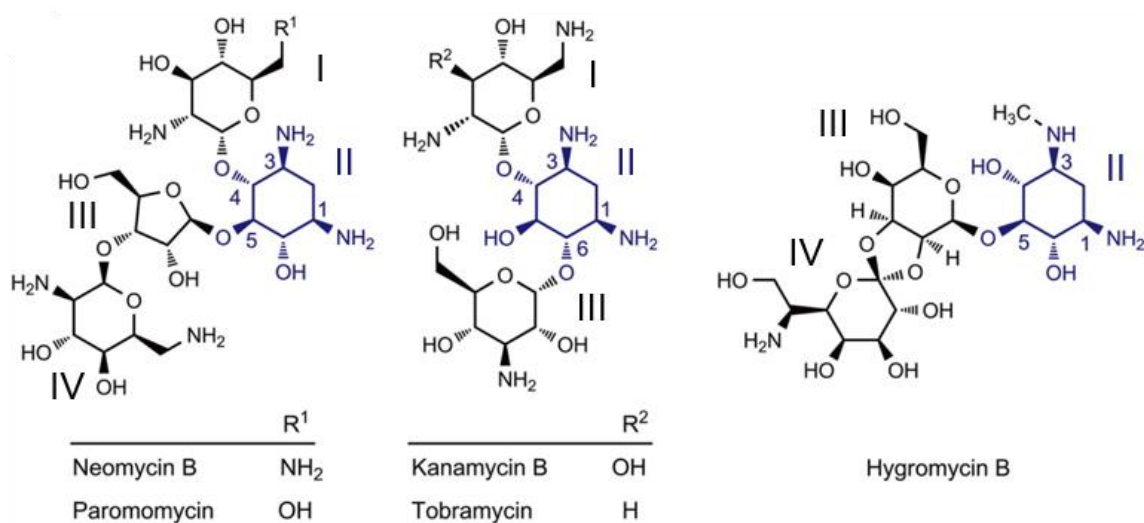


Figure 4.13: Aminoglycosides which bind the ribosomal decoding site (A-site). The conserved 2-deoxystreptamine (2-DOS) ring is highlighted in blue. Figure is adapted from ref¹¹⁵, 2010, Oxford University Press.

Attempts were made to crystallize both ^{Se}U-modified RNA model constructs. Crystallization trials for the ^{Se}U-modified HCV Ila switch RNA construct resulted in only poorly diffracting crystals. On the contrary, crystallization trials for the ^{Se}U-modified A-site RNA construct successfully yielded crystals diffracting to 2.14Å. The 3D structure of this construct was determined by molecular replacement (Figure 4.14)⁶¹ using a model from a similar previously determined structure,¹¹⁴ and the structure statistics are shown in Table 4.1.^{61,114} This structure unambiguously reveals the selenophene-modified uridine which was incorporated in this RNA construct (Figure 4.15). A superimposition of the ^{Se}U-modified and non-modified (PDB: 1T0D) A-site RNA structures reveals this ^{Se}U to be minimally perturbing of native free RNA structure (Figure 4.16), and the ^{Se}U residue stacks nicely between neighboring residues in the RNA helix (Figure 4.17). A superimposition of ^{Se}U-modified and aminoglycoside-bound non-modified A-site RNA

structures reveals this ^{Se}U to be minimally disturbing of A-site ligand binding interactions for neomycin, paramomycin, and hygromycin B (Figures 4.18, 4.19, 4.20), but not kanamycin A or tobramycin (Figures 4.20, 4.21), in good agreement with fluorescence detection of A-site ligand binding in solution.^{60,61} Attempts to co-crystallize the ^{Se}U -modified A-site RNA with neomycin, paramomycin, and hygromycin B resulted in highly diffracting crystals which lacked any bound aminoglycoside. Overall, these studies show this ^{Se}U to be a useful tool for fluorescently monitoring RNA-ligand binding and determining novel nucleic acid crystal structures with minimal disruption of native RNA structure.

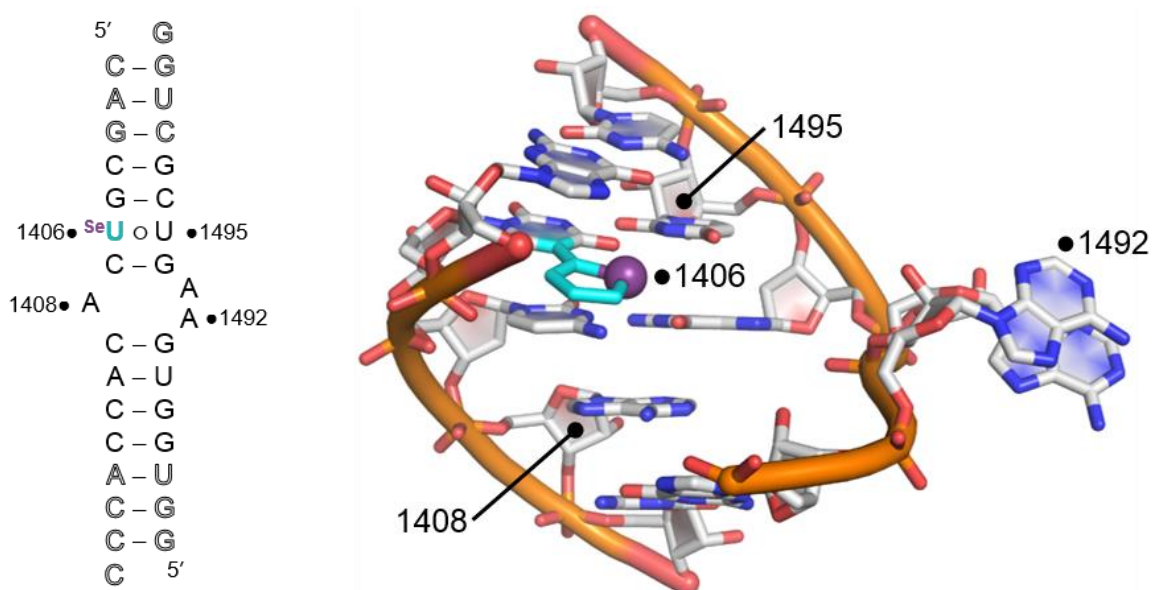


Figure 4.14: Structure of ^{Se}U -modified A-site RNA construct. The selenophene ring is shown in cyan with the selenium atom shown as an indigo sphere. Nucleotides deviating from the bacterial decoding site wild-type sequence are shown in outlined font.

Table 4.1: Crystallographic data collection and refinement statistics for ^{Se}U-modified A-site RNA.

Data Collection	
Wavelength (Å)	1.54
High-resolution limit (Å)	2.14
Low-resolution limit (Å)	26.02
Completeness (%) ^a	77.6 (46.7)
Unique reflections	7483
Refinement	
Space group	P2 ₁
Cell dimensions (Å)	
<i>a</i>	31.55
<i>b</i>	86.96
<i>c</i>	32.56
α	90.00
β	94.14
γ	90.00
<i>R</i> _{work} / <i>R</i> _{free}	0.197 / 0.247
No. atoms	
RNA atoms	1408
Solvent atoms	60
Metal ions	10 Mg ²⁺
Mean <i>B</i> factors (Å ²)	
RNA	32.6
Solvent	32.2
Metal	37.5
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.366
Dihedral angles (°)	20.892

^aNumbers in parentheses are for the highest-resolution shell.

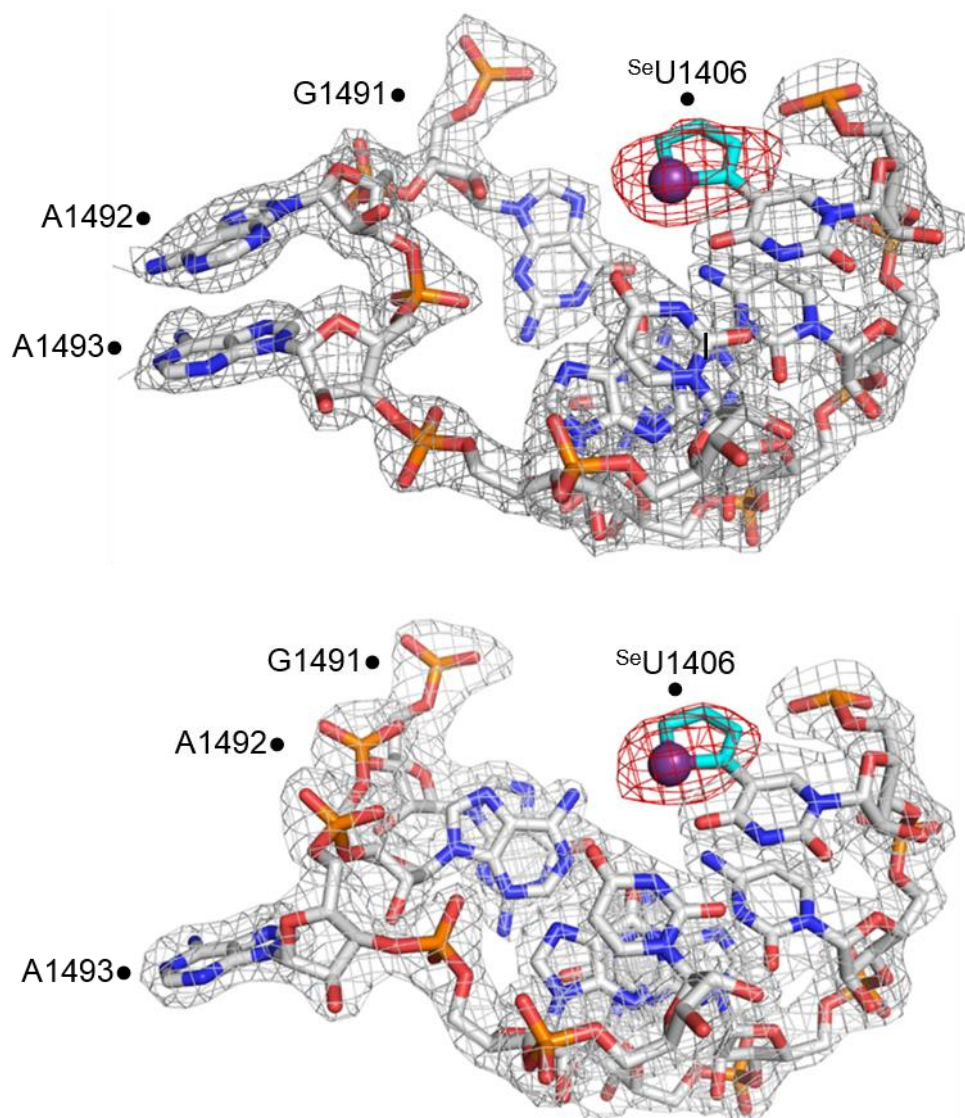


Figure 4.15: Electron density map around the ^{Se}U-modified A-site RNA construct with A1492 flipped out (top) or in (bottom) the RNA helix, as is seen in a previous A-site model RNA structure.¹¹⁴ The selenophene ring is shown in cyan with the selenium atom shown as an indigo sphere. The 2F_o-F_c electron density map, contoured at 1.0σ, is shown in gray, and the F_o-F_c electron density map, contoured at 3.0σ, is shown in red. These electron density maps were calculated using a model which lacked the selenophene ring. When the selenophene ring is deleted from the ^{Se}U-modified A-site RNA structure model, crystal structure refinement statistics are negatively affected $R_{work}=0.20 \rightarrow 0.22$ and $R_{free}=0.24 \rightarrow 0.27$.

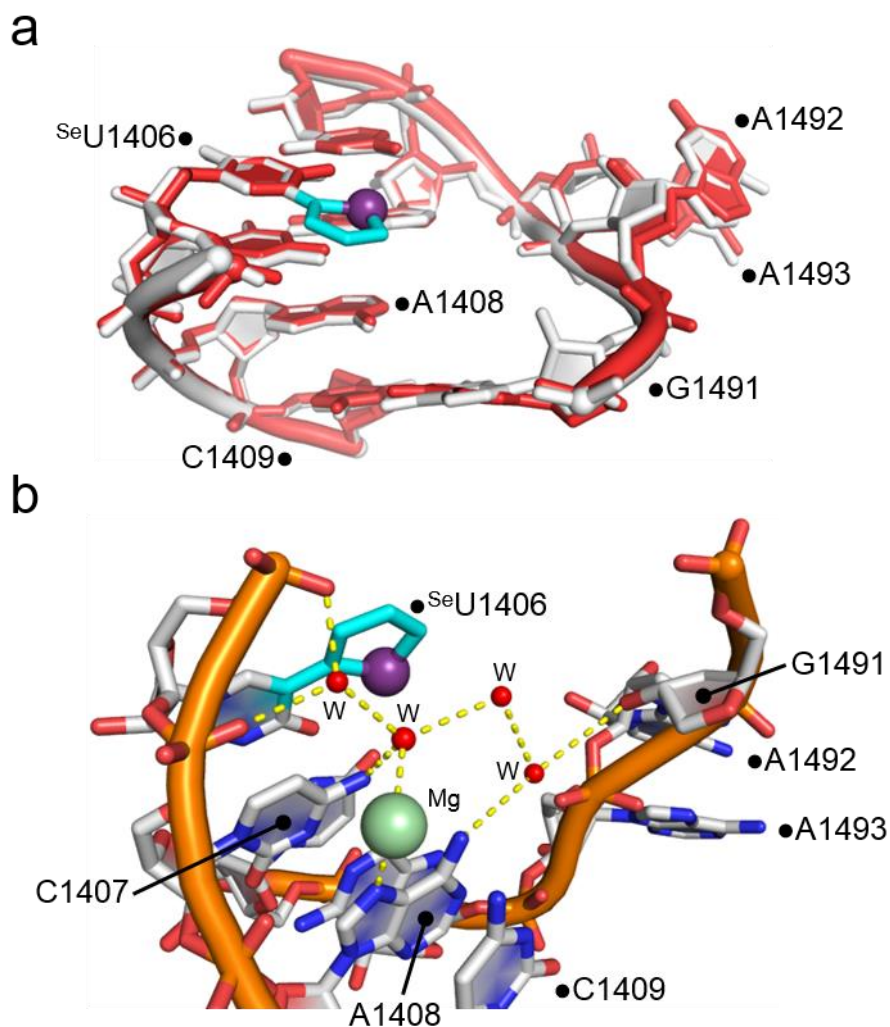
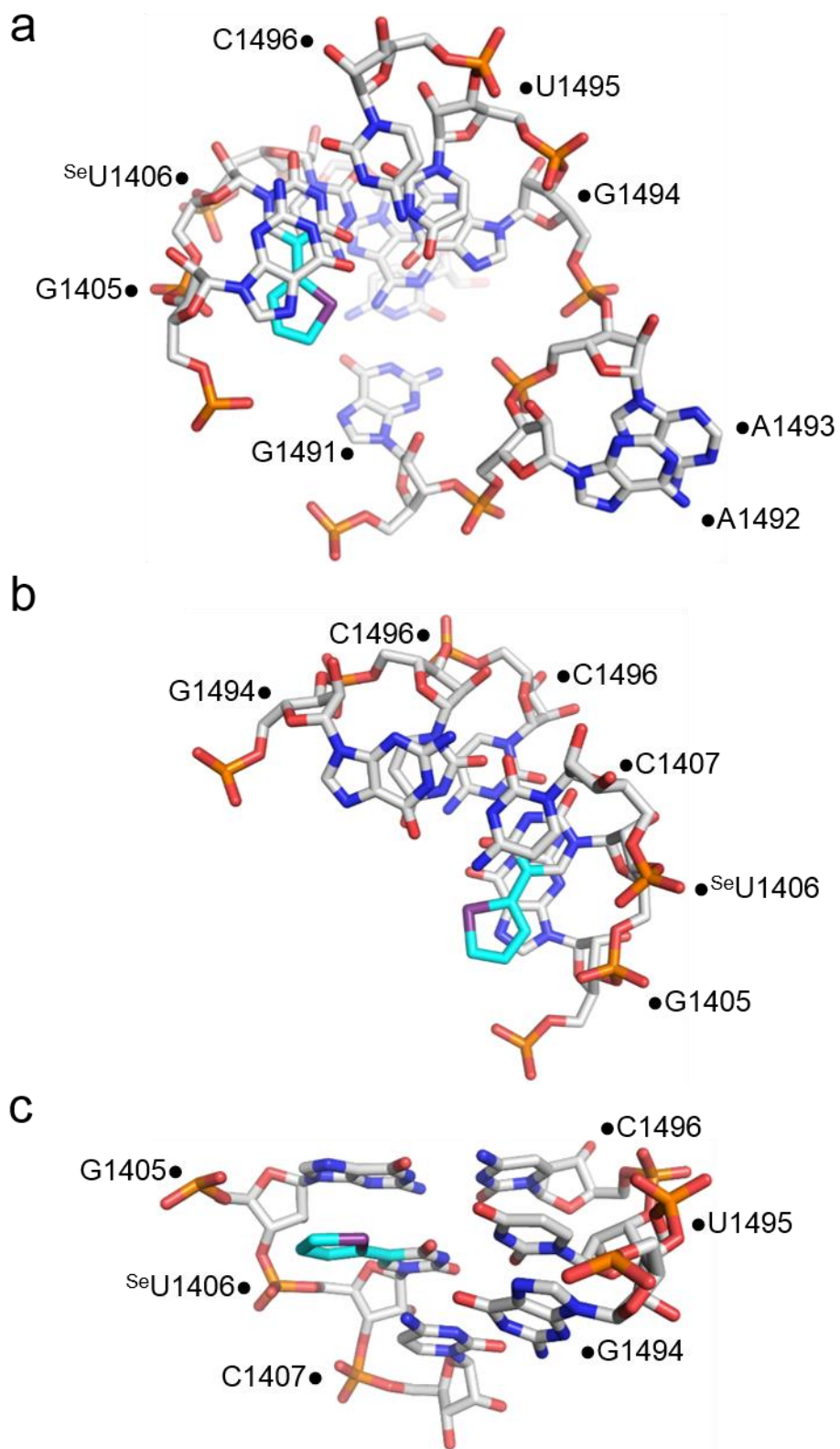


Figure 4.16: a) Superimposition of the ^{Se}U-modified (white) and non-modified (red; PDB: 1T0D) A-site RNA constructs. The RMS distance is .367Å for 1305 identical atoms (out of 1362 total atoms) between entire models for ^{Se}U-modified and non-modified RNA containing one molecule in each of the A1492 in and out conformations (out conformation shown here). The 2'-OH of the G1491 sugar is drawn into the interior of the decoding site in the A1492-out conformation of the ^{Se}U-modified RNA, whereas this is not seen for the non-modified RNA. b) This G1491 2'-OH interacts with a bridging water molecule to A1408. This bridging water also interacts with a chain of water molecules beneath the selenophene ring and above a Mg²⁺ ion.

Figure 4.17: Stacking of neighboring residues G1405 and C1407 on the selenophene-modified U1406. a) Top-down view with G1405 in the foreground. b) Bottom-up view with C1407 in the foreground. c) Side view with G1405 above and C1407 below the selenophene-modified U1406.



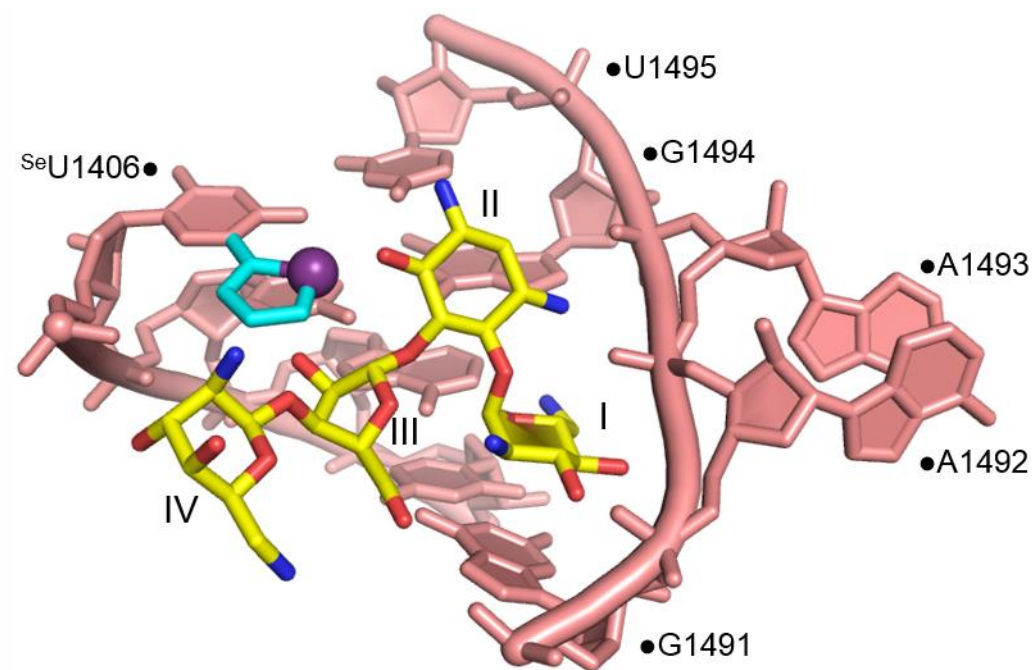


Figure 4.18: Superimposition of the ^{76}SeU -modified (cyan/indigo) and non-modified neomycin-bound (yellow/pink, PDB: 2A04) A-site RNA constructs. In this superimposition, the 6-hydroxy group on ring II of neomycin is 2.4\AA from the selenium atom, and the $2''$ -hydroxy group on ring III of neomycin is 2.6\AA from the selenium atom.

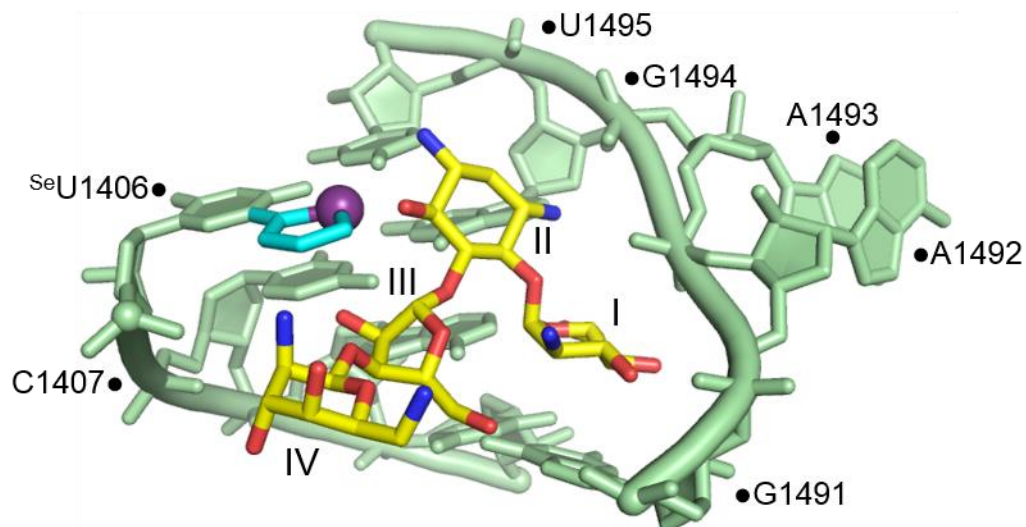


Figure 4.19: Superimposition of the ^{76}SeU -modified (cyan/indigo) and non-modified **paramomycin**-bound (yellow/green, PDB: 1J7T) A-site RNA constructs. In this superimposition, the 6-hydroxy group on ring II of paramomycin is 2.2Å from the selenium atom, and the 2''-hydroxy group on ring III of paramomycin is 3.4Å from the selenium atom.

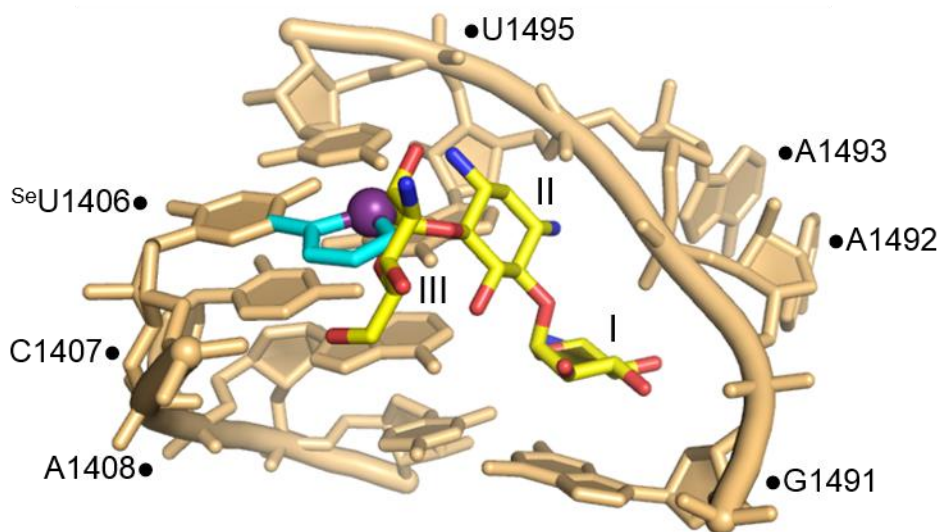


Figure 4.20: Superimposition of the ^{76}SeU -modified (cyan/indigo) and non-modified **kanamycin** A-bound (yellow/brown, PDB: 2ESI) A-site RNA constructs. Due to the clashes shown between the selenophene and ring III, it is predicted that kanamycin would not bind this SeU-modified RNA.

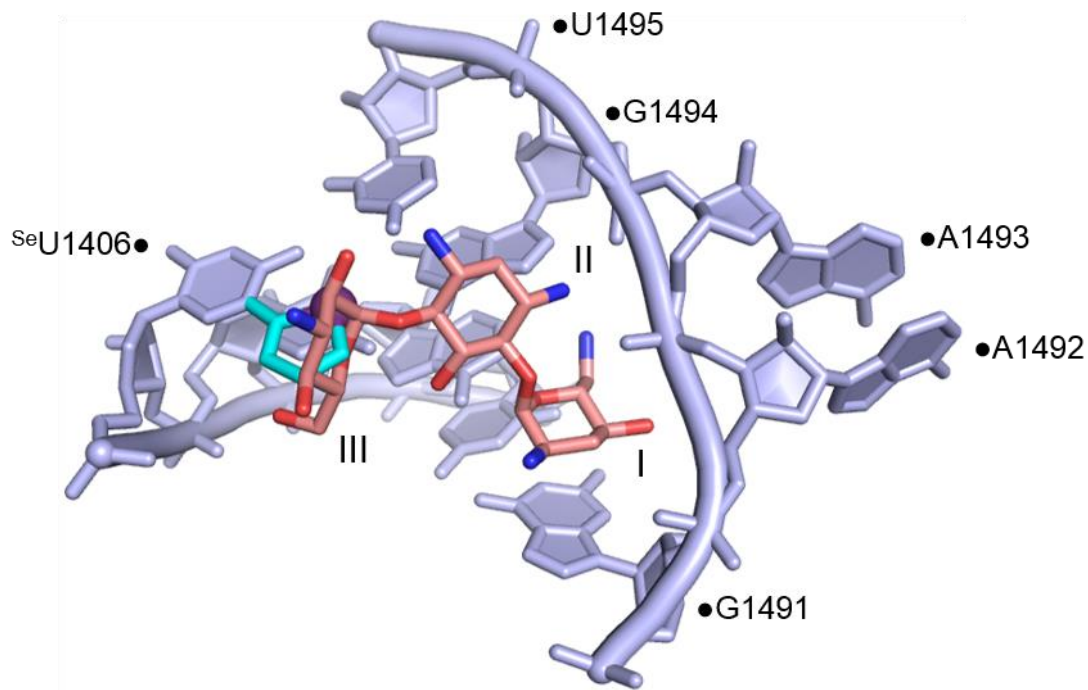


Figure 4.21: Superimposition of the ^{Se}U-modified (cyan/indigo) and non-modified tobramycin-bound (yellow/blue, PDB: 1LC4) A-site RNA constructs. Due to the clashes shown between the selenophene and ring III, it is predicted that tobramycin would not bind this SeU-modified RNA.

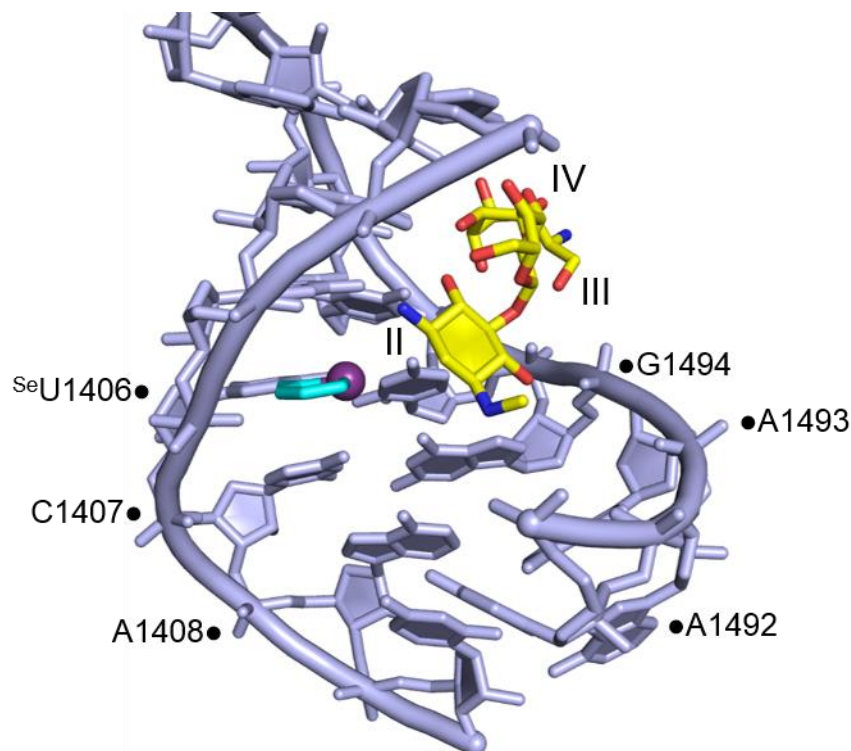


Figure 4.22: Superimposition of the ^{76}SeU -modified (cyan/indigo) and non-modified **hygromycin B**-bound (yellow/blue, PDB: 1HNZ) A-site RNA constructs. In this superimposition, the 1-amino group on ring II of hygromycin B is 3.5\AA from the selenium atom.

Conclusions

Ligand-responsive RNA switches, which are distinct from bacterial riboswitches, have recently been discovered in the IRES elements of positive strand RNA viruses (Chapters 2-3).^{39,58} These viral switches constitute a new class of conformationally flexible RNA modules which, unlike conventional metabolite-sensing riboswitches, adopt distinct three-dimensional folds in the free and ligand-bound states while retaining the same secondary structure. X-ray crystallographic studies of the ligand-free and bound HCV switch as well as the ligand free SVV switch described here have been instrumental in characterizing the architecture of viral RNA switches and in elucidating molecular

recognition by ligands that capture the extended state. Future crystallographic efforts will focus on determining the structure of an RNA switch which is located on the 3' proximal side of domain II, such as those found in BVDV, CSFV, AEV, GPV, or BDV, to reveal the fashion by which these dissimilar switches function in an analogous way to HCV. Additionally, the structure determination of a non-HCV RNA switch in complex with benzimidazole inhibitor **1**²¹ would provide a useful comparison of this ligand's molecular recognition of diverse RNA switches. This would also provide useful insights and comparisons of the architecture of a second ligand-captured RNA switch in an elongated conformation. The structure determination of the switch from HCV or another virus in complex with a guanine-like molecule (see Chapter 2), would also provide useful insights into the specific binding of this class of molecules and direct future development of unique viral inhibitors with guanine-like scaffolds.

Investigation of RNA switch conformational changes in real time requires dynamic techniques applied in solution. FRET experiments, outlined in this chapter, have been used to monitor folding and ligand-captured switching action of the viral RNA modules. A new approach has been described that uses fluorescently labeled modular oligonucleotide model constructs representing the switch from HCV and which are readily adapted for FRET studies of various other RNA switches. This new approach may be used to characterize the dynamics of newly identified switches in response to magnesium as well as guanosine and other HCV Ila inhibitors to determine HCV-like switch function. Switches functioning similarly to HCV may be interrogated for binding with known HCV Ila binding molecular scaffolds. As structure activity relationships emerge for ligands targeting each individual switch, general trends may be realized for

the inhibition of guanosine-binding RNA switches. These principles will inform future development of HCV IRES inhibitors as well as inhibitors for other potential human pathogenic RNA switch-containing viruses yet to be discovered.

A heavy atom containing fluorescent nucleoside analog, outlined in this chapter, has been used to detect RNA-ligand binding in the HCV Ila switch as well as the ribosomal A-site RNA, and a crystal structure of a ^{Se}U-modified A-site model RNA reveals this ^{Se}U to be minimally structure perturbing and useful for obtaining phase information in crystal structure determination. This dual functional probe which maintains the ability to form Watson-Crick base pairs and only minimally modulates native structure may be incorporated into countless other functional RNAs to aid in the detection of RNA-ligand binding, conformational dynamics, and structure determination.

Materials and Methods

Preparation of RNA constructs for FRET experiments

Cyanine dye labeled and unlabeled RNA and DNA oligonucleotides were obtained by chemical synthesis (Integrated DNA Technologies, Coralville, IA). Unlabeled RNA oligonucleotides were purified by standard desalting while dye labeled RNA and DNA oligonucleotides were purified by HPLC to remove any excess free dye. Lyophilized single stranded oligonucleotides were dissolved in 10 mM sodium cacodylate buffer, pH 6.5, to prepare stock solutions.

Working solutions for conventional model FRET constructs were prepared at 1 μ M concentration from terminally Cy3/Cy5-labeled single stranded RNA

oligonucleotides diluted with 10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer, pH 7.0.

Working solutions for modular FRET constructs were prepared at 1 μ M concentration from single stranded RNA oligonucleotides and 5' terminally Cy3/Cy5-labeled single stranded DNA oligonucleotides diluted with 10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer, pH 7.0.

FRET titration experiments monitoring folding and ligand binding

FRET folding experiments at increasing concentration of Mg^{2+} were prepared in PCR tubes with conventional or modular FRET construct at 100 nM concentration in 10 mM HEPES buffer, pH 7.0, in a total volume of 110 μ L. RNA concentration and volume may be increased to achieve larger absolute fluorescence signals and minimize pipetting error. FRET constructs were annealed from oligonucleotides and folded in the presence of Mg^{2+} by heating PCR tube samples at 65°C for 5 minutes followed by snap cooling and incubating on ice for 15 minutes. 100 μ L of each sample was then transferred to a 96-well plate and incubated in the dark at 25°C for 10 minutes. FRET measurements were performed on a Spectra Max Gemini XS monochromator plate reader (Molecular Devices, Sunnyvale, CA) at 25°C by exciting the Cy3 label at 520 nm and reading the transferred fluorescence as Cy5 emission at 670 nm. Emission filters were set at 550 and 665 nm. Mg^{2+} dependent FRET folding was calculated and analyzed as described previously.⁷²

FRET ligand binding experiments were performed as above, except a 2 mM Mg^{2+} concentration was held constant while increasing the concentration of compound **1**.

Ligand binding occurred in the annealing and folding steps as described above. Following FRET measurements, Cy3 fluorescence was recorded by exciting the Cy3 label at 520 nm and reading the emission at 570 nm with an emission filter set at 550 nm. Cy5 fluorescence was recorded by exciting the Cy5 label at 620 nm and reading the emission at 670 nm with an emission filter set at 665 nm. Ligand induced FRET changes were calculated and analyzed as described previously.⁷²

Crystallization and Data Collection

RNA constructs (Figures 4.2-4.6) were annealed from stoichiometric amounts of the single strands in 10mM sodium cacodylate, pH 6.5 in the presence or absence of 5mM MgCl₂ by heating to 65°C for 4min followed by slow cooling to room temperature or snap cooling on ice for 10min followed by slow warming to room temperature. After annealing, the RNA was crystallized at 6, 12, 16, or 22°C by hanging drop vapor diffusion. For crystallization, 1μL of 0.2mM RNA was mixed with an equal volume of crystallizing solution. Crystals grew after equilibration against 700μL of well solution containing crystallizing solution.

The ^{Se}U-modified A-site RNA crystal construct (Figure 4.14) was annealed from stoichiometric amounts of the single strands in 10mM sodium cacodylate, pH 6.5 by heating to 65°C for 4min followed by snap cooling on ice for 10min followed by slow warming to room temperature. After annealing, the RNA was crystallized at 22°C by hanging drop vapor diffusion. For crystallization, 1μL of 0.2mM RNA was mixed with an equal volume of crystallizing solution containing 2.40-2.55M ammonium sulfate, 10mM magnesium acetate, and 40mM 2-(N-morpholino)ethanesulfonic acid (MES)

buffer, pH 5.6. Crystals grew over the course of four to seven months after equilibration against 700 μ L of well solution containing crystallizing solution.

Crystals were flash-cooled in liquid nitrogen. Crystal screening and X-ray diffraction data were collected at 110K on a Rigaku rotating anode X-ray generator ($\lambda = 1.54 \text{ \AA}$) equipped with a MAR345 imaging plate detector system. Datasets were processed, integrated, and scaled with the HKL2000 package.⁹¹

Structure Solution and Refinement

The three-dimensional structures of the SVV subdomain IIa RNA constructs (Chapter 2), the ^{Se}U-modified A-site RNA construct, and the RNA nanotriangle construct (Chapter 5) were solved by molecular replacement with the program Phaser⁹² using A-form RNA duplexes or previously determined structures,^{21,39,47} as search models and refined by the program Refmac⁹³ both within the CCP4 package.⁹⁴ Subsequent iterative rounds of manual building and refinement, alternating between Refmac and manual rebuilding in Coot,⁹⁵ were based on the obtained 2Fo-Fc and Fo-Fc maps. Final refinement was carried out in PHENIX⁹⁶ with individual isotropic atomic displacement parameters and water picking.

Acknowledgements

The material in Chapter 4, in part, has been published in *Methods* (2015, 91, 35-39) titled *Conformational flexibility of viral RNA switches studied by FRET* with Thomas Hermann as co-author. The dissertation author was the primary investigator and author of this material.

The material in Chapter 4, in part, is currently being prepared for submission for publication titled *Fluorescence detection of RNA-ligand binding and crystal structure determination of ribosomal decoding site RNA using a heavy atom containing fluorescent ribonucleoside* with Ashok Nuthanakanti, Seergazhi Srivatsan, and Thomas Hermann as co-authors. The dissertation author was the co-primary investigator and author of this material.

Thanks to Jing Gu for help setting up several AEV crystal screens. Thanks to Andrew Bergdorf for help setting up several BVDV crystal screens. Thanks to Lourdes Myjak for help setting up several SPV-C1 crystal optimization trials, as well as ^{Se}U-modified A-site RNA co-crystallization screens with aminoglycoside ligands.

Chapter 5:

Crystal structure-guided design of self-assembling RNA nanotriangles

Parts of this work are published in:

“Crystal structure-guided design of self-assembling RNA nanotriangles”
Angewandte Chemie International Edition, 55, 4097–4100 (2016).

Parts of this work are in preparation for submission to be published in:

“Design and Crystallography of Self-Assembling RNA Nanostructures,” (2016).

Abstract

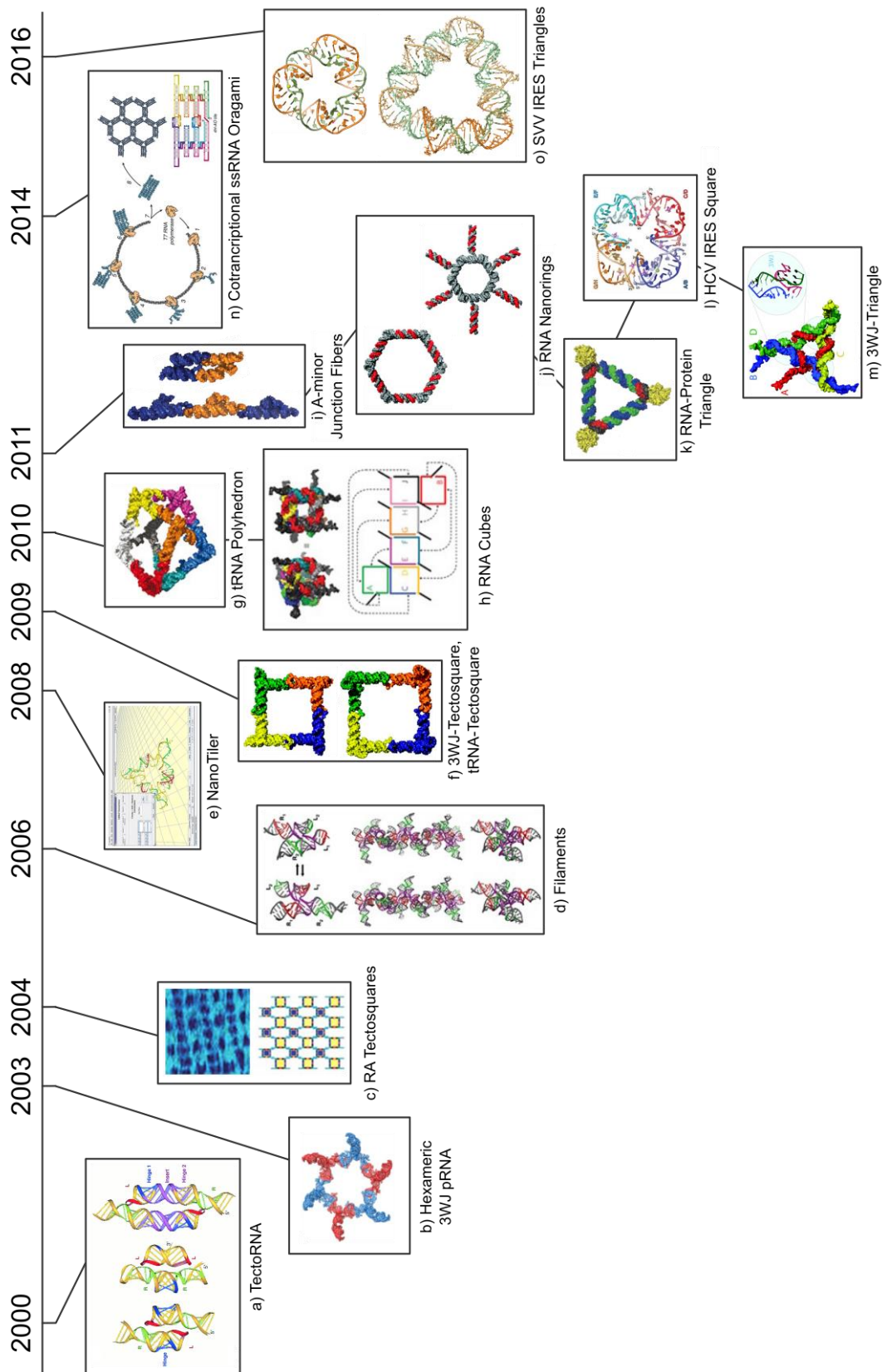
RNA nanotechnology harnesses RNA structural motifs to build nano-sized architectures which assemble through selective base pair interactions. In this chapter, the crystal structure-guided design of highly stable RNA nanotriangles that self-assemble cooperatively from short oligonucleotides is reported. The crystal structure of an 81 nucleotide nanotriangle determined at 2.6 Å reveals the yet smallest circularly closed nano-object made entirely of double-stranded RNA. Assembly of the nanotriangle architecture involved RNA corner motifs that were derived from ligand-responsive RNA switches which offer the opportunity to control self-assembly and dissociation.

Introduction

Nucleic acid has been used extensively to build nano-sized objects by controlling assembly through designed base pair interactions. Complex nano-objects have been obtained by recursive folding of long nucleic acid sequences through inclusion of alternating double- and single-stranded regions, junctions and helper oligonucleotides. The design of nano-objects has exploited structural motifs observed in crystal structures (Figure 4.1).^{104,116–128} Hermann and co-workers have previously used short oligonucleotides to construct a self-assembling RNA nanosquare of 100 nucleotides.¹⁰⁴ Here, the crystal structure-guided design of RNA nanotriangles that self-assemble in a cooperative process from multiple copies of short oligonucleotides is described. Crystal structure analysis of an RNA triangle containing 81 nucleotides reveals the yet smallest circularly closed nano-object made entirely of double-stranded RNA. Nanotriangle self-

assembly and dissociation is sequence-dependent and may be modulated by ligands which bind recognition motifs incorporated into the RNA architectures.

Figure 5.1: Several significant structural achievements in the history of RNA nanotechnology. a) TectoRNA dimers constructed using the GAAA tetraloop-receptor from a group I self-splicing intron.^{116,129} b) Hexameric pRNA monomers containing a three-way junction (3WJ) from the bacteriophage Phi29 DNA-packaging motor used to construct nanoscale dimers and trimers.^{117,130,131} c) Tectosquares constructed using tetraloop-receptors, ribosomal right angle (RA) motifs, and kissing loop (KL) complexes from the dimerization initiation site of human immunodeficiency virus (HIV) RNA.¹¹⁸ d) RNA filaments constructed in a controlled fashion using tectoRNA-containing four-way junctions (4WJ) from the hairpin ribozyme.^{116,119} e) NanoTiler, the first computational RNA nanostructure design tool, is created and uses RNA three dimensional structural motifs.¹²⁰ f) Square-shaped RNA particles constructed using a 3WJ (UA_h_3WJ) from ribosomal RNA (rRNA) and a five-way junction (5WJ) from a transfer RNA (tRNA).¹²¹ g) RNA Polyhedron constructed from a tRNA.¹²² h) RNA cube constructed based on computational design using NanoTiler.¹²³ i) RNA fibers and particles constructed using tectoRNA-containing A-minor 4WJ motifs from rRNA and HIV KL.¹²⁴ j) RNA nanorings constructed using RNA I/II inverse kissing complexes found in ColeE1 plasmid replication regulating RNA transcripts in *Escherichia coli*.¹²⁵ k) RNA-protein triangle constructed using a kink-turn (K-turn) motif from the Archaeal C/D box small nucleolar RNA (snoRNA) and a K-turn binding protein L7Ae.¹²⁶ l) Self-assembling RNA square constructed using subdomain IIa switch motifs from the HCV IRES.¹⁰⁴ m) RNA triangle constructed using 3WJ from rRNA.¹²⁷ n) Cotranscriptional ssRNA origami constructed using hairpins, KL complexes, and dovetail seams.¹²⁸ o) Self-assembling RNA triangles constructed using subdomain IIa switch motifs from the SVV IRES.⁵⁹ Inspiration for figure came from ref¹³². Images reproduced from a) ref¹²⁹, 2001, Oxford University Press. b) ref¹³¹, 2013, Cold Spring Harbor Laboratory Press. c) ref¹¹⁸, 2014, The American Association for the Advancement of Science. d) ref¹¹⁹, 2006, Oxford University Press. e) ref¹²⁰, 2008, Elsevier. f) ref¹²¹, 2009, American Chemical Society. g) ref¹²², 2010, Nature Publishing Group. h) ref¹²³, 2010, Nature Publishing Group. i) ref¹²⁴ 2011, Oxford University Press. j) ref¹²⁵, 2011, American Chemical Society. k) ref¹²⁶, 2011, Nature Publishing Group. l) ref¹⁰⁴, 2011, National Academy of Sciences, USA. m) ref¹²⁷, 2011, American Chemical Society. n) ref¹²⁸, 2014, The American Association for the Advancement of Science. o) ref⁵⁹, 2016, Wiley.



Recently, the simplest known ligand-responsive RNA switches have been discovered in the internal ribosome entry sites (IRES) of positive strand RNA viruses of the *Flavi-* and *Picornaviridae* families. These RNA switch motifs are located in subdomain IIa of IRES elements and regulate viral protein synthesis through a ligand-dependent conformational transition.^{39,58} Unlike traditional riboswitches, viral RNA switches do not undergo secondary structure changes but rather a purely mechanical switching between two distinct and stable conformations. The viral RNA motifs adopt a bent conformation in the absence of a ligand while an elongated conformation is captured by a bound ligand. Structures of ligand-free IIa switches from hepatitis C virus (HCV) and Seneca Valley virus (SVV) IRES elements as well as the ligand-bound switch from the HCV IRES were previously determined by X-ray crystallography.^{21,39,47} The ligand-free IIa switch motif from HCV has previously been used as a corner building block to design and construct a self-assembling nanosquare and a nanoprism.^{104,105} It was concluded that other RNA nano-objects may be obtained from viral IIa switch motifs by rational design including variation in length of the helices flanking each corner unit and adjusting the orientation of corners relative to a common plane.

Results and Discussion

Design of RNA Nanotriangles

Crystal structure analysis of the IIa corner motif from the SVV IRES provided design directions to construct triangular nano-objects (Figure 5.2). The three-dimensional structure of the SVV motif was obtained from crystals of a short and long RNA construct, both of which showed identical corner structures (see Chapter 2, Figures 2.10-2.12).³⁹

Crystal packing of both constructs revealed circularly closed triangles involving pseudo-continuous stacking and intermolecular base pair formation between a 3' terminal overhanging nucleotide on both ends of each of three identical corner units (Figure 5.2b). Unique crystal structure packing elements for these two constructs will be discussed in greater detail in the *Discussion of Crystal Structures used in Design* later in this chapter.

Self-assembling nanotriangle constructs were designed guided by the triangular arrangement seen in the SVV IIa RNA crystal packing. To promote self-assembly of corner motifs into a triangle, constructs were designed with the same total number of nucleotides and base identity but varying 3' terminal overhang length. Secondary structure models of exemplary self-assembling constructs containing four nucleotide overhangs are shown in Figure 5.2c. Other self-assembling constructs will be discussed in greater detail in the *Discussion of Self-Assembling Constructs* later in this chapter. A small triangle is constructed from two oligonucleotides, including an inner and outer strand of 11 and 16 residues, respectively (Figure 5.2c, top). Three copies of each the inner and outer strands are designed to self-assemble as a single small triangle of 81 nucleotides. Similarly, a large triangle is constructed from strands of 20 and 26 residues to form a nano-object of 138 nucleotides (Figure 5.2c, bottom).

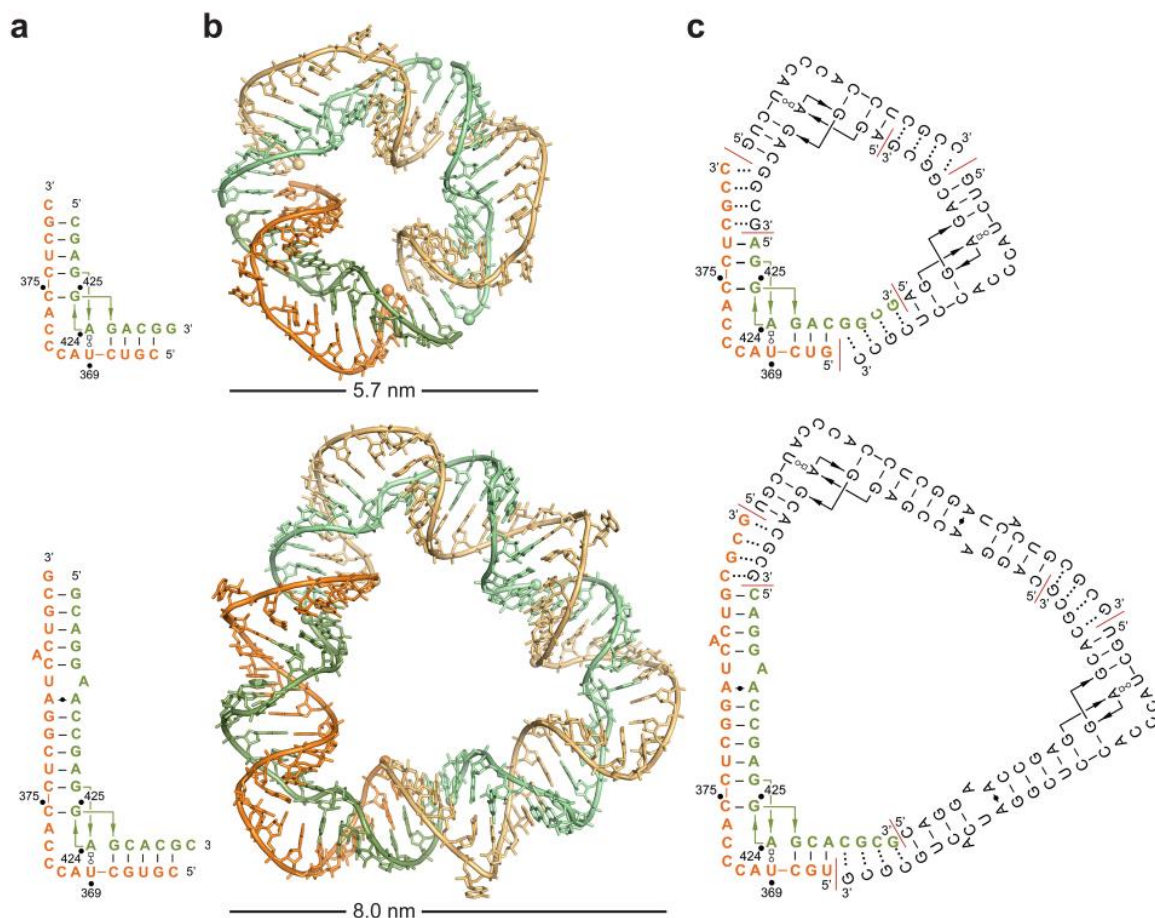


Figure 5.2: Design of self-assembling RNA nanotriangles. a) Secondary structures of short and long oligonucleotide constructs representing the subdomain IIa motif from the SVV IRES. Single nucleotide overhangs aided crystallization by facilitating RNA packing. Residue numbering refers to the SVV genome. b) Circularly closed triangles seen in the packing of both short and long IIA RNA crystal structures (PDB ID: 4P97 and 4PHY). c) Secondary structure models of self-assembling RNA triangles containing four nucleotide overhangs and designed using the crystal packing of short and long IIA constructs. Red lines indicate oligonucleotide termini.

Self-Assembly of RNA Nanotriangles

SVV short and long IIA corner motif constructs (Figure 5.2a) which contain single 3' nucleotide overhangs were analyzed by native polyacrylamide gel electrophoresis (PAGE) and found to migrate as single bands consistent with their respective size (Figure

5.3a, 1nt-overhang both). In contrast, nanotriangle RNA constructs carrying four overhanging nucleotides (Figure 5.2c) to promote self-assembly were found to migrate slower and consistent with the size of symmetrical triangles comprised of three identical corners (Figure 5.3a; 4nt-overhang both). The small triangle construct migrated as a single band, while the large triangle construct gave rise to a faster moving major band and a slower moving minor band, the latter consistent with the size of a dimer of triangles. This slower moving minor band is present in a greater proportion at higher Mg^{2+} concentrations (Figure 5.4). The faster moving major band was confirmed as a single triangle of three corners by comparative analysis of an identically sized programmable triangle that contains three distinct corners (A, B, C), each with a unique single-stranded overhang sequence which allows formation exclusively of the designed triangle with an A-B-C configuration but not other assemblies (Figure 5.3c, d; see also *Discussion of Programmable Large Triangle* later in this chapter).

The slower moving minor band is formed by both the symmetrical and programmable large triangles (see Figures 5.3, 5.4, 5.7), and is likely a dimer of triangles or a nonplanar hexagon with A-B-C-A-B-C configuration. Face-to-face dimer formation of pseudo-continuously closed large triangles involving 12 total base pair interactions was seen in the crystal structure packing of the long SVV IIa construct (Figure 5.15), and may be similar to the interactions causing the slower moving minor band for the large triangle in solution.

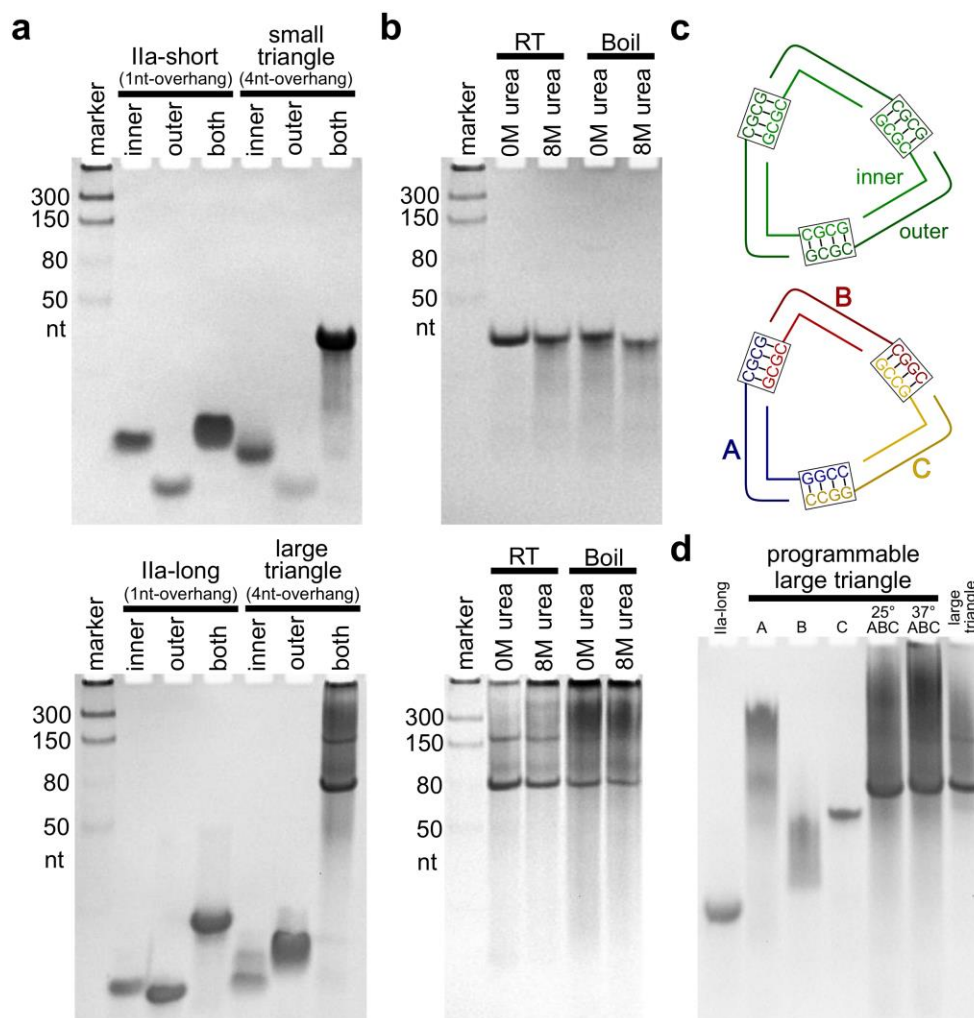


Figure 5.3: Nanotriangle assembly and stability. Structures were analyzed by native PAGE in the presence of 2.5mM MgCl₂. a) Top: Assembly of subdomain Ila short crystal construct containing single nucleotide (nt) overhangs (secondary structure shown in Figure 5.2a, top) and self-assembling small triangle construct containing four nucleotide overhangs (secondary structure shown in Figure 5.2c, top). Bottom: Assembly of subdomain Ila long crystal construct containing single nucleotide overhangs (secondary structure shown in Figure 5.2a, bottom) and self-assembling large triangle construct containing four nucleotide overhangs (panel c, top; secondary structure shown in Figure 5.2c, bottom). b) Stability of small triangle (top) and large triangle (bottom) when treated with or without 8M urea at room temperature (RT) or boiling. c) Diagram of large triangle (top) and programmable large triangle (bottom). d) Assembly of programmable large triangle (ABC) from corners A, B, and C.

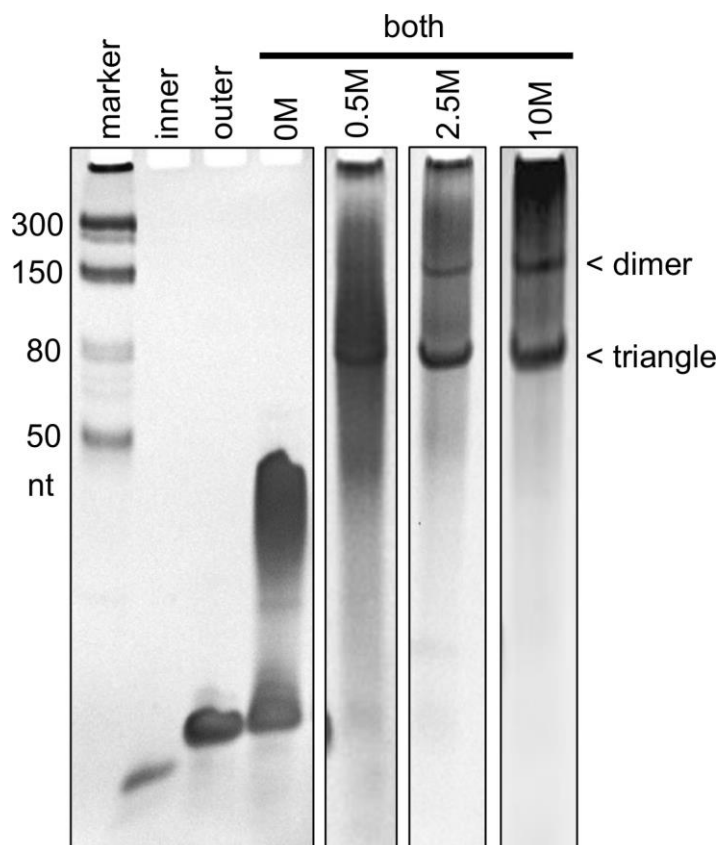


Figure 5.4: Magnesium ion concentration dependence of large nanotriangle self-assembly and formation of major and minor band species analyzed by 13% native polyacrylamide gel electrophoresis (PAGE) with 0, 0.5, 2.5, or 10mM MgCl_2 . Each box shows results from separate gel experiments. The corresponding magnesium concentration used in the reaction, loading and running buffers, as well as the gel are indicated in the legend on top. A single-stranded RNA ladder was loaded on each gel as a reference marker but shown only for the left-most gel. At 0M MgCl_2 major or minor band formation is not observed. At 0.5M MgCl_2 the major and minor band formed at a 10:1 ratio. At 2.5M MgCl_2 the major and minor band formed at a 1.7:1 ratio. At 10M MgCl_2 the major and minor band formed at a 1.3:1 ratio.

Stability of RNA Nanotriangles

Both RNA nanotriangles were evaluated for chemical and thermostability, and shown to resist boiling temperature as well as incubation with 8M urea (Figure 5.3b). After 5 minutes incubation at 100°C, the assembled small triangle was recovered largely intact (~70%; Figure 5.3b, top, boil, 0M urea), with slightly less recovery (~65%; Figure

5.3b, bottom, boil, 0M urea) of the assembled large triangle. Similarly, after 5 minutes incubation with 8M urea at room temperature, the small and large triangles were mostly intact (~60 and ~75% recovery, respectively; Figure 5.3b, RT, 8M urea). Boiling in the presence of 8M urea led to lower recovery of the small and large triangles (~40% and ~50%, respectively; Figure 5.3b, boil, 8M urea). Additional stability studies are looked at in the *Discussion of Triangle Stability* section at the end of this chapter.

Effect of Switch-Binding Ligand on RNA Nanotriangle Assembly

Preliminary studies were performed to determine effects of switch-binding ligands on RNA nanotriangle assembly and dissociation. Native PAGE analysis revealed a decrease in large triangle assembly efficiency in the presence of 500uM ligand⁵³ (Figure 5.5a, c) but not with control compounds (Figure 5.6). As a consequence of the ligand capture of elongated Iia RNA switches in the self-assembling constructs, end-to-end association of straightened corner units of the large triangle led to multimer formation. This was not observed when ligand was added to the small triangle, likely due to the lability of the construct which has only 7 base pairs flanking an internal loop of 5 unpaired bases and forms by all-or-nothing assembly from single strands. Dissociation of already assembled large triangles, which are less compact and stabilized by 15 base pairs flanking the internal loop, was observed upon incubation with ligand for the symmetrical (Figure 5.5b) as well as the programmable construct (Figure 5.7). These findings suggest that the RNA corner units retain their function as ligand-responsive switches when incorporated into nanotriangles. The ligand-triggered dissociation of nanotriangles and assembly of alternate structures opens potential avenues for the construction of RNA

nano-objects that respond to environmental signals. Additional nanostructure ligand binding studies are examined in the *Discussion of Ligand-Responsive Nanostructures* section at the end of this chapter.

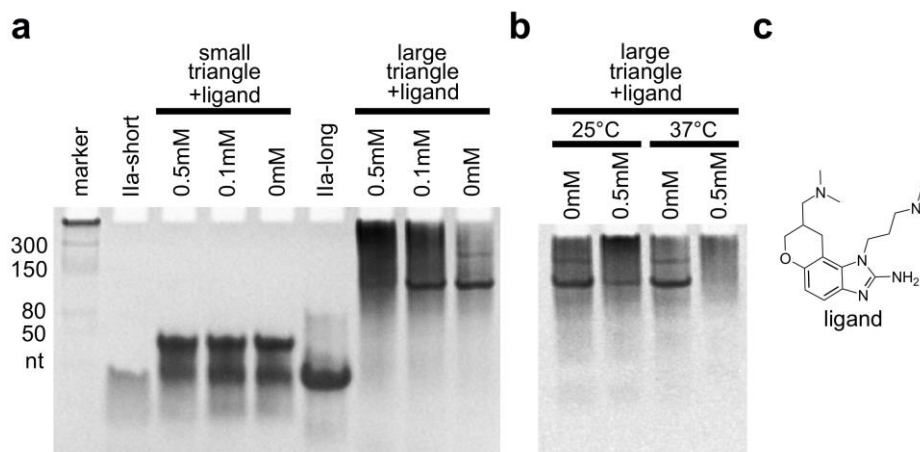


Figure 5.5: Nanotriangle self-assembly efficiency and dissociation in the presence of ligand. Ligand binding of the Ila switch captures an elongated conformation of the RNA.³⁹ a) Assembly of the small triangle is not affected in the presence of a binding ligand. Assembly of the large triangle is partially prevented in the presence of a binding ligand. Single corner units of the large triangle construct, captured by the ligand in an elongated conformation, prevent triangle formation and promotes end-to-end multimerization into longer species. b) Dissociation of large triangles was observed when incubated post-assembly with a binding ligand. c) Structure of the benzimidazole ligand which binds to the Ila RNA switch element of the nanotriangles.

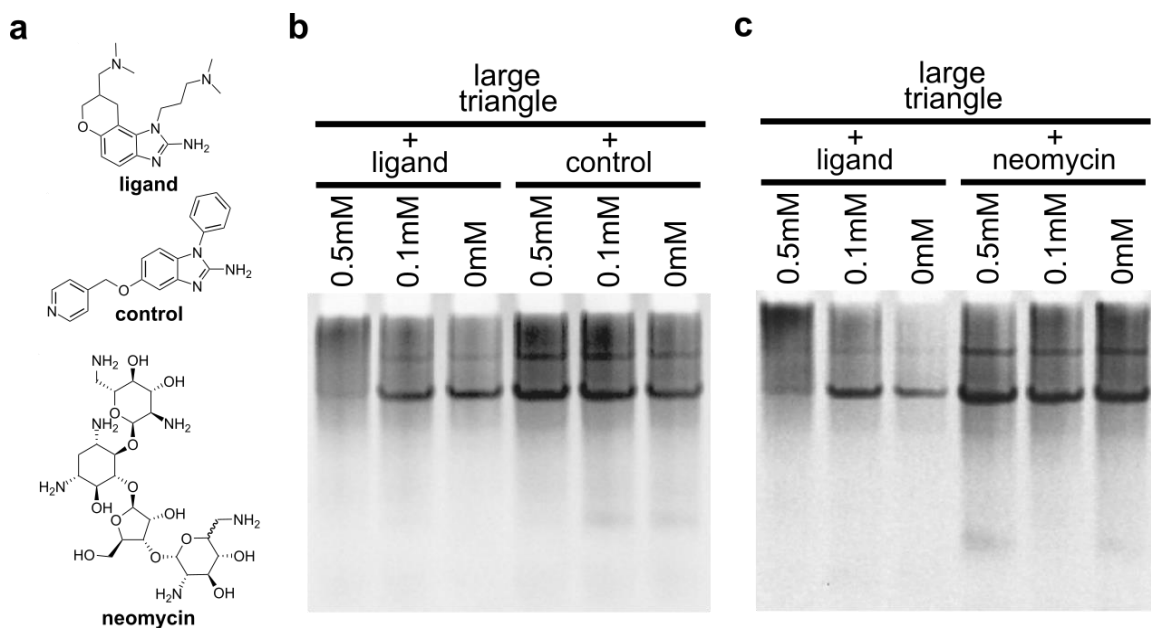


Figure 5.6: Large nanotriangle self-assembly efficiency in the presence of a binding ligand and control compounds. a) Structures of the benzimidazole ligand which binds to the IIA RNA switch element of the nanotriangles, the structurally related but non-binding aryl-substituted benzimidazole control, and the non-specific RNA binder neomycin. b) Large triangle assembly was disrupted only by specific IIA-switch binding benzimidazole ligand and not by a related control compound which was found to be inactive in a target-specific FRET binding assay.⁶⁵ c) Neomycin, an aminoglycoside which binds structured RNA promiscuously, also showed no effect on large triangle assembly.

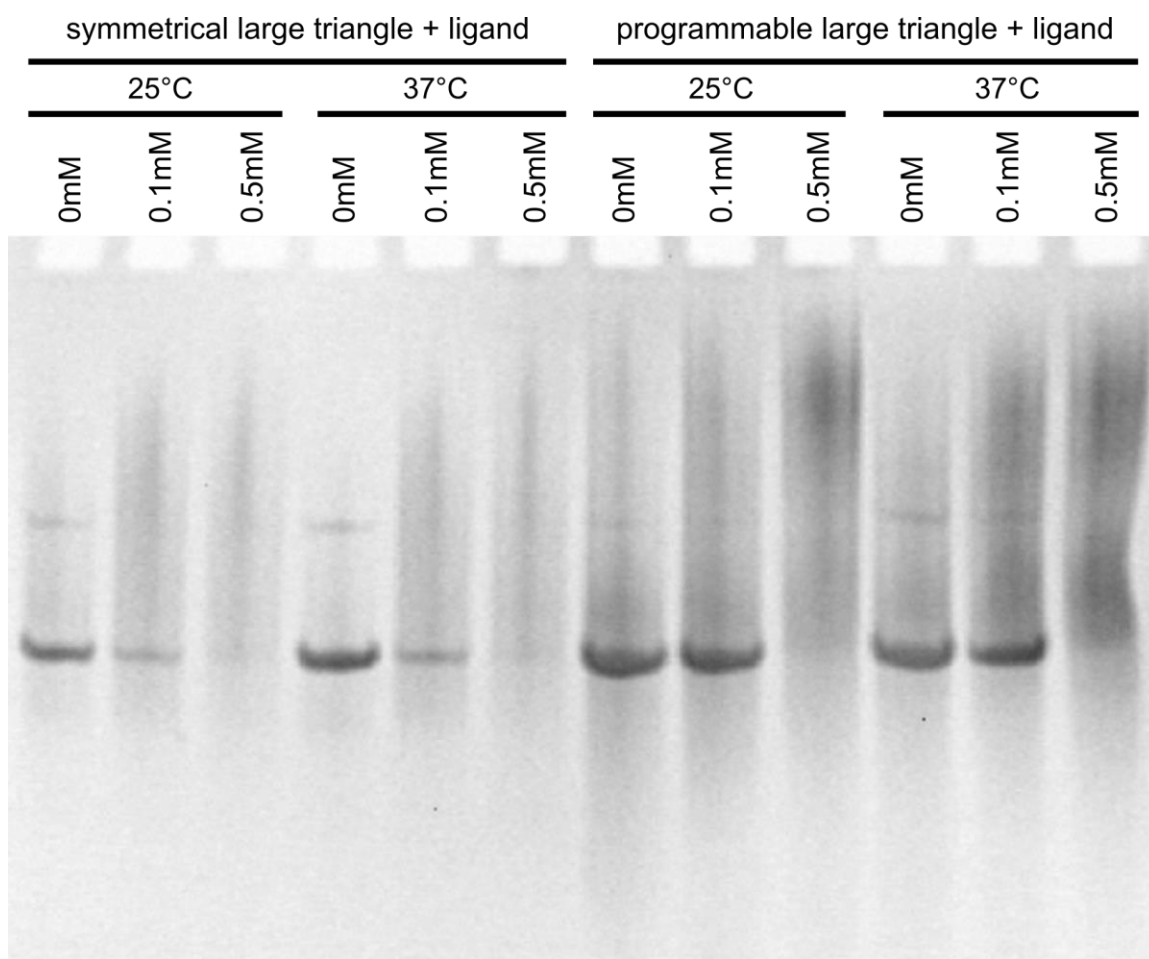


Figure 5.7: Dissociation of symmetrical and programmable large nanotriangles was observed when incubated post-assembly with 100 or 500 μ M benzimidazole ligand (Figure 5.5c) for 20 minutes at 25°C or 37°C.

Crystal Structure of Small RNA Nanotriangle

To investigate the three-dimensional structure of the small and large RNA nanotriangles, we crystallized both constructs. Well-diffracting crystals were obtained for the small triangle. Structure determination by X-ray diffraction at 2.6 Å resolution (Figures 5.8-5.12 and Table 5.1) revealed a circularly closed and continuously double-stranded RNA which exhibited an architecture similar to the pseudo-continuously closed

small triangle seen in the crystal packing of the short IIa construct (Figures 5.2b, 5.8). As designed, the nanotriangle comprises three identical and symmetrical corner units, each forming from an inner and outer RNA strand with four overhanging nucleotides which hybridize with neighboring corner strands. Triangle sides are composed of 11 base pairs and measure ~5 nm in length, while corners contain the IIa internal loop of 5 bases. The overall structure appears hexagonal at first glance as it is not planar but has distorted sides that twist to accommodate three 90° corner motifs in a closed triangular architecture. The resulting structure is more compact than a planar triangle of comparable side length. The 12 termini of the 6 single strands constituting the RNA triangle are located on the same face of the nanostructure (Figures 5.8-5.9). This feature offers an opportunity to build more complex structures and to functionalize the nanotriangle for sensor and materials applications.

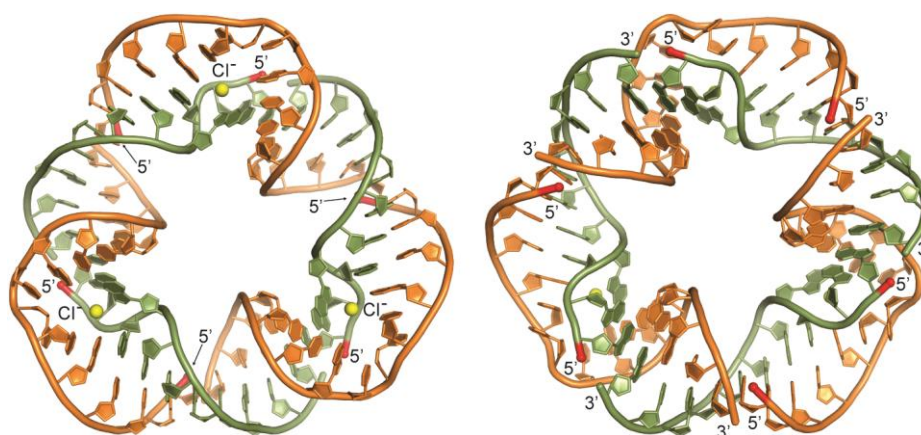


Figure 5.8: Crystal structure of the self-assembling RNA nanotriangle. Views from both sides of the triangle plane are shown. The back view (left) reveals three Cl⁻ ions (yellow spheres) bound at the Watson-Crick edge of A374 and C375. The terminal residues of all constituting oligonucleotides reside on one face of the triangle (front view, right). 5' termini are highlighted in red. Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB ID: 5CNR).

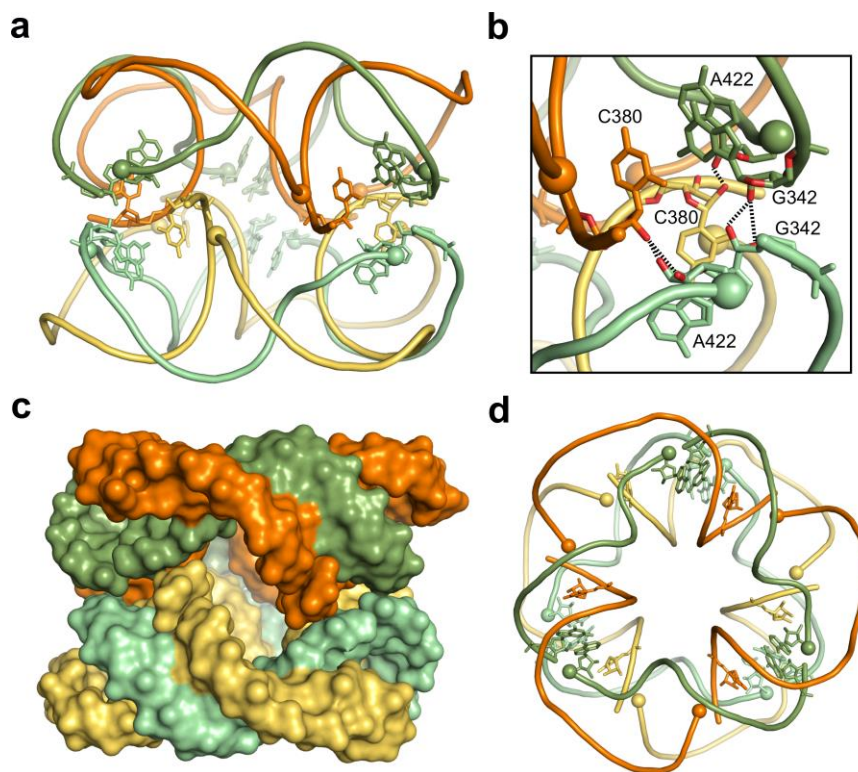


Figure 5.9: In the intermolecular packing of the crystal structure, the RNA nanotriangle forms face-to-face dimers. These triangle packing dimers form 15 intermolecular hydrogen bond contacts in addition to having complementary surface shapes that maximize packing. a) Two RNA triangles pack facing each other with the side that carries the oligonucleotide termini. Residues that are involved in hydrogen bonding contacts between triangles are shown in stick representation. 5' ends are marked by spheres. b) Hydrogen bond interactions between the sugars of two different C380-A422 pairs separated by the hydrogen bond interactions between the sugars of two G342 residues at the strand termini of the upper (orange, green) and lower (yellow, teal) triangle at the dimer interface. In total, 15 hydrogen bonds form across the packing interface of two triangles. c) Shape complementarity at the interface of RNA triangle dimers maximizes packing surface contacts. The sugar phosphate backbone at the termini of the outer strands in the upper triangle (orange) is fitted into the major groove of the bend in the outer strand of the lower triangle (yellow). d) Top down view of a dimer showing the backbone-into-groove fitting between outer strands of the upper and lower RNA triangles.

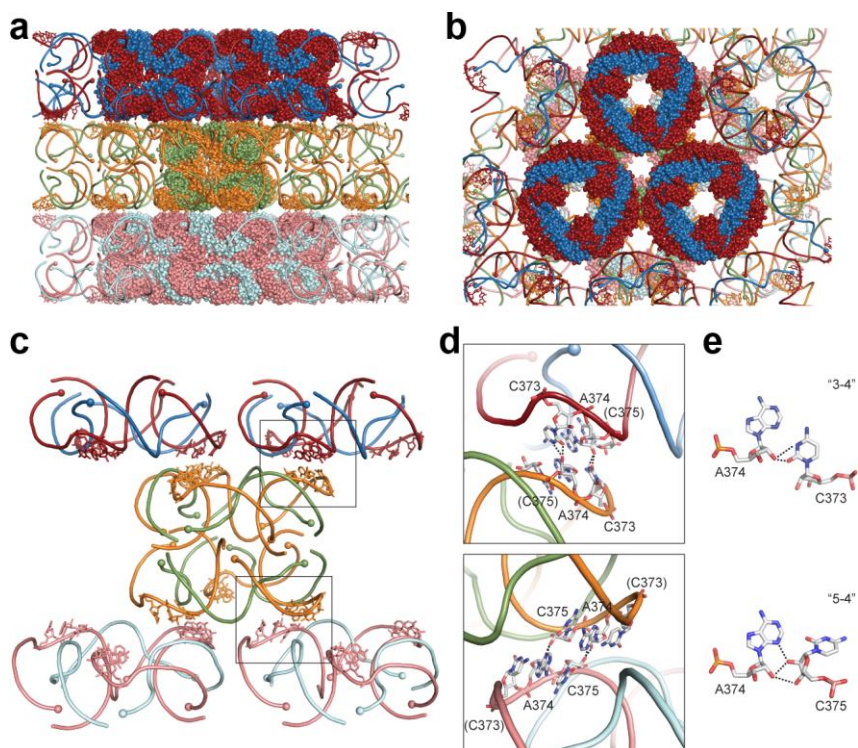


Figure 5.10: In the extended crystal packing, triangle dimers form layers that stack in alternating orientations. While no direct lateral hydrogen bond interactions are formed within layers, alternating dimer layers form hydrogen bond contacts with neighboring layers in one of two distinct ways which involve residues C373, A374, and C375 and alternate with each additional layer. a) Face-to-face triangle dimers form layers. Outer strands are shown in red/orange hues and inner strands in blue/green color. b) Consecutive layers of face-to-face triangle dimers stack in an alternating fashion. c) Packing contacts occur between dimer layers and involve RNA outer strands only. Lateral contacts between triangles within layers are not observed. d) Residues C373, A374 and C375 participate in packing interactions mediated by two distinct patterns of hydrogen bonding (indicated by boxes in panel c). e) The “3-4” pattern consists of two symmetrical C373○A374 base pairs in which the Watson-Crick edge of the cytidine forms two hydrogen bonds with the 2'OH group of the adenosine. The “5-4” motif includes two symmetrical C375○A374 pairs in which the cytosine ribose engages in three hydrogen bonds with the sugar edge of the adenosine.

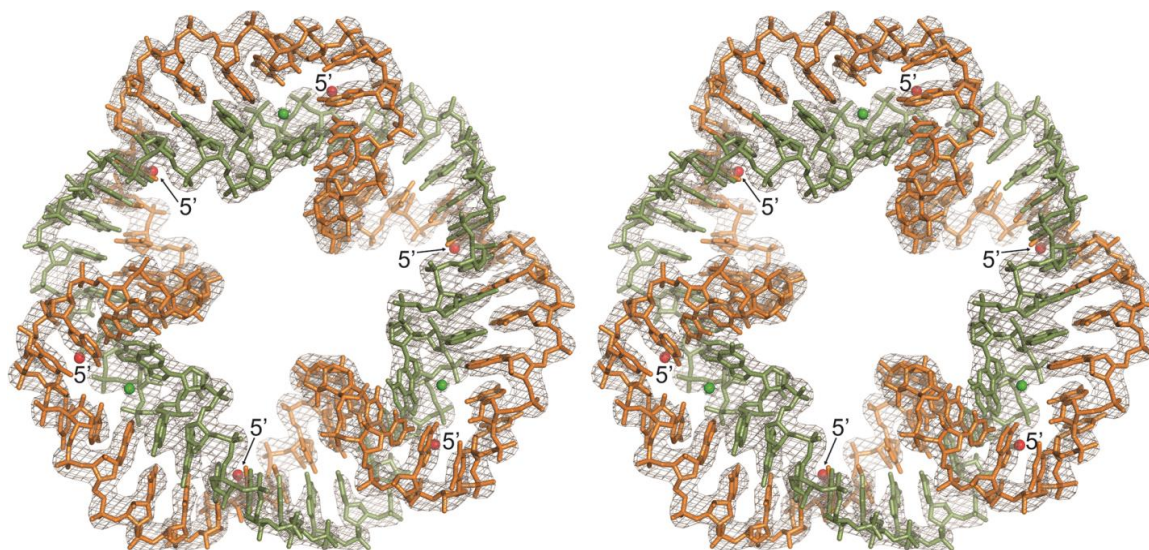


Figure 5.11: Stereo view of the crystal structure and electron density map ($2F_o - F_c$, contoured at 2.0σ) of the self-assembling RNA nanotriangle. Positions of oligonucleotide 5' termini are indicated by red spheres. While triangle self-assembly in solution was dependent on magnesium ion concentration, cations were not observed in the crystal structure, but three chloride anions per triangle were found at the Watson-Crick edge of A374 and C375 (green spheres) and supported by electron density and coordination environment. Similarly, magnesium ions were not observed in the short SVV IIa construct crystal that shows the pseudo-continuously closed small triangle packing, although magnesium was required for its folding in solution.³⁹

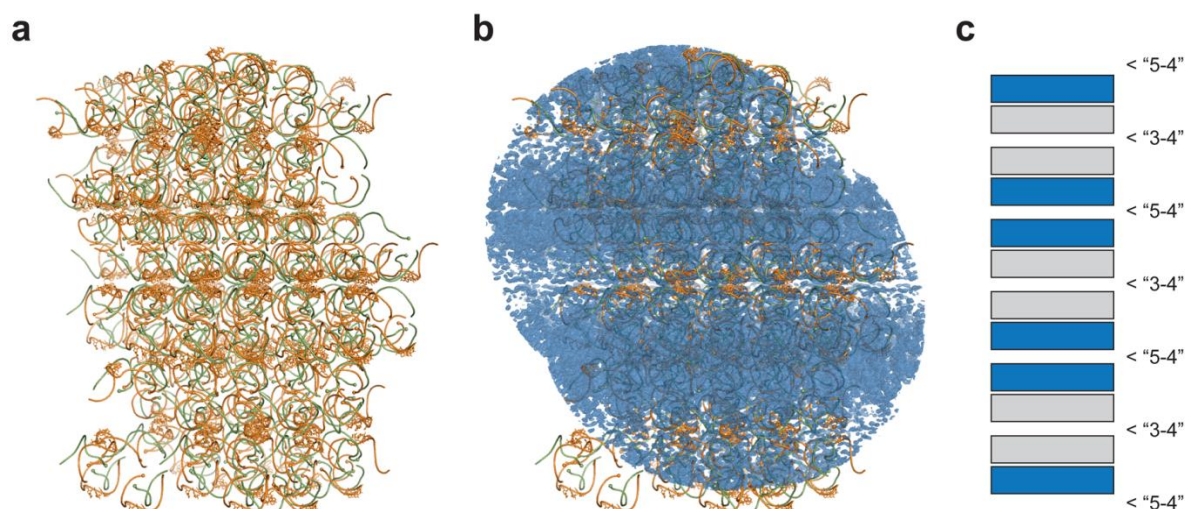


Figure 5.12: One-dimensional twinning disorder of crystal packing layers of the self-assembling RNA nanotriangle. a) Side view of six layers of RNA triangle face-to-face dimers. b) Electron density map ($2F_o-F_c$) around the six layers of RNA triangle dimers shown in panel a. Disordered layers are visible as stripes of low density (see discussion below). c) Arrangement of RNA triangle layers indicating disordered regions. Pairs of rectangles indicate face-to-face dimers of RNA triangles. Packing interactions between dimers are classified as “3-4” or “5-4” type, depending on the hydrogen bonding pattern (see Figure 5.10d, e). RNA triangles represented by blue rectangles are well ordered and completely covered by electron density while those indicated by grey rectangles are within disordered regions of non-continuous density.

The one-dimensional twinning disorder observed in the crystal packing gives rise to a regular pattern of weak electron density layers interspersed between well-defined regions. Stacking of consecutive layers of nanotriangles allows for several possible arrangements (see Figure 5.10) such that each successive layer may be displaced laterally relative to the one below it. Such twinning disorder occurs when successive planar crystal layers shift laterally into any of several alternative, energetically similar configurations. Weak stacking interactions between layers facilitate this type of twinning which has been investigated in detail by Trame and McKay¹³³ and was observed in other crystal structures, both of proteins (for example, Benirschke and coworkers)¹³⁴ and small molecules (for example, Schwarzenbach and coworkers).¹³⁵ As these authors pointed out, the regularity of the electron density in the ordered layers allows for structure refinement to acceptable parameters. This is confirmed by the quality of parameters obtained for the RNA nanotriangle crystal structure (see Table 5.1).

Table 5.1: Crystallographic data collection and refinement statistics for the self-assembling RNA nanotriangle.

Data Collection	
Wavelength (Å)	1.54
High-resolution limit (Å)	2.60
Low-resolution limit (Å)	19.40
Redundancy ^a	17.6 (2.0)
Completeness (%) ^a	83.2 (11.7)
$I/\sigma(I)$ ^a	35.53 (1.83)
Total reflections	88115
Unique reflections	5011
Refinement	
Space group	R32
Cell dimensions (Å)	
<i>a</i>	59.93
<i>b</i>	59.93
<i>c</i>	262.62
α	90
β	90
γ	120
R_{work}/R_{free}	0.19 / 0.24
No. atoms	
RNA atoms	1136
Ligand	Cl ⁻
Mean <i>B</i> factors (Å ²)	
RNA	152.1
Ligand	124.7
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.445
Dihedral angles (°)	14.08

^aNumbers in parentheses are for the highest-resolution shell.

Discussion of Crystal Structures used in Design

The short SVV IIa RNA crystal structure (PDB ID: 4P97) displays unique intermolecular interactions and packing arrangements (Figure 5.13). This oligonucleotide construct adopts an L-shaped conformation orienting an upper and lower helix perpendicular to each other (Figure 5.2a; top). The upper and lower helices consisting of 5 base pairs each, are connected by an internal loop of 5 unpaired nucleotides, and contain single 3' nucleotide overhangs. These overhangs allow for intermolecular base pair formation. The crystal packing reveals this L-shaped RNA forming each of three identical corners of a triangular structure (Figure 5.2b-c, top). While the secondary structure model for this pseudo-continuously closed triangle appears hexagonal (Figure 5.2c, top), its three dimensional structure when viewed top-down appears circular (Figure 5.2b, top; Figure 5.13b). In order to accommodate three 90° corner motifs into a closed equilateral triangle architecture, the corners distort and twist out of the plane. This results in a structure that is more compact (shorter sides, taller height) than would be a planar equilateral triangle comprising an equal number of base pairs per side and corner angles of 60°. Its sides of 11 base pairs are roughly 5.7 nm long and its height is roughly 2.3 nm. The interior of the triangle measures 1.1 nm from the innermost reaching parts of two outer strands and 1.9 nm from these outer strands to the innermost reaching part of the inner strands. The pseudo-continuously closed triangles pack in a lateral pattern which repeats every three layers (Figure 5.13a, b). Identical lateral contacts within each layer involve each triangle forming six total hydrogen bond interactions with six neighboring triangles. Layers contain triangles facing in the same direction. A total of 21 hydrogen bond contacts form on each triangle face with three overlaying triangles (Figure 5.13c, d)

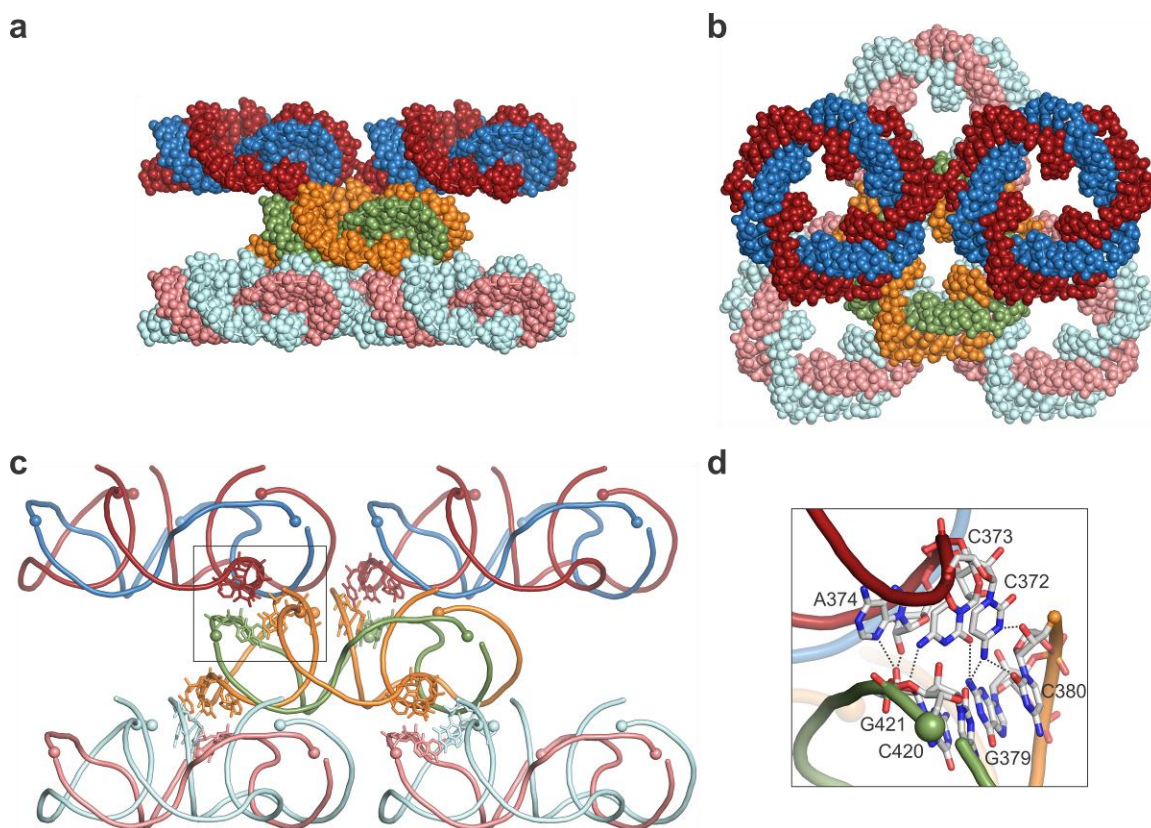


Figure 5.13: Crystal packing of pseudo-continuously closed triangles in the SVV subdomain IIa short oligonucleotide construct (see Figure 5.2a, top). a) Side view of triangle layers. b) Top-down view reveals triangles arranged in a pattern that repeats every three layers. c) Residues that are involved in hydrogen bonding contacts between layers are shown in stick representation. 5' termini of oligonucleotides are indicated by spheres. d) Hydrogen bond interactions between both strands in one triangle (orange, green) and the outer strand (red) of the overlaying triangle. In total, 21 hydrogen bonds form across the interface of one triangle with three overlaying triangles.

The long SVV IIa RNA crystal structure (PDB ID: 4PHY) also displays unique intermolecular interactions and packing arrangements (Figures 5.14-5.16). This extended oligonucleotide construct adopts an L-shaped conformation identical to the short SVV IIa structure (Figure 5.2a). The upper and lower helices of this construct consist of 13 and 6 base pairs respectively, and also contain single 3' nucleotide overhangs allowing for intermolecular base pair formation. The crystal packing reveals this extended L-shaped

RNA forming each of three identical corners of a triangular structure (Figure 5.2b, c, bottom). This pseudo-continuously closed large triangle appears more triangular than the pseudo-continuously closed small triangle (Figure 5.2b, c) due to its longer sides of 20 base pairs. In order to accommodate three 90° corner motifs into a closed equilateral triangle architecture, the corners also distort and twist out of the plane (Figure 5.14). This results in a structure that is more compact (shorter sides, taller height) than would be a equilateral triangle comprising an equal number of base pairs per side and corner angles of 60° . This large triangle's sides are roughly 8.0 nm long on their outer edge and 4.5 nm long on their inner edge. Its height is roughly 2.7 nm. The centermost point of each side's interior measures 2.7 nm to the symmetrical point on each neighboring side.

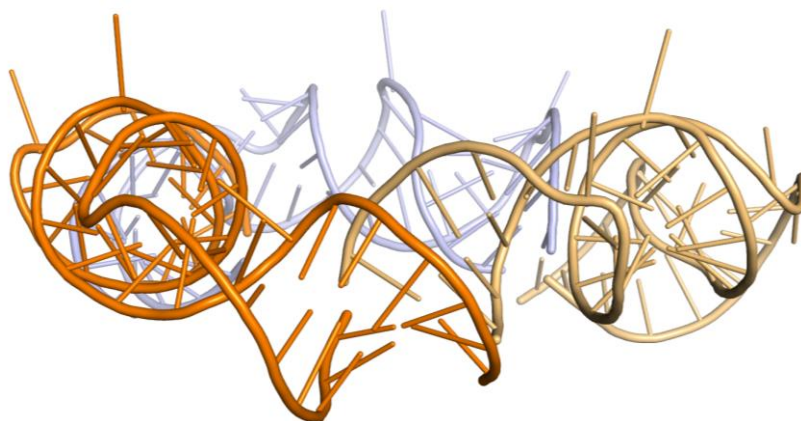


Figure 5.14: Side view of pseudo-continuously closed large triangle in crystals of the SVV subdomain IIa long oligonucleotide construct (see Figure 5.2a, b, bottom). This perspective shows into the upper helix of the front left corner motif (orange) viewed from the internal loop. The lower helix begins at the lowest point from this perspective and gradually slopes upwards to where it connects to the upper helix of the neighboring corner motif (light orange) where it continues sloping upwards until it reaches the highest point at the internal loop. This helical orientation and upward slope are symmetrical for each of the three triangle sides.

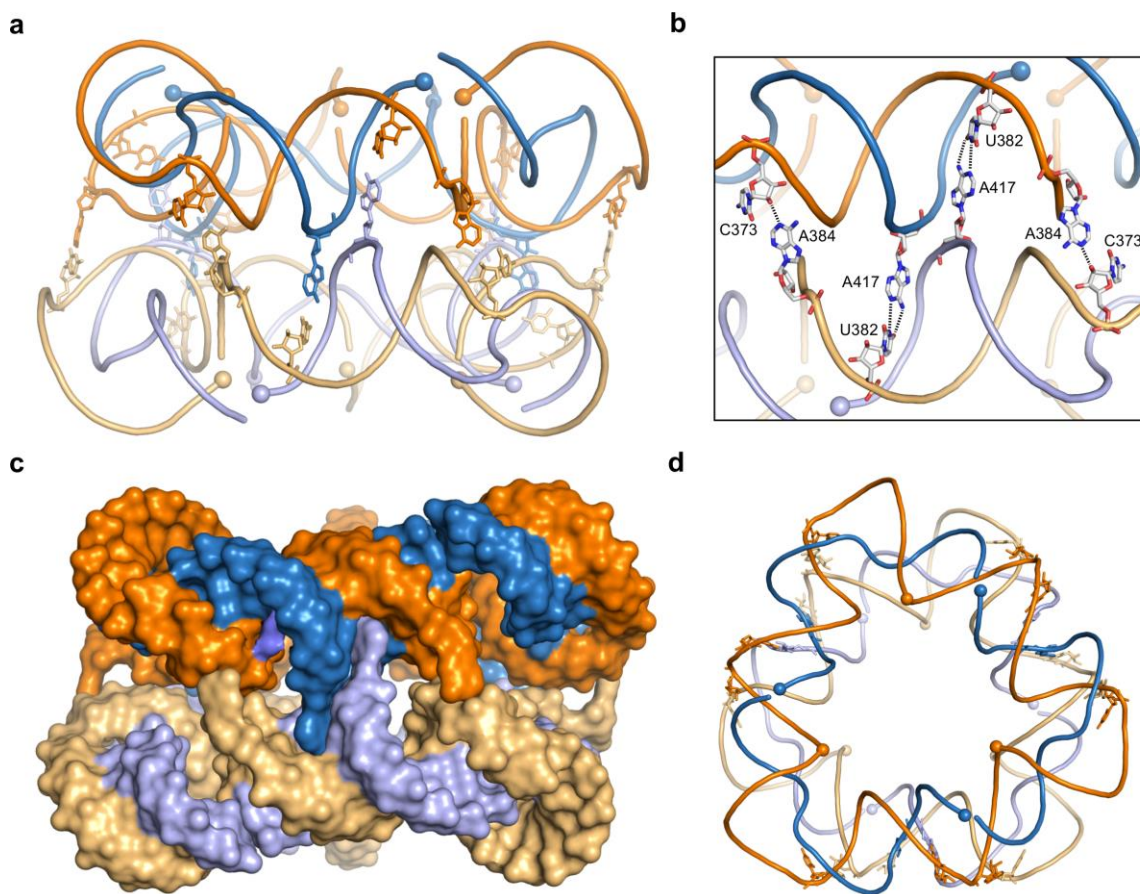
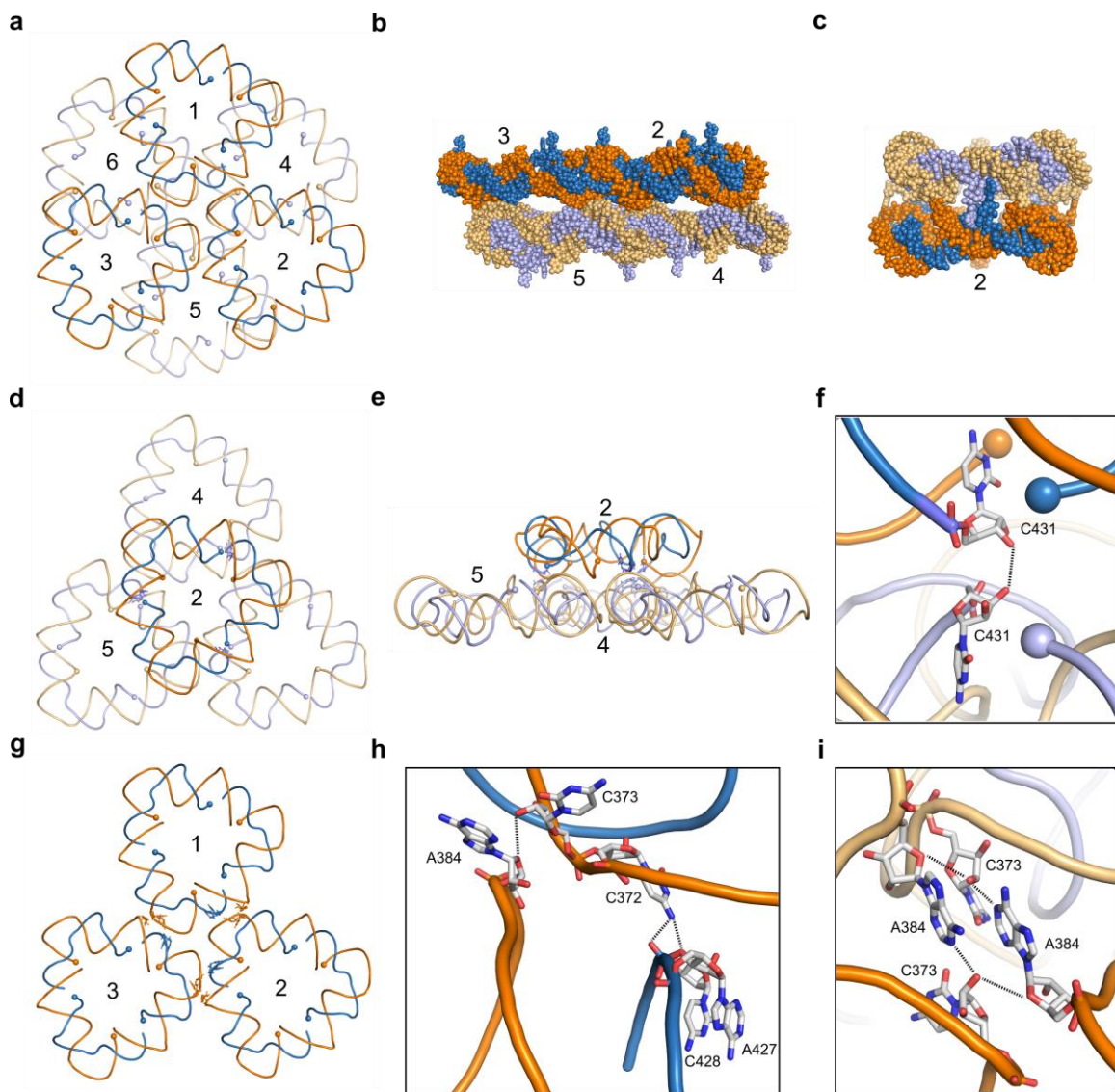


Figure 5.15: Face-to-face dimer formation of pseudo-continuously closed large triangles in crystals of the subdomain IIa long oligonucleotide construct (see Figure 5.2a, b, bottom). a) Side view of triangle dimer packing interactions which involves 12 total base pair interactions. Outer strands are shown in orange hues and inner strands are shown in blue/violet hues. b) The packing of each triangle side consists of two symmetrical cross-over A417-U382 Watson-Crick base pairs and two symmetrical A384-C373 base pairs in which the cytosine 2'OH group forms a hydrogen bond with the Watson-Crick edge of the adenosine. c) Side-view surface representation of dimer showing unique cross-over base pairing interactions between triangles. d) Top-down view of triangle dimers.

Figure 5.16: End-to-end and side-to-side packing of pseudo-continuously closed large triangles in crystals of the long subdomain IIa oligonucleotide construct (see Figure 5.2a, b, bottom). Outer strands are shown in orange hues and inner strands are shown in blue/violet hues. a) Top down view of two layers of end-to-end crystal packing not involved in dimer formation. b) Side view of end-to-end packing shown in panel a. Nucleotides involved in intermolecular base pairing between dimers are extend out from the top and bottom of these layers c) Side view of triangle dimer (Figure 5.15c) involving triangle 2 and a triangle in the layer above it. d) Shifted top down view of triangle 2 interacting in an end-to-end fashion with the layer beneath it. Triangles 4 and 5 can also be seen in panel a. e) Side view of end-to-end packing shown in panel d. f) Single hydrogen bond formation between one of three C431 residues of triangle 2 and a C431 residue of one of the 3 triangles beneath 2. g) Top down view of triangles packing side-to-side. h) Hydrogen bonds formed between triangles within a layer. Six symmetrical interactions involving the 3 hydrogen bonds shown (18 total) are formed for each triangle with the surrounding six triangles within a layer i) Hydrogen bond and packing interactions involving two A384○C373 pairs involved in dimer formation (see panel c and Figure 5.15b) and side-to-side interactions (see panel h).



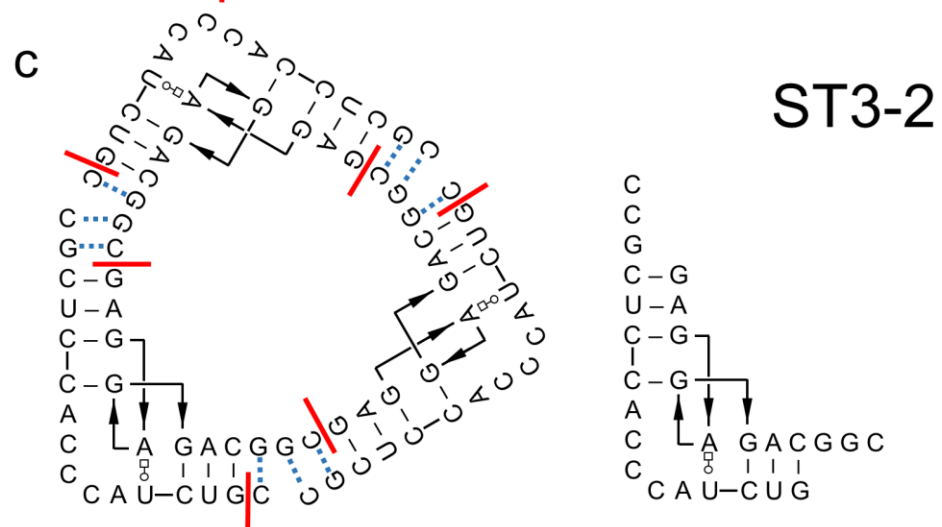
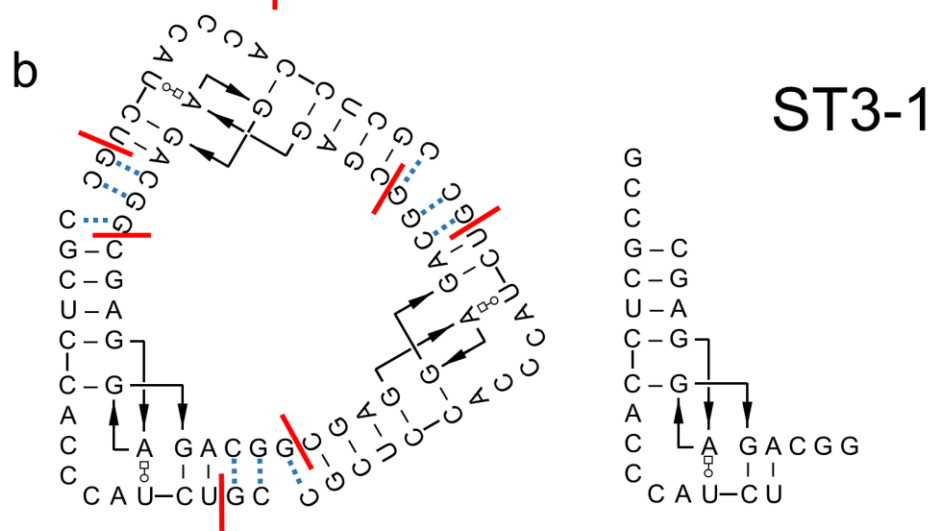
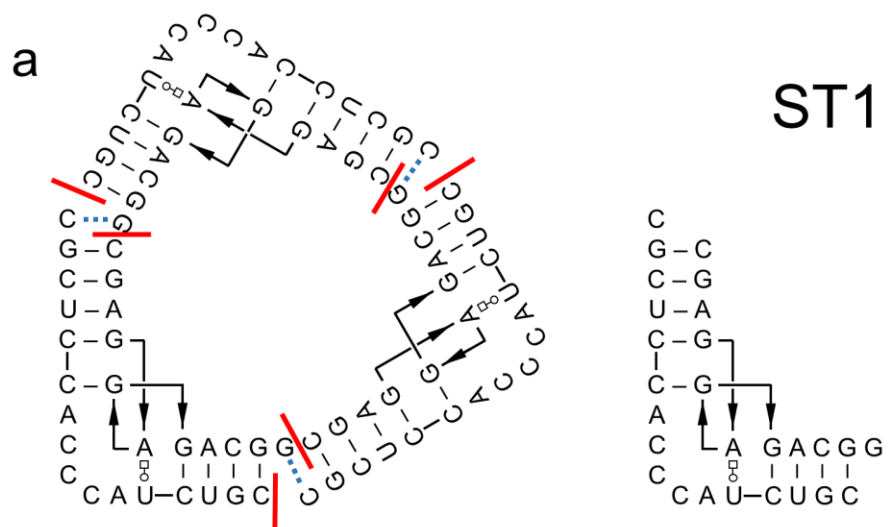
Pseudo-continuously closed large triangles form face-to-face dimers in the SVV extended Iia crystal structure (Figure 5.15). Dimer interactions involve 3 pairs of symmetrical A417-U382 Watson-Crick base pairs and 3 pairs of symmetrical A384-C373 hydrogen bonding interactions (Figure 5.15a, b). In the A417-U382 interaction, the adenosine of one triangle replaces the dimer triangle's adenosine by reaching across to form a Watson-Crick base pair with the dimer triangle's uridine and stacks in between the neighboring base pairs without interrupting the RNA helix. A384 flips out and its Watson-Crick edge forms a hydrogen bond interaction with the 2'OH group of the unpaired C373 corner residue. This results in a very compact dimer structure stabilized by both base pairing and stacking interactions (Figure 5.15c). As dimers pack face-to-face, each dimer triangle's nucleotide numbering moves in opposite directions, and each is slightly rotated in respect to the other (Figure 5.15d).

The pseudo-continuously closed large triangles pack in a lateral pattern which repeats every two layers (Figure 5.16). These lateral layers interact with neighboring layers in a face-to-face dimer fashion (Figures 5.15, 5.16c) or end-to-end fashion (Figure 5.16a, b, d-f). End-to-end triangle layers have some small shape complementarity (Figure 5.16a, b), and each triangle makes only three total end-to-end hydrogen bonds, one to each of three triangles in this neighboring layer (Figure 5.16d-f). Side-to-side lateral interactions involve 18 total hydrogen bond interactions with 6 neighboring triangles (Figure 5.16g-h). A384-C373 pairs involved in dimer layers, also participate in these side-to-side interactions (Figure 5.16i).

Discussion of Self-Assembling Constructs

In addition to exemplary self-assembling constructs containing four nucleotide overhangs (Figure 5.2c) which have been discussed in great detail, other self-assembling constructs were studied to determine optimal triangle overhang size and location. Self-assembling small triangle constructs of identical size but varying oligonucleotide termini and number of overhanging nucleotides were constructed and explored for self-assembling ability (Figures 5.17-5.19). Construct ST4-2 (four overhanging nucleotides; small triangle discussed previously and studied by crystallography) migrated as a stable single band consistent with its size and hydrodynamic radius, and displaying a high intensity of ethidium bromide staining consistent with dsRNA (Figures 5.3a-b, 5.18-5.19). ST5-1 also displayed a major band with high staining intensity, but it migrated faster than ST4-2 and showed greater smearing (Figures 5.18-5.19).

Figure 5.17: Self-assembling small RNA triangle constructs of identical size but varying oligonucleotide termini and number of overhanging nucleotides. a) Construct ST1 (small triangle, one-nucleotide overhang) is the same as the short SVV IIa crystal construct (Figure 5.2a, b, top; Chapter 2, Figure 2.10) and is shown here for comparison even though it does not self-assemble in solution. b) Construct ST3-1 (small triangle, three-nucleotide overhang, construct 1). c) Construct ST3-2 (small triangle, three-nucleotide overhang, construct 2). d) Construct ST3-3 (small triangle, three-nucleotide overhang, construct 3). e) Construct ST4-1 (small triangle, four-nucleotide overhang, construct 1). f) Construct ST4-2 (small triangle, four-nucleotide overhang, construct 2). g) Construct ST5-1 (small triangle, five-nucleotide overhang, construct 1).



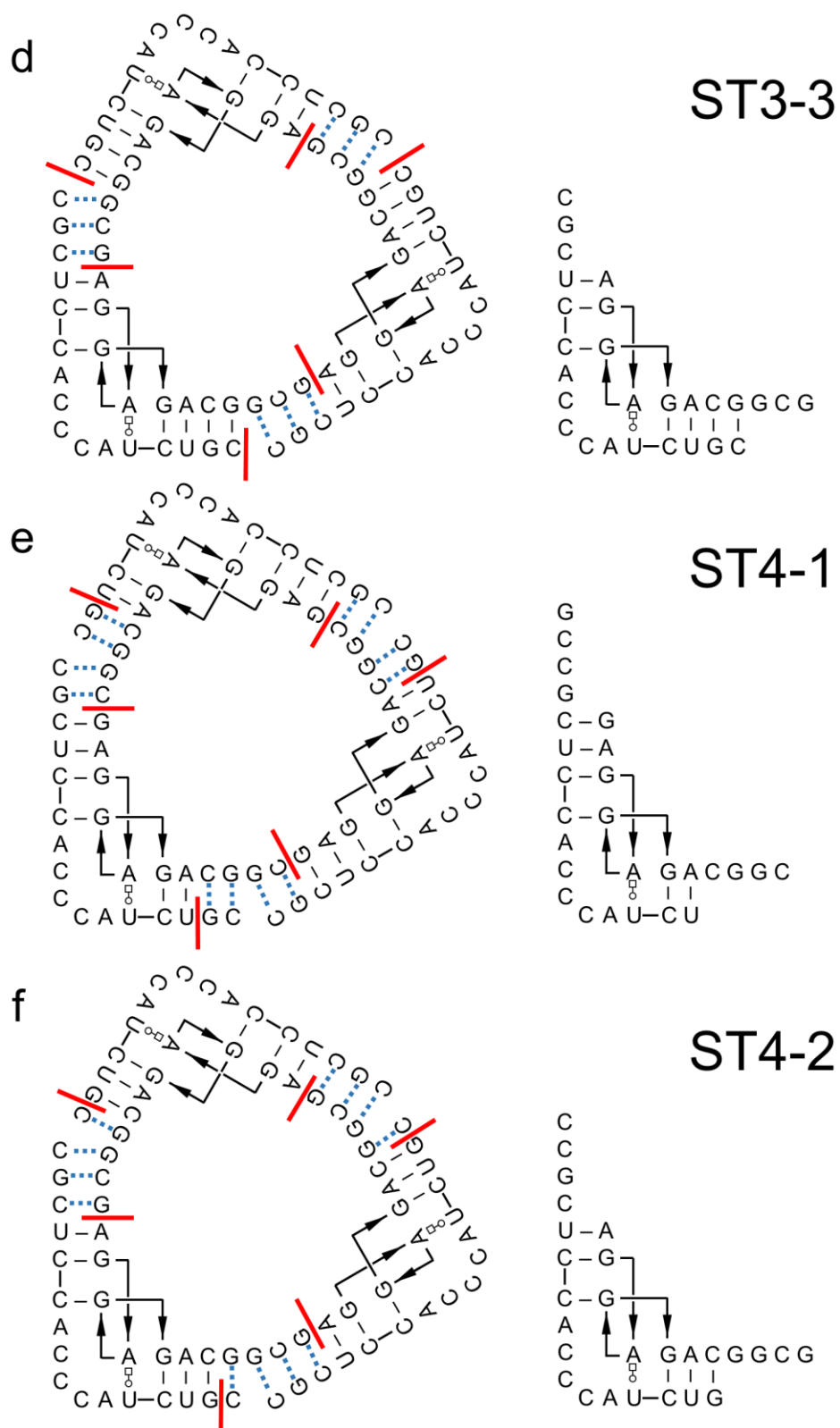


Figure 5.17: Self-assembling small RNA triangle constructs of identical size but varying oligonucleotide termini and number of overhanging nucleotides. *Continued.*

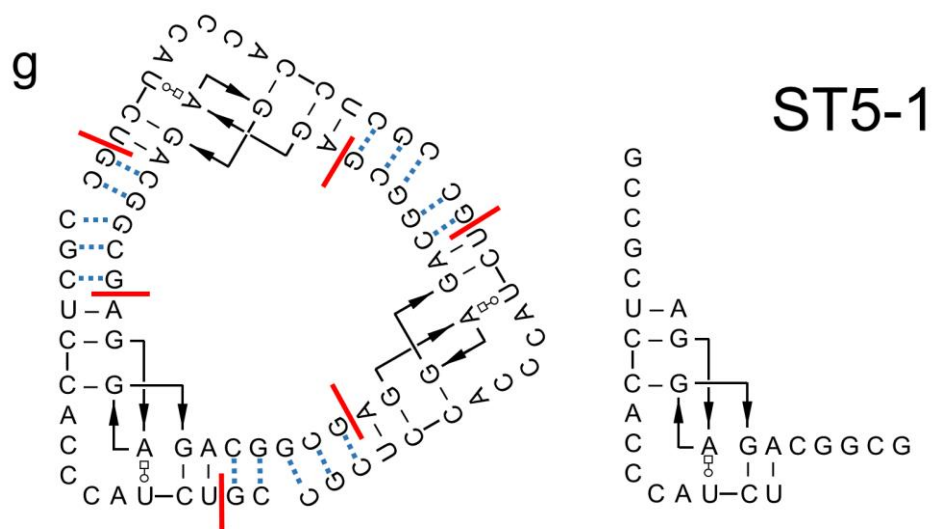


Figure 5.17: Self-assembling small RNA triangle constructs of identical size but varying oligonucleotide termini and number of overhanging nucleotides. *Continued.*

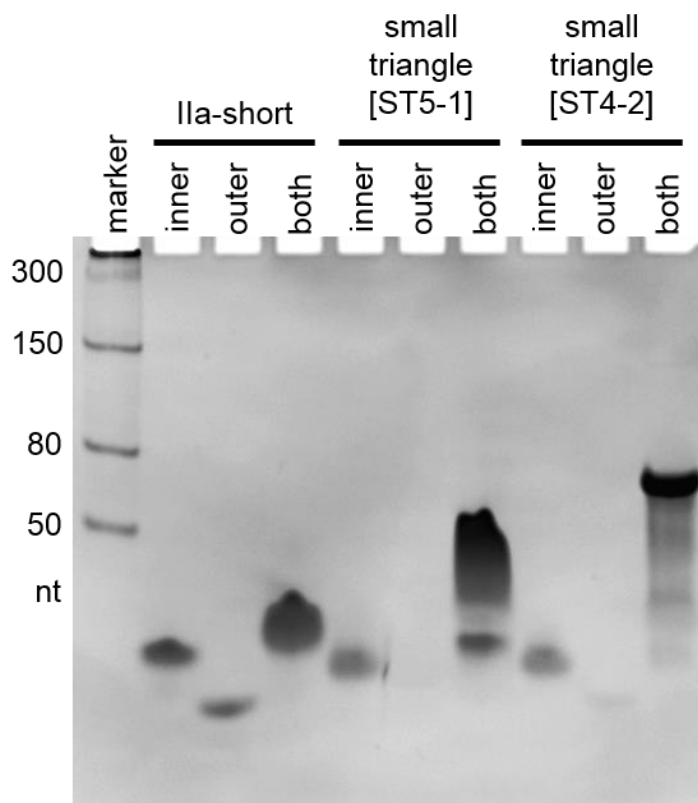


Figure 5.18: Native PAGE gel in 5mM MgCl₂ of self-assembling small RNA triangle constructs ST5-1 (5 base pair overlapping sequence) and ST4-2 (4 base pair overlapping sequence) with lanes showing inner or outer ssRNA or both mixed to assemble nanostructure. The Ila-short construct is shown for a size reference.

3nt, 4nt, and 5nt overhangs are all tolerated to varying degrees in the self-assembly of the small triangle. 4nt overhang variants ST4-1 and ST4-2 form the cleanest single band species consistent with the small triangle's size. 3nt and 5nt overhang variants form less stable single species with greater smearing. Hybrid RNA/DNA variants of ST4-2 do not form high ordered assembly products as these bands migrate faster than the single Ila corner unit. The DNA only ST4-2 migrates as a single band faster than the hybrid constructs.

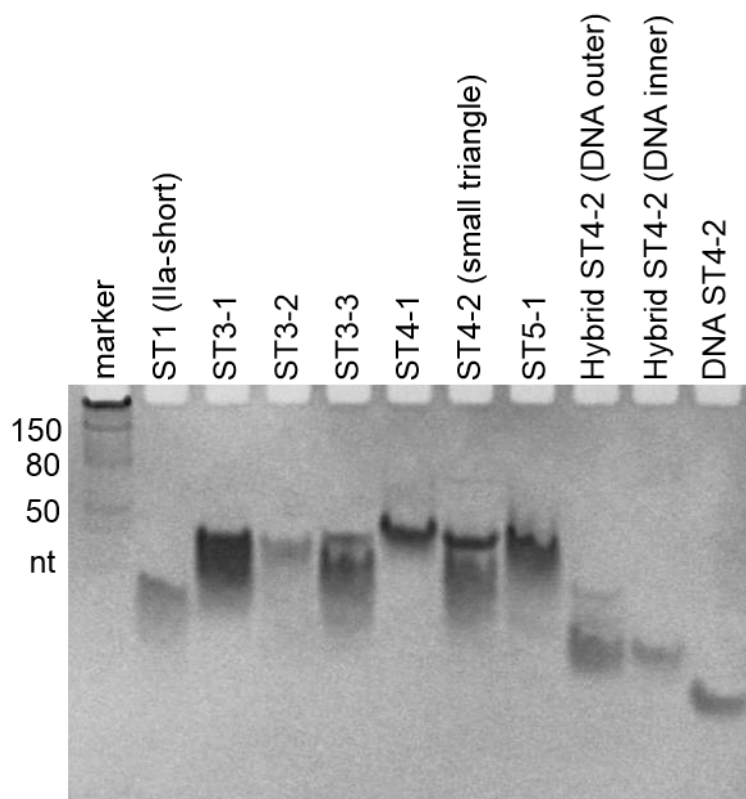


Figure 5.19: Native PAGE gel in 5mM MgCl₂ of self-assembling small triangle constructs ST1 (IIa-short), ST3-1, ST3-2, ST3-3, ST4-1, ST4-2 (small triangle), ST5-1, both DNA/RNA and RNA/DNA inner/outer strand hybrid ST4-2 (small triangle), and DNA ST4-2.

Self-assembling small triangle constructs of various size and varying numbers of overhanging or underhanging nucleotides were also constructed and explored for self-assembling ability (Figure 5.20-5.23). A small triangle (+1bp) construct was designed by inserting one C–G pair per triangle side adjacent to the overhanging nucleotides and above the internal loop bent motif of the exemplary small triangle construct (Figure 5.20a). The G was inserted on the 5' terminus of the inner strand and the corresponding C inserted into the appropriate interior location of the outer strand. This construct is designed to be an 87 nucleotide triangle (6 more nucleotides than the small triangle) and

indeed self-assembles into a species which migrates as a single band consistent with its size and hydrodynamic radius on a native PAGE gel (Figure 5.21, small triangle +1bp, both). It migrates slightly slower than the small triangle of 81 nucleotides, but slightly faster than the square of 100 nucleotides (Figure 5.22, small triangle +1bp).

A small triangle (+2bp) construct was designed by inserting two C–G pairs per triangle side adjacent to the overhanging nucleotides and above the internal loop bent motif of the exemplary small triangle construct (Figure 5.20b). Two G's were inserted on the 5' terminus of the inner strand and the corresponding two C's inserted into the appropriate interior location of the outer strand. This construct is designed to be a 93 nucleotide triangle (12 more nucleotides than the small triangle). This construct does not self-assemble into a 93nt triangle as designed, but likely forms multiple structures of various sizes as its band appears as a large smear migrating much slower than the small or large triangles (Figures 5.22-5.23). High Mg^{2+} concentration ($\geq 10mM$) appears to partially stabilize a slower migrating band species roughly twice the size of a dimer of large triangles (Figure 5.24).

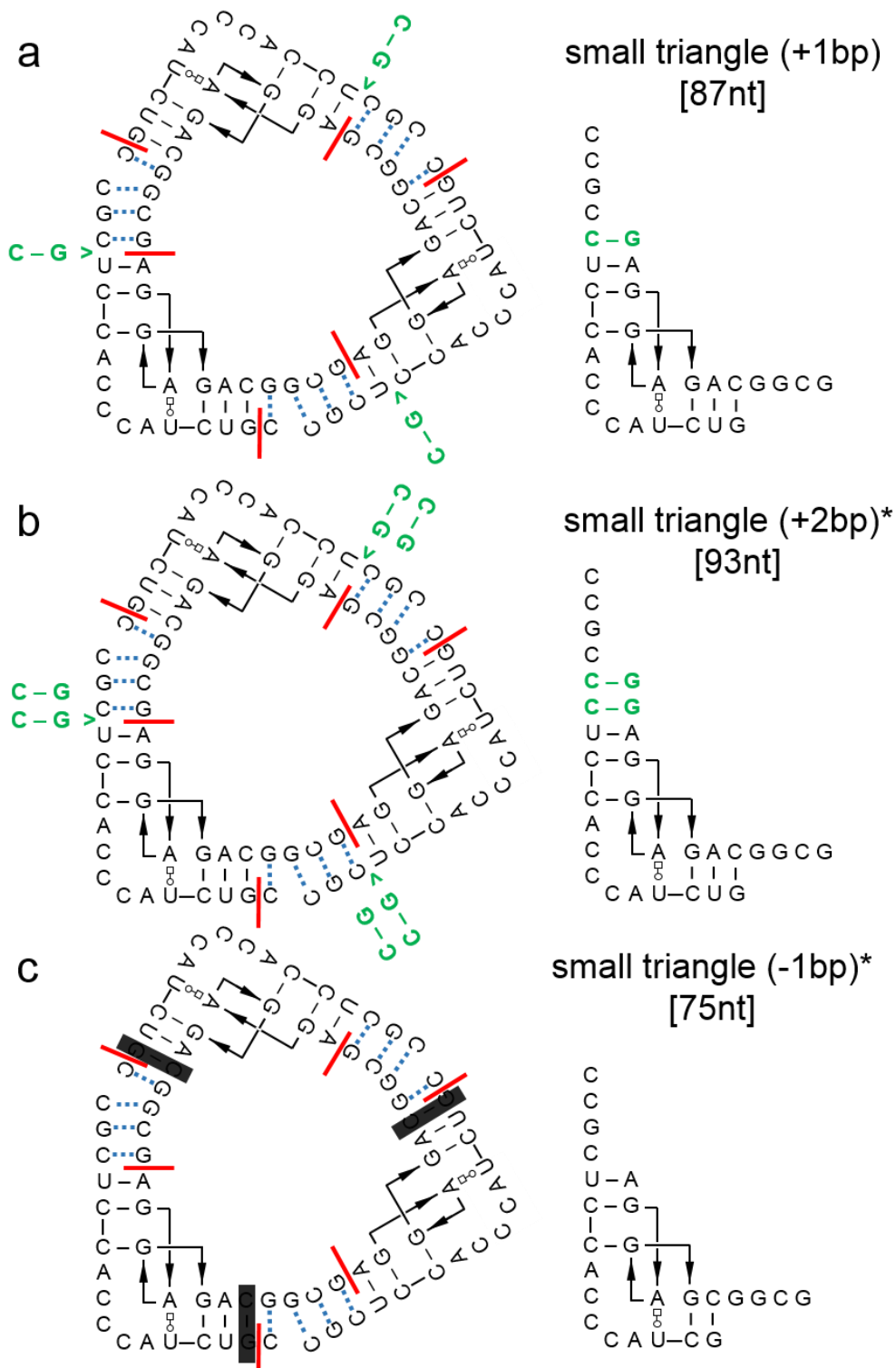
A small triangle (-1bp) construct was designed by deleting one G–C pair per triangle side adjacent to the overhanging nucleotides and below the internal loop bent motif of the exemplary small triangle construct (Figure 5.20c). The G residue was deleted from the 5' terminus of the outer strand and the corresponding C residue was deleted from the appropriate interior location of the inner strand. This construct is designed to be a 75 nucleotide triangle (6 less nucleotides than the small triangle). This construct does not self-assemble into a 75nt triangle as designed, likely due to steric restraints in forming a triangle from three 90° corner units (Figure 5.22). The corner

units of the small triangle of 81 nucleotides must already twist out of a planar structure to be accommodated into a triangular structure, and the shorter sides of this construct would only make this more difficult to accommodate.

A derivative of the small triangle was designed with two 5'-dangling G residues off the inner strand of each triangle side by using the small triangle outer strand and the small triangle (+2bp) outer strand (Figure 5.20d). This construct was used to explore the small triangle's toleration of modifications to the 5' terminus of the inner strand. This construct does not self-assemble into a species consistent with the size of an 87 nucleotide triangle (Figure 5.22, small triangle, 5'-inner 2nt dangling end).

Derivatives of the small triangle (+1bp) and the small triangle (+2bp) were designed with one unpaired C residue per triangle side (Figure 5.20e, f). The small triangle (+1bp) derivative was constructed with the small triangle (+1bp) outer strand and the small triangle inner strand but did not self-assemble into a species consistent with the size of an 84 nucleotide triangle (data not shown). The small triangle (+2bp) derivative was constructed with the small triangle (+2bp) outer strand and the small triangle (+1bp) inner strand but did not self-assemble into a species consistent with the size of a 90 nucleotide triangle (Figure 5.22, small triangle +2bp, 1nt unpaired per side). These constructs demonstrate that these triangles do not tolerate flexible single stranded nucleotides in their double stranded triangle sides, but require the stability of completely double stranded sides.

Figure 5.20: Self-assembling small triangle constructs of various size and varying numbers of overhanging or underhanging nucleotides. a) Small triangle (+1bp) construct designed to be an 87nt triangle. b) Small triangle (+2bp) construct designed to be an 93nt triangle. c) Small triangle (+1bp) construct designed to be a 75nt triangle. d) Small triangle (5'-inner+2nt dangling end) construct designed to be an 81nt triangle with 2 nucleotides added to the 5' terminus of each inner strand. e) Small triangle (+1bp, 1nt unpaired per side) construct designed to be an 84nt triangle with a single unpaired C residue per side. e) Small triangle (+2bp, 1nt unpaired per side) construct designed to be a 90nt triangle with a single unpaired C residue per side. *These constructs do not assemble as designed.



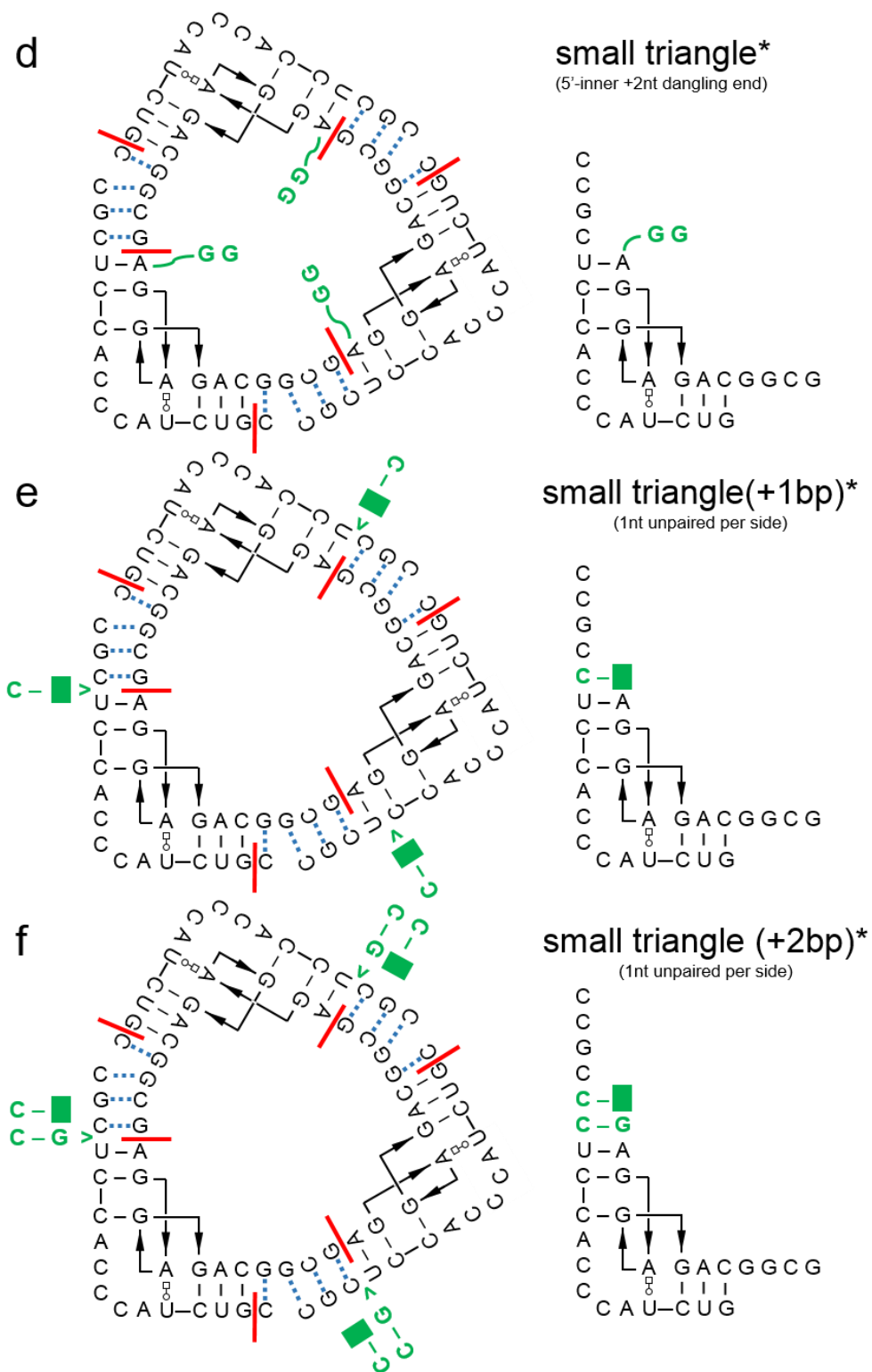


Figure 5.20: Self-assembling small triangle constructs of various size and varying numbers of overhanging or underhanging nucleotides. *Continued.*

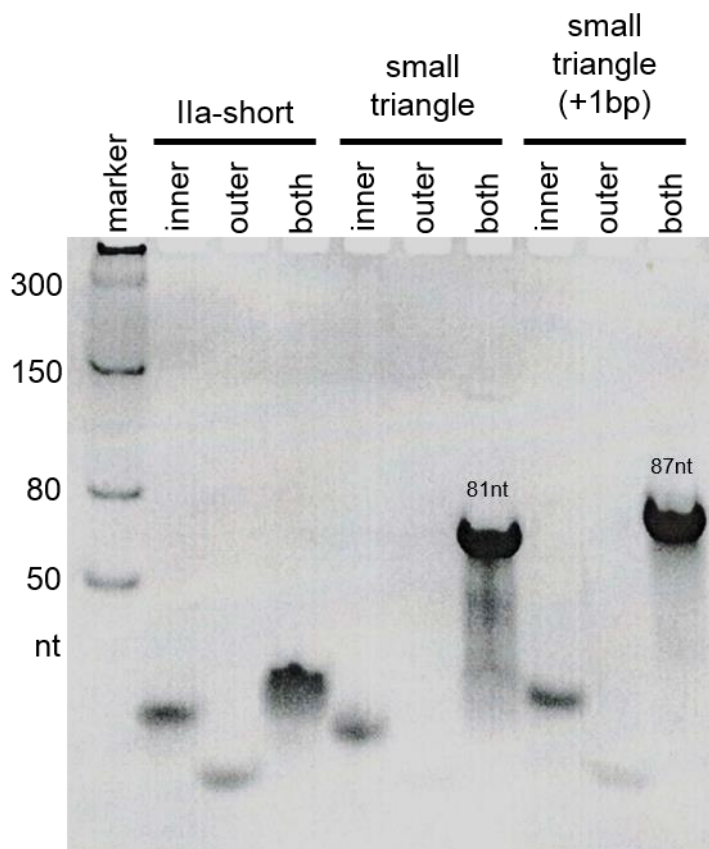


Figure 5.21: Native PAGE gel in 5mM MgCl₂ of self-assembling small triangle and small triangle (+1bp) constructs. Triangle sizes are indicated on the gel. The Ila-short construct is shown for a size reference.

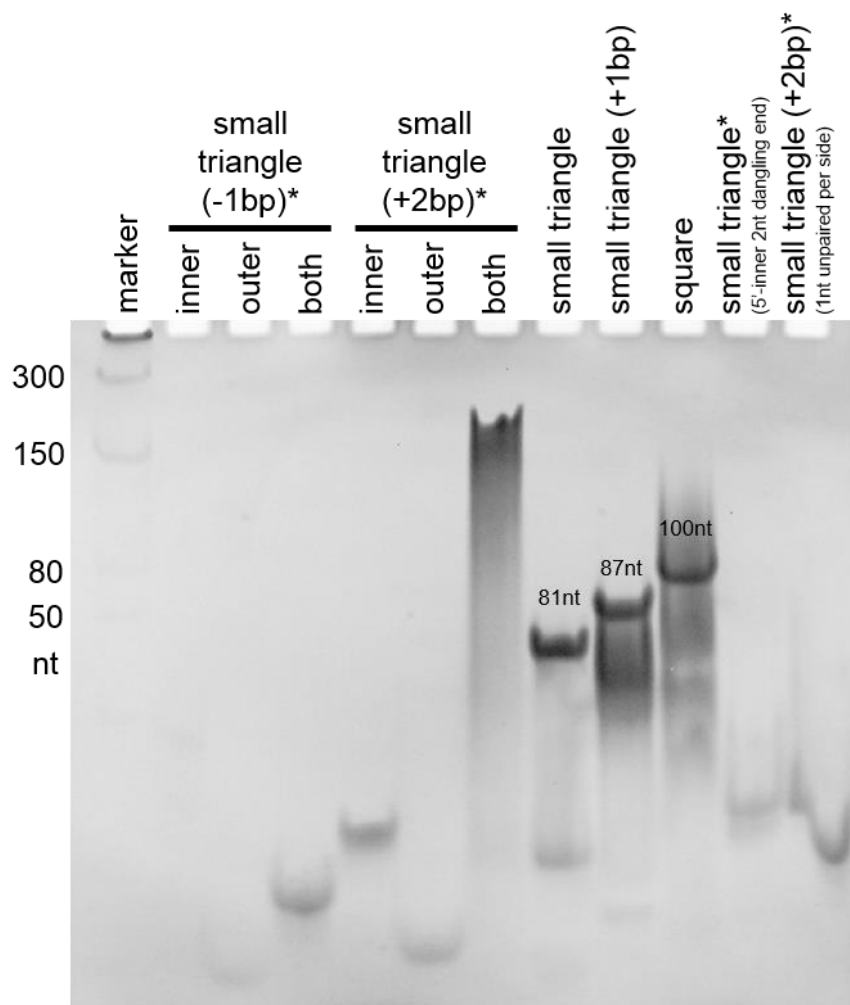


Figure 5.22: Native PAGE gel in 5mM MgCl₂ of self-assembling small RNA triangle constructs of various sizes described left to right. The small triangle (-1bp) does not self-assemble into a 75nt triangular structure as designed (Figure 5.20c), likely due to steric restraints in forming a triangle from three 90° corner units. The small triangle (+2bp) does not self-assemble into a 93nt triangular structure as designed (Figure 5.20b), but likely forms multiple structures of various sizes as its band appears as a large smear. The small triangle of 81nt (Figure 5.2c, top), small triangle (+1bp) of 87nt (Figure 5.20a), and square¹⁰⁴ (Figure 5.38) of 100nt all assemble as expected and migrate at speeds consistent with their size. Two additional dangling nucleotides placed at the 5' end of the inner strand of the small triangle prevent its assembly (Figure 5.20d). removing a single 5' inner strand nucleotide prevents leaving one unpaired outer nucleotide prevents the formation the small triangle (+2bp) large smeared band (Figure 5.20e). *Does not form as shown in designed secondary structure schematics (Figure 5.20).

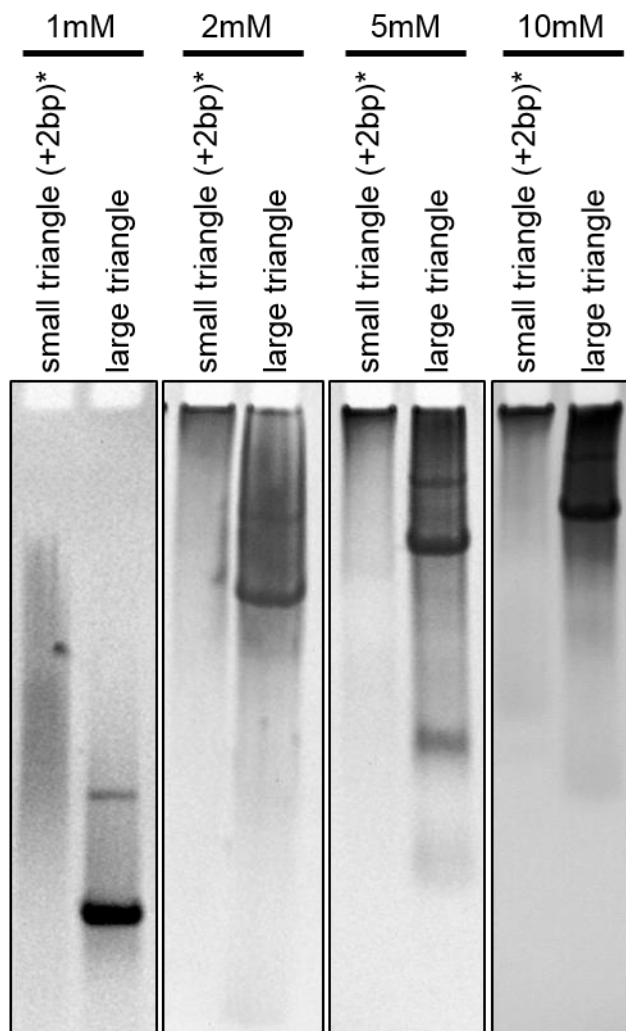


Figure 5.23: Native PAGE gels in 1, 2, 5 or 10mM MgCl_2 (left to right) run for various lengths of time for the small triangle (+2bp) construct (Figure 5.20b). The symmetrical large triangle is shown as a size reference. Mg^{2+} concentration does not affect the stability of any single band species. This small triangle (+2bp) runs as a smeared band slower than the large triangle. *Does not form as shown in designed secondary structure schematic (Figure 5.20b).

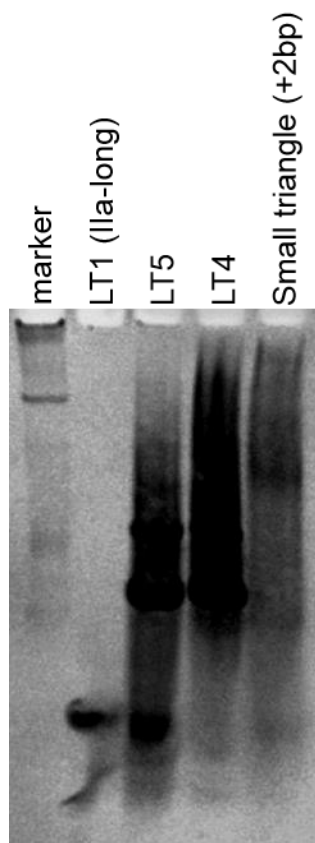


Figure 5.24: Native PAGE gel in 10mM MgCl₂ for the small triangle (+2bp) construct. Gel image contrast is increased compared to Figure 5.27 to show the darker band regions of this construct at the expense of clearly seeing the large triangle constructs LT4, LT5, and IIa-long for a size comparison. Amidst the smearing, a slower migrating band is seen which corresponds to about twice the size of a dimer of large triangles [(138×2)×2].

Two different large triangle constructs were explored (Figure 5.25). LT4 has 4 nucleotide overhangs and has been discussed in great detail. LT5, the 5 nucleotide overhang variant forms very similar bands to the LT4, including a major faster migrating triangle band and a slower migrating dimer band (Figures 5.26-5.27). One additional minor faster migrating band is seen for LT5 and corresponds to the size of a single corner unit. LT4 was chosen as the exemplary construct over LT5 as it gave cleaner major triangle and minor dimer bands without any minor faster migrating bands as seen for LT5. Additionally, the four nucleotide overhang of LT4 made this a more comparable construct to the ST4-2 small triangle construct.

Figure 5.25: Self-assembling large RNA triangle constructs of identical size but varying oligonucleotide termini and number of overhanging nucleotides. a) Construct LT1 (large triangle, one-nucleotide overhang) is the same as the extended SVV IIa crystal construct (Chapter 2, Figure 2.12; Figure 5.2a, b, bottom) and is shown here for comparison even though it does not self-assemble in solution. b) Construct LT4 (large triangle, four-nucleotide overhang) which is the exemplary construct used for the majority of studies. c) Construct LT5 (large triangle, five-nucleotide overhang).

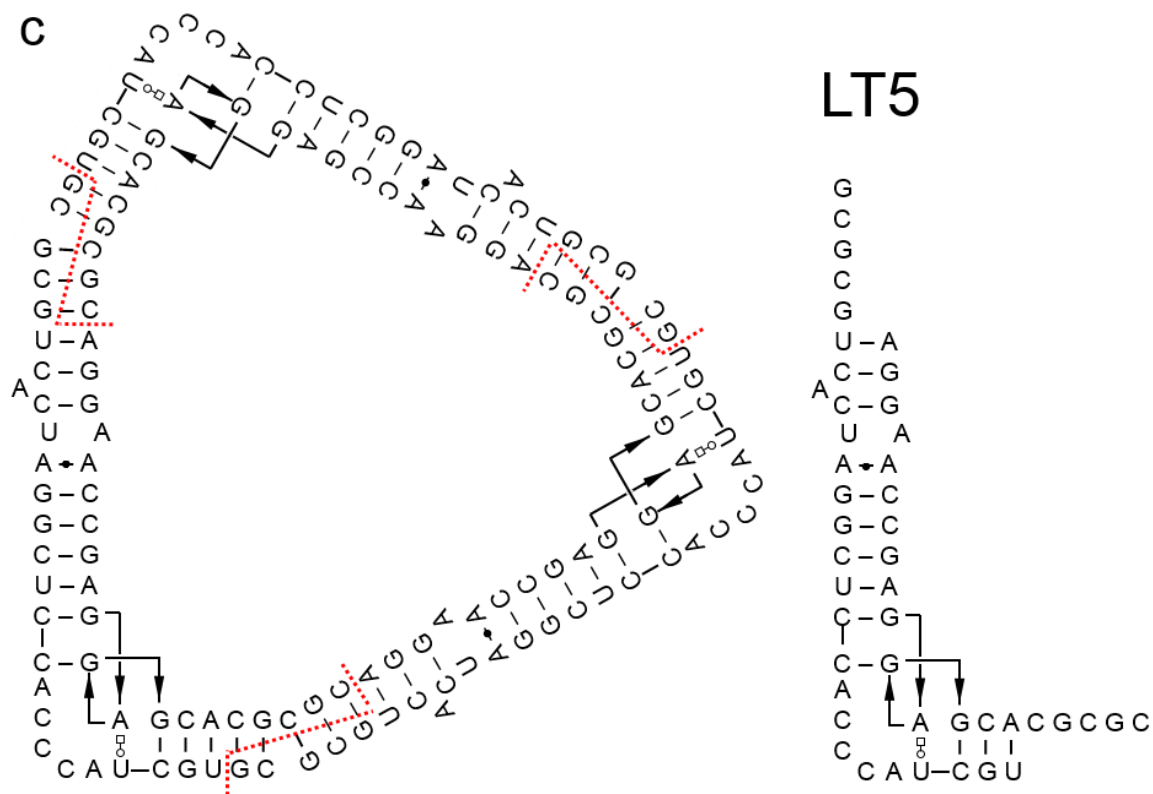


Figure 5.25: Self-assembling large RNA triangle constructs of identical size but varying oligonucleotide termini and number of overhanging nucleotides. *Continued.*

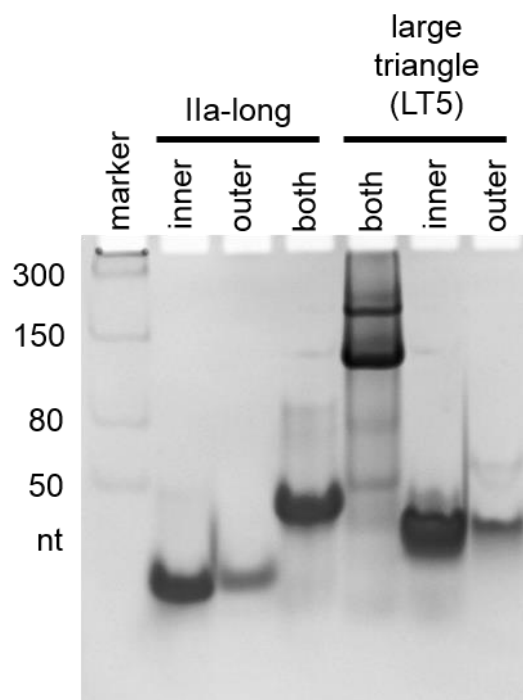


Figure 5.26: Native PAGE gel in 5mM MgCl₂ of self-assembling large RNA triangle construct LT5 (5 residue overlapping sequence) with lanes showing inner or outer ssRNA or both mixed to assemble nanostructure. The Ila-short construct is shown for a size reference. Lanes indicate inner or outer ssRNA or both ssRNA mixed to assemble nanostructure. This large triangle LT5 construct forms very similar bands to the LT4 construct studied in greater detail. A major faster migrating triangle band is seen with a slower migrating dimer band.

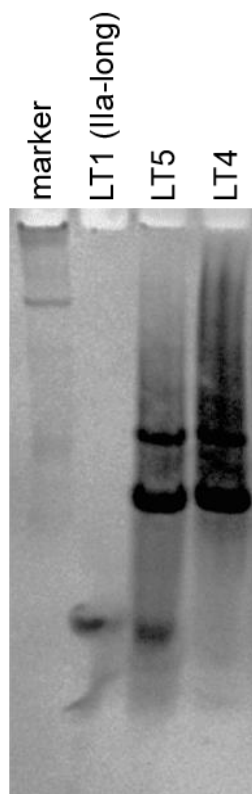


Figure 5.27: Native PAGE gel in 10mM MgCl₂ of self-assembling large RNA triangle constructs LT4 (4 residue overlapping sequence) and LT5 (5 residue overlapping sequence). The Ila-short construct is shown for a size reference. A major faster migrating triangle band and a slower migrating dimer band are seen for both LT5 and LT4 constructs. LT5 has one additional faster migrating band which corresponds to the size of a single corner unit.

Discussion of Programmable Large Triangle

The Programmable Large Triangle (PLT) assembly was attempted in a variety of Mg²⁺ concentrations and for a variety of assembly times at 25° and 37° (Figures 5.28-5.32). Assembly results were analyzed by native PAGE and the Mg²⁺ concentration for each assembly was mirrored in the gel, loading buffer, and running buffer. First, the PLT was assembled stepwise in 5mM MgCl₂ for 10 minutes (Figure 5.28), and its band migration matched that of the symmetrical large triangle (SLT). Due to band smearing in

the gel, the Mg^{2+} concentration was lowered to 2mM, and the assembly time was increased to 20 minutes (Figure 5.29). These conditions gave very clear results for an identically sized PLT and SLT. Mg^{2+} concentration was increased to 10mM and assembly time increased to 30 minutes to see the effect of these changes on the assembly of PLT and its intermediates (Figure 5.29). PLT assembly results appear similar, though dual corner intermediates AB, BC, and AC bands were better defined. An increase to 20mM $MgCl_2$ gave similar results to 10mM but with increased smearing observed (Figure 5.31). As 2mM $MgCl_2$ gave the cleanest results, an assembly time of 30 minutes (versus 20 minutes) was performed at 2mM $MgCl_2$, and gave similarly clean results (Figure 5.32). Overall, these experiments gave very clear results for an identically sized PLT and SLT. The best assembly conditions determined for this construct were in 2mM $MgCl_2$ for 20-30minutes at 25-37°.

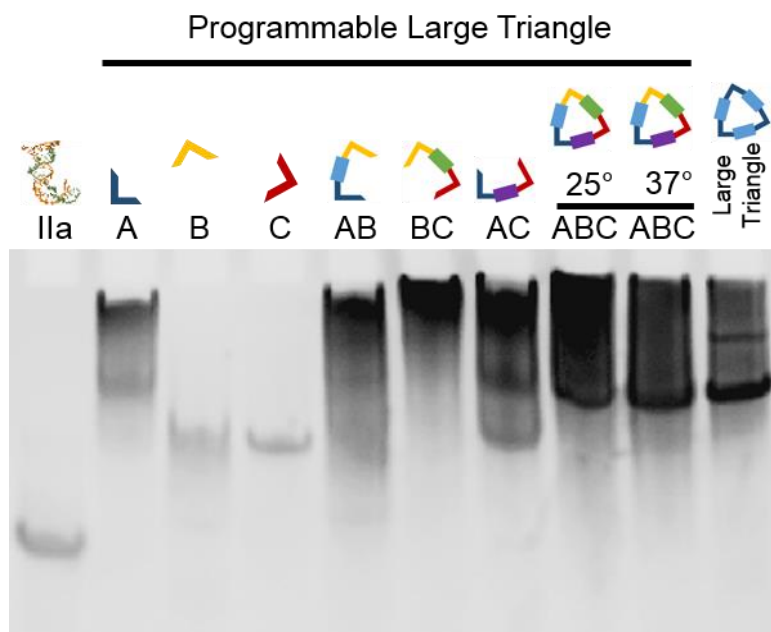


Figure 5.28: Programmable large triangle (PLT) assembled in **5mM MgCl₂** for **10min** at 25° or 37°. PLT band migration matches that of the symmetrical large triangle (SLT), and its assembly appears more efficient at 37° in these conditions. The faint slower migrating band present in the SLT lane appears to be absent from the PLT lanes. Individual A, B, and C corners (4 nucleotide overhangs of varying sequence) migrate differently from each other and differently from the Ila corner (1 nucleotide overhang). Dual intermediates AB, BC, and AC migrate differently from A, B, and C, but do not migrate consistent with the size 2/3 of the programmable triangle. Individual and dual intermediates may not form simple and stable structures as indicated by cartoons above each well; the stability gained upon formation of the entire triangle may exclude alternate conformations and intermediates.

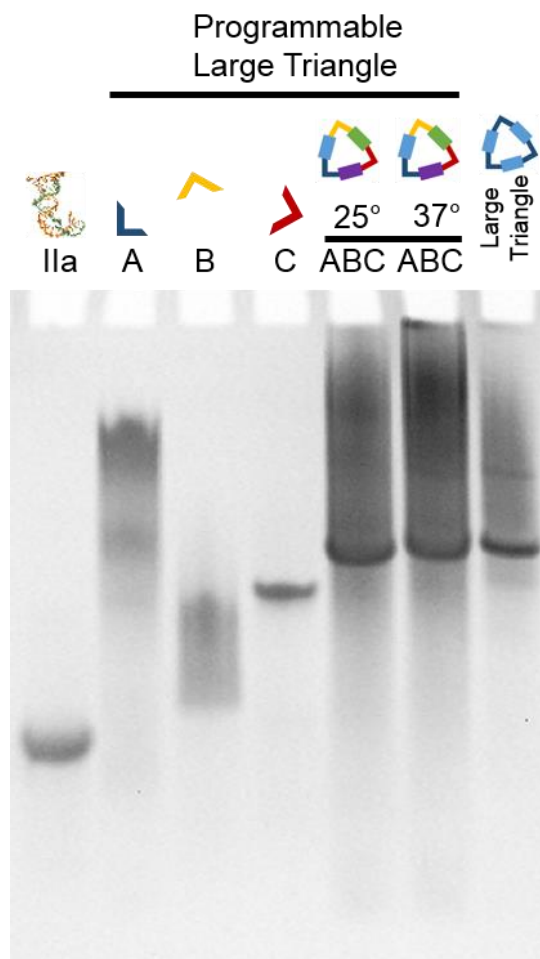


Figure 5.29: PLT assembled in **2mM MgCl₂** for **20min** at 25° or 37° (lower Mg²⁺ concentration and longer assembly time due to smearing seen in Figure 5.28), enlarged from Figure 5.3d for comparison here. PLT assembly appears efficient at both temperatures in these conditions. The faint slower migrating band present in the SLT lane may also be present in the 37° PLT lane. Individual A, B, C, and IIa corners are included for reference.

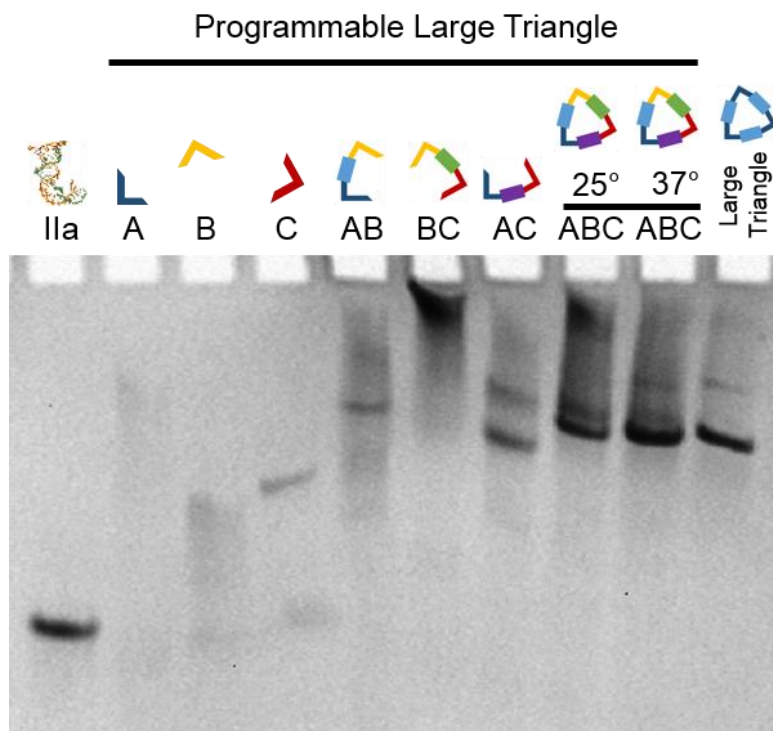


Figure 5.30: PLT assembled in **10mM MgCl₂** for **30min** at 25° or 37° (higher Mg²⁺ concentration and longer assembly time than Figures 5.28-5.29). PLT assembly appears slightly more efficient at 37° in these conditions (higher Mg²⁺, longer assembly time). Dual intermediates AB, BC, and AC bands are better defined (less smearing) at this high Mg²⁺ concentration, and are partially present in the 25° PLT lane, indicating incomplete assembly at this temperature in these conditions. The faint slower migrating band present in the SLT lane appears to be present in the 37° PLT lane. Individual A, B, C, and Ila corners and dual intermediates AB, BC, and AC are included for reference.

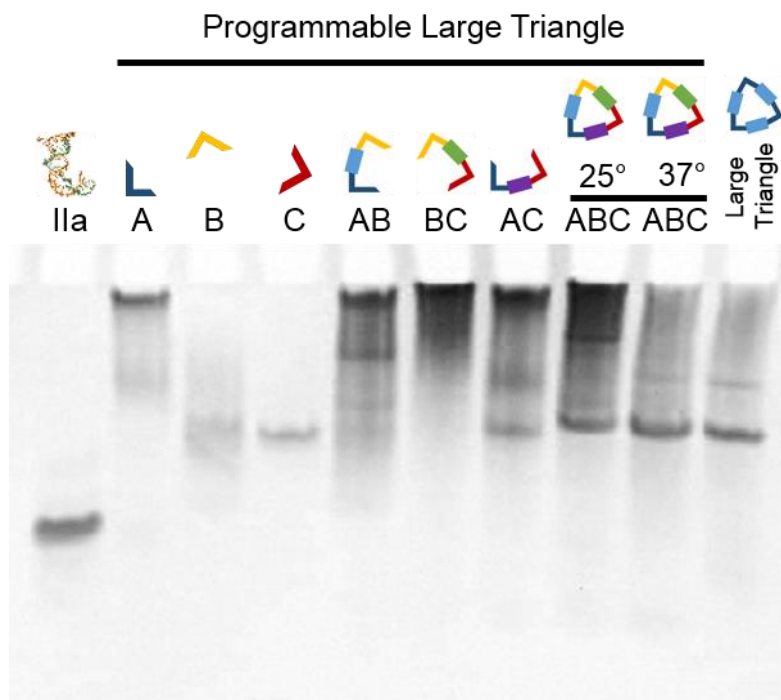


Figure 5.31: PLT assembled in **20mM MgCl₂** for **30min** at 25° or 37° (higher Mg²⁺ concentration than Figures 5.28-5.30). Results are similar to the experiment in 10mM MgCl₂ (Figure 5.30), with increased smearing seen here at higher Mg²⁺ concentration.

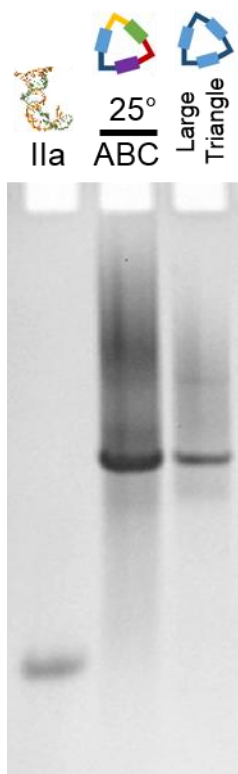


Figure 5.32: PLT assembled in **2mM MgCl₂** for **30min** at **25°** (same Mg²⁺ concentration but longer assembly time than Figure 5.28). PLT assembly results are similar to a shorter assemble time (Figure 5.28).

Discussion of Triangle Stability

The small triangle's resistance to endonuclease degradation was tested (Figures 5.33-5.35). The small triangle showed greater exposure to RNase A degradation than tRNA (Figure 5.33). The small triangle does not show any greater resistance to RNase degradation than the individual Ila corner unit (Figures 5.34-5.35)

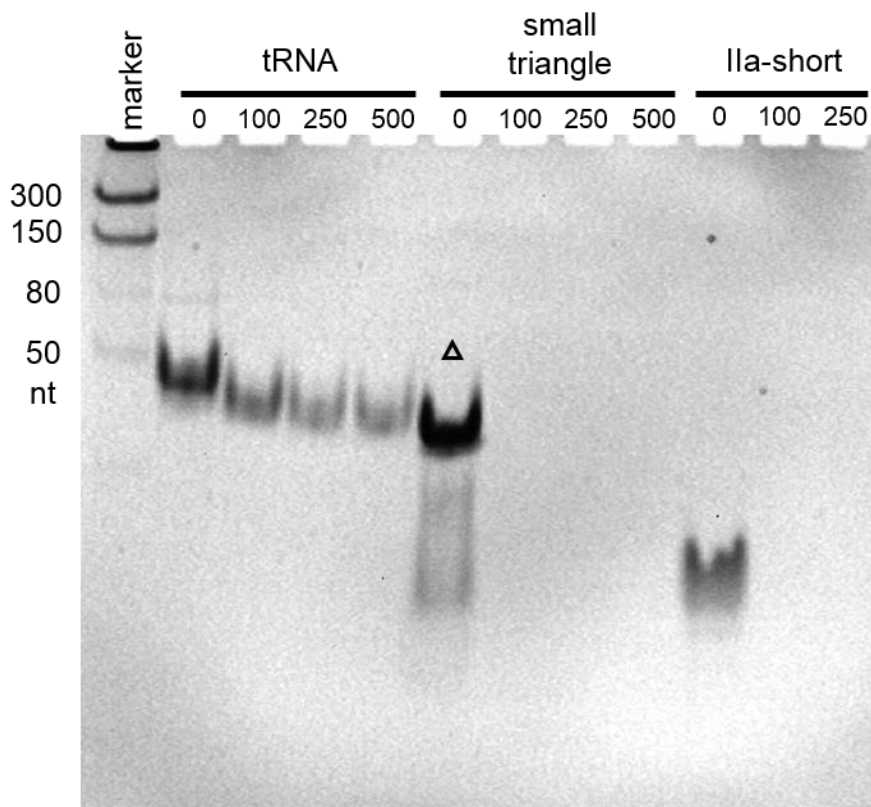


Figure 5.33: Comparison of the small triangle's resistance to RNase A (endoribonuclease) degradation compared to tRNA and the Ila-short construct. Concentrations shown are in $\mu\text{g/mL}$. Degradation is seen for all three RNAs. tRNA, a longer ssRNA which adopts complex folds, shows degradation products which appear as faster migrating bands. The small triangle and Ila-short constructs do not show degradation products as they are composed of short ssRNA building blocks.

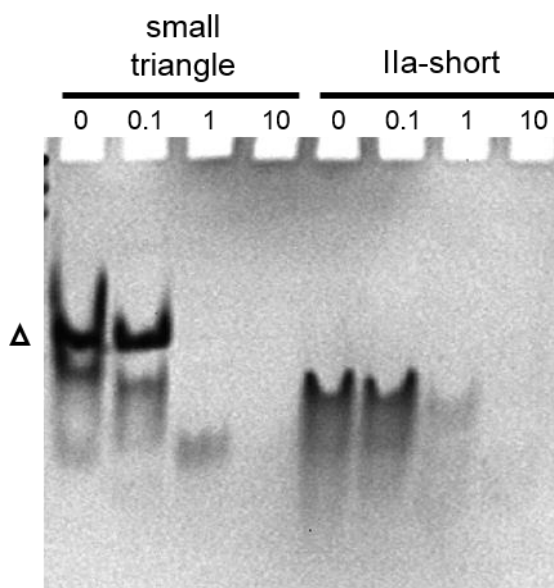


Figure 5.34: Small triangle and Ila-short construct stability when treated with RNase A (endoribonuclease). Concentrations shown are in $\mu\text{g/mL}$. The small triangle does not appear to have added resistance to RNase degradation versus single corner dsRNA (Ila-short).

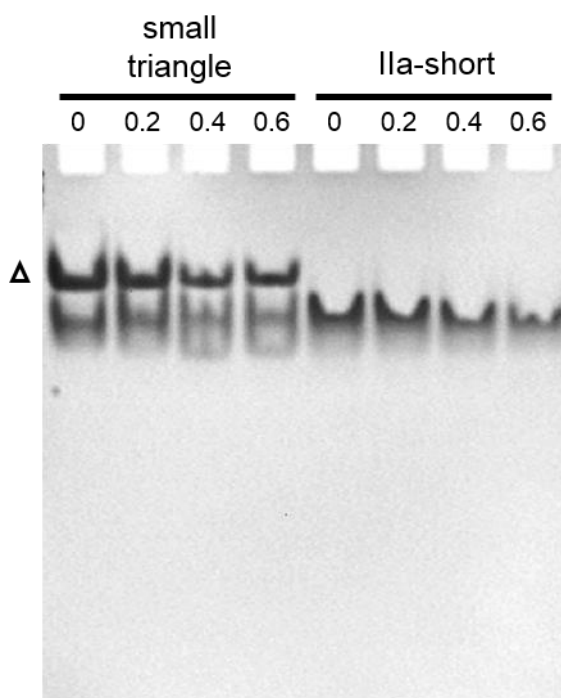
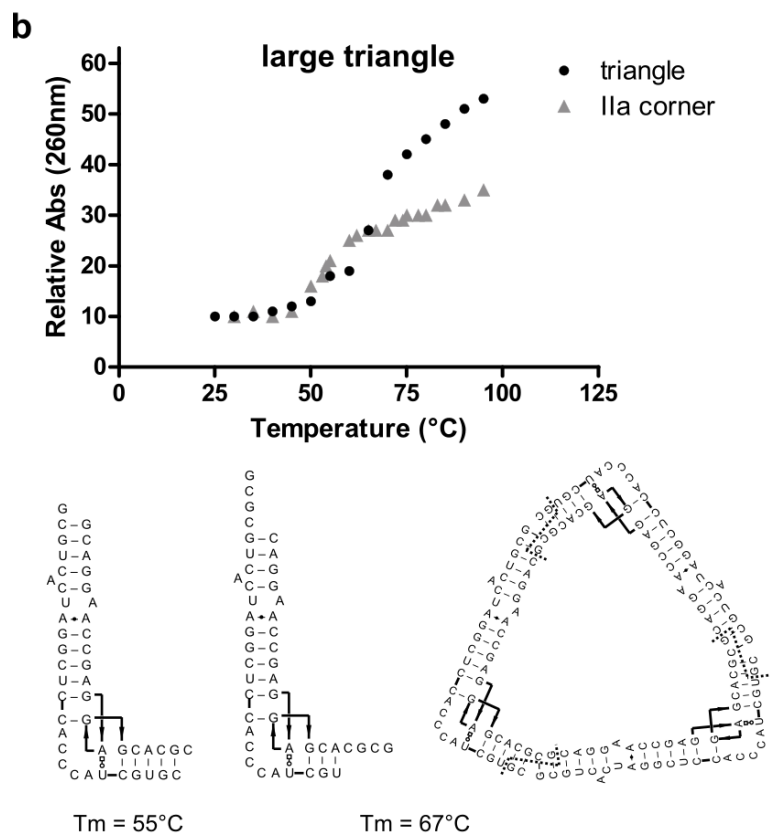
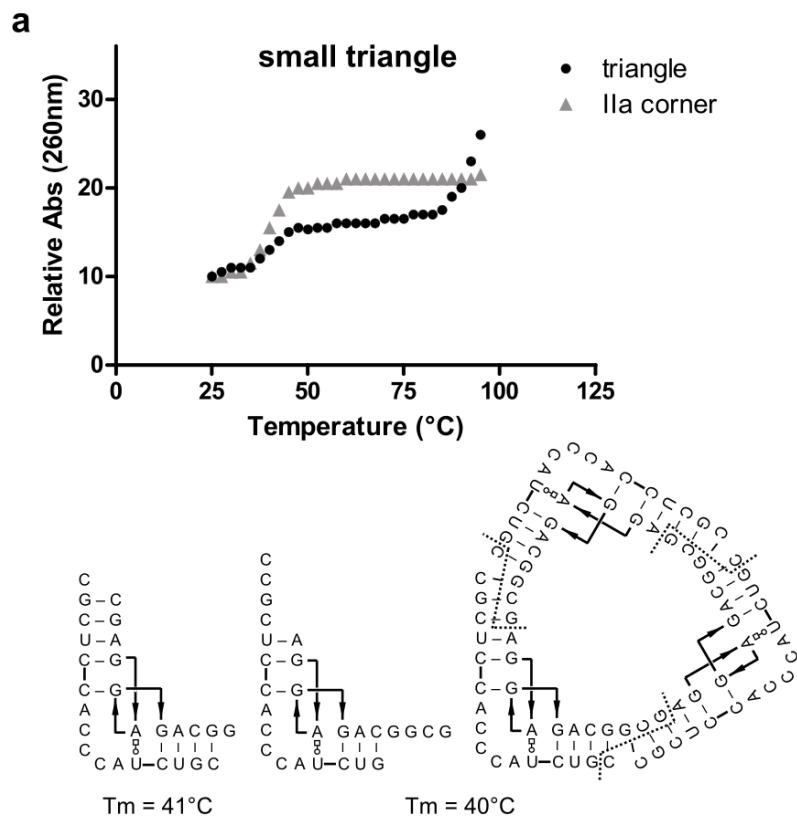


Figure 5.35: Small triangle and Ila-short construct stability when treated with RNase A (endoribonuclease). Concentrations shown are in $\mu\text{g/mL}$. The small triangle does not appear to have added resistance to RNase degradation versus single corner dsRNA (Ila-short).

Small and large triangle thermostability was measured by UV absorption during thermal melting experiments (Figure 5.36). The small triangle was found to have a T_m of 40°C , which is similar to the Ila-short construct ($T_m=41^\circ\text{C}$) and higher than the predicted T_m of 31°C for an isolated self-assembly corner unit. The large triangle was found to have a T_m of 67°C , which is higher than the Ila-long construct ($T_m=55^\circ\text{C}$) and higher than the predicted T_m of 54°C for an isolated self-assembly corner unit.

Figure 5.36: UV thermal melting experiments of small and large triangles as well as subdomain IIa motifs representing single corners of triangles as seen in crystal structure packing. Melting temperatures (T_m) values shown are calculated from experiments shown in graphs. Predicted melting temperatures in figure caption are calculated using mfold.¹³⁶ a) UV thermal melting experiment of the small nanotriangle with T_m transitions at 40°C and 95°C and the short subdomain IIa motif (IIa corner) with a T_m of 41°. The predicted T_m of a single self-assembling corner unit not in the context of a full triangle (center of three secondary structures) is 31°. (b) UV thermal melting experiment of large nano triangle with a T_m of 67°C and long subdomain IIa motif (IIa corner) with a T_m of 55°. The predicted T_m of a single self-assembling corner unit not in the context of a full triangle (center of three secondary structures) is 54°.



Discussion of Ligand-Responsive Nanostructures

Large triangle assembly is not affected by guanine at concentrations up to 2mM (Figure 5.37). Guanine does bind the SVV Ila switch element and captures an elongated conformation with an EC_{50} value of $\sim 1\text{mM}$ in the context of an *in vitro* FRET assay.³⁹ Ila switch elements are captured in a bent conformation in the context of a triangle and there is less elongated switch conformations in equilibrium to be captured by guanine. Guanosine monophosphate at 10mM and 2-aminoquinazolin-4(3H)-one at 2mM similarly showed no effect on large triangle assembly (data not shown).

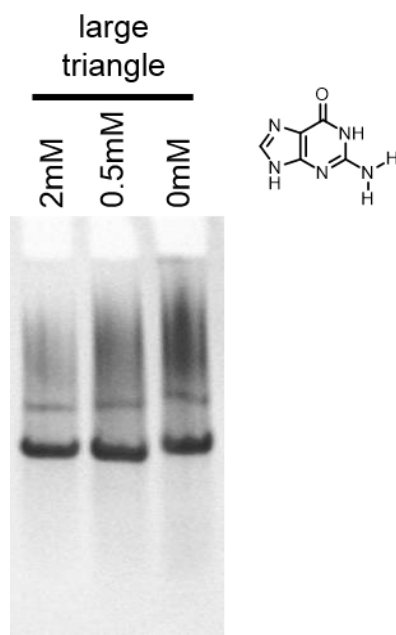


Figure 5.37: Native PAGE analysis in 5mM MgCl_2 of SLT assembled in the presence of 0, 0.5, or 2 mM guanine. At these concentrations, guanine has no effect on SLT assembly.

The ligand has a similar effect in both pre and post assembly methods on the self-assembling RNA square previously constructed by Hermann and coworkers (Figure 5.38).¹⁰⁴ In both methods, there is a reduction of the single square band, and the

appearance of multiple slower migrating bands with increased smearing in this region (Figure 5.38a). This is very similar to what is seen for the large triangle (Figure 5.5). Square assembly is partially prevented by the ligand, and incubation with ligand partially dissociates the square. In both instances the ligand captures square single corner units in an elongated conformation, preventing square formation and promoting end-to-end multimerization into longer species. The square also partially survives boiling water temperatures but is partially dissociated as seen by a decrease in the size and intensity of the square band (Figure 5.38b). A second, faster migrating band appears after boiling and corresponds to the size of a single corner.

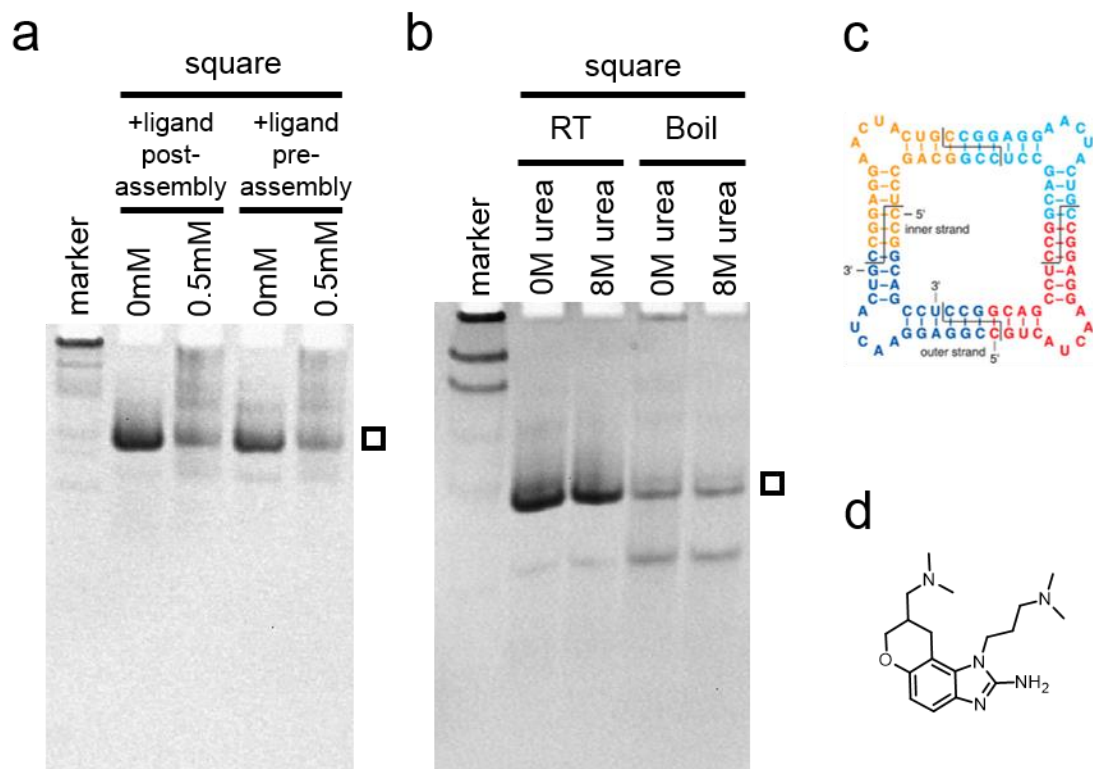


Figure 5.38: Native PAGE analysis in 5mM MgCl₂ of RNA square.¹⁰⁴ a) Square assembled in the presence of 0 or 0.5 mM ligand. In lanes labelled “pre,” ligand was added to the reaction before the assembly protocol. In lanes labelled “post,” ligand was added to the reaction after square assembly, and incubated at 37° for 15 minutes before loading on the gel. b) Stability of square when treated with or without 8M urea at room temperature (RT) or in boiling water. c) Self-assembling RNA square secondary structure as shown in Reference¹⁰⁴. d) Structure of the benzimidazole ligand which binds to the IIa RNA switch element of the square’s corners.

As discussed previously in Figures 5.3d and 5.7 and also shown here in Figure 5.39, the programmable large triangle is responsive to ligand binding. Individual corners A, B, and C were also assembled in the presence of 500μM ligand and analyzed by PAGE. No difference is seen for individual corner bands in the presence or absence of ligand. The individual corner units are captured in an elongated conformation by the

ligand versus a bent conformation captured by Mg^{2+} , but these PAGE conditions are not able to resolve a band migration difference between these two conformations.

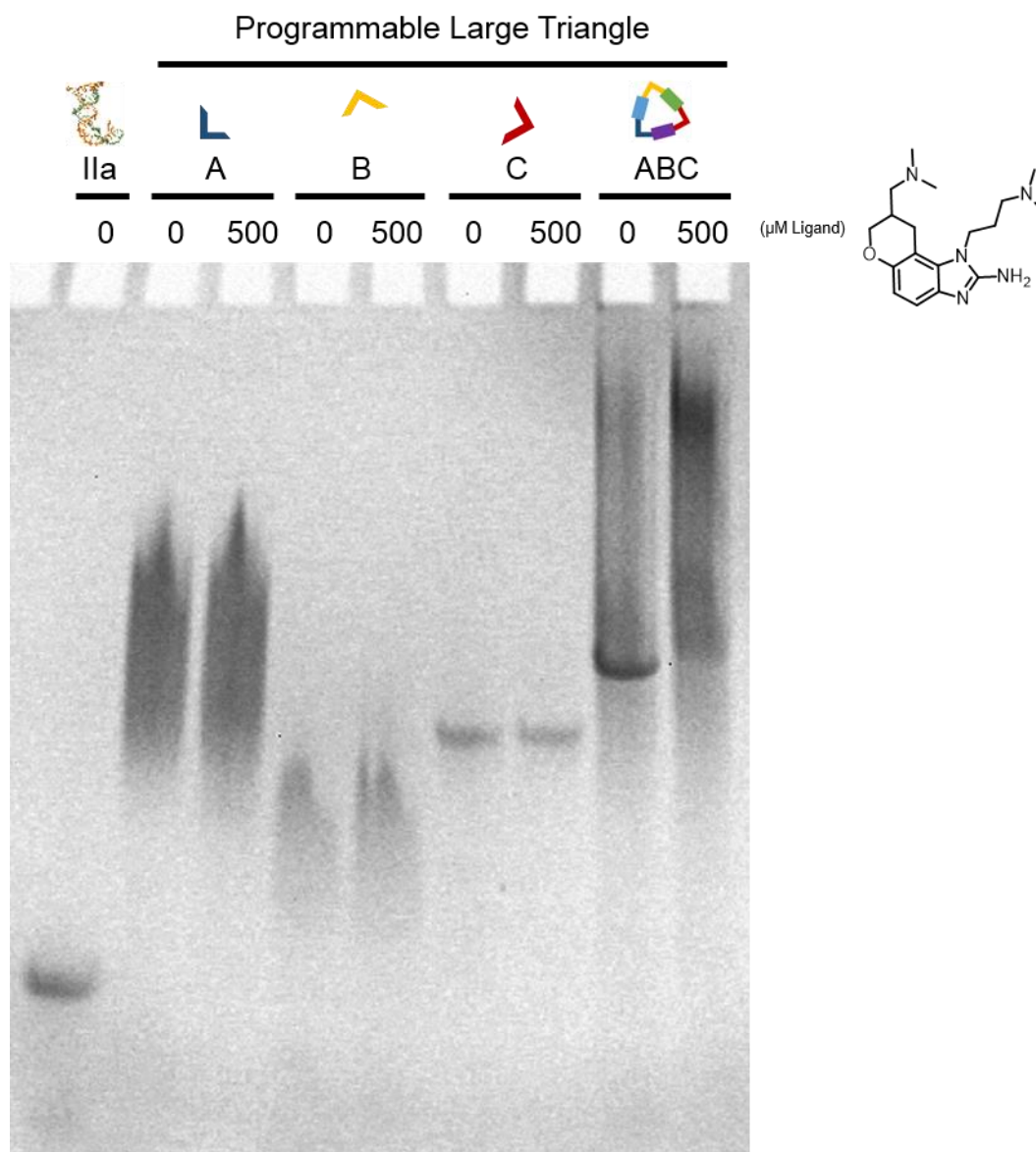


Figure 5.39: PLT assembled in the presence of 0 or 500 μM ligand in ideal assembly conditions of **2mM $MgCl_2$** for **20min** at 25° . Individual corner bands A, B, and C appear identical with or without the ligand present, but PLT assembly is inhibited by the presence of 500 μM ligand.

As discussed previously in Figure 5.5 and also shown here in Figure 5.40, the small triangle is not responsive to ligand binding. In 0.5mM $MgCl_2$ the small triangle

does not assemble efficiently (0mM ligand), and increased smearing is seen with increased ligand concentration. The small triangle assembles well at 2 or 5mM MgCl_2 (0mM ligand), but does not appear affected by increased ligand concentrations.

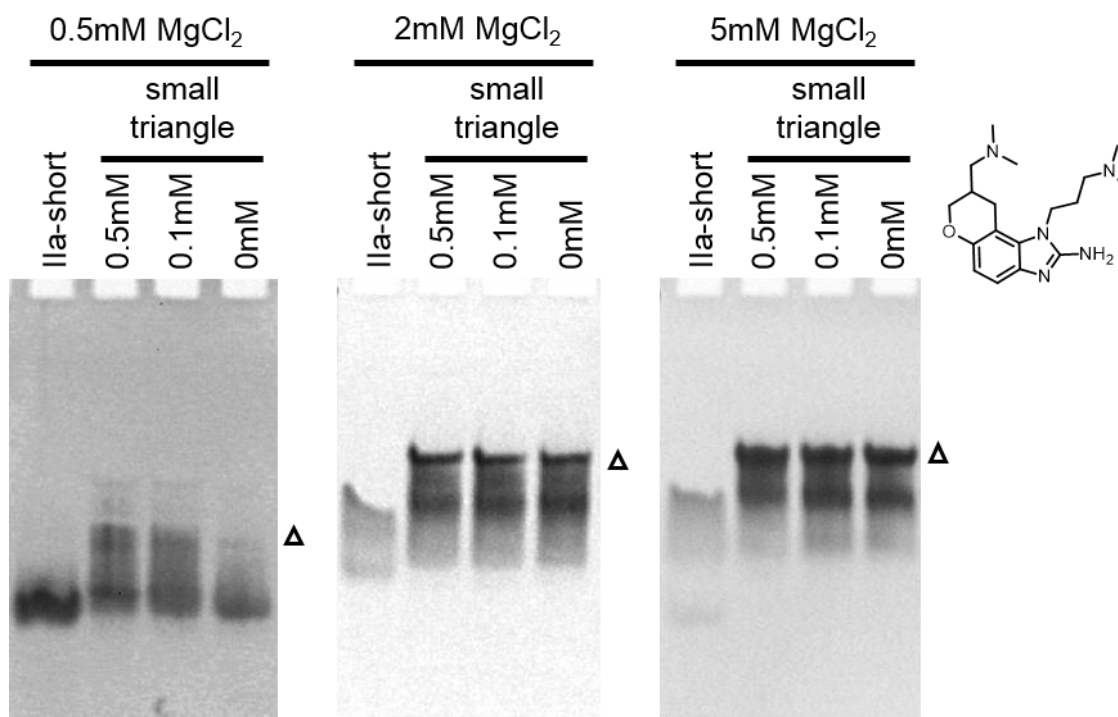


Figure 5.40: Native PAGE analysis in 0.5, 2, or 5 mM MgCl_2 (left to right) of the small triangle assembly in the presence of 0, 0.1, or 0.5 mM ligand. The triangle to the right of each gel indicates the small triangle band.

The small triangle (+1bp) that forms an 87 nt triangle (Figures 5.20a, 5.21, 5.22), was also tested for its ability to respond to ligands. This construct has one less base pair per triangle side, and is assumed to form a triangle which is slightly less distorted than the 81nt small triangle. Similar to this 81nt small triangle, the ligand has negligible effect on the small triangle (+1bp) of 87nt (Figure 5.41).

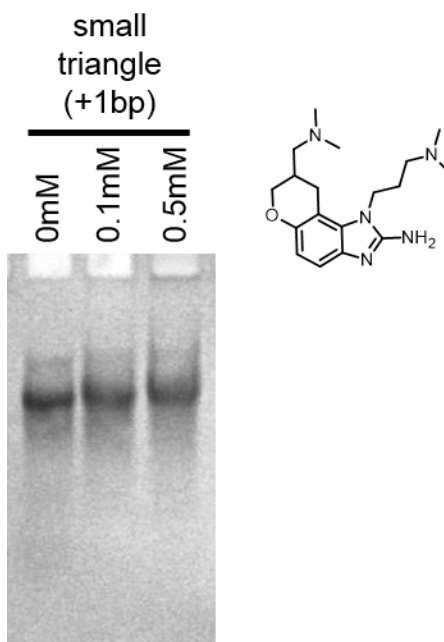


Figure 5.41: Native PAGE in 2mM Mg of the small triangle (+1bp) (Figure 5.20a) assembled in the presence of 0, 0.1, or 0.5 mM ligand. The ligand appears to have negligible effect on this small triangle (+1bp) of 87nt.

Discussion of RNA Nanostructure Crystallography

Attempts were made to crystallize numerous self-assembling RNA triangle constructs, both large and small. Most constructs yielded no highly diffracting crystals (Table 5.2). ST1 and LT1 crystal structures are discussed in Chapter 2 and the *Discussion of Crystal Structures used in Design* section found previously in this chapter. The ST4-2 crystal structure is discussed in the *Crystal Structure of Small RNA Nanotriangle* section earlier in this chapter. ST4-2 small triangle crystals used to collect the 2.6Å dataset can be seen in Figure 5.42a.

Table 5.2: RNA triangle crystal constructs

Crystal Construct	Structure (Figure)	Description	Results
ST1	Fig. 5.17	Small SVV IIa construct which packs into pseudo-continuous small triangles in its crystal structure with 1nt overhangs	1.86Å dataset PDB ID: 4P97
ST3-1	Fig. 5.17	Small triangle of 81nt with 3nt overhangs.	No diffracting crystals
ST3-2	Fig. 5.17	Small triangle of 81nt with 3nt overhangs.	No diffracting crystals
ST3-3	Fig. 5.17	Small triangle of 81nt with 3nt overhangs.	No diffracting crystals
ST4-1	Fig. 5.17	Small triangle of 81nt with 4nt overhangs.	No diffracting crystals
ST4-2	Fig. 5.17	Small triangle of 81nt with 4nt overhangs.	2.6Å dataset PDB ID: 5CNR
Small triangle (+1bp)	Fig. 5.20	Small triangle of 87nt with 4nt overhangs (1bp add to each triangle side of ST4-2)	No diffracting crystals
Small triangle (+2bp)	Fig. 5.20	Small triangle of 93nt with 4nt overhangs (2bp add to each triangle side of ST4-2)	No diffracting crystals
Small triangle (-1bp)	Fig. 5.20	Small triangle of 93nt with 4nt overhangs (2bp add to each triangle side of ST4-2)	No diffracting crystals
LT1	Fig. 5.20	Long SVV IIa construct which packs into pseudo-continuous large triangles in its crystal structure.	3.2Å dataset PDB ID: 4PHY
LT4	Fig. 5.20	Large triangle with 4nt overhangs.	No diffracting crystals
LT5	Fig. 5.20	Large triangle with 5nt overhangs.	No diffracting crystals
Programmable large triangle	Fig. 5.3c	Programmable large triangle with 4nt overhangs of varying sequence.	No diffracting crystals

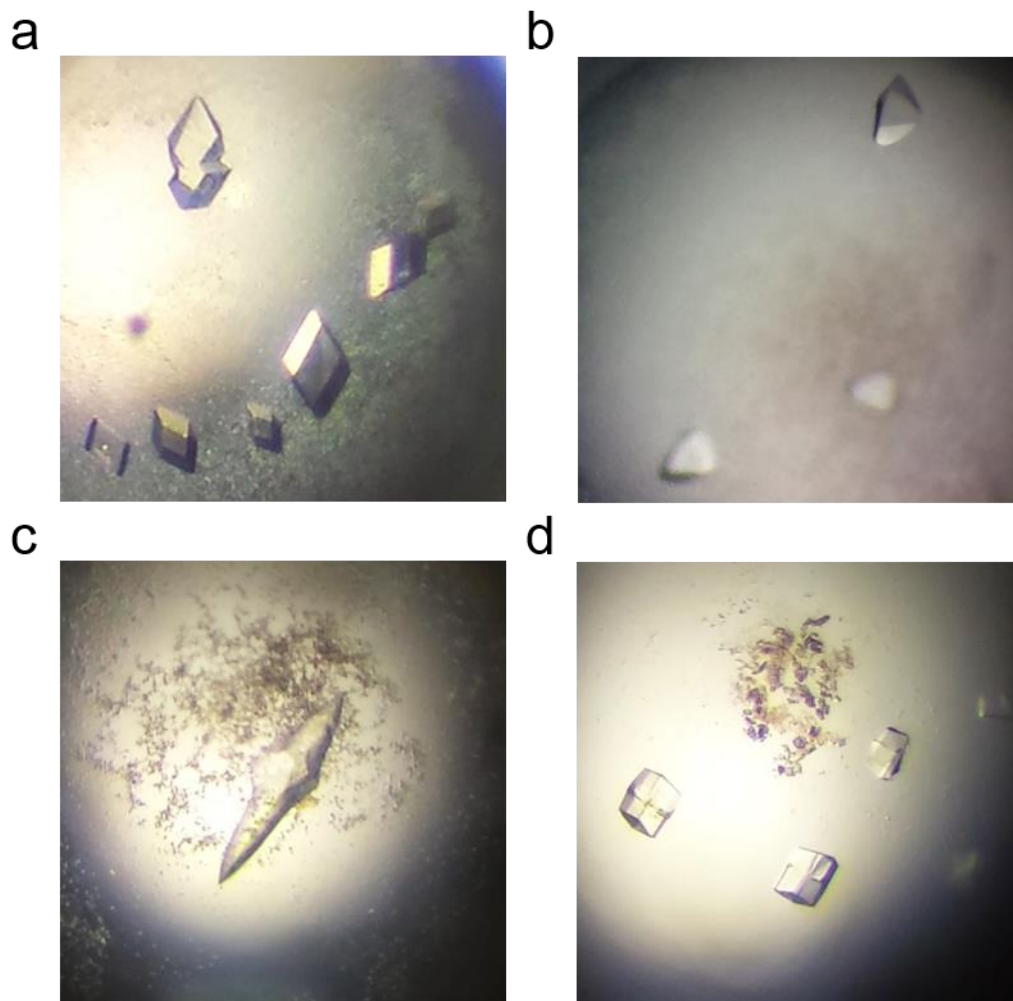


Figure 5.42: Photos of nanostructure crystals. a) ST4-2 construct crystals used to collect the 2.6Å dataset used to solve the small triangle crystal structure. b) D-ST5-1 construct crystals used to collect the 2.7Å dataset which was not successfully phased and solved. c) D-ST5-1-Br1 construct crystals used to collect a 3.3Å dataset which did not index. d) H1-ST5-1 construct crystals used to collect a 2.3Å dataset which contained twinning defects and was not successfully phased and solved.

Attempts were also made to crystallize numerous self-assembling DNA or DNA/RNA hybrid triangle constructs, both large and small. Most constructs yielded no highly diffracting crystals (Table 5.3). The all DNA version of ST5-1 (D-ST5-1; DNA thymidine residues substituted for RNA uridine residues) yielded nice crystals diffracting

to 2.7 Å resolution. No solution was determined by molecular replacement methods. In order to obtain initial phase estimates of D-ST5-1, a derivative construct was designed to contain a brominated U residue replacing the T residue at position 377 of the outer strand (SVV numbering, Figure 5.2). Crystals of D-ST5-1-Br1 did not diffract below 3.3Å resolution, and a 3.3Å dataset did not index into a spacegroup. A hybrid DNA/RNA version of ST5-1 (H1-ST5-1; RNA outer strand, DNA inner strand) yielded nice crystals diffracting to 2.3Å. No solution was determined by molecular replacement methods. In order to obtain initial phase estimates of H1-ST5-1, a derivative construct was designed to contain a brominated U residue replacing the C residue at position 420 of the inner strand (SVV numbering, Figure 5.2). A C residue was chosen for replacement as the inner strand contains no T residues, and a DNA inner strand is more cost effective to modify than an RNA outer strand. The bromine-modified hybrid construct as well as a second site of bromine modification in the DNA construct (H1-ST5-1-Br1, D-ST5-1-Br2) yielded no diffracting crystals for either construct.

Table 5.3: DNA and DNA/RNA hybrid triangle crystal constructs.

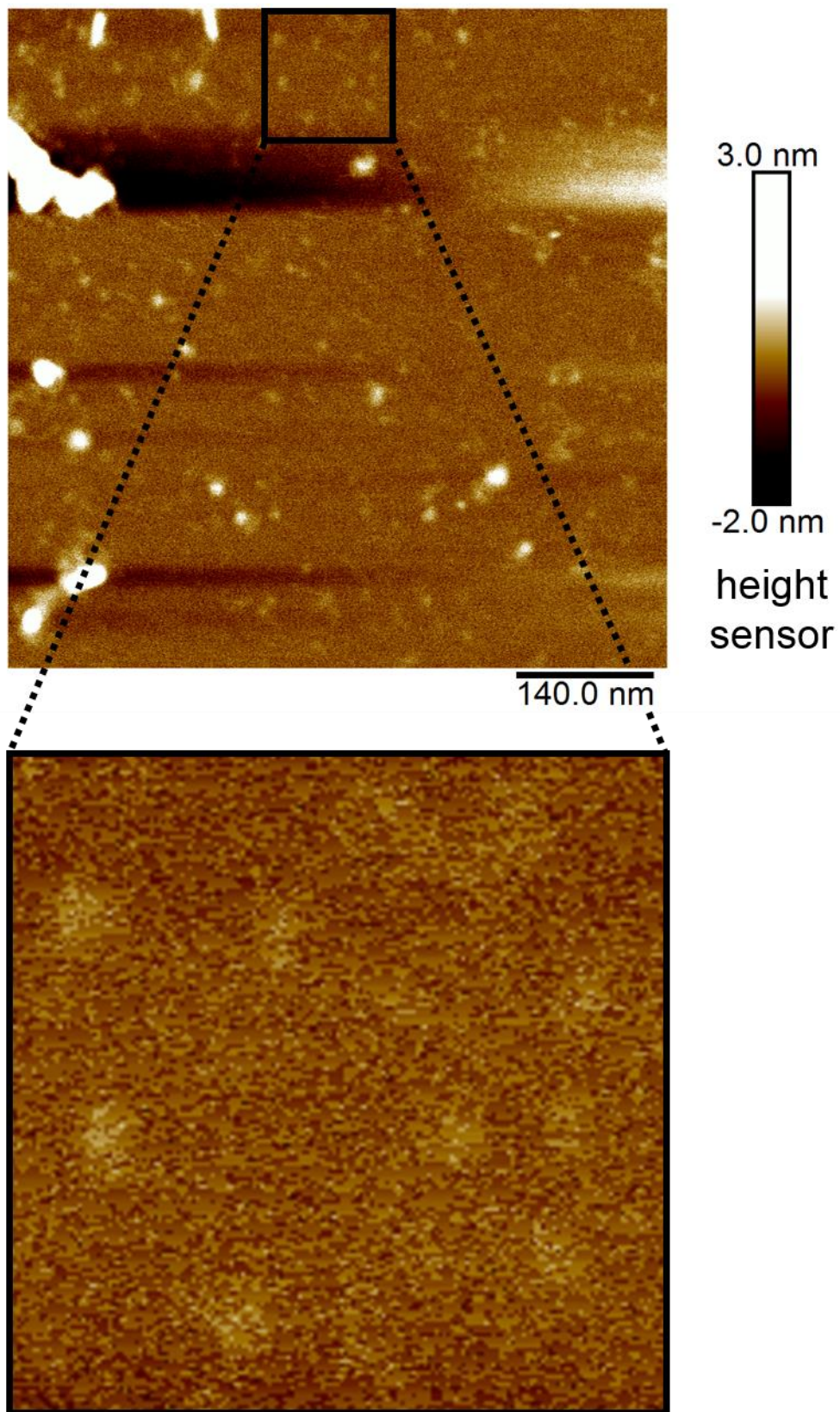
Crystal Construct	Description	Results
D-ST1	DNA small SVV IIa construct	No diffracting crystals
D-ST2	DNA small triangle of 81nt with 2nt overhangs.	No diffracting crystals
D-ST3-2	DNA small triangle of 81nt with 3nt overhangs.	No diffracting crystals
D-ST4-2	DNA small triangle of 81nt with 4nt overhangs.	No diffracting crystals
D-ST5-1	DNA small triangle of 81nt with 5nt overhangs.	2.7Å dataset No solution
D-ST5-1-Br1	DNA small triangle of 81nt with 5nt overhangs.	≥ 3.3Å diffracting crystals
D-ST5-1-Br2	DNA small triangle of 81nt with 5nt overhangs.	No diffracting crystals
D-ST6-1	DNA small triangle of 81nt with 6nt overhangs.	No diffracting crystals
D-ST6-2	DNA small triangle of 81nt with 6nt overhangs.	No diffracting crystals
D-ST7	DNA small triangle of 81nt with 7nt overhangs.	No diffracting crystals
H1-ST5-1	Hybrid RNA/DNA small triangle of 81nt with 5nt overhangs. RNA-outer strand, DNA-inner strand.	2.3Å dataset No solution (twinning issues)
H2-ST5-1	Hybrid DNA/RNA small triangle of 81nt with 5nt overhangs. DNA-outer strand, RNA-inner strand.	No diffracting crystals
D-LT1	DNA long SVV IIa construct	No diffracting crystals
D-LT4	DNA large triangle with 4nt overhangs.	No diffracting crystals
D-LT5	DNA large triangle with 5nt overhangs.	No diffracting crystals

Discussion of RNA Small Triangle Atomic Force Microscopy

Attempts were made to structurally characterize the small triangle construct by atomic force microscopy (AFM) to determine if this method was useful for non-

crystallizing constructs. This small triangle crystal structure was measured to have side lengths of roughly 6nm (Figure 5.2b, top). AFM experiments revealed particles between 5-7nm in diameter (Figure 5.43), consistent with crystal structure measurements.

Figure 5.43: AFM images of small RNA triangle. Top image shows a zoomed out image with a 140nm scale bar beneath it. Color scale of image is indicated by height sensor on right. Boxed area of 132nm^2 is enlarged on the bottom. Triangle particles shown measure 5-7nm in diameter, consistent with crystal structure measurements.



Conclusions

The design of nano-scale objects that self-assemble from short oligonucleotides remains a key challenge in the emerging field of RNA nanotechnology. While larger nano-architectures have previously been constructed from structurally complex RNA motifs or long nucleic acid sequences in conjunction with helper oligonucleotides, we aimed at creating minimalist RNA nano-objects by efficient assembly of short sequences which by themselves do not adopt stable structures. Detailed insight from X-ray crystallography was exploited to design and construct two different RNA nanotriangles that self-assembled from six oligonucleotides in solution and were crystallized for structural studies. The nanotriangles display remarkable stability towards denaturation and their composition offers unique structural features that promise applications in medicine, nanomaterials engineering, and as tools to test nano-scale phenomena. As a consequence of the hierarchical self-assembly from six short oligonucleotides, the RNA nanotriangles will be readily modified by conjugation at any of the 12 component strand termini to introduce additional functionality. Since the corner units used in the construction of the nanotriangles are also ligand-dependent conformational switches, association and dissociation of the resulting RNA architectures is tunable which will enable the design of nanodevices sensitive to environmental or cellular milieus.

Materials and Methods

Preparation of RNA

RNA oligonucleotides for solution and crystallography studies were obtained by chemical synthesis (Integrated DNA Technologies, Coralville, IA) and purified by desalting. Stock solutions were prepared by dissolving lyophilized single stranded oligonucleotides in 10mM sodium cacodylate buffer, pH 6.5.

Double-stranded RNA for solution studies was annealed from stoichiometric amounts of single-stranded stock solutions in the presence of 5mM MgCl₂ by heating to 65°C for 5 minutes followed by snap cooling and incubation at 4°C for 5 minutes. Double-stranded RNA for crystallographic studies was annealed in the same way but in the absence of MgCl₂ and by slow cooling to room temperature over 45 minutes. For programmable large triangle solution studies, corners A, B, and C were assembled individually by annealing stoichiometric amounts of single-stranded stock solutions in the presence of 5mM MgCl₂ by heating to 65°C for 5 minutes followed by snap cooling and incubation at 4°C for 5 minutes. Equal volume amounts of corners A, B, and C were then mixed and incubated at 25°C or 37°C for 20 minutes to assemble ABC programmable large triangles.

Gel Electrophoresis

RNA solution structures were analyzed on 13% native polyacrylamide gels in 2X MOPS buffer, pH 7.0 (40mM 3-morpholinopropane-1-sulfonic acid, 10mM sodium acetate) and corresponding MgCl₂ concentration at 4°C. RNA was visualized by staining with ethidium bromide and illumination under UV. Single-stranded RNA (inner, outer

alone) was loaded at 1.5X concentration of double-stranded RNA (both; Figures 5.3, 5.4) for greater band staining intensity with ethidium bromide. Size reference markers were a low range single-stranded RNA ladder (New England Biolabs, Ipswich, MA).

Thermo- and Chemical Stability

To test for resistance to boiling temperatures and chemical denaturation, assembled RNA nanotriangles were incubated at 25°C or 100°C in the presence or absence of 8M urea for 5 minutes. Samples were then immediately frozen on dry ice to prevent the reassembly of dissociated RNA nanotriangles, where they remained until loaded on a gel at 4°C.¹³⁷

RNA Assembly and Dissociation in Presence of Switch-Binding Ligand

RNA nanotriangle assembly in the presence of switch-binding or control ligands was measured by annealing stoichiometric amounts of single-stranded RNA in the presence of 5mM MgCl₂, with 0, 100, or 500uM switch-binding benzimidazole ligand⁵³ or control ligands by heating to 65°C for 5 minutes followed by snap cooling and incubation at 4°C for 5 minutes. Dissociation of already assembled RNA nanotriangles was measured when incubated at 25°C or 37°C for 20 minutes in the presence of 0, 100, or 500uM switch-binding ligand.

Crystallization, Data Collection, and Structure Determination

SVV IIa short and long RNA crystal structures were obtained as described earlier.³⁹ Assembled RNA nanotriangles were crystallized at 6°C by hanging drop vapor

diffusion. In crystallization setups, 1 μ L of 0.2mM RNA was mixed with an equal volume of precipitant solution containing 55mM NaCl, 12mM spermine, 40mM Tris hydrochloride buffer, pH 7.5, and 20% methyl-2,4-pentanediol (MPD). Crystals grew over the course of one month of equilibration against 700 μ L of well solution containing 120mM NaCl and 18% MPD.

Crystals were flash-cooled in liquid nitrogen. X-ray diffraction data were collected at 110K on a Rigaku rotating anode X-ray generator ($\lambda = 1.54 \text{ \AA}$) equipped with a MAR345 imaging plate detector system. Datasets were processed, integrated, and scaled with the HKL2000 package.¹³⁸ The three-dimensional structure of the RNA nanotriangle was solved by molecular replacement with the program Phaser⁹² using search models derived from the short and long SVV Ila RNA crystal structures (RCSB Protein Data Bank accession codes 4P97 and 4PHY), and refined by the program Refmac⁹³ both within the CCP4 package.⁹⁴ Subsequent iterative rounds of manual building and refinement, alternating between Refmac and manual rebuilding in Coot,⁹⁵ were based on the obtained 2Fo-Fc and Fo-Fc maps. Chloride ions were assigned based on electron density, B-factor refinement and distance as well as geometry of coordinating ligands. Final refinement was carried out in PHENIX⁹⁶ with individual isotropic atomic displacement parameters and water picking (Table 5.1). Coordinates and structure factors have been deposited in the RCSB Protein Data Bank under accession code 5CNR.

Atomic Force Microscopy

Samples were prepared for AFM experiments as follows: Small RNA triangles were assembled in 5mM MgCl₂ and 10mM sodium cacodylate, pH 7.0 at 65° for 5

minutes and snap cooled to 4° for 5 minutes. Mica was glued to a silicon wafer solid support using clear nail polish and allowed to dry overnight. Immediately before deposition, mica was freshly cleaved using standard clear Scotch tape. 5µL of 5nM assembled RNA small triangle was deposited on mica surface and incubated at room temperature for 2 minutes. Mica was then held at a 45° angle with forceps and gently tapped to shake off 5µL RNA sample. Mica was then washed with 10uL 2mM MgSO₄ to dilute and remove excess magnesium chloride and sodium cacodylate buffer salt. Washing consisted of pipetting 10uL of 2mM MgSO₄ onto mica surface and then held at a 45° angle with forceps and gently tapped to shake off 10µL wash. Mica with deposited RNA small triangle was allowed to dry overnight in air and was imaged in air on a Dimension Icon Atomic Force Microscope (Bruker Corporation, Billerica, MA) operating in PeakForce Tapping® mode.

Acknowledgements

The material in Chapter 5, in part, has been published in *Angewandte Chemie International Edition* (**2016**, 55, 4097-4100) titled *Crystal structure-guided design of self-assembling RNA nanotriangles* with Sergey Dibrov and Thomas Hermann as co-authors. The dissertation author was the primary investigator and author of this material.

The material in Chapter 5, in part, is currently being prepared for submission for publication titled *Design and Crystallography of Self-Assembling RNA Nanostructures* with Thomas Hermann as co-author. The dissertation author was the primary investigator and author of this material.⁸⁹

Thanks to Clare LeGuyader for her expertise and assistance with AFM experiments. Thanks to Andrew Bergdorf for initial help setting up crystal screens.⁶

References

1. Cech, T. R. & Steitz, J. A. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
2. Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. *Nat. Rev. Genet.* **15**, 423–437 (2014).
3. Hermann, T. Drugs targeting the ribosome. *Current Opinion in Structural Biology* **15**, 355–366 (2005).
4. McCoy, L. S., Xie, Y. & Tor, Y. Antibiotics that target protein synthesis. *Wiley Interdiscip. Rev. RNA* **2**, 209–232 (2011).
5. Le Grice, S. F. J. Targeting the HIV RNA genome: High-hanging fruit only needs a longer ladder. *Curr. Top. Microbiol. Immunol.* **389**, 147–169 (2015).
6. Dibrov, S. M., Parsons, J., Carnevali, M., Zhou, S., Rynearson, K. R., Ding, K., Garcia Segal, E., Brunn, N. D., Boerneke, M. A., Castaldi, M. P., Hermann, T. Hepatitis C Virus Translation Inhibitors Targeting the Internal Ribosomal Entry Site. *J. Med. Chem.* **57**, 1694–1707 (2014).
7. Howe, J. A., Wang, H., Fischmann, T. O., Balibar, C. J., Xiao, L., Galgoci, A. M., Malinverni, J. C., Mayhood, T., Villanfanía, A., Nahvi, A., Murgolo, N., Barbieri, C. M., Mann, P. A., Carr, D., Xia, E., Zuck, P., Riley, D., Painter, R. E., Walker, S. S., Sherborne, B., de Jesus, R., Pan, W., Plotkin, M. A., Wu, J., Rindgen, D., Cummings, J., Garlisi, C. G., Zhang, R., Sheth, P. R., Gill, C. J., Tang, H., Roemer, T. Selective small-molecule inhibition of an RNA structural element. *Nature* **526**, 672–677 (2015).
8. Matzner, D. & Mayer, G. (Dis)similar analogues of riboswitch metabolites as antibacterial lead compounds. *J. Med. Chem.* **58**, 3275–3286 (2015).
9. Edlin, B. R., Eckhardt, B. J., Shu, M. A., Holmberg, S. D. & Swan, T. Toward a more accurate estimate of the prevalence of hepatitis C in the United States. *Hepatology* **62**, 1353–1363 (2015).
10. Denniston, M. M., Jiles, R. B., Drobeniuc, J., Klevens, M., Ward, J. W., McQuillan, G. M., Holmberg, S. D. Chronic Hepatitis C Virus Infection in the United States, National Health and Nutrition Examination Survey 2003 to 2010. *Ann. Intern. Med.* **160**, 293–300 (2014).
11. Lavanchy, D. Evolving epidemiology of hepatitis C virus. *Clinical Microbiology and Infection* **17**, 107–115 (2011).

12. Ly, K. N., Xing, J., Klevens, M., Jiles, R. B., Ward, J. W., Holmberg, S. D. The increasing burden of mortality from viral hepatitis in the United States between 1999 and 2007. *Ann. Intern. Med.* **156**, 271–8 (2012).
13. World Health Organization. *Hepatitis C Fact Sheet no. 164*. (2015).
14. Center for Disease Control and Prevention. *Hepatitis C FAQs for Health Professionals*. (2016). at <http://www.cdc.gov/hepatitis/hcv/hcvfaq.htm#section4>
15. Feld, J. J. & Hoofnagle, J. H. Mechanism of action of interferon and ribavirin in treatment of hepatitis C. *Nature* **436**, 967–972 (2005).
16. American Association For The Study Of Liver Diseases. *Recommendations for Testing, Managing, and Treating Hepatitis C*. (2016).
17. Enserink, M. First Specific Drugs Raise Hopes for Hepatitis C. *Science* **332**, 159–160 (2011).
18. Chung, R. T., Davis, G. L., Jensen, D. M., Masur, H., Saag, M. S., Thomas, D. L., Aronsohn, A. I., Charlton, M. R., Feld, J. J., Fontana, R. J., Ghany, M. G., Godofsky, E. W., Graham, C. S., Kim, A. Y., Kiser, J. J., Kottlil, S., Marks, K. M., Martin, P., Mitruka, K., Morgan, T. R., Naggie, S., Raymond, D., Reau, N. S., Schooley, R. T., Sherman, K. E., Sulkowski, M. S., Vargas, H. E., Ward, J. W., Wyles, D. L. Hepatitis C guidance: AASLD-IDSAs recommendations for testing, managing, and treating adults infected with hepatitis C virus. *Hepatology* **62**, 932–954 (2015).
19. Luetkemeyer, A. F. & Wyles, D. L. CROI 2015: Highlights of Viral Hepatitis Therapy. *Top Antivir Med* **23**, 66–76 (2015).
20. Robinson, M., Tian, Y., Delaney, W. E. & Greenstein, A. E. Preexisting drug-resistance mutations reveal unique barriers to resistance for distinct antivirals. *Proc. Natl Acad. Sci. USA* **108**, 10290–10295 (2011).
21. Dibrov, S. M., Ding, K., Brunn, N. D. Parker, M. A., Bergdahl, B. M., Wyles, D. L., Hermann, T. Structure of a hepatitis C virus RNA domain in complex with a translation inhibitor reveals a binding mode reminiscent of riboswitches. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5223–8 (2012).
22. Honda, M., Ping, L. H., Rijnbrand, R. C. A., Amphlett, E., Clarke, B., Rowlands, D., Lemon, S. M. Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. *Virology* **222**, 31–42 (1996).

23. Plank, T.-D. M. & Kieft, J. S. The structures of nonprotein-coding RNAs that drive internal ribosome entry site function. *Wiley Interdiscip. Rev. RNA* **3**, 195–212 (2012).
24. Hellen, C. U. T. & Pestova, T. V. Translation of hepatitis C virus RNA. *J. Viral Hepat.* **6**, 79–87 (1999).
25. Ji, H., Fraser, C. S., Yu, Y., Leary, J. & Doudna, J. A. Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16990–5 (2004).
26. Otto, G. A. & Puglisi, J. D. The Pathway of HCV IRES-Mediated Translation Initiation. *Cell* **119**, 369–380 (2004).
27. Fraser, C. & Doudna, J. Structural and mechanistic insights into hepatitis C viral translation initiation. *Nat. Rev. Microbiol.* **5**, 29–38 (2007).
28. Kieft, J. S., Zhou, K. H., Jubin, R., Murray, M. G., Lau, J. Y. N., Doudna, J. A. The hepatitis C virus internal ribosome entry site adopts an ion-dependent tertiary fold. *J. Mol. Biol.* **292**, 513–529 (1999).
29. Friebe, P., Lohmann, V., Krieger, N. & Bartenschlager, R. Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication. *J. Virol.* **75**, 12047–57 (2001).
30. Kieft, J. S., Zhou, K., Jubin, R. & Doudna, J. A. Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* **7**, 194–206 (2001).
31. Berry, K., Waghray, S., Mortimer, S., Bai, Y. & Doudna, J. Crystal structure of the HCV IRES central domain reveals strategy for start-codon positioning. *Structure* **19**, 1456–1466 (2011).
32. Berry, K. E., Waghray, S. & Doudna, J. A. The HCV IRES pseudoknot positions the initiation codon on the 40S ribosomal subunit. *RNA* **16**, 1559–1569 (2010).
33. Quade, N., Boehringer, D., Leibundgut, M., van den Heuvel, J. & Ban, N. Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-Å resolution. *Nat. Commun.* **6**, 7646 (2015).
34. Spahn, C. M., Kieft, J. S., Grassucci, R. A., Penczek, P. A., Zhou, K., Doudna, J. A., Frank, J. Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science* **291**, 1959–62 (2001).
35. Filbin, M. E. & Kieft, J. S. HCV IRES domain IIb affects the configuration of coding RNA in the 40S subunit's decoding groove. *RNA* **17**, 1258–73 (2011).

36. Pestova, T. V., Shatsky, I. N., Fletcher, S. P., Jackson, R. J. & Hellen, C. U. T. A. prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev.* **12**, 67–83 (1998).
37. Terenin, I. M., Dmitriev, S. E., Andreev, D. E. & Shatsky, I. N. Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. *Nat. Struct. Mol. Biol.* **15**, 836–41 (2008).
38. Kim, J. H., Park, S. M., Park, J. H., Keum, S. J. & Jang, S. K. eIF2A mediates translation of hepatitis C viral mRNA under stress conditions. *EMBO J.* **30**, 2454–64 (2011).
39. Boerneke, M. A., Dibrov, S. M., Gu, J., Wyles, D. L. & Hermann, T. Functional conservation despite structural divergence in ligand-responsive RNA switches. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15952–15957 (2014).
40. Hussain, T., Llácer, J. L., Fernández, I. S., Munoz, A., Martin-Marcos, P., Savva, C. G., Lorsch, J. R., Hinnebusch, A. G., Ramakrishnan, V. Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell* **159**, 597–607 (2014).
41. Yamamoto, H., Collier, M., Loerke, J., Ismer, J., Schmidt, A., Hilal, T., Sprink, T., Yamamoto, K., Mielke, T., Bürger, J., Shaikh, T. R., Dabrowski, M., Hildebrand, P. W., Scheerer, P., Spahn, C. M. T. Molecular architecture of the ribosome- bound Hepatitis C Virus internal ribosomal entry site RNA. *EMBO J.* **34**, 3042–3058 (2015).
42. Yamamoto, H., Unbehaun, A., Loerke, J., Behrmann, E., Collier, M., Bürger, J., Mielke, T., Spahn, C. M. T. Structure of the mammalian 80S initiation complex with initiation factor 5B on HCV-IRES RNA. *Nat. Struct. Mol. Biol.* **21**, 721–7 (2014).
43. Kolupaeva, V. G., Pestova, T. V. & Hellen, C. U. Ribosomal binding to the internal ribosomal entry site of classical swine fever virus. *RNA* **6**, 1791–1807 (2000).
44. Locker, N., Easton, L. & Lukavsky, P. HCV and CSFV IRES domain II mediate eIF2 release during 80S ribosome assembly. *EMBO J.* **26**, 795–805 (2007).
45. Fraser, C. S., Hershey, J. W. B. & Doudna, J. A. The pathway of hepatitis C virus mRNA recruitment to the human ribosome. *Nat. Struct. Mol. Biol.* **16**, 397–404 (2009).

46. Filbin, M. E., Vollmar, B. S., Shi, D., Gonen, T. & Kieft, J. S. HCV IRES manipulates the ribosome to promote the switch from translation initiation to elongation. *Nat. Struct. Mol. Biol.* **20**, 150–8 (2013).
47. Dibrov, S. M., Johnston-Cox, H., Weng, Y.-H. & Hermann, T. Functional architecture of HCV IRES domain II stabilized by divalent metal ions in the crystal and in solution. *Angew. Chem. Int. Ed. Engl.* **46**, 226–9 (2007).
48. Lukavsky, P., Kim, I., Otto, G. & Puglisi, J. Structure of HCV IRES domain II determined by NMR. *Nat. Struct. Mol. Biol.* **10**, 1033–1038 (2003).
49. Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J. D. & Stark, H. Structure of the hepatitis C virus IRES bound to the human 80S ribosome: Remodeling of the HCV IRES. *Structure* **13**, 1695–1706 (2005).
50. Fukushi, S. *et al.* Ribosomal Protein S5 Interacts with the Internal Ribosomal Entry Site of Hepatitis C Virus. *J. Biol. Chem.* **276**, 20824–20826 (2001).
51. Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J. D. & Stark, H. Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. *Structure* **13**, 1695–706 (2005).
52. Fuchs, G. Fuchs, Gabriele, Petrov, A. N., Marceau, C. D., Popov, L. M., Chen, J., O'Leary, S. E., Wang, R., Carette, J. E., Sarnow, P., Puglisi, J. D. Kinetic pathway of 40S ribosomal subunit recruitment to hepatitis C virus internal ribosome entry site. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 319–25 (2015).
53. Parsons, J., Castaldi, M. P., Dutta, S., Dibrov, S. M., Wyles, D. L., Hermann, T. Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. *Nat. Chem. Biol.* **5**, 823–825 (2009).
54. Seth, P. P. Seth, P P, Miyaji, A., Jefferson, E. A., Sannes-Lowery, K. A., Osgood, S. A., Propp, S. S., Ranken, R., Massire, C., Sampath, R., Ecker, D. J., Swayze, E. E., Griffey, R. H. SAR by MS: Discovery of a new class of RNA-binding small molecules for the hepatitis C virus: Internal ribosome entry site IIA subdomain. *J. Med. Chem.* **48**, 7099–7102 (2005).
55. Carnevali, M., Parsons, J., Wyles, D. L. & Hermann, T. A modular approach to synthetic RNA binders of the hepatitis C virus internal ribosome entry site. *Chembiochem* **11**, 1364–1367 (2010).
56. Hellen, C. U. T. & de Breyne, S. A distinct group of hepacivirus/pestivirus-like internal ribosomal entry sites in members of diverse picornavirus genera: evidence for modular exchange of functional noncoding RNA elements by recombination. *J. Virol.* **81**, 5850–63 (2007).

57. Asnani, M., Kumar, P. & Hellen, C. U. T. Widespread distribution and structural diversity of Type IV IRESs in members of Picornaviridae. *Virology* **478C**, 61–74 (2015).
58. Boerneke, M. A. & Hermann, T. Ligand-responsive RNA mechanical switches. *RNA Biol.* **12**, 780–786 (2015).
59. Boerneke, M. A., Dibrov, S. M. & Hermann, T. Crystal-Structure-Guided Design of Self-Assembling RNA Nanotriangles. *Angew. Chemie Int. Ed.* **55**, 4097–4100 (2016).
60. Pawar, M. G., Nuthanakanti, A. & Srivatsan, S. G. Heavy atom containing fluorescent ribonucleoside analog probe for the fluorescence detection of RNA-ligand binding. *Bioconjug. Chem.* **24**, 1367–77 (2013).
61. Boerneke, M. A., Nuthanakanti, A., Srivatsan, S. G. & Hermann, T. *Fluorescence detection of RNA-ligand binding and crystal structure determination of ribosomal decoding site RNA using a heavy atom containing fluorescent ribonucleoside.* (2016).
62. Boerneke, M. A. & Hermann, T. Conformational flexibility of viral RNA switches studied by FRET. *Methods* **91**, 35–39 (2015).
63. Pestova, T. V, de Breyne, S., Pisarev, A. V, Abaeva, I. S. & Hellen, C. U. T. eIF2-dependent and eIF2-independent modes of initiation on the CSFV IRES: a common role of domain II. *EMBO J.* **27**, 1060–72 (2008).
64. Parker, M. a, Satkiewicz, E., Hermann, T. & Bergdahl, B. M. An efficient new route to dihydropyranobenzimidazole inhibitors of HCV replication. *Molecules* **16**, 281–90 (2011).
65. Ding, K., Wang, A., Boerneke, M. A., Dibrov, S. M. & Hermann, T. Aryl-substituted aminobenzimidazoles targeting the hepatitis C virus internal ribosome entry site. *Bioorg. Med. Chem. Lett.* **24**, 3113–7 (2014).
66. Rynearson, K. D., Charrette, B., Gabriel, C., Moreno, J., Boerneke, M. A., Dibrov, S. M., Hermann, T. 2-Aminobenzoxazole ligands of the hepatitis C virus internal ribosome entry site. *Bioorg. Med. Chem. Lett.* **24**, 3521–5 (2014).
67. Puglisi, J. D., Chen, L., Frankel, A. D. & Williamson, J. R. Role of RNA structure in arginine recognition of TAR RNA. *Proc Natl Acad Sci U S A* **90**, 3680–3684 (1993).
68. Kondo, J. & Westhof, E. Classification of pseudo pairs between nucleotide bases and amino acids by analysis of nucleotide-protein complexes. *Nucleic Acids Res.* **39**, 8628–8637 (2011).

69. Walberer, B. J., Cheng, A. C. & Frankel, A. D. Structural diversity and isomorphism of hydrogen-bonded base interactions in nucleic acids. *J. Mol. Biol.* **327**, 767–780 (2003).
70. Abu Almakarem, A. S., Petrov, A. I., Stombaugh, J., Zirbel, C. L. & Leontis, N. B. Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic Acids Res.* **40**, 1407–1423 (2012).
71. Yarus, M. A specific amino acid binding site composed of RNA. *Science* (80-.). **240**, 1751–1758 (1988).
72. Zhou, S., Rynearson, K. D., Ding, K., Brunn, N. D. & Hermann, T. Screening for inhibitors of the hepatitis C virus internal ribosome entry site RNA. *Bioorg. Med. Chem.* **21**, 6139–44 (2013).
73. Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenko, E. V., Shatsky, I. N., Agol, V. I., Hellen, C. U. Molecular mechanisms of translation initiation in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 7029–36 (2001).
74. Willcocks, M. M., Locker, N., Gomwalk, Z., Royall, E., Bakhshesh, M., Belsham, G. J., Idamakanti, N., Burroughs, K. D., Reddy, P. S., Hallenbeck, P. L., Roberts, L. O. Structural features of the Seneca Valley virus internal ribosome entry site (IRES) element: a picornavirus with a pestivirus-like IRES. *J. Virol.* **85**, 4452–4461 (2011).
75. Leontis, N. B. & Westhof, E. A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J. Mol. Biol.* **283**, 571–83 (1998).
76. Reusken, C. B. E. M., Dalebout, T. J., Eerligh, P., Bredenbee, P. J., Spaan, W. J. M. Analysis of hepatitis C virus/classical swine fever virus chimeric 5'NTRs: sequences within the hepatitis C virus IRES are required for viral RNA replication. *J. Gen. Virol.* **84**, 1761–1769 (2003).
77. Blight, K. J., McKeating, J. A., Marcotrigiano, J. & Rice, C. M. Efficient replication of hepatitis C virus genotype 1a RNAs in cell culture. *J Virol* **77**, 3181–3190 (2003).
78. Shi, S. T. & Lai, M. M. C. HCV 5' and 3' UTR: When Translation Meets Replication. *Hepat. C Viruses Genomes Mol. Biol.* 49–87 (2006).
79. Romero-Lopez, C. & Berzal-Herranz, A. A long-range RNA – RNA interaction between the 5' and 3' ends of the HCV genome. *RNA* **15**, 1740–1752 (2009).

80. Wyles, D. L., Kaihara, K. A., Korba, B. E., Schooley, R. T., Beadle, J. R., Hostetler, K. Y. The octadecyloxyethyl ester of (S)-9-[3-hydroxy-2-(phosphonomethoxy) propyl]adenine is a potent and selective inhibitor of hepatitis C virus replication in genotype 1A, 1B, and 2A replicons. *Antimicrob. Agents Chemother.* **53**, 2660–2 (2009).
81. Traut, T. W. Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem.* **140**, 1–22 (1994).
82. Seth, P. P., Jefferson, E. A., Griffey, R. H. & Swayze, E. E. Benzimidazoles and analogs thereof as antivirals. **1**, (2012).
83. Agarwal, A., Banerjee, A. & Banerjee, U. C. Xanthine oxidoreductase: a journey from purine metabolism to cardiovascular excitation-contraction coupling. *Crit. Rev. Biotechnol.* **31**, 264–80 (2011).
84. De Clercq, E. & Field, H. J. Antiviral prodrugs - the development of successful prodrug strategies for antiviral chemotherapy. *Br. J. Pharmacol.* **147**, 1–11 (2006).
85. Nichol, C. A., Smith, G. K. & Duch, D. S. Biosynthesis and metabolism of tetrahydrobiopterin and molybdopterin. *Annu Rev Biochem* **54**, 729 (1985).
86. Thöny, B., Auerbach, G. & Blau, N. Tetrahydrobiopterin biosynthesis, regeneration and functions. *Biochem. J.* **347 Pt 1**, 1–16 (2000).
87. Kis, K., Volk, R. & Bacher, a. Biosynthesis of riboflavin. Studies on the reaction mechanism of 6,7-dimethyl-8-ribityllumazine synthase. *Biochemistry* **34**, 2883–92 (1995).
88. Ding, K. Structure-Guided Design and Synthesis of Small Molecules Targeting Hepatitis C Virus Internal Ribosome Entry Site RNA. (2014).
89. Charrette, B., Boerneke, M. A. & Hermann, T. *Aminobenzodiazepinones targeting the hepatitis C virus internal ribosome entry site.* (2016).
90. Brunn, N. D., Garcia Segal, E., Kao, M. B. & Hermann, T. Targeting a regulatory element in human thymidylate synthase mRNA. *Chembiochem* **13**, 2738–44 (2012).
91. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **276**, 307–326 (1997).
92. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

93. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255 (1997).
94. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **50**, 760–3 (1994).
95. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
96. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Terwilliger, T. C. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948–1954 (2002).
97. Wyles, D. L., Kaihara, K. A., Vaida, F. & Schooley, R. T. Synergy of small molecular inhibitors of hepatitis C virus replication directed at multiple viral targets. *J. Virol.* **81**, 3005–8 (2007).
98. Serganov, A. & Nudler, E. A decade of riboswitches. *Cell* **152**, 17–24 (2013).
99. Garst, A. D., Edwards, A. L. & Batey, R. T. Riboswitches: structures and mechanisms. *Cold Spring Harb. Perspect. Biol.* **3**, a003533 (2011).
100. Roth, A. & Breaker, R. R. The Structural and Functional Diversity of Metabolite-Binding Riboswitches. *Annu. Rev. Biochem.* **78**, 305–334 (2009).
101. Jaeger, L., Verzemnieks, E. J. & Geary, C. The UA_handle: a versatile submotif in stable RNA architectures. *Nucleic Acids Res.* **37**, 215–30 (2009).
102. Wadley, L. M. & Pyle, A. M. The identification of novel RNA structural motifs using COMPADRES: an automated approach to structural discovery. *Nucleic Acids Res.* **32**, 6650–9 (2004).
103. Nomenclature Committee of the International Union of Biochemistry (NC-IUB). Nomenclature for incompletely specified bases in nucleic acid sequences. *J. Biol. Chem.* **229**, 281–286 (1985).
104. Dibrov, S. M., McLean, J., Parsons, J. & Hermann, T. Self-assembling RNA square. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6405–8 (2011).
105. Yu, J., Liu, Z., Jiang, W., Wang, G. & Mao, C. De novo design of an RNA tile that self-assembles into a homo-octameric nanoprism. *Nat. Commun.* **6**, 5724 (2015).
106. PDB search criteria- experimental method: X-ray; molecule type: protein; 05/18/2016. at <<http://www.rcsb.org/pdb/search/advSearch.do?search=new>>

107. PDB search criteria- experimental method: X-ray; molecule type: nucleic acid; 05/18/2016. at <<http://www.rcsb.org/pdb/search/advSearch.do?search=new>>
108. Honda, M., Brown, E. A. & Lemon, S. M. Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *RNA* **2**, 955–968 (1996).
109. Edwards, A. L., Garst, A. D. & Batey, R. T. Determining structures of RNA aptamers and riboswitches by X-ray crystallography. *Methods Mol. Biol.* **535**, 135–163 (2009).
110. Petrov, A. & Puglisi, J. D. Site-specific labeling of *Saccharomyces cerevisiae* ribosomes for single-molecule manipulations. *Nucleic Acids Res.* **38**, 1–9 (2010).
111. Blanchard, S. C. Single-molecule observations of ribosome function. *Curr. Opin. Struct. Biol.* **19**, 103–109 (2009).
112. Ogle, J. M., Carter, A. P. & Ramakrishnan, V. Insights into the decoding mechanism from recent ribosome structures. *Trends in Biochemical Sciences* **28**, 259–266 (2003).
113. Moazed, D. & Noller, H. F. Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature* **327**, 389–394 (1987).
114. Shandrick, S., Zhao, Q., Han, Q., Ayida, B. K., Takahashi, M., Winters, G. C., Simonsen, K. B., Vourloumis, D., Hermann, T. Monitoring molecular recognition of the ribosomal decoding site. *Angew. Chemie - Int. Ed.* **43**, 3177–3182 (2004).
115. Dibrov, S. M., Parsons, J. & Hermann, T. A model for the study of ligand binding to the ribosomal RNA helix h44. *Nucleic Acids Res.* **38**, 4458–4465 (2010).
116. Jaeger, L. & Leontis, N. Tecto-RNA: One-Dimensional Self-Assembly through Tertiary Interactions. *Angew. Chem. Int. Ed. Engl.* **39**, 2521–2524 (2000).
117. Shu, D., Huang, L. P., Hoepflich, S. & Guo, P. Construction of phi29 DNA-packaging RNA monomers, dimers, and trimers with variable sizes and shapes as potential parts for nanodevices. *J. Nanosci. Nanotechnol.* **3**, 295–302 (2003).
118. Chworos, A., Severcan, I., Koyfman, A. Y., Weinkam, P., Oroudjev, E., Hansma, H. G., Jaeger, L. Building programmable jigsaw puzzles with RNA. *Science* **306**, 2068–72 (2004).
119. Nasalean, L., Baudrey, S., Leontis, N. B. & Jaeger, L. Controlling RNA self-assembly to form filaments. *Nucleic Acids Res.* **34**, 1381–92 (2006).

120. Bindewald, E., Grunewald, C., Boyle, B., O'Connor, M. & Shapiro, B. A. Computational strategies for the automated design of RNA nanoscale structures from building blocks using NanoTiler. *J. Mol. Graph. Model.* **27**, 299–308 (2008).
121. Severcan, I., Geary, C., Verzemnieks, E., Chworos, A. & Jaeger, L. Square-shaped RNA particles from different RNA folds. *Nano Lett.* **9**, 1270–1277 (2009).
122. Severcan, I., Geary, C., Chworos, A., Voss, N., Jacovetty, E., Jaeger, L. A polyhedron made of tRNAs. *Nat. Chem.* **2**, 772–9 (2010).
123. Afonin, K. A., Bindewald, E., Yaghoubian, A. J., Voss, N., Jacovetty, E., Shapiro, B. A., Jaeger, L. In vitro assembly of cubic RNA-based scaffolds designed in silico. *Nat. Nanotechnol.* **5**, 676–682 (2010).
124. Geary, C., Chworos, A. & Jaeger, L. Promoting RNA helical stacking via A-minor junctions. *Nucleic Acids Res.* **39**, 1066–80 (2011).
125. Grabow, W. W., Zakrevsky, P., Afonin, K. A., Chworos, A., Shapiro, B. A., Jaeger, L. Self-assembling RNA nanorings based on RNAI/II inverse kissing complexes. *Nano Lett.* **11**, 878–87 (2011).
126. Ohno, H., Kobayashi, T., Kabata, R., Endo, K., Iwasa, T., Yoshimura, S. H., Takeyasu, K., Inoue, T., Saito, H. Synthetic RNA–protein complex shaped like an equilateral triangle. *Nat. Nanotechnol.* **6**, 116–120 (2011).
127. Bindewald, E., Afonin, K., Jaeger, L. & Shapiro, B. a. Multistrand RNA secondary structure prediction and nanostructure design including pseudoknots. *ACS Nano* **5**, 9542–51 (2011).
128. Geary, C., Rothmund, P. W. K. & Andersen, E. S. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science.* **345**, 799–804 (2014).
129. Jaeger, L., Westhof, E. & Leontis, N. B. TectoRNA: modular assembly units for the construction of RNA nano-objects. *Nucleic Acids Res.* **29**, 455–63 (2001).
130. Guo, P., Zhang, C., Chen, C., Garver, K. & Trottier, M. Inter-RNA Interaction of Phage ϕ 29 pRNA to Form a Hexameric Complex for Viral DNA Transportation. *Mol. Cell* **2**, 149–155 (1998).
131. Zhang, H., Endrizzi, J. A., Shu, Y., Haque, F., Sauter, C., Shlyakhtenko, L. S., Lyubchenko, Y., Guo, P., Chi, Y. Crystal structure of 3WJ core revealing divalent ion-promoted thermostability and assembly of the Phi29 hexameric motor pRNA. *Rna* **19**, 1226–1237 (2013).

132. Jabbari, H., Aminpour, M. & Montemagno, C. Computational approaches to nucleic acid origami. *ACS Comb. Sci.* (2015). doi:10.1021/acscombsci.5b00079
133. Trame, C. B. & McKay, D. B. Structure of Haemophilus influenzae HslU protein in crystals with one-dimensional disorder twinning. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57**, 1079–1090 (2001).
134. Benirschke, R. C., Thompson, J. R., Nominé, Y., Wasielewski, E., Juranić, N., Macura, S., Hatakeyama, S., Nakayama, K. I., Botuyan, M. V., Mer, G. Molecular basis for the association of human E4B U box ubiquitin ligase with E2-conjugating enzymes UbcH5c and Ubc4. *Structure* **18**, 955–65 (2010).
135. Schwarzenbach, D., Kirschbaum, K. & Pinkerton, A. A. Order-disorder twinning model and stacking faults in alpha-NTO. *Acta Crystallogr. B.* **62**, 944–8 (2006).
136. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
137. Khisamutdinov, E. F., Jasinski, D. L. & Guo, P. RNA as a boiling-resistant anionic polymer material to build robust structures with defined shape and stoichiometry. *ACS Nano* **8**, 4771–81 (2014).
138. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).