ABSTRACT

Title of Document: NEW CHEMICAL TOOLS TO INVESTIGATE RNA FUNCTIONS

Yiling Luo, Doctor of Philosophy, 2013

Directed By: Professor Kwaku Dayie and
Professor Herman Sintim
Department of Chemistry and Biochemistry

Ribonucleic acid (RNA), as one of the essential macromolecules of life, plays an active role in gene regulation, catalysis, and signaling because of its ability to adopt complex 3D structures that can exist in multiple conformations. Until now, RNA preparation methods devised by most investigators utilized partially denatured the RNA. The mis-folding caused by denaturing – renaturing can seriously affect RNA structure and functional activity. To test this hypothesis, in PART I of this dissertation, we presented a simple strategy using 'click' chemistry to couple biotin to a 'caged' photocleavable guanosine monophosphate to synthesize native RNAs that are properly folded. We demonstrated that RNA ribozymes, ranging in size from 27 to 527 nt, prepared by our non-denaturing method form a homogenous population with superior catalytic activity than those prepared by traditional refolding methods.

Having developed a method for *in vitro* RNA synthesis, we studied the riboswitch family of RNAs that require remolding their structure for function in

PART II. C-di-GMP riboswitch, as the only currently known secondary messenger riboswitch, senses c-di-GMP using its aptamer domain to modulate the expression of genes which affects biofilm formation and virulence factors production in bacteria. To specifically target pathogenic bacteria in polymicrobial systems that control the RNA-mediated c-di-GMP signaling pathway through riboswitch regulation, different c-di-GMP analogs have been used as chemical tools to investigate the structure-activity relationship (SAR) of c-di-GMP binding to different c-di-GMP riboswitches. We demonstrated that different 2'-position modified c-di-GMP analogs could differentiate between two different classes of c-di-GMP riboswitches and even bind to one particular riboswitch in class I with different affinities.

Specifically, *Clostridium tetani* (Ct-E88) RNA in class I c-di-GMP riboswitch was used for (structural dynamic) study to understand how ligand binding drives the conformational change to regulate the downstream genes within the expression platform. Our preliminary data obtained via different biochemical and biophysical tools – such as selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE), small-angle X-ray scattering (SAXS) and nuclear magnetic resonance (NMR) spectroscopy – demonstrated that Mg^{2+} ions accelerate ligand recognition by pre-organizing the RNA, and then rapid ligand binding folds the RNA into a compact structure for likely downstream gene regulation.

NEW CHEMICAL TOOLS TO INVESTIGATE RNA FUNCTIONS

By

Yiling Luo

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Professor Kwaku Dayie, Chair/Advisor
Professor Herman Sintim, Advisor
Professor David Fushman
Professor Lyle Isaacs
Professor Jason Kahn
Professor Jonathan Dinman

Acknowledgements

This endeavor would never have been possible without help and support from my wonderful mentors, collaborators, colleagues, friends and family.

I would specially thank my two advisors, Professor Herman O. Sintim and Professor Kwaku T. Dayie for their long-term support and professional guidance throughout my graduate career at the University of Maryland. Their assistance and advice was vital in finishing my PhD and their passion for knowledge inspire and motivate me through my life.

I would like to acknowledge both Professor Jonathan Dinman and Professor Nicole LaRonde-LeBlanc for kindly allowing me to perform radioactive experiments in their laboratories.

I thank all the members of the Sintim laboratory and the Dayie laboratory, past and present. Thanks for their assistance, encouragement and engaging discussions. It has been my pleasure working with all of them. I am especially indebted to my collaborators from Sintim's group, Jie Zhou and Jingxin Wang, for their generously and continuously supplying me all the synthesized c-di-GMP analogs studied in this dissertation and Dr. Shizuka Nakayama, for helping me make mutant WspR, and my collaborator from Dayie's group, Dr. Bin Chen, for taking so much time to help me process the SAXS data and patiently explain all the details to me and Dr. Eldho Nadukkudy for his help on native RNA preparation method development. I would thank Dr. Xiaobin Zuo and Dr. Yun-Xing Wang (NCI) for SAXS data collection. I would like to acknowledge the invaluable assistance of Dr. Ashton Trey Belew and Dr. Chandar Thakur for the SHAPE experiment. I would also like to express

appreciation to Dr. Daoning Zhang for his help and support with the NMR experiments. I thank Professor Marc Dreyfus for giving us the source of T7 RNA polymerase and Professor Vincent Lee for giving us the *E. Coli* bacteria containing wild type WspR plasmid.

I appreciate my committee members, Professor David Fushman, Professor Lyle Isaacs and Professor Jason Kahn for their support, useful suggestion and valuable critique during the course of my PhD study. My appreciation also goes to Professor Jonathan Dinman for agreeing to be the Dean's representative on my committee.

I sincerely appreciate the Sih family and the Department of Chemistry and Biochemistry for honoring me with the Dean's Summer Fellowship and Jenny T. Sih Endowed Student Award, which covered tuition and stipend for last semester and summer of my graduate studies and RNA Society Travel Award for offering me the opportunity to attend the RNA Society Meeting 2012.

In addition, I would like to add my special thanks to Luigi Alvarado, Rachel Brown, Regan Lebanc and Andrew Longhini for volunteering their time to assist me in proofreading my dissertation and independent proposal.

Finally, I would like to acknowledge my parents Yanjun Han and Xinhua Luo, and my boyfriend John Hui for their love, understanding and constant emotional support through the challenges of research and dissertation writing.

Table of Contents

List of Tables

List of Figures

List of Schemes

List of Abbreviations

BTCM	Bromotrichloromethane
C-di-GMP	Cyclic diguanylate
DCM	Dichloromethane
DMSO	Dimethyl sulfoxide
DMAP	4-Dimethylaminopyridine
DMT	Dimethyltryptamine
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetracetic acid
EMSA	Electrophoretic mobility shift assay
ETT	Ethyl-thiol-tetrazol
GMP	Guanosine monophosphate
HBTU	O-Benzotriazole-N, N’ N’-tetramethyl-uronium-hexafluorophosphate
HOBt	1-hydroxybenzotriazole
HSQC	Heteronuclear single quantum correlation
mRNA	Messenger RNA
NMIA	N-methylisotoic anhydride
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
NSD	Normalized spatial discrepancy
nt	Nucleotide

PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PC tag	Photocleavable tag
PDDF	Pair distribution function
R_g	Radius of gyration
RNA	Ribonucleic acid
RNAP	RNA polymerases
rRNA	Ribosomal RNA
SAR	Structure-activity relationship
SAXS	Small-angle X-ray scattering
SER	Sliced extron reopening
SHAPE	Selective 2'-hydroxyl acylation analyzed by primer extension
smFRET	Single molecule fluorescence resonance energy transfer
TEA	Triethanolamine
TCA	Trichloroacetic acid
TL	Tetraloop
TLR	Tetraloop receptor
TPP	Triphenylphosphite
tRNA	Transfer RNA
UV	Ultraviolet
WAXS	Wide-angle X-ray scattering

PART I – Native RNA preperation

Chapter 1: Introduction to Part I

1.1 Introduction to ribonucleic acids (RNAs)

The transfer of genetic information from deoxyribonucleic acids (DNAs) to proteins is one of the essential processes in living organisms. For a long time, ribonucleic acids (RNAs) had only been viewed as messengers that translate the genomic information, stored in DNAs, into proteins via mRNA transcripts. However, it is now appreciated that RNAs, including messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear RNA (snRNA) and other non-coding RNA (ncRNA), partake in a broad range of cellular processes, such as gene regulation and expression[1-3], enzymatic catalysis[4, 5] and signaling[6] in the cell. In fact, recent genomic sequencing efforts have indicated that protein coding sequences account for less than 2% of the human genomic output, whereas non-protein coding RNA sequences constitute at least 80% of that output.[7-10]

Like DNA, ribonucleic acid (RNA) is also made up of a long chain of nucleotide blocks. Each nucleotide consists of a nucleobase, a ribose sugar, and a phosphate group (**Figure 1.1**). An important structural feature of RNA distinct from DNA is that RNA has a hydroxyl group at the 2'-position of the ribose sugar, which facilitates the degradation of RNA via an intramolecular in-line attack at the adjacent phosphate moiety to cleave the backbone. In addition, the 2'-OH in RNA promotes a C2'-endo

sugar puckering and makes B-form helix formation unfavorable due to steric reasons.[11]

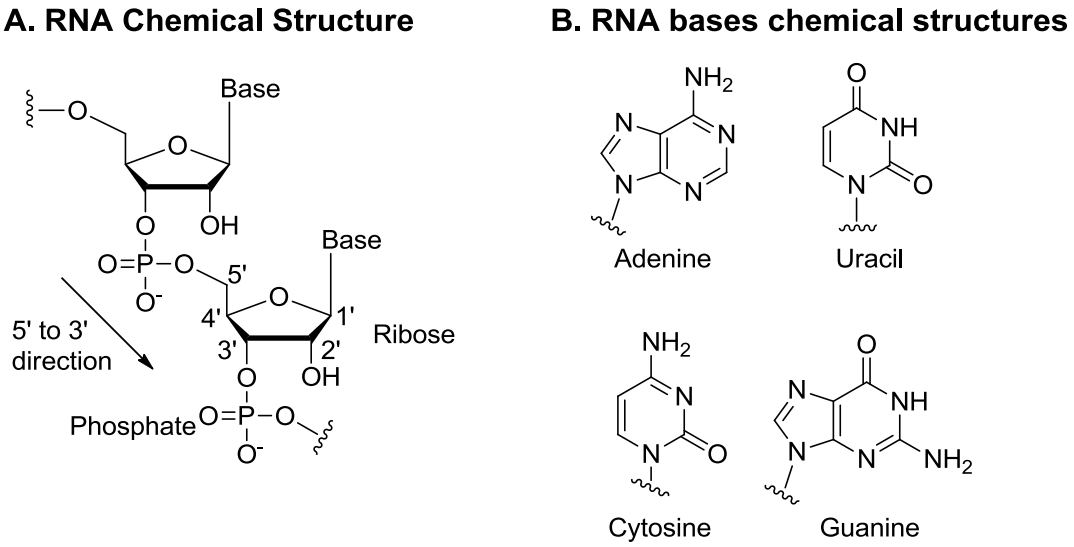


Figure 1.1. (A) Chemical Structure of RNA. **(B)** Chemical Structures of RNA bases.

1.2 Traditional method to prepare RNA in vitro

Given these peculiar and important features of RNA structure, it is not surprising that biochemical and biophysical studies of RNA are performed *in vitro* to gain valuable insights into RNA behavior *in vivo*. However, the *in vitro* studies of RNA structures and functions rely on the ability to prepare large amounts of RNA. *In vitro* transcription is a conventional RNA preparation procedure that allows for the synthesis of RNA molecules of any sequence from short oligonucleotides of less than 100 nucleotides (nts) to several kilobases in microgram to milligram amounts from DNA templates.[12, 13] The average rate of *in vitro* transcription is 200-260 nucleotide/second.[14-16] Traditionally, the DNA templates used for preparing RNA are plasmids that contain an insert of interest, double – stranded DNA (dsDNA) obtained via PCR or a single – stranded DNA with an annealed oligonucleide. All of these

templates contain a duplex T7 promoter site next to the transcription start site (**Figure 1.2**). The most frequently used RNA polymerases (RNAP) for *in vitro* transcription are SP6, T3 and T7.[17] The T7 RNAP, derived from the T7 bacteriophage of *Escherichia coli* (*E. coli*), is routinely used for transcription because it is efficient.[13] The T7 RNAP associates with the T7 class III promoter and initiates transcription with GTP or modified guanosine. The resulting RNA is then traditionally purified by phenol/chloroform extraction to remove associated proteins and denaturing polyacrylamide gel electrophoresis (PAGE) to separate RNAs of different lengths. The DNA templates, enzymes, nucleotides, and incomplete transcripts are removed efficiently using this purification protocol. However, because the T7 RNAP often appends one or even several extra nucleotides to the RNA 3'-ends, the full-length RNA transcripts are usually heterogeneous.[13, 18, 19] Such micro-heterogeneous full-length transcripts are difficult to separate efficiently by denaturing PAGE separation. To eliminate 5' and 3' end heterogeneity, catalytic ribozyme fragments acting in *cis* or *trans*, such as hammerheads, can be placed at both ends of the RNA. The 5' and 3' flanking self-cleaving ribozymes will cut at a specific position during transcription and release the RNA of interest with homogeneous ends.[20-22]

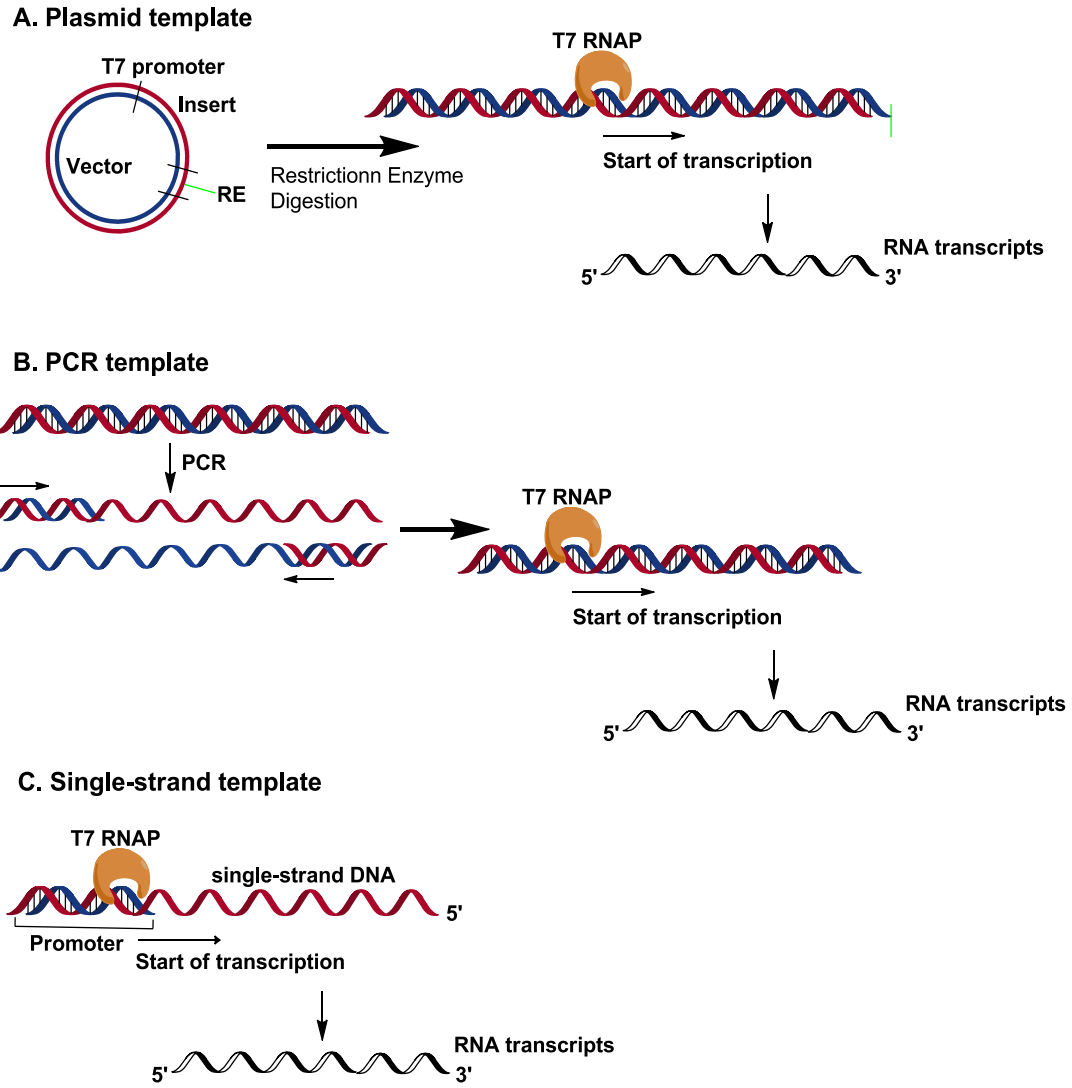


Figure 1.2. Three different types of DNA templates can be used for *in vitro* transcription. **(A)** A circular plasmid DNA with an insert coding for the RNA of interest cloned between a T7 promoter and a unique restriction enzyme (RE) site is linearized and transcribed from the promoter. **(B)** A dsDNA template (plasmid, genomic DNA, cDNA or a cloned fragment) containing a T7 promoter acts as the template in PCR with a 5'-primer that is not complementary to the template and 3'-primer that is complementary to the template. The resulting PCR-product is transcribed into RNA. **(C)** A short T7 promoter DNA oligonucleotide is annealed to an oligo that has the complimentary sequence fused to a DNA template sequence of interest. The partial dsDNA oligos can be transcribed into short RNAs.

1.3 Problems associated with traditional RNA purification methods

Clearly, these traditional RNA preparation and purification methods can be problematic. That is preparing RNA using phenol/chloroform extraction of associated

proteins, ethanol precipitation, heating in urea, denaturing PAGE separation based on size, and gel elution retrieval requires extensive denaturing-refolding cycles. However, recent studies indicate that this method generally creates heterogeneous active (native) or inactive (non-native) populations of conformers[23, 24] and the heterogeneous conformation caused by misfolding from denaturation and renaturation cycles could seriously affect RNA structures and functions.[23-25] For example, it has been demonstrated that *in vitro* denaturing and refolding process introduced severe conformational heterogeneity into the Varkud satellite (VS) ribozyme.[26] In addition, denaturing PAGE purification of RNA also often leaves water-soluble acrylamide oligomers in the desired RNA product, which cannot be completely removed. The acrylamide oligomers bound to RNA could lead to an increase in molecular weight of the RNA and catalysis of base pair exchange through groove interactions to seriously affect NMR studies of RNA structure and dynamics.[27]

The awareness of this problem is not new in the RNA field. What is new is sustained efforts to find a widely acceptable solutions to the problem of mis-folding caused by denaturing – renaturing.[26-36] To date, several different methods, including affinity-capture release and chromatographic methods using size exclusion or weak anion-exchange methods based on size or charge, have been proposed to overcome this problem.[26, 27, 30-32, 35, 37, 38] The affinity-capture release method uses high affinity tag ligands, which are immobilized on chromatography resins, to isolate the tag labeled RNA.[26, 28-32] However, the most important factor to consider for RNA affinity tags is that the bound RNA can be eluted under native conditions without co-eluting nonspecifically bound contaminants. Use of weak anion exchange chromatography

often require very high concentration of monovalent salts (500-1000 mM) for eluting high molecular weight RNAs (>400 nt), but these non-physiological high salt concentrations may still partially misfold some RNAs.[33, 34] Most chromatographic methods allow the efficient recycling of unincorporated NTPs, but they generally lack the high resolution necessary to obtain RNAs of very high purity. For example, HPLC is only appropriate for RNA of up to about 35 nt long because size exclusion requires a large size differences for separation. Furthermore, the RNAs produced from *in vitro* transcription reactions contain proteins that need to be removed by phenol/chloroform extraction before loading it on the column; otherwise the column bed may be coated with proteins and will lose its loading capacity over time. However, this phenol/chloroform extraction procedure also denatures RNA. More details on affinity-capture release methods and chromatographic methods have also been discussed in our recent publication (**Chapter 2, Figure 2.1**).[38]

Due to the mis-folding problem in denaturing PAGE purification and the limitations of existing methods, our aim is to develop new methods that could improve native folded RNA preparation and aid in the studies of RNA structure and function.

1.4 Newly developed native RNA preparation methods

Affinity-capture release is a promising RNA preparation method, but it is desirable to obtain an untagged biomolecule after it has been isolated via affinity chromatography without altering the structure of the molecule. Biotin is often used as a tag for isolation and detection because of its exceptionally tight binding to avidin

with a dissociation constant of ~10^{-15} M.[39] The complex formed between biotin and avidin is essentially irreversible under a wide variety of conditions.[40] Accordingly, biotin tags can be used to immobilize RNA to avidin solid supports and provide a means to separate desired labeled RNA from undesired unlabeled molecules. However, conventional methods that use biotin-affinity tags are limited by the difficulty of removing the affinity label from the target RNA. There are several methods of dissociation of biotinylated molecules from avidin capture probes. Firstly, biotin-avidin complex can be dissociated with 6-8 M guanidinium hydrochloride.[41] Secondly, biotin analogs with decreased affinity toward avidin (e.g., iminobiotin) can be utilized.[42] Thirdly, biotin derivatives can be coupled with chemically cleavable spacer arms.[43-46] However, none of these methods allow the release of target molecule in an unaltered form because the biotin or part of the cleavable spacer arm remains attached. In addition, these methods require harsh conditions which can denature the target RNA. In **Chapter 2**, we present a simple strategy using biotin coupled to photocleavable tags (PC tags) that circumvents many of the difficulties associated with breaking the very strong avidin-biotin interaction and, in a single step, yields natively purified RNA.[38]

Our method uses Cu catalyzed 1, 3-dipolar Huigen's cycloaddition reaction ('click' chemistry) to couple the biotin to a guanosine monophosphate with a photolabile nitrophenyl group[47] (see **Chapter 2**). The polyethylene glycol (PEG) spacer is chosen to position the biotin moiety optimally from the GMP and to increase the water solubility of the modified GMP. On the basis of a previous finding of the guanosine-initiated transcription system[48], we demonstrate that the T7 RNAP can

accept this ‘caged’ biotin guanosine monophosphate (referred to biotin-PC-GMP). RNAs with biotin-PC-GMP attached to their 5’ ends can be prepared in a straightforward manner by transcription priming.[49] During transcription with the T7 RNAP, nucleotides lacking a 5’-triphosphate cannot be incorporated into an elongating RNA chain, but they can be used to initiate a transcript, so transcription priming can produce a mixed population of modified and unmodified RNAs. To maximize analog incorporation, it is essential to include a large excess of biotin-PC-GMP over the corresponding unmodified nucleotide triphosphate (GTP). While this has the effect of increasing the fraction of population with the desired 5’ modification, it also tends to lower transcription efficiency. It is therefore generally necessary to empirically determine the appropriate modified GMP concentration to balance these factors needed for high yield of biotin-PC-GMP-RNA. Then only the RNA with biotin-PC-GMP tag binds to the avidin or streptavidin column and all the enzymes, NTPs and unlabeled RNA are washed away. Long wavelength UV (365 nm), which does not damage RNA, can then be used to release the photocleavable biotin affinity tags. Since the RNA does not go through unfolding and refolding steps during photocleavable affinity purification, the purified RNA remains in its native form. Also, the RNA transcripts that are cleaved by UV have guanosine monophosphates at their 5’-ends, which can allow further ligations to other RNA fragments.

The hypothesis that natively prepared RNA is likely catalytically more active than RNA prepared using denaturing methods that require heat-cooling and a refolding process was tested by employing minimal group II intron ribozyme splicing

model constructs (D1235 – 527 nt or D123 – 493 nt with D5 – 34 nt) from *Pylaiella littoralis* (PL).[34, 50] Our results suggest that denaturing-refolding and native RNA preparation method makes a dramatic difference for the catalytic function of the D123 or D1235 RNA ribozyme.[38] Moreover, the results are in agreement with previous observations that RNA can be kinetically trapped in misfolded states which are not easy to overcome completely even by heat annealing.[24, 26]

Chapter 2: RNA synthesized using photocleavable biotinylated nucleotides have dramatically improved catalytic efficiency[38]

2.1 Introduction

The fundamental ability of RNA to switch between structures is a great asset for signaling, yet it can also become a liability when studying complex RNAs *in vitro.* Standard methods for RNA preparation typically denature the RNA, giving rise to heterogeneous populations of conformers.[23-25] Preparation of RNA using these traditional protocols typically involves phenol chloroform extraction of associated proteins away from the RNA, RNA concentration by ethanol precipitation, heating in urea or formamide, and denaturing polyacrylamide gel electrophoresis (PAGE) fractionation. These established methods, while robust, may partially misfold the RNA and the resultant RNAs may retain only marginal activity.[23-25]

Several methods recently proposed to overcome this problem have their own limitations.[26-35, 51] These are summarized in **Figure 2.1**. The first group involves affinity capture-cleavage methods.[26, 28-32] The second group of methods uses traditional protein chromatography methods of either size-exclusion[27, 35] or weak anion exchange chromatography.[33, 34] The size-exclusion methods allow efficient recycling of unincorporated NTPs and separation of RNA oligomers. However T7 RNA polymerase (T7 RNAP) needs to be removed prior to chromatography using tedious phenol-chloroform extraction and desalting.[26, 35] The weak anion exchange method dispenses with the phenol-chloroform step, but requires very high

concentration of monovalent salts (500-1000 mM) for eluting high molecular weight RNAs (> 400 nucleotides).[33, 34]

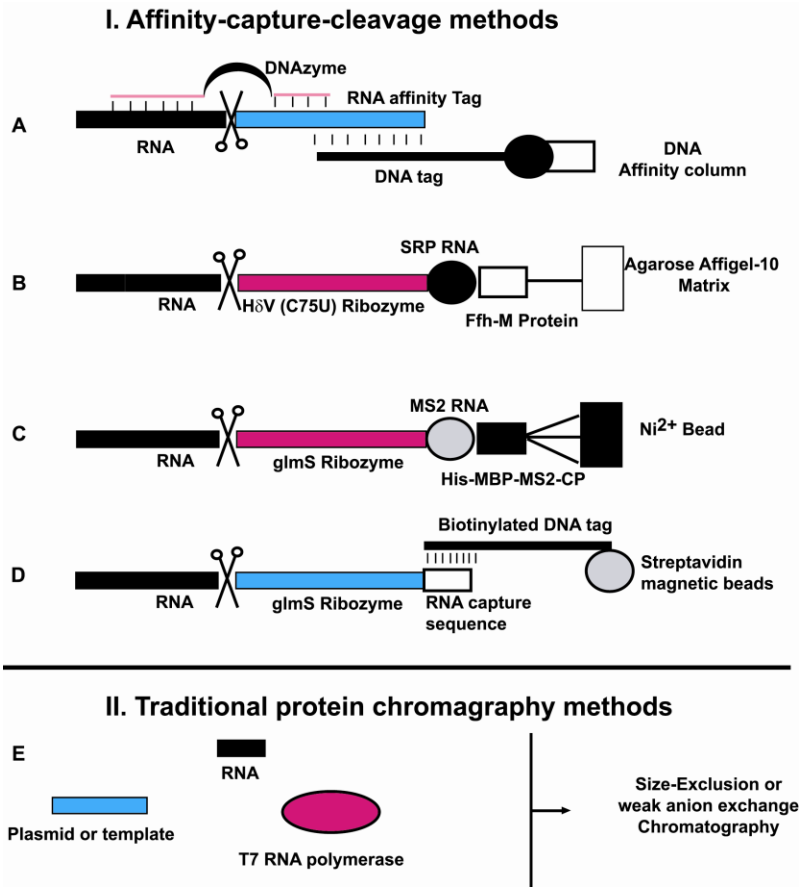
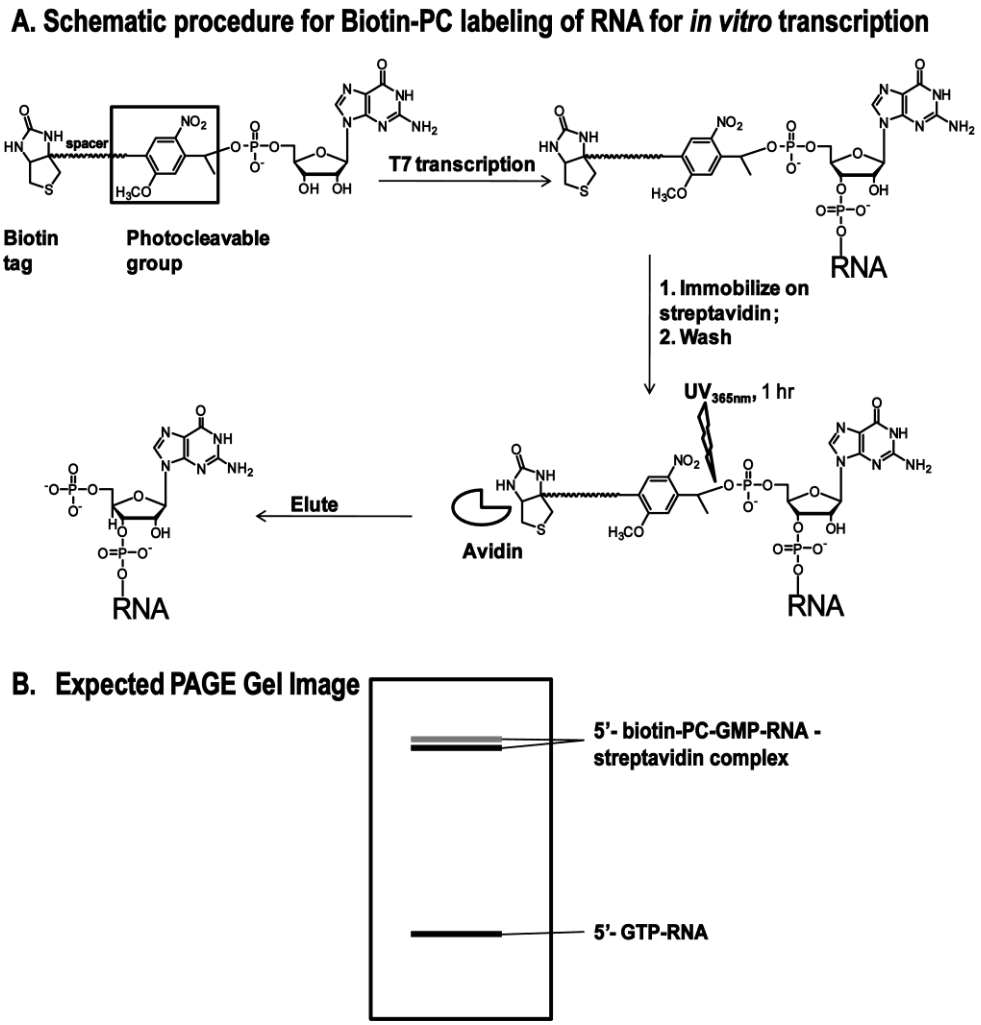


Figure 2.1. Two broad methods for native RNA purification.[51] **(I) Affinity-capture release methods.** (**A**) Native purification of RNA using self-cleaving DNAzyme affinity chromatography.[32] The DNAzyme comprises a core structure of 15-nt flanked by arms of 7-13 nt which can hybridize to the target RNA, and three-quarters of the DNA affinity tag is complementary to the target RNA.[32, 51] (**B**) Schematic of protein-RNA affinity column.[31, 51] The RNA of interest is appended to a self-cleaving hepatitis delta virus (HDV) ribozyme **(**~68-nt) and a stem loop RNA motif **(**~100-nt) from a signal recognition particle (SRP) that binds tightly to the SRP protein, the Ffh-M domain. The transcribed RNA is run over an affiGel affinity column to which the Ffh protein has been immobilized. Addition of imidazole, ~200 mM, cleaves the target RNA and enables elution of the RNA. (**C)** Schematic of protein-RNA affinity column.[29, 51] The RNA of interest is appended to a self-cleaving *glmS* Ribozyme **(**~140-nt) and a 43-nt MS2 coat protein binding RNA stem-loops upstream of the *glmS* Ribozyme. The transcribed RNA is run over the widely used Ni^{2+} affinity column to which a His-tagged form of the maltose binding protein (MBP)-MS2-CP coat fusion protein has been immobilized. Addition of 1mM Glc6NP cleaves the target RNA and enables elution of the RNA. (**D**) Schematic of an RNA affinity column.[30] The RNA of interest is immediately downstream of a self-cleaving *glmS* Ribozyme **(**~211-nt) and a 20-nt RNA sequence that is complementary a biotinylated DNA capture sequence. The transcribed RNA complexed to the biotinylated DNA capture sequence is incubated with streptavidin coated magnetic beads.

Addition of 1mM Glc6NP cleaves the target RNA and enables elution of the RNA. (**II) Chromatographic methods using size exclusion or weak anion-exchange methods based on size or charge.**

Development of new methods that facilitate the preparation of RNAs under native conditions would complement existing methods and aid in studies that investigate the impact of natively folded RNA structure on function. To this end, we have developed a new method that uses ‘click’ chemistry[52, 53] to couple biotin[43, 45, 46, 54] to a "caged" or photolabile[55] guanosine (hereafter referred to as biotin-PC-GMP) in high yield, and used this GMP analog as initiator nucleotide for T7 RNAP transcription[56] (**Scheme 2.1**).



Scheme 2.1. Preparation of native RNA using biotin-PC-GMP as initiator
(A) Schematic procedure for biotin-PC labeling of RNA during *in vitro* transcription. (**B)** Expected gel image for analysis of RNA labeling efficiency by PAGE (Two bands for biotin-PC-RNA- streptavidin complex shown are due to multivalent nature of streptavidin, i.e. one streptavidin molecule is able to bind to up to four biotinylated RNA)[57]

We demonstrate the methodology on four RNAs ranging in size from 27-nt RNA fragment[58] to a 530-nt ribozyme[50, 59](**Figure 2.2**). Importantly, we show that the biotin-PC-GMP-synthesized ribozymes rapidly cleave fluorescently labeled substrates (in < 0.5 h) whereas the same ribozymes prepared using the conventional methods and refolded are unable to cleave the substrate after 2 h. These results underscore an obvious but often overlooked fact: RNAs prepared using traditional denaturing

methods can have reduced function making conclusions based on results from such preparations precarious.

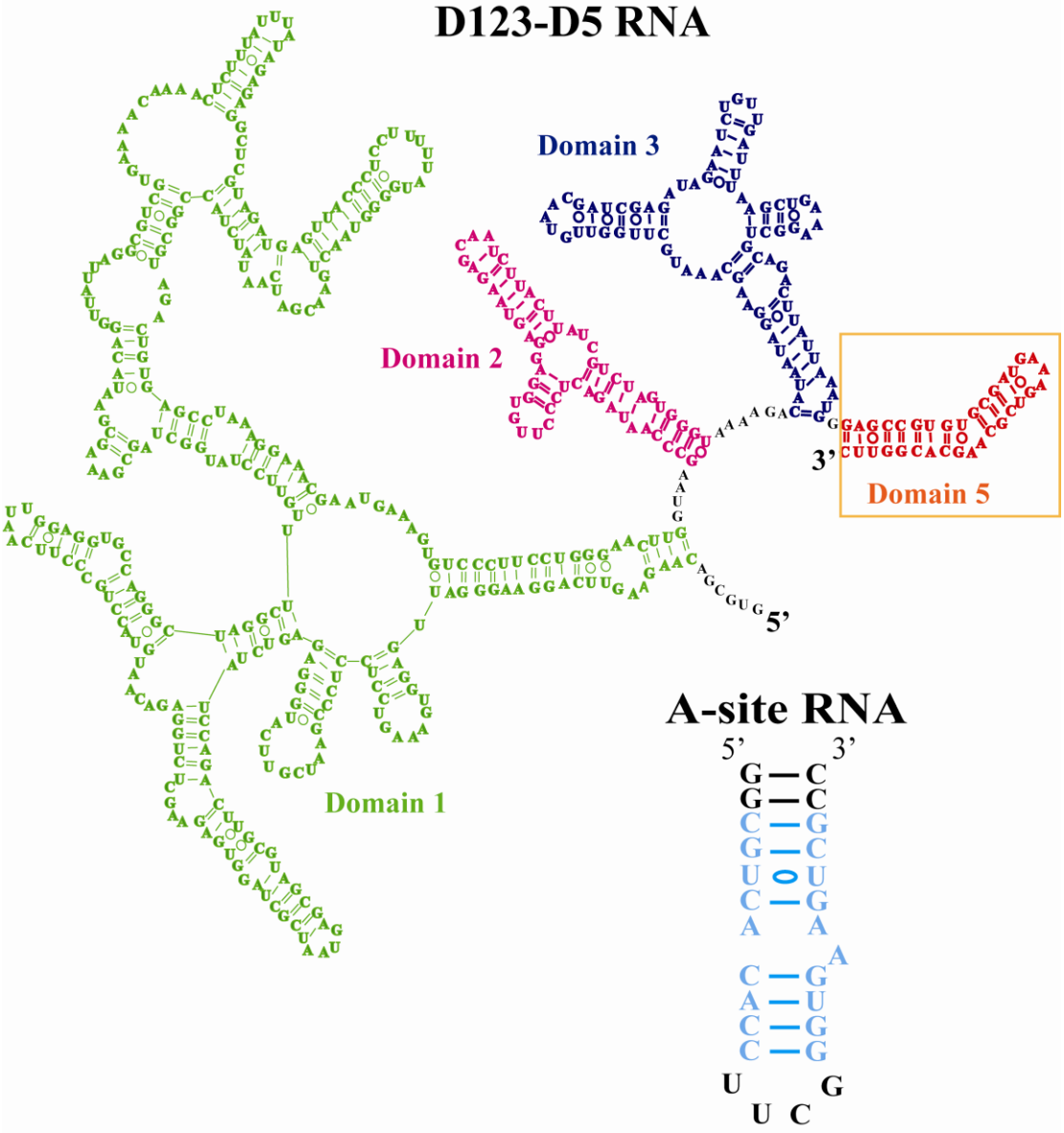


Figure 2.2. Secondary structures for the four RNAs studied. (**A**) The 494-nt D123 RNA. (**B**) The 530-nt D123-D5 RNA (36-nt D5 RNA connected in *cis* to 494-nt D123 RNA). (**C**) The 27-nt A-site RNA.

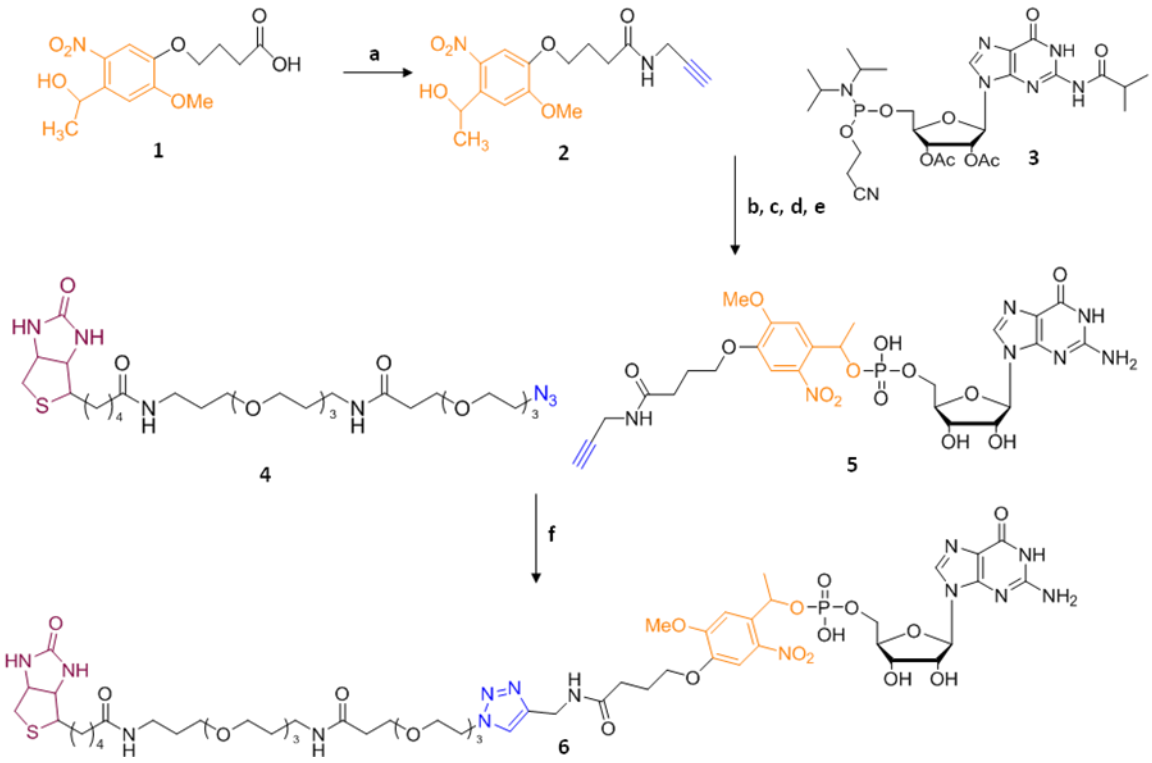
2.2 Materials and Methods

Materials

All synthesis reactions were performed in oven-dried glassware under N_2 atmosphere and covered with aluminum foil, if not stated otherwise. Dimethylformamide (DMF) was dried using the PureSolv™ solvent purification system and MeCN was distilled from CaH_2 and used immediately. All reactions were followed by TLC, unless stated otherwise. TLC analysis was performed by illumination with UV (254nm) or staining with *p*-anisaldehyde and heating. Purification of products by gel column chromatography was performed using 40-63 μm silica gel. 2',3'-Diacetyl Guanosine (n-iBu) 5'-CED phosphoramidite (**3**) was obtained from Chemgenes Corporation and Biotin-dPEG™ 3+4 – azide (**4**) was obtained from Quanta BioDesign. 4-[4-(1-hydroxyethyl)-2-methoxy-5-nitrophenoxyl] butyric acid (**1**) and propargylamine were obtained from Sigma Aldrich. All NMR spectra were referenced to TMS (tetramethyl silane) and recorded on a Bruker 400 MHz spectrometer at 25˚C.

General procedures

Chemical synthesis procedures are shown in **Scheme 2.2**.



Scheme 2.2. Synthesis of biotin-PC-GMP using 'click' chemistry
Reagents and conditions: (a) HOBt, HBTU, Et3N, DMF (s); (b) ETT, MeCN (s); (c) I_2 (THF/pyridine/H_2O 7:2:1); (d) $Na_2S_2O_3$; (e) 28-30% NH_3 H_2O; (f) $CuSO_4$, isoascorbic acid, THF/water (s). The Photocleavage group is colored orange and the biotin group is colored purple.

2.2.1 Synthesis of Photocleavable (PC) alkyne

To a solution of 4-[4-(1-hydroxyethyl)-2-methoxy-5-nitrophenoxyl] butyric acid (**1**) (300 mg, 1 mmol., 1 eq) in dry DMF (6 mL) was added distilled propargylamine (0.26 mL, 4 mmol., 4 eq), HBTU (758 mg, 2 mmol, 2 eq), HOBt (67.5 mg, 0.5 mmol, 0.5 eq) and Et_3N (280 µL, 2 mmol, 2 eq). The mixture was stirred under nitrogen, at room temperature overnight. After completion of the reaction (as judged by TLC analysis), the solvent was removed under vacuum and the mixture was purified by silica gel column chromatography with a step gradient of MeOH (0.5-2.5%) in $CHCl_3$. The solvent was evaporated, and the residue was dried under vacuum to yield the

product **2** as a yellow solid (235 mg, 70%): ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.49 (s, 1H), 7.28 (s, 1H), 6.45 (s, 1H), 6.45 (br, 1H), 5.50 (q, J = 6.4 Hz, 1H), 4.05 (t, J = 4.0 Hz, 2H), 4.00 (t, J = 2.4 Hz, 2H), 3.93 (s, 3H), 3.27 (br, 1H), 2.42 (t, J = 7.2 Hz, 2H), 2.20 (d, J = 2.4 Hz, 1H), 2.17 (q, J = 6.8 Hz, 2H), 1.49 (d, J = 6.4, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 172.6, 154.3, 147.0, 139.7, 138.0, 109.4, 109.1, 80.0, 71.9, 68.8, 65.9, 56.7, 33.0, 29.5, 25.1, 24.9. ESI-MS：$[M+H]^+$ calcd 337.1, found 337.1. (See **Appendix 1**)

2.2.2 Synthesis of PC alkyne GMP

a) Mixture of compound **(2)** (198 mg, 0.59 mmol, 1.5 eq) and ETT (153 mg, 1.18 mmol, 3 eq) was dried under high-vacuum. Then, a solution of 2′,3′-Diacetyl Guanosine (n-iBu) CED phosphoramide (**3**) (250 mg, 0.39 mmol, 1 eq) in anhydrous CH_3CN (3 mL) was added dropwise, under nitrogen. The reaction mixture was stirred at room temperature overnight. After completion of the reaction, the phosphite derivative was obtained as yellow oil after evaporation of the solvent and was directly used in the next step without purification.

b) The crude phosphite derivative (2 eq) obtained from step **a)** was dissolved in the "oxidation" solvent (27 mL, 0.02 M I_2 in 7:2:1 of THF:pyridine: H_2O, 3 eq). After stirring at room temperature for 10 min, the reaction was quenched with $Na_2S_2O_3$ (86 mg, 0.54 mmol., 3 eq) in H_2O (1 mL). Stirring was continued at room temperature for another 10 min during which the mixture turned from brown to a yellow color. The oxidized product (phosphate derivative) was obtained as yellow oil after evaporation of the solvent and was directly used in the next step (deprotection).

c) A solution of 28-30% aqueous ammonia (35 mL) was added to the crude product obtained from step **b**. The reaction mixture was stirred at 65^{o}C for 8 h. After 8 h, NH_3 was evaporated at room temperature for 3 h. The reaction mixture was subjected to HPLC purification.

The mixture from step c was purified using a Varian HPLC fitted with a reverse phase C18 column (4.6 mm×250 mm, 5μm particle size). The reaction mixture components were eluted using linear gradients of solvent A (water) and B (100% acetonitrile) as follow: 5% to 15% B for 15 min, 15% to 40% B for 5 min, 40% to 90% for 4 min, 90% B hold for 4 min, 90% to 5% B for 4 min and 90% B washing for 3 min. The eluent was monitored using UV-Vis absorption at 254 nm. Data was collected and analyzed by galaxy software. Compound **5** eluted at 18-20 min.

The collected fractions were lyophilized to yield the product **5** (66.39 mg, overall yield of 25% from compound 2): ^{1}H NMR (400 MHz, D_2O) δ 7.76 (s, 1H), 7.37 (d, J = 24.0 Hz, 1H), 7.25 (d, J = 9.6 Hz, 1H), 5.87 (m, 1H), 5.70 (s, 1H), 4.46-4.30 (m, 2H), 4.14 (s, 1H), 4.03-3.93 (m, 8H), 2.56 (d, J = 2.0 Hz, 1H), 2.40-2.39 (m, 2H), 2.07-2.05 (m, J = 5.2 Hz, 2H), 1.97 (s, 1H), 1.52-1.50 (m, 3H); ^{13}C NMR (100 MHz, D_2O) δ 176.02, 175.98, 158.87, 153.94, 153.87, 153.78, 151.65, 151.34, 146.40, 146.35, 138.46, 138.39, 137.71, 137.47, 135.40, 135.34, 109.50, 109.23, 108.82, 108.41, 88.34, 87.51, 83.56, 83.18, 79.92, 74.45, 74.05, 72.06, 70.93, 70.52, 70.34, 70.12, 68.51, 65.87, 56.50, 47.03, 32.53, 32.50, 29.12, 24.81, 24.11, 24.06, 8.60; ^{31}P NMR (162 MHz, D_2O) δ 0.14, 0.46; ESI-MS: $[M+H]^+$ calcd 682.2, found 682.2. (See **Appendix 2**)

2.2.3 Synthesis of biotin-PC- GMP

To a solution of PC alkyne GMP **(5)** (46 mg, 0.0675 mmol, 1 eq) in THF/water (1:1, v/v, 3.5 mL), the following were added: biotin-dPEG^{TM} 3+4 – azide (48.6 mg, 0.0675 mmol, 1 eq), $CuSO_4$ $5H_2O$ (3.37 mg, 0.0135 mmol, 0.2 eq), and isoascorbic acid (23.8 mg, 0.135 mmol, 2 eq). The reaction mixture was stirred at room temperature for 5 h, and then subjected to HPLC purification.

The mixture that was obtained after the azide alkyne Huisgen cycloaddition ('click reaction') was purified using HPLC C18 reverse phase column. Reaction mixture components were eluted using linear gradient of solvent A (water) and B (100% acetonitrile) as follow: 5% to 20% B for 8 min, 20% to 40% B for 12 min, 40% to 90% B for 2 min, 90% B hold for 2 min, 90% to 5% B for 2 min and 5% B washing for 2 min. The eluent was monitored using UV-Vis absorption at 254 nm. Data was collected and analyzed by galaxy software. Compound **6** eluted at 16-17 min.

The collected fractions were lyophilized to yield the product **6** (80.32 mg, 85%). ^{1}H NMR (400 MHz, D_2O) δ 7.93 (s, 1H), 7.81 (br, 1H), 7.41-7.26 (m, 4H), 5.97-5.88 (m, 1H), 5.73 (t, J = 4.6 Hz, 1H), 4.61-4.33 (m, 7H), 4.15-3.89 (m, 9H), 3.77-3.54 (m, 27H), 3.31-3.21 (m, 6H), 2.97 (dd, J = 13.0, 5.0 Hz, 1H), 2.78 (d, J = 13.0 Hz, 1H), 2.52-2.45 (m, 4H), 2.24 (t, J = 7.2 Hz, 2H), 2.13-2.08 (m, 2H), 1.94 (s, 2H), 1.82-1.52 (m, 11H), 1.39-1.36 (m, 2H), 1.31 (t, J = 7.32 Hz, 2H); ^{13}C NMR (100 MHz, D_2O) δ 176.86, 175.90, 175.87, 174.14, 165.61, 158.90, 154.00, 153.87, 153.76, 146.50, 146.44, 138.58, 138.53, 135.41, 124.55, 109.57, 109.32, 108.79, 108.39, 88.35, 87.54, 83.33, 74.55, 74.14, 70.39, 69.98, 69.94, 69.88, 69.86, 69.77, 69.12, 68.84, 68.73, 68.51, 67.71, 62.45, 60.61, 56.59, 56.57, 55.78, 50.38, 47.07, 40.12, 36.77, 36.72,

36.46, 35.91, 34.79, 32.49, 28.69, 28.36, 28.10, 25.59, 25.00, 24.22, 8.66; ^{31}P NMR (162 MHz, D_2O) δ -0.03, -0.35; ESI-MS: $[M+H]^+$ calcd 1401.57, found 1401.53. (See **Appendix 3**)

2.2.4 Optimization of ribosomal A-site RNA and D5 RNA transcription

RNAs were synthesized by *in vitro* transcription with T7 RNAP from a synthetic DNA template (Integrated DNA Technologies)[56]. The T7 RNAP was expressed in *Escherichia coli* BL21 (DE3) and purified on a Ni-chelating Sepharose column (Pharmacia). The T7 RNAP promoter sequence was 5′ - CTA ATA CGA CTC ACT ATA G - 3′. The template strand of ribosomal A-site[58] was 5′- GmGmC GAC TTC ACC CGA AGG TGT GAC GCC TAT AGT GAG TCG TAT TAG- 3′ and the template strand of the D5 RNA[50, 59] was 5′- g AAC CGT ACG TGC GAC TTT CAT CGC ATA CGG CTC c TAT AGT GAG TCG TAT TAG -3′. Two terminal 2′-O-methyl modifications in the template strand indicated by "m" were introduced to substantially reduce the amount of transcripts with extra nucleotides at the 3′-end.[60] The lower case letters represent additional nucleotides introduced to improve transcription yield.[61] All the purchased DNA strands were purified by denaturing PAGE. The transcription conditions for each RNA were optimized by varying the [A/C/UTPs] and ratio of [GTP]/[biotin-PC-GMP] in transcription buffer B [40 mM Tris-HCl (pH 8.1), 1 mM spermidine, 10 mM dithiothreitol (DTT), 0.01% Triton X-100, 80 mg/mL PEG 8000, 0.2 units of RNase inhibitor (New England Biolabs), 0.2 units of inorganic pyrophosphatase (New England Biolabs)] containing 15 mM Mg^{2+}, 300 nM of each DNA strand, and 1 μL of 4 mg/mL T7 RNAP per 20 μL of

transcription volume. The reactions were incubated for 4 h at 37˚C. After 4 h, 0.2 units of Turbo DNase (Ambion) were added and the samples were incubated for another 15 min. The RNA transcripts were immediately analyzed on a 12% denaturing PAGE/8M urea/1 ×TBE gel and stained with ethidium bromide (EtBr).

2.2.5 Streptavidin-biotin-PC-D5-RNA gel shift experiment

After the removal of unused biotin-PC-GMP, the purified biotin-PC-D5-RNA was incubated with three-fold molar excess of streptavidin (Promega) (with respect to RNA) at 37 ℃ for 1 h. After incubation, the sample was split into two equal parts, one was irradiated with UV light at 365 nm for 1 h and the other half was left on the benchtop for 1 h. Each set was loaded into a tube containing a 0.5 mL ultra spin filter with a molecular weight cut-off (MWCO) of 10 kDa (Millipore-Amicon), followed by washing with 500 μL of double-distilled water. This wash step was repeated three times; the flow-throughs were combined, concentrated, and analyzed through a 12% denaturing PAGE.

2.2.6 Purification of biotinylated RNA with Neutravidin-agarose resin and photocleavage

A 20 μL transcription mixture was loaded into a tube containing a 0.5 mL ultra spin filter with a MWCO of 3 kDa (Millipore-Amicon), followed by washing with 500 μL of potassium phosphate buffer (0.1 M phosphate, 0.15 M potassium chloride, pH

6.4). This wash step was repeated three times to completely remove unincorporated free biotin-PC-GMP from the transcription reaction.

The transcription mixture was added to a suspension of NeutrAvidin Agarose Resin solution (Thermo Scientific, IL) and pre-equilibrated in potassium phosphate buffer (0.1 M phosphate, 0.15 M potassium chloride, pH 6.4). After 1 h incubation, the column was washed with 8-10 column volumes of potassium phosphate buffer to remove all unbound molecules. The bound biotin-PC-RNA on the resin was subsequently irradiated with UV light at 365 nm for 1 h. After irradiation, the sample in the column was eluted with 8-10 column volumes of potassium phosphate buffer. All the fractions were pooled, concentrated and analyzed by UV absorption spectroscopy at 260 nm. The pooled fractions were then separated through 8M urea denaturing PAGE.

2.2.7 Native D123 (GTP-D123) RNA sample preparation

D123 RNA was transcribed from template DNA produced by PCR amplification from the original plasmid containing domains 1-3 (D1 to D3).[50] D123 template was amplified using an upstream primer containing the T7 RNAP promoter sequence and the first 19 nt from the 5′-end of the D1 sequence and a downstream primer containing 19 nt of the 3′-end of the intron domain 3. The RNA was transcribed in transcription buffer B containing 3.33 mM each ATP, CTP and UTP, 2 mM GTP, 8 mM biotin-PC-GMP, 24 mM Mg^{2+}, 1.5 μg of gel purified PCR template, and 10 μL of T7 RNAP (4 mg/mL) per 200 μL of reaction volume. After 4 h, 0.2 units of Turbo DNase (Ambion) was added and the samples were incubated for another 15 min. Then the

D123 RNA was affinity purified as described above. UV absorbance at 260 nm was used to calculate the concentration based on the extinction coefficients of the individual nucleotides.

2.2.8 Native D1235 (GTP-D1235) RNA sample preparation

D1235 RNA was transcribed from template DNA produced by PCR amplification from the original plasmid containing domains 1, 2, 3, and 5 as described.[34] The RNA was transcribed and affinity purified as described above. RNAs produced from PCR transcripts were determined to be identical to those prepared from plasmids and the two methods gave similar RNA yields.

2.2.9 Gel purified D5 RNA sample preparation

The D5 RNA was synthesized as described above in transcription buffer B containing 10 mM total NTPs and 15 mM Mg^{2+}, and was purified by 12% denaturing PAGE. The product band was detected by brief UV-shadowing, excised and electro-eluted in an Elutrap electro-separation system (Schleicher and Schuell). The purified D5 RNA was precipitated with three volumes of absolute ethanol and 0.3 M sodium acetate pH 5.2. The RNA pellet was then dissolved in water and dialyzed in a Biodialyzer (Nestgroup) with a 500 Da MWCO membrane (Nestgroup). Each dialysis step was performed for 12 h against dialysis buffer A (100 mM potassium phosphate at pH 6.4, 10 mM EDTA, 0.5 M KCl), B (100 mM potassium phosphate at pH 6.4, 1 mM EDTA, 0.1 M KCl), and C (100 mM potassium phosphate at pH 6.4, 0.1 mM

EDTA, 0.1 M M KCl), and finally against two changes of double-distilled H_2O. Note that the denaturing preparation method just described may take about 2-3 days which contrasts with our native preparation method described earlier that is completed 4-6 h after the transcription reaction. After dialysis, the RNA was lyophilized.

2.2.10 Gel Purified D123 RNA (GTP-D123 RNA) and D1235 RNA (GTP-D1235 RNA) sample preparation

D123 RNA was transcribed from DNA produced by PCR amplification as described above using transcription buffer B containing 15 mM total NTPs and 24 mM Mg^{2+}.[50] Similarly D1235 RNA was transcribed from DNA produced by PCR amplification as described above using transcription buffer B containing 10 mM total NTPs and 18 mM Mg^{2+}. Each resulting RNA was gel-purified as described above for D5 RNA, except that a denaturing 6% gel was used.

2.2.11 Anion exchange purified D123 RNA (GTP-D123 RNA) and D1235 RNA (GTP-D1235 RNA) sample preparation.

The *in vitro* transcription reaction mixture was incubated with 0.2 U of Turbo DNase at 37 ℃ for 0.5 h. A 1 mL reaction mixture was pelleted to remove inorganic pyrophosphate and the supernatant was loaded into a 1 mL loop of an AKTA FPLC purifier system (GE Healthcare) connected to a 5 mL HiTrap Q column (GE Healthcare). The column was equilibrated with 4 column volumes of buffer A1 (20 mM KH_2PO_4, 25 mM KCl, 100 M EDTA, pH 6.50). The anion exchange

chromatography was performed with buffer A2 (20 mM KH_2PO_4, 2 M KCl, 100 M EDTA, pH 6.50), generating a gradient by simultaneously decreasing the concentration of buffer A1 and increasing the concentration of buffer A2. The separations were collected as 5 mL fractions with the following gradients of elution: 0-30 mL for injection and loop washing (0% A2), 30-80 mL for washing off rNTPs and proteins (0-17% A2 at 2 mL/min), 80-110 mL for washing off small abortive transcripts (17-27% A2 at 2 mL/min), 110-175 mL for eluting RNA off the column (27-60% A2 at 2 mL/min). The column was further washed with 50 mL of buffer A2 for the next cycle of purification. Fractions were analyzed by 8% PAGE containing 8M urea.

2.2.12 Labeling E1E2 substrate RNAs with Alexa 647 fluorophore

For the fluorescence studies, Alexa 647 carboxylic acid succinimidyl ester (Invitrogen/Molecular Probes Inc.) was site-specifically incorporated at the 5′-end of E1E2 RNA, which comprises the last 16 nt of the 5′-exon (E1) and the first 6 nt of the 3′-exon (E2), via a primary amine C-6 linker (A647-C6 linker-5′-GAC UGU UUA UUA AAA A↕CAC AGG-3′ (Dharmacon Inc); the cleavage site is indicated by ↕).[50] This labeled RNA (22 nt) is hereafter referred to as A647-E1E2 RNA. The RNA was labeled using 10-15 molar excess of the dye in 0.1 M sodium tetraborate (pH 8.5) at room temperature for 12 h with gentle mixing. The resulting A647-E1E2 RNA was purified using a denaturing 15% gel prior to electroelution. The eluted RNA was ethanol-precipitated and dissolved in 500 μL of MOPS buffer (40 mM MOPS at pH 7.5, 100 mM KCl), and dialyzed extensively against buffers A, B, and C at pH 6.0 as

described above, and several times in double-distilled H_2O before lyophilizing and dissolving in MOPS storage buffer (40 mM MOPS at pH 7.5, 100 mM KCl).

2.2.13 RNA refolding protocol

Previous work on group II introns indicated the importance of properly folding D123 and D1235 RNAs to obtain efficient catalytic cleavage as described above.[34, 50] For each model system, the following RNA refolding protocol was used as described previously[34, 50]: Each component RNA in buffer D (40 mM MOPS-K, pH 7.5, and 100 mM KCl) was heated to 90 ℃ for 1 min, allowed to cool to 42 ℃ for 0.5 h, prefolded for various times (0, 5, 10, 20, 30, 60, 120 min) in 1M KCl and 50 mM $MgCl_2$, and then the separately folded RNAs were combined to initiate the SER reaction.[34, 50] We found that 2 h was necessary for prefolding and so all subsequent experiments were performed using 2 h of folding. Another set of SER reactions were carried out without this elaborate heat denaturation and refolding protocol. Instead each component RNA in buffer D was incubated at 42 ℃ for 0.5 h, and then all the RNA components were combined in high salt buffer (40 mM MOPS-K, pH 7.5, 1 M KCl, and 50 mM $MgCl_2$) to initiate the spliced extron reopening (SER) reaction.

2.2.14 Cleavage of A647-E1E2 RNA substrate using gel, anion exchange or affinity purified D123 RNAs monitored by an in-gel shift experiment

The spliced exon reopening reaction (SER)[50, 62] was carried out at 42˚C with D123 RNA purified by three different methods: denaturing gel, anion exchange, and

biotin-PC purified. The reaction used 1.46 μM D123, 2 μM D5, and 16 nM A647-E1E2 RNA in 50 mM $MgCl_2$, 40 mM MOPS-K (pH 7.5) and 1M KCl. For one set of experiments, GTP-D123, anion exchange purified GTP-D123, and biotin-PC-GMP-D123 RNAs were prefolded using the RNA refolding protocol described above (with heating and cooling) in 50 mM $MgCl_2$ at 42˚C for 2 h before adding D5 and substrate RNAs to initiate the SER reaction. For another set of experiments, GTP-D123, anion exchange purified GTP-D123, and biotin-PC-GMP-D123 RNAs were not prefolded (i.e. no heating and cooling), but allowed to incubate at 42˚C for 30 min before adding D5 and substrate RNAs to initiate the SER reaction. In either case, each reaction was incubated at 42˚C for various times (0, 2, 4, 8, 30, 60, 120, 240 min), and 5 μL of the reaction mixture was aliquoted into a stop buffer (50% glycerol and 100 mM EDTA), and loaded onto a denaturing 12% polyacrylamide gel. The Alexa-Fluor 647 RNA band was then quantified using a fluorescence scanner (Molecular Dynamics Storm 860) equipped with 635- and 450- nm diode pumped solid-state lasers.

2.2.15 Cleavage of A647-E1E2 RNA substrate using gel, anion exchange or affinity purified D1235 RNAs monitored by an in-gel shift experiment

The SER reactions[50, 62] were carried out at 42˚C with gel, anion exchange, or biotin-PC purified D1235 (1.5 μM) using 16 nM A647-E1E2 RNA substrate in buffer D. The A647-E1E2 RNA substrate was heated at 90˚C for 2 min separately, and incubated at 42˚C in a water bath for 30 min.[50] Before adding the substrate to initiate the SER reaction, GTP-D1235 and biotin-PC-GMP-D1235 RNAs were either

prefolded using the RNA refolding protocol or not refolded as described above for D123. Again the reaction was incubated at 42˚C for various times (0, 2, 4, 8, 30, 60, 120, 240 min), 5 μL of the reaction mixture was aliquoted into a stop buffer (50% glycerol and 100 mM EDTA) and loaded onto a denaturing 12% polyacrylamide gel. The Alexa-Fluor 647 RNA band was then quantified using a fluorescence scanner as described above. The fraction of the A647-E1E2 RNA substrate cleaved at each time point was quantified and fitted to an exponential function using SigmaPlot 11.0.

2.2.16 D123 RNA analyzed on native gel

The gel purified, anion exchange purified and biotin-PC purified D123 were analyzed on 6 % native PAGE with 0.5×TME (12.5 mM Tris/MOPs, 0.1 mM EDTA) as the running buffer at 4˚C. To ensure the accurate determination of the RNA concentrations loaded on the gel, a small amount of sample was completely hydrolyzed in 1.0 M sodium hydroxide at 65℃ for 2 hours, and the UV absorbance was used to calculate the concentration. To rule out overloading on the gel, two concentrations of RNAs were used: 1 μM and 3 μM for each RNA.

2.3 Results and Discussion

2.3.1 General Concept

The traditional method of synthesizing RNAs *in vitro* typically involves denaturation during purification followed by a refolding step. This method implicitly assumes these RNAs are properly folded and that they retain their function. While

several methods have been developed recently to purify RNAs without denaturation, very few studies have examined how the manner of RNA preparation affects both the folding and the function of the RNA. To show that the manner of RNA preparation can remarkably affect function such as catalysis, we proceeded to develop a simple strategy to make biotinylated GMP **Scheme 2.1** which can be used to prime RNA synthesis. This strategy also allows for complete removal of the biotin moiety without leaving any non-natural chemical entity appended to the RNA. Importantly we show for the first time that large complex RNAs that are prepared by denaturing methods, and subsequently refolded, are catalytically less active than those prepared using native methods.

2.3.2 Synthesis of biotin-PC-GMP

Rationale: Results from previous studies have indicated that nucleoside monophosphates modified with an affinity label such as biotin can be used to prime RNA synthesis.[57, 63-65] Once made, the RNA has to be cleaved from the biotin label. None of these previous methods provide a means to remove the biotin tag without leaving any trace. Therefore we have devised a facile strategy that allows efficient biotinylation of RNA, provides a traceless means to remove the biotin after the purification, and allows preparation of natively folded RNAs.

The synthesis of the biotin-PC-GMP **6** is depicted in **Scheme 2.2**. The photosensitive precursor **1** was coupled with propargylamine to produce the PC-alkyne **2**. Then compound **2** and protected GMP **3** were reacted to produce an intermediate phosphite, and oxidation of the resulting phosphite yielded the phosphate.

The PC-alkyne GMP **5** was then coupled to biotin-dPEG^{TM} azide **4** through azide alkyne Huisgen cycloaddition 'click' chemistry using Cu(I) catalyst,[34, 53] and the final product biotin-PC-GMP **6** was obtained in 85% overall yield starting from **5**. The polyethylene glycol (PEG) spacer was chosen to increase the water solubility of biotin-PC-GMP and also to position the RNA further away from the biotin as the RNA could interfere with the biotin-neutravidin interactions during the immobilization step.

2.3.3 T7 RNAP accepts biotin-PC-GMP

Biotin-PC-GMP **6** was designed for direct incorporation into RNA by *in vitro* transcription catalyzed by T7 RNAP. To demonstrate that T7 RNAP accepts biotin-PC-GMP, we transcribed four different RNAs of varying size: a 27-nt A-site RNA fragment[56] as well as three group II intron-derived RNAs[34, 50, 59], i.e. a 36-nt D5 RNA, a 493-nt D123 RNA comprising domains 1, 2, 3, and a 530-nt D1235 RNA comprising domains 1, 2, 3, and 5 (**Figure 2.2**).

Visualization of the reaction product by 12% denaturing PAGE demonstrated that the initiator GMP was incorporated into the 36-nt D5 RNA. Lane 1 (**Figure 2.3A**) shows the D5 RNA synthesis without **6,** and Lane 2 (**Figure 2.3A**) shows the transcription with 4 mM of initiator **6**, demonstrating that T7 RNAP is able to initiate transcripts with the photolabile biotinylated nucleotide monophosphate (**Figure 2.3A**: Lanes 2) to make the photolabeled D5 RNA. A maximum incorporation of ~17% was achieved using 4 mM of the initiator **6** and 1 mM GTP. Interaction of this photolabeled D5 RNA with the streptavidin column and its photocleavage were

evaluated by mixing streptavidin agarose beads with the RNA, separating the beads from solution by filtration, and irradiating the beads with UV light at 365 nm. Lane 3 (**Figure 2.3A**) demonstrates that the wash from the streptavidin column is identical to 5′-GTP-RNA loaded on the column (**Figure 2.3A**: Lane 1). Lane 4 indicates that the 5′-biotin-PC-GMP-RNA is retained on the column (cf. Lane 2) and only the unlabeled 5′-GTP-RNA is washed off the column (**Figure 2.3A**: Lanes 4). The 5′-biotin-PC-GMP-RNA bound to the column (**Figure 2.3A**: Lanes 4) was released in the wash following photocleavage (**Figure 2.3A**: Lane 6). As expected, 5′-GTP-RNA was not present in the wash following photocleavage (**Figure 2.3A**: Lane 5). These results also show that photocleavage does not degrade the RNA (**Figure 2.3A**: Lanes 6; **Figure 2.3B**: Lane 4).

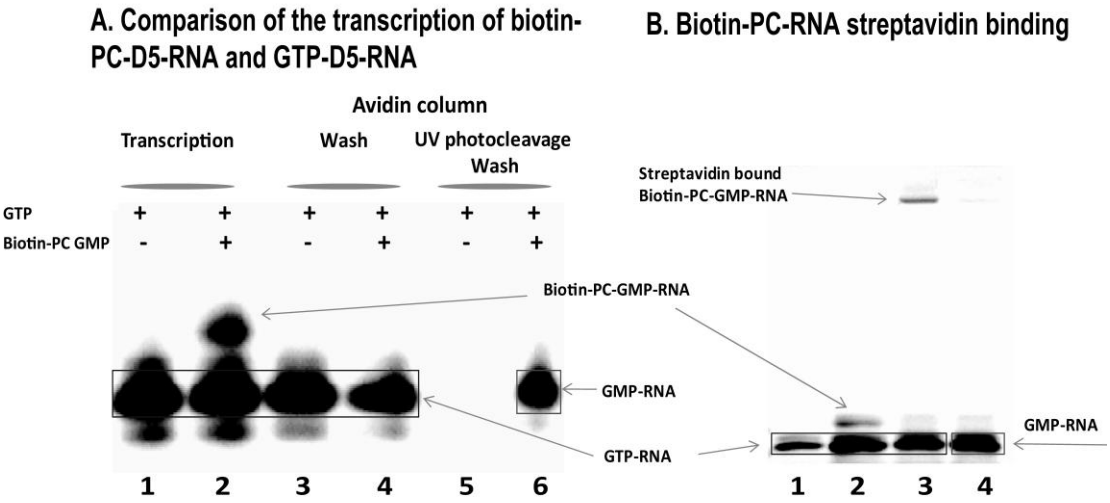


Figure 2.3. Denaturing PAGE assay for biotin-PC modified D5 RNA. (A) Transcription of D5 RNA with (+) or without biotin-PC-GMP (-) followed by purification of the D5 RNA (GTP-D5 or biotin-PC-GMP-D5 RNA) with an avidin column after photocleavage. Lane 1: Transcription of D5 with unmodified GTP (I); Lane 2: Transcription of D5 with biotin-PC-GMP/GTP mixture (II); Lane 3: Streptavidin column flow-through of unlabeled GTP-D5 RNA; Lane 4: Streptavidin flow-through of biotin-PC labeled D5 RNA. As expected unlabeled GTP-D5 is washed off the column. Lane 5: Elution of the RNA on the streptavidin column following photo cleavage of unlabeled D5 RNA. As expected there is no unmodified D5 RNA left on the streptavidin column following the initial wash shown in lane Lane 3. Lane 6: Elution of the RNA on the streptavidin column following photo cleavage of biotin-PC labeled D5 RNA. Note that as expected photo cleavage releases biotin-PC modified D5 RNA

as GMP-D5 RNA that now runs at almost the same position as the unmodified GTP-D5 RNA. The RNAs are depicted as: GTP transcribed RNA (I), biotin-PC transcribed RNA (II). **(B)** RNA biotinylation assay with streptavidin. Lane 1: Unlabeled D5 RNA; Lane 2: Biotin-PC labeled D5 RNA; Lane 3: Biotin-PC labeled D5 RNA incubated with excess streptavidin (three times number of moles to RNA) for 1 h. Lane 4: GMP-RNA, flow-through using 10K MWCO spin filter after photocleavage of streptavidin-biotin-PC-RNA

To rule out any subtle UV damage effects, we compared the FPLC UV chromatogram of a mixture of all four nucleotide triphosphates that were either irradiated with UV light at 365 nm or left on the desktop for 2 h. The FPLC traces for the irradiated and non-irradiated nucleotide mixture are completely superimposable with identical retention times, lineshapes, and intensity (**Figure 2.4A**). As a second test to rule out UV induced cleavage, we compared the mass spectrum of an RNA tetramer that was either irradiated with UV light or left on the desktop for 2 h; no change in the mass of the tetramer was observed (**Figure 2.4B and 2.4C**).

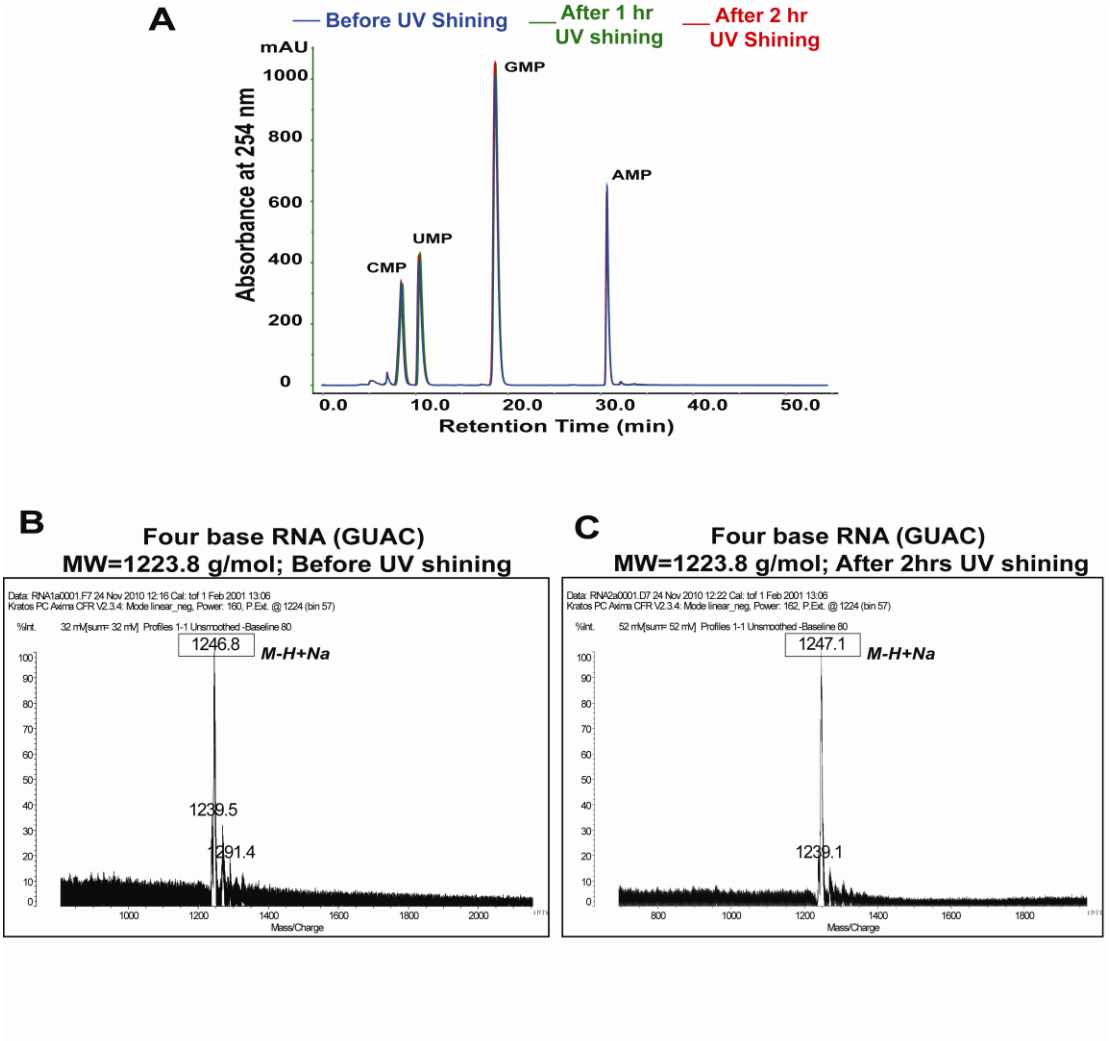


Figure 2.4. No UV-induced RNA or NMPs damage through irradiation at 365 nm for 1-2 hrs. **(A)** HPLC traces of an NMP mixture (AMP/GMP/UMP/CMP) using the gradient MeCN%. NMPs mixture (AMP/GMP/UMP/CMP) was divided into three equal amounts. First portion was loaded on HPLC without UV shining (shown in blue); Second portion was irradiated by UV_{365nm} for 1 hr before being loaded on HPLC (shown in green); Third portion was shined by UV_{365nm} for 2 hrs before loaded on HPLC (shown in red). Three traces were exactly overlap, which indicated that there was no mutation with UV shining. **(B)** Linear negative mode MALDI-MS for four bases RNA (GUAC) before and after 2 hrs UV_{365} irradiation. Matrix: CHCA (α-cyano-4-hydroxycinnamic acid); $[M-H]^-$ calcd 1223.8; before UV found $[M-H+Na]^-$ at 1246.8; after UV found $[M-H+Na]^-$ at 1247.1

To further ascertain that the D5 RNA was indeed labeled with biotin-PC, the fractions were complexed to streptavidin and analyzed by denaturing PAGE (**Figure 2.3B**). Given the strong affinity between biotin and avidin complexes, one of the strongest known in nature, (K_d of ~10^{-15} M), denaturing PAGE is often used to

visualize RNA-biotin-Avidin complexes.[57, 63, 64] As expected, only the 5′-biotin-PC labeled RNA underwent a gel shift after incubation with streptavidin (**Figure 2.3B**: compare lanes 2 and 3). After photocleavage of these streptavidin-biotin-PC-GMP D5 RNA complexes (irradiated for 1 h using UV light at 365 nm), the band migrated at almost the same position as the unmodified 5′-GTP-D5 RNA (**Figure 2.3B**, Lane 4).

2.3.4 Transcription optimization of ribosomal A-site RNA and D5 RNA using unmodified GTP and biotin-PC-GMP

As the modified GMP competes with unmodified GTP for transcription initiation, it is important to optimize the different concentrations of the three nucleotides ATP, CTP and UTP, as well as the different ratios of biotin-PC-GMP **6** to unmodified GTP in order to increase the transcription efficiency of biotin-PC-GMP **6.** To quantify the incorporation efficiency of **6** by PAGE, a 27-nt A-site RNA and a 36-nt D5 RNA were used, as we had earlier observed that the electrophoretic mobility of the 5′-biotin-PC-RNA transcripts differs from that of the unmodified 5′-GTP-RNA transcripts of these RNAs. Therefore, we reasoned that it would be straightforward to quantify the amounts of 5′-biotin-PC-GMP-RNA and the unlabeled RNA (i.e. 5′-GTP-RNA) using these two RNAs. As the concentration of the biotin-PC-GMP was increased, a new band corresponding to the 5′-biotin-PC-RNA appeared that exhibited decreased mobility on PAGE. For quantification, the volumes of each of the peak positions on the gel were measured using the Quantity One software (Bio-Rad). Global background correction was applied using similar-sized rectangles for regions devoid of signal. The results of the 27-nt ribosomal A-site transcription priming

efficiency trials are shown in **Figure 2.5** and **Figure 2.6**; those for the optimization of 36-nt D5 RNA are shown in **Figure 2.7**.

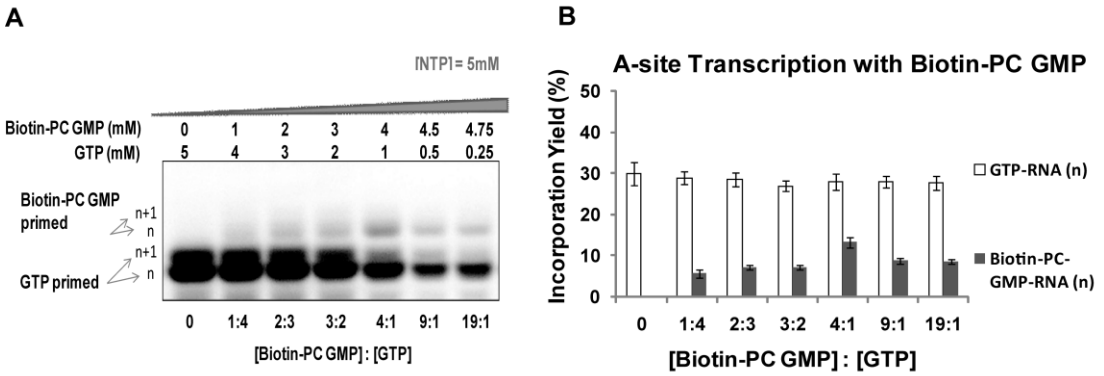


Figure 2.5. Enzymatic incorporation of the initiator biotin-PC-GMP with different ratios of biotin-PC-GMP: unmodified GTP using the ribosomal A-site DNA template. **(A)** Dependence of transcription priming of A-site RNA with biotin-PC-GMP on the ratios of biotin-PC-GMP : GTP in the presence of a total NTP concentration of 5 mM (i.e. 1.67 mM each ATP, CTP and UTP); n and n+1 products appeared for transcripts initiated with GTP and biotin-PC analog. **(B)** Relative amounts of GTP- and biotin-PC-primed Asite RNA products transcribed in the presence of different ratios of biotin-PC-GMP : GTP at 1.67 mM each ATP, CTP and UTP. Each bar is based on an average of three experiments and the error bar is based on the standard deviation from 3 experimental repeats. Note: Incorporation yield is the amount of initiated transcripts divided by the total amount of transcripts for each band. (See **Figure 2.6** for details of incorporation yield for n+1 and abortive yield)

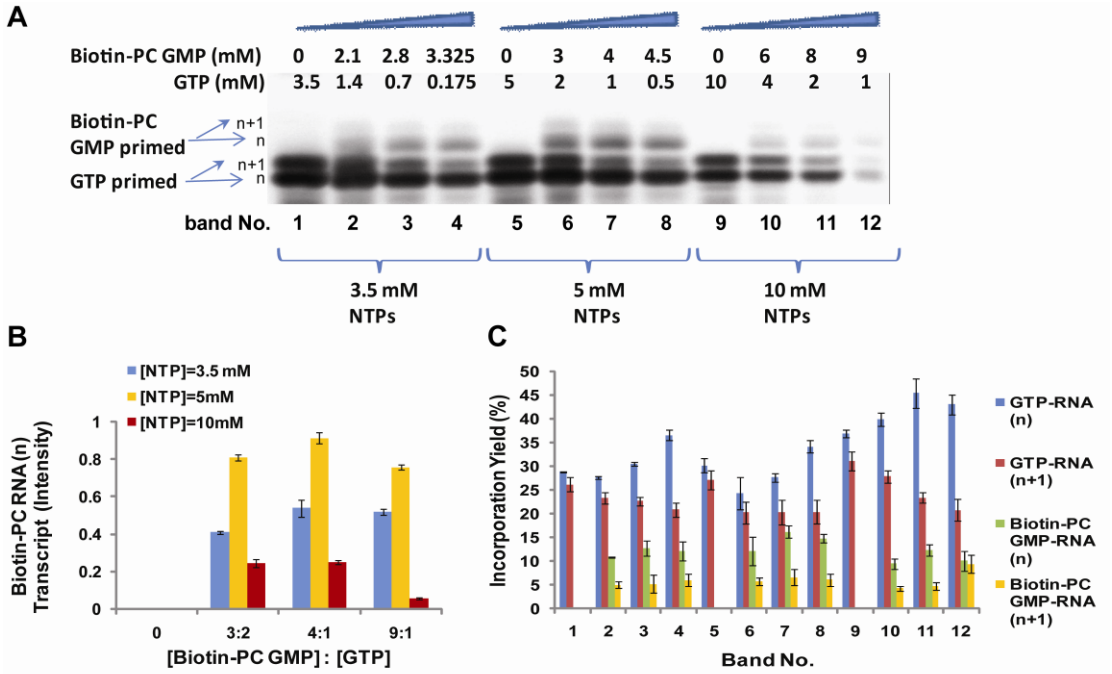


Figure 2.6. Enzymatic incorporation of the initiator biotin-PC-GMP with different total concentrations of NTPs (ATP, CTP and UTP) using the ribosomal A-site DNA template. **(A)** Dependence of transcription priming of the A-site RNA with biotin-PC-GMP on the ratio of biotin-PC-GMP : GTP in the presence of total NTP concentrations of 3.5 mM (i.e. 1.167 mM each ATP, CTP and UTP), 5 mM (i.e. 1.67 mM each ATP, CTP and UTP) or 10 mM (i.e. 3.33 mM each ATP, CTP and UTP). The double band in lane 1 represents n and n+1 GTP A-site RNA. The biotin-PC-RNA also migrated as two bands. **(B)** Trends of biotin-PC-RNA formed with the different concentrations of NTPs (ATP, CTP and UTP). The transcripts yield is based on the gel intensity. **(C)** Incorporation yield (the amount of initiated transcripts divided by the total amount of transcripts) of 5′-GTP-RNA and 5′- biotin-PC-GMP-RNA for each band in (A). Each bar is generated based on the average of three experiments and the error bar based on standard deviation from 3 experiment repeats.

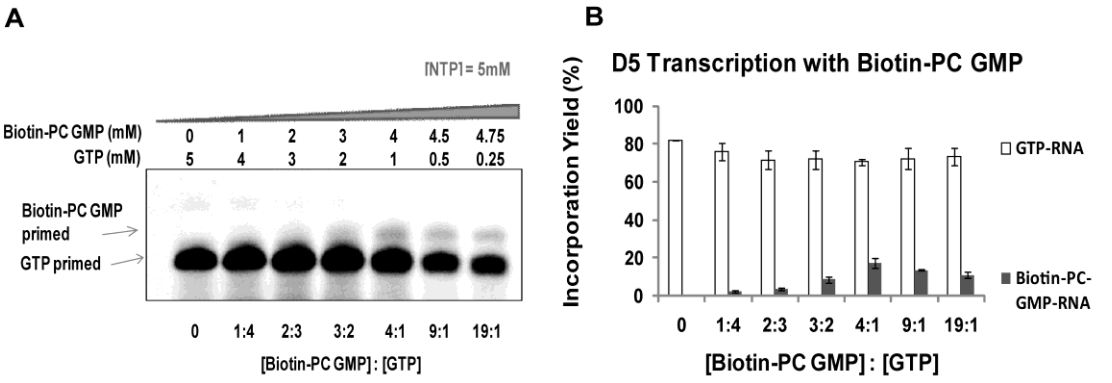


Figure 2.7. Enzymatic incorporation of the initiator biotin-PC-GMP with different ratios of biotin-PC-GMP: unmodified GTP using the D5 template. **(A)** Dependence of transcription

priming of D5 RNA with biotin-PC-GMP on the ratio of biotin-PC-GMP : GTP in the presence of 1.67 mM each ATP, CTP and UTP; **(B)** Relative amounts of GTP- and biotin-PC-primed D5 RNA products transcribed in the presence of different ratios of biotin-PC-GMP : GTP at 1.67 mM each ATP, CTP and UTP. Each bar is based on an average of three experiments and the error bar is based on the standard deviation from 3 experimental repeats. Note: Incorporation yield is the amount of initiated transcripts divided by the total amount of transcripts for each band.

The overall yield of transcribed 'biotin-PC primed' A-site RNA was higher (~16 % incorporation yield) at 5 mM total NTP (ATP, CTP and UTP) concentration (**Figure 2.5**) than at 3.5 mM (~11 % incorporation yield) or 10 mM (~10% incorporation yield) total [NTP] concentrations (**Figure 2.6**). At the optimal 5 mM total concentration of NTP for 5′-biotin-PC-GMP-RNA incorporation, 28% of the transcribed RNAs were full-length 5′-GTP-RNA whereas abortive transcripts accounted for ~29% (**Figure 2.6C**). Note that the n+1 transcripts for the 5′-GTP-RNA and the 5′-biotin-PC-GMP-RNA were 20% and 7% respectively. The efficiency of transcription initiation was also analyzed using the DNA template for 36-nt D5 RNA. Again, as the concentration of biotin-PC-GMP was increased, a new band corresponding to 5′-biotin-PC-GMP D5 RNA appeared and migrated with decreased mobility on PAGE. A 17% yield (incorporation into the transcripts) of 5′-biotin-PC-GMP D5 RNA transcripts was obtained at [biotin-PC-GMP] : [unmodified GTP] ratio of 4:1 and at 5 mM total NTP concentration (**Figure 2.7**). The results for A-site and D5 RNA transcription suggest that it is essential to optimize the NTP concentrations and to include a fourfold excess of biotin-PC-GMP over unmodified GTP in order to maximize analog incorporation.

2.3.5 Purification of biotin labeled RNA with avidin affinity column and photocleavage

To utilize the biotin-PC-RNA in functional or biophysical studies, the tag needs to be removed. To prevent the free biotin-PC-GMP from competing with the biotin labeled RNA for the Neutravidin-agarose resin binding, the unincorporated free biotin-PC-GMP was removed from the mixture using 3 kDa MWCO spin filters. The 5′-tagged RNA was then immobilized on the Neutravidin-agarose resin and the unbound RNA (primarily unlabeled 5′-GTP-RNA) was washed off with phosphate buffer. The RNA, bound to the resin, was then irradiated with UV light at 365 nm for 1 h to liberate the bound RNA (**Figure 2.3A**: Lane 6).

2.3.6 Biotin-PC synthesized minimal group II intron ribozyme has enhanced catalytic activity compared to gel purified GTP-RNA in the SER reaction

While a number of methods have been proposed to synthesize RNA under native conditions, it remains unclear how native RNA preparations affect function. To test the hypothesis that natively prepared RNA are more catalytically active than RNA prepared using denaturing methods that require heat-cooling refolding cycles, two variants of the minimal group II intron ribozyme splicing model system from *Pylaiella littoralis* (PL) were employed.[34, 50] The first variant is a three-piece ribozyme system comprising D123, D5, and a substrate reflecting the product of exon ligation (A647-E1E2 RNA, 22-nt).[50] In contrast, the second variant is a two piece ribozyme system comprising D1235 and the same substrate used with the three-piece system.[34] The bipartite model is expected to be more catalytically active than the tripartite

model because D5 is connected *in cis* rather than added *in trans*. For each model system, denatured GTP-RNA, anion-exchanged purified RNA, and 'natively' prepared biotin-PC-GMP-RNA were used in the spliced exon reopening reaction to test catalytic cleavage of the A647-E1E2 RNA.[34, 50, 62]

The effect of heat denaturation and refolding on the D123 RNA to cleave the A647-E1E2 RNA substrate in the presence of D5 RNA were tested first (**Figure 2.8A**). In the absence of heat annealing, the natively prepared biotin-PC-GMP-D123 RNA was able to catalyze cleavage of the A647-E1E2 RNA substrate to near completion in ~ 30 min (**Figure 2.8B**). Similarly natively prepared GTP-D123 RNA purified using anion exchange was able to catalyze near complete cleavage of the substrate in ~ 30 min (**Figure 2.8C**), again underscoring the importance of native RNA preparation. In contrast, gel-purified GTP-D123 RNA required at least 180 min for complete cleavage (**Figure 2.8D, H**). These results suggest that the method of the RNA preparation makes a dramatic difference in D123 RNA ribozyme catalytic function; whereas cleavage by natively purified RNA was detectable within 2 min and was complete in less than 60 min, substantial cleavage by RNA purified by denaturing PAGE was not observed before 30 min and only reached completion after approximately 240 min (**Figure 2.8H**).

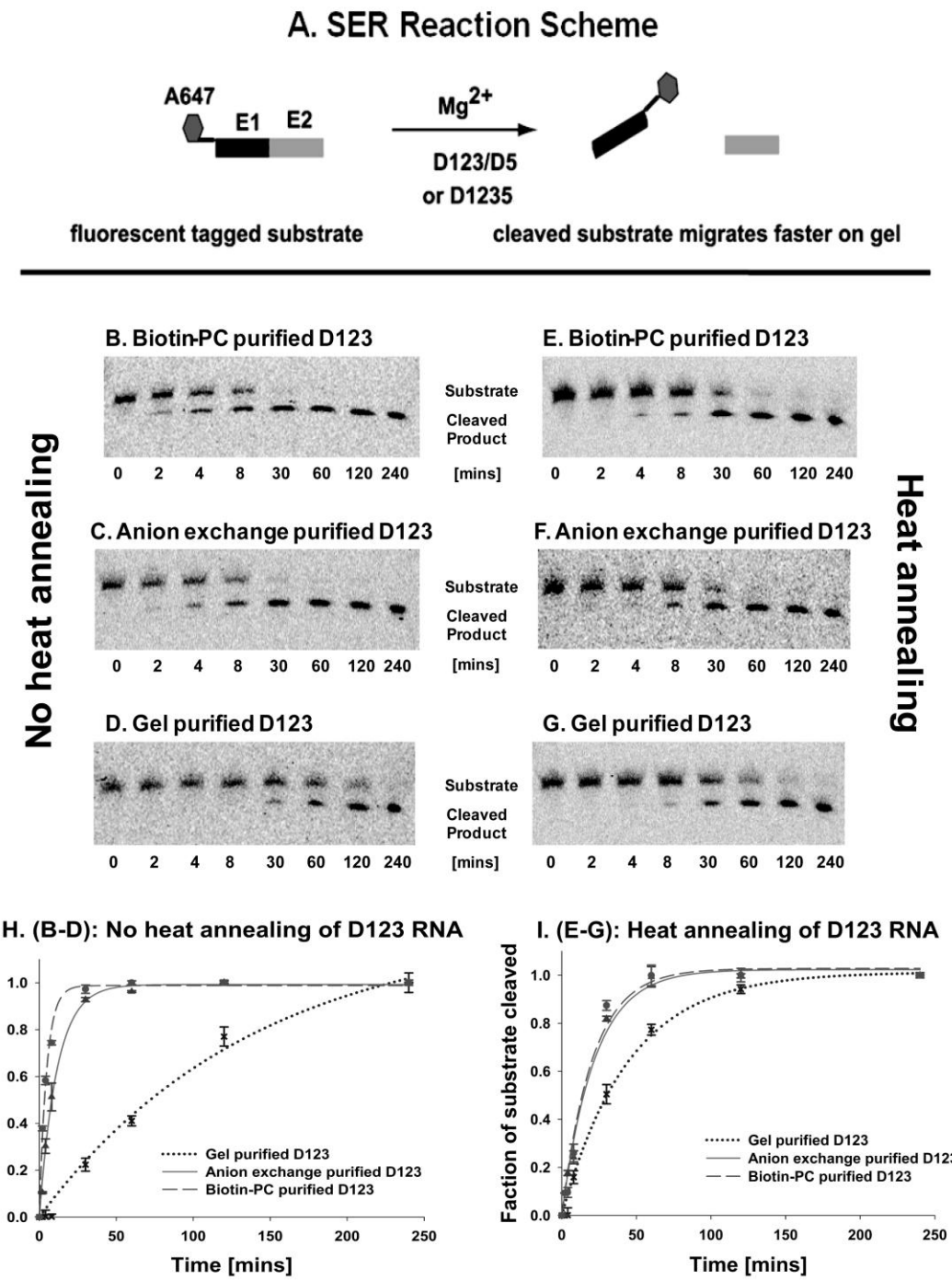


Figure 2.8. (A) In-gel fluorescent assay for catalytic cleavage of a 22-nt spliced exon substrate (E1E2) by a minimal group II intron ribozyme comprising domains 1, 2, and 3 (D123) and domain 5 (D5) or a ribozyme comprising domains 1, 2, 3, and 5 (D1235) in the presence of 50 mM Mg^{2+} at 42 ℃ (SER reaction, for details, see Materials and methods); **(B-D)** Comparison of the catalytic activity of gel purified/anion exchange purified and (photo-cleaved) biotin-PC purified D123 and D5 RNA at no heat annealing (native) conditions; **(E-G)** Comparison of the catalytic activity of gel purified/anion exchange purified and (photo-cleaved) biotin-PC purified D123 and D5 RNA at heat annealing (refolding) conditions respectively. Error bars are generated based on standard deviation between 2-3

experiment repeats. **(H)** Rates of buildup of cleaved A647-E1 product for no heat annealing SER reaction of D123 RNA were fitted into single-exponential function y=a*(1-exp(-b*x) using non-linear square minimization in SigmaPlot 11.0 sofeware. **(I)** Rates of buildup of cleaved A647-E1 product for heat annealing SER reaction of D123 RNA were fitted into single-exponential function. Error bars are generated based on standard deviation between 2-3 experiment repeats.

After heat annealing and refolding, the natively prepared biotin-PC-GMP-D123 RNA cleaved the substrate in less than 60 min (**Figure 2.8E**), slightly more slowly than without the heat annealing procedure (**Figure 2.8B, H**). Similarly natively prepared GTP-D123 RNA purified using anion exchange cleaved the RNA substrate in ~ 60 min (**Figure 2.8F**). Here again the gel-purified GTP-D123 RNA required at least 2 h for complete cleavage (**Figure 2.8G, I**). In this case, cleavage began at ~ 8 min instead of 30 min. These findings suggest that under conditions typically used in making folded RNAs, even heat annealing does not completely rescue the catalytic activity of the D123 RNA ribozyme. As shown below, folding is incomplete for the gel purified RNA.

The effect of heat denaturation and refolding on the D1235 RNA to cleave the A647-E1E2 RNA substrate was employed as a second example of an RNA mediated catalysis. For this second model system the catalytic cofactor D5 RNA was connected *in cis*, and thus it was anticipated that this ribozyme would be catalytically more active and more robust in response to heat denaturation and refolding than D123. As expected, in the absence of heat annealing and refolding, the natively prepared biotin-PC-GMP-D1235 RNA was able to catalyze complete cleavage of the substrate in almost 30 min (**Figure 2.9A**). As observed for D123 RNA, natively prepared GTP-D123 RNA purified using anion exchange completely cleaved the substrate in ~ 1 h (**Figure 2.9B**). Once again, the gel-purified GTP-D1235 RNA that had been heat

denatured and refolded for 2 h now required approximately 4 h for complete cleavage (**Figure 2.9F**). When the heating step was omitted, the endpoint of the reaction was reached at a comparable time period of ~4 h, but unlike the gel-purified GTP-D123 RNA the onset of the reaction was only slightly retarded (**Figure 2.8D, G, H, I** versus **Figure 2.9C, F, G, H**). Altogether, these results suggest that conditions that are traditionally used in folding RNAs may not be as effective as envisioned especially in rescuing the catalytic activity of ribozymes. As shown below, folding is also incomplete for the gel purified RNA.

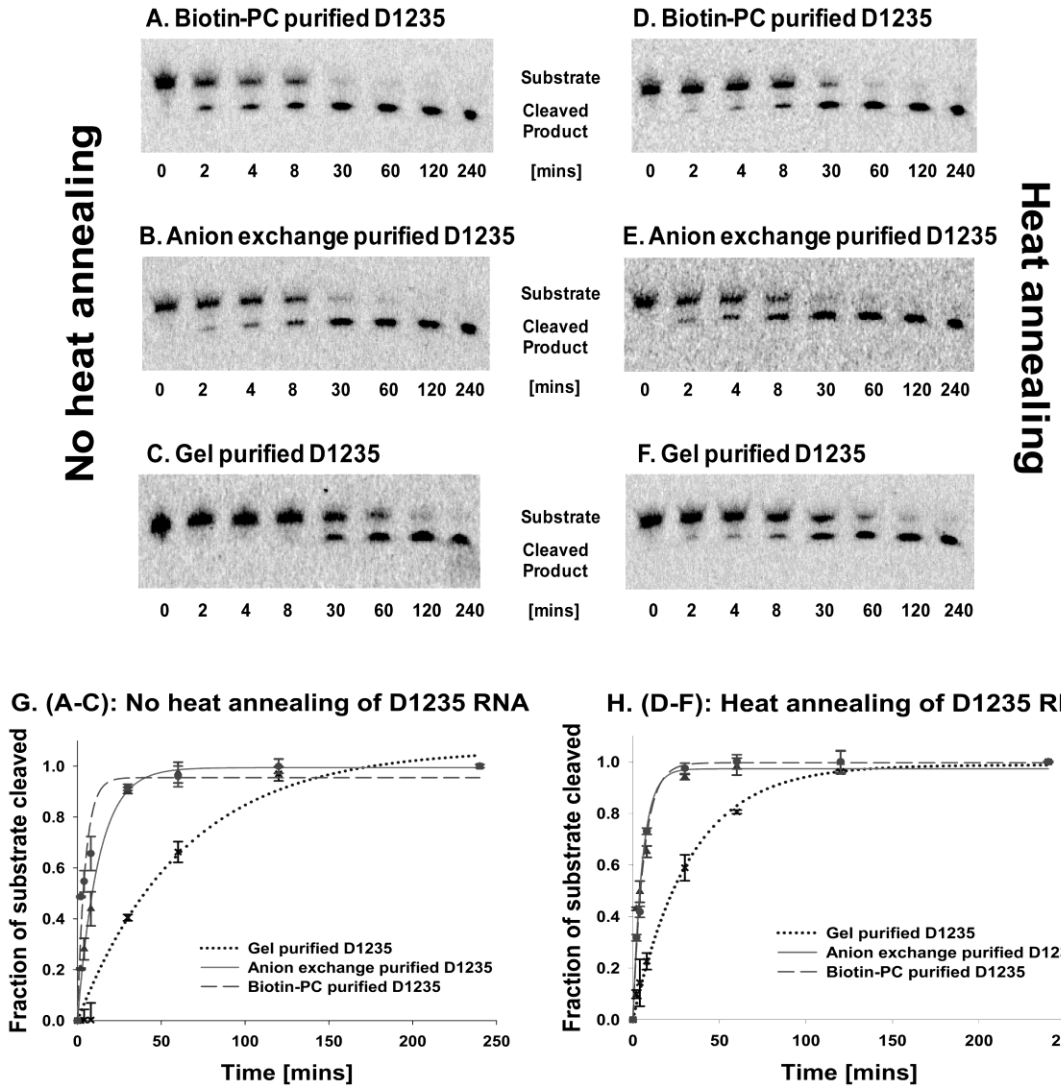


Figure 2.9. (A-C) Comparison of the catalytic activity of gel purified/anion exchange purified and (photo-cleaved) biotin-PC purified D1235 RNA at no heat annealing (native) conditions; **(D-F)** Comparison of the catalytic activity of gel purified/anion exchange purified and (photo-cleaved) biotin-PC purified D1235 RNA at heat annealing (refolding) conditions respectively. **(G)** Rates of buildup of cleaved A647-E1 product for no heat annealing SER reaction of D123 RNA were fitted into single-exponential function y=a*(1-exp(-b*x) using non-linear square minimization in SigmaPlot 11.0 sofeware. **(H)** Rates of buildup of cleaved A647-E1 product for heat annealing SER reaction of D123 RNA were fitted into single-exponential function. Error bars are generated based on standard deviation between 2-3 experiment repeats.

2.3.7 The population of biotin-PC synthesized minimal group II intron ribozyme conformers is more homogenous than that of gel purified GTP-RNA

Using the native preparation method, the question of how different RNA preparation methods affects folding was examined. To test the hypothesis that natively prepared RNA should be more conformationally homogenous than RNA prepared using denaturing methods that require heat-cooling refolding cycles, samples of D123 RNA prepared by urea PAGE (denatured and heat annealed as described earlier), anion-exchanged purified D123 RNA, and 'natively' prepared biotin-PC-GMP-D123 RNA were analyzed by non-denaturing PAGE. Strikingly, the denaturing preparation had a significant impact on the migration pattern of the gel purified D123 RNA: even after refolding, >86% of the RNA was in alternative conformations (**Figure 2.10**). In contrast, ~93% of the 'natively' prepared biotin-PC-GMP-RNA, which had not been denatured, was in a homogenous conformation (**Figure 2.10**). Similarly ~85% of natively prepared GTP-D123 RNA purified using anion exchange was in homogenous conformation, with ~15% of the population in alternative slower migrating conformations (**Figure 2.10**). To rule out overloading of the gel purified D123 RNA, the experiment shown in **Figure 2.10** was executed on 6% native PAGE using two different dilutions of all three RNAs tested. The concentrations of the RNAs were determined by UV absorbance measurements of RNAs hydrolyzed using NaOH to remove any hyperchromicity effects. As can be seen in **Figure 2.10**, the gel purified D123 RNA was consistently smeared, while the native preparations consistently produce yields of > 90% as a single band.

Altogether, these results suggest that although the D123 or D1235 RNAs purified by denaturing PAGE conditions were heat annealed under conditions typically used in making folded RNAs, their catalytic activities were not completely rescued and most of the RNA existed in alternative conformations. These findings are in agreement with previous observations that RNA can be kinetically trapped in misfolded states which are not easy to efficiently correct, even by heat annealing.[24, 26, 66]

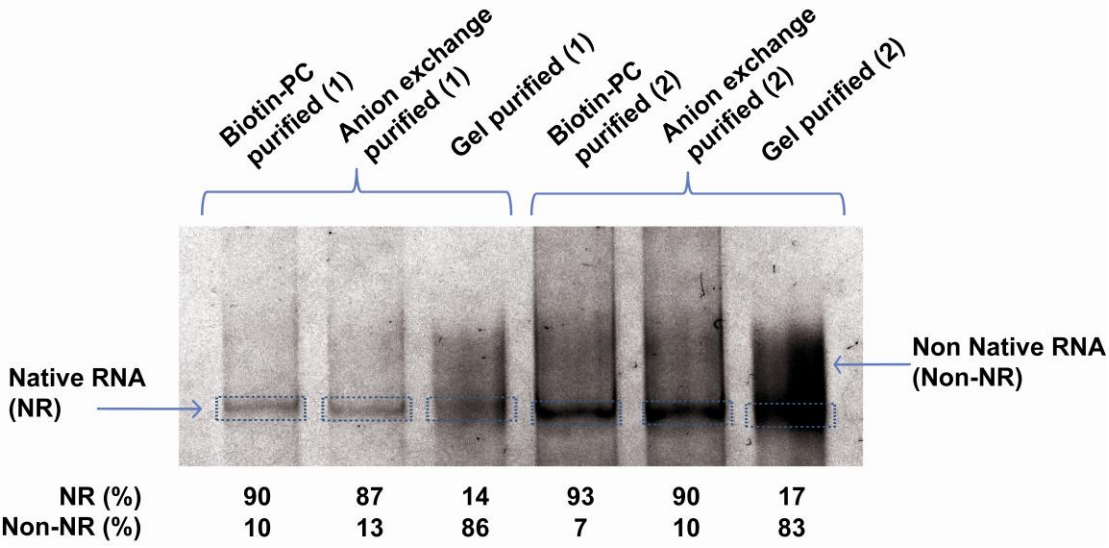


Figure 2.10. Comparison of gel purified, anion exchange purified and biotin-PC purified D123 on the 6% native PAGE gel with 0.5×TME（12.5 mM Tris/MOPs, 0.1 mM EDTA）as the running buffer at 4˚C. The boxed bands in blue represent the native RNAs and the smears part represent the non-native RNAs with heterogeneous conformations. There are two different amounts loaded on the native PAGE gel, (2) is three times more than (1).

2.4 Conclusions

Experimental evidence is presented demonstrating that RNA enzymes not only fold differently but also have compromised catalytic activities depending on whether they were prepared using denaturing or nondenaturing purification methods. Although the RNA community has believed that large RNAs can become trapped into what has been described as the "alternative conformer hell"[24, 66], no systematic investigations to

reveal the true extent of the problem have been heretofore performed. Standard methods for RNA preparation, while robust and well-established, can create not only heterogeneous populations of conformers,[24, 26] but also produce RNAs that may become marginally active. The goal of this work was to develop a photolabile biotinylated nucleotide monophosphate strategy to facilitate native RNA preparation and to address the question of how the nature of the preparation can affect RNA function. This designed and synthesized biotin-PC-GMP was shown to be utilized by T7 RNAP as an initiator nucleotide to produce RNA labeled at the 5′-end with biotin-PC.

This photoremovable affinity tag confers several desirable advantages for use in numerous RNA biophysical applications. First the ability to prepare "native" RNA without having to refold should minimize misfolding and thereby preserve the functionality of these RNAs. As proof of concept, we have shown that the two ribozymes prepared by our new method retain greater catalytic activity than when they were prepared by denaturing methods and have to be refolded. To determine if other native preparation methods would also give enzymatically more proficient RNAs, anion exchange was used to purify the RNAs under native conditions. As anticipated, RNAs purified via the native anion exchange method were almost as active as those prepared via the biotin-PC-GMP method. The biotin-PC-GMP native preparation is therefore a nice complementary method to other native RNA preparation methods such as anion exchange FPLC, and there is no panacea for preparing native RNAs. Importantly, this biotin-PC-GMP method does not require expensive instrumentation and is amenable to high throughput native RNA preparation. Our results are also

comparable to those reported recently which showed quite unambiguously that heating and subsequent cooling of a 164-nt Varkud satellite (VS) ribozyme led to ~70% slowly migrating non-native RNA folds, whereas a VS ribozyme purified under nondenaturing conditions appeared to be ~93% fast migrating single band representing the native RNA fold. These observations are consistent with the expectation that heat annealing cannot adequately guide RNAs across their rugged free energy folding landscape towards the native states.[26] Rather native purification methods are increasingly vital to perform meaningful RNA research in general, and specifically in *in vitro* RNA folding and structural RNA studies. Our work argues for caution in interpretating kinetic data involving RNA prepared using heat-denaturation-refolding protocols. It also provides an important framework for future studies regarding RNA catalysis and recognition.

Chapter 3: Conclusions and future work – Native RNA preperation

The first part of this dissertation connected the development of chemical tools to solve an important problem in RNA structural biology, which is how to circumvent the denaturing process in RNA preparation to avoid the adverse effect on function caused by misfolding during the refolding process? Using this method, we demonstrated that the RNA 'folding/misfolding problem' encountered by many researchers, when preparing RNA for biological studies, is more serious than originally thought.[38] The traditional enzymatic method uses denaturing polyacrylamide gel electrophoresis (PAGE) purification with required cycles of denaturation and renaturation. Using that method, not only is it difficult to completely refold the denatured RNA into its native form, but it is also problematic to create a heterogeneous population of conformers.[23, 24] However, our new RNA preparation method, using 'click' chemistry (Huisgen 1, 3-dipolar cycloaddition reaction) to couple biotin to a 'caged' photocleavable (PC) guanosine monophosphate (GMP) to synthesize RNAs, avoids the denaturing and refolding steps. The native RNA is easily washed off an avidin column after on-column photocleavage and elution under native preparation conditions. We successfully applied this method to transcribe RNAs ranging in size from 27 to 530 nucleotides, demonstrated that each of the RNA ribozymes D123 (493 nt) and D1235 (527 nt), which were prepared by our non-denaturing method, forms a homogenous population with higher catalytic activity than those prepared by denaturing methods.[38]

Although our method is easy and efficient for synthesizing native RNAs, the low yield of RNA (~17%) when primed with biotin-PC-GMP leads to large amounts of mixture of unlabeled and labeled RNA. To improve the yield of biotin-PC labeled RNA, a second generation of labeling methodology will have to be developed.

PART II – Cyclic-di-GMP (c-di-GMP) riboswitch

Chapter 4: Introduction to Part II

4.1 The Importance of cyclic-di-GMP (c-di-GMP) in bacteria

Bis-(3'-5')-cyclic guanosine monophosphate (c-di-GMP) was first discovered in 1987 by Moshe Benziman, who demonstrated that it allosterically activated the membrane bound cellulose synthase enzyme in *Gluconoacetobacter xylinus*.[67] However, it was not until a decade later that c-di-GMP was proven to be a ubiquitous second messenger in bacteria that regulates the transition between sessile and motile lifestyles by interfering with cell motility at a low concentration and promoting the expression of an adhesive matrix component and biofilm formation at a high concentration (**Figure 4.1A**)[68-71]. The importance of c-di-GMP in controlling bacterial pathogenesis is now universally accepted.[72-75] Recent studies have indicated that c-di-GMP is not only associated with cell-cell communication, pilus assembly[69], and biofilm formation, but also has essential role in virulence or acts as an effective immunomodulator.[74, 76] The link between c-di-GMP and virulence was initially demonstrated in the pathogenic organism *Vibrio cholerae.*[75] It was shown that the production of virulent cholera toxin, which is responsible for the massive and watery diarrhea produced during a cholera infection, was increased with low intracellular levels of c-di-GMP.[75] Recently, the critical function of c-di-GMP in regulating virulence was also demonstrated in some other known human pathogens, including the causative agents of gastroenteritis[73] and Lyme disease.[72, 73]

4.2 Biosynthesis and degradation of c-di-GMP

The intracellular level of c-di-GMP in the bacterial cell is influenced by an enzymatic synthesis and degradation process. C-di-GMP is generated from two guanosine-5'-triphosphate (GTP) molecules by diguanylate cyclase (DGC) enzymes containing GGDEF (Gly-Gly-Asp-Glu-Phe) or GGEEF (Gly-Gly-Glu-Glu-Phe) domain and degraded selectively by two different phosphodiesterase (PDE) enzymes that contain either EAL (Glu-Ala-Leu) or HD-GYP (His-Asp-Gly-Tyr-Pro) amino acid domains (**Figure 4.1B**).[77, 78] EAL domain PDEs cleave c-di-GMP into a linear dinucleotide (pGpG) intermediate and then further into two GMP molecules,[79] whereas HD-GYPs hydrolyze c-di-GMP directly into GMP in one step.[80] Interestingly, there are some hybrid proteins that contain both DGC and PDE domains within the same polypeptide.[69-71] Based on genomic sequencing, it is shown that 50-70% of all GGDEF and EAL domain proteins contain an additional accessory domain that function as a sensor of environmental factor to produce an output.[81] These well-recognized regulatory motifs, such as phosphorylation receiver domains (REC)[82, 83], gas sensor domains (haemerythrin)[84], oxygen-redox potential sensor domains (PAS)[85] and blue light sensor domains (BLUF)[86] fuse to GGDEF or EAL domains to regulate the enzymatic activities of the PDEs and DGCs for further c-di-GMP concentration regulation. Herein, the fact that GGDEF, EAL and HD-GYP domains are widespread in bacteria suggests that c-di-GMP is utilized as a second messenger by many different bacterial species to control cell signaling.[87]

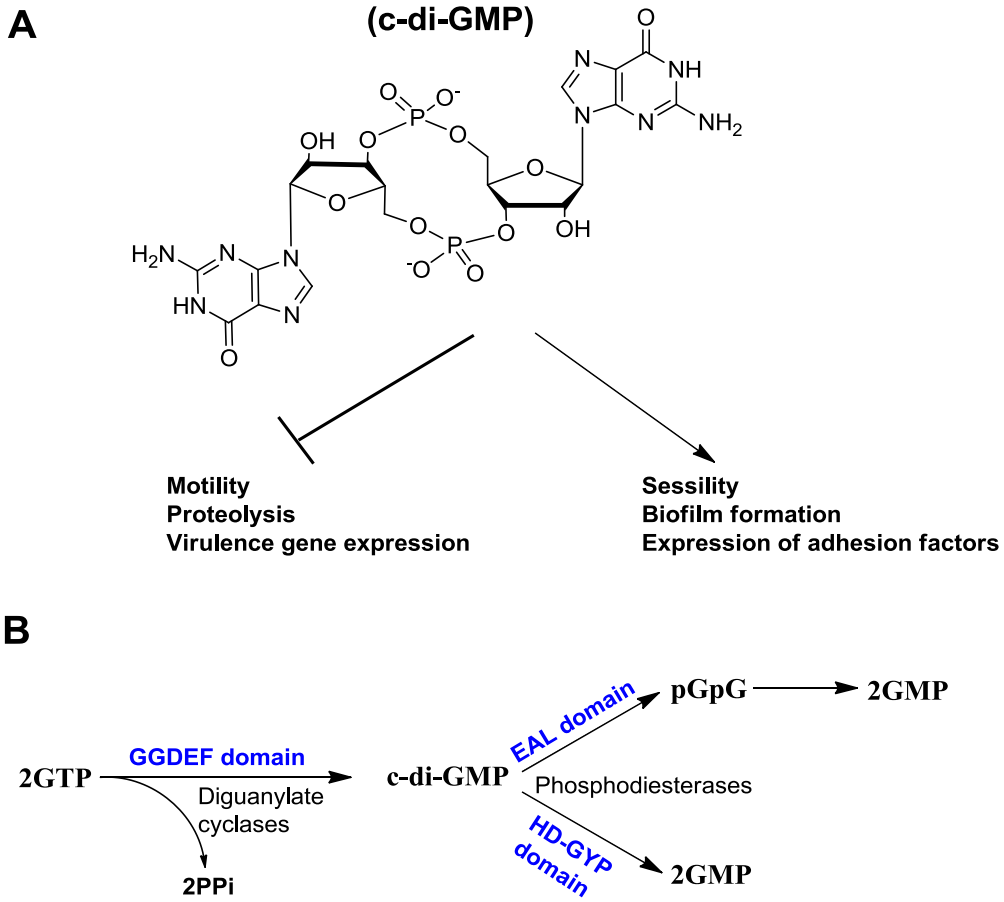


Figure 4.1. (A) Phenotypes and processes controlled by c-di-GMP. High concentration levels of c-di-GMP promote sessility and biofilm formation via the production of adhesive extracellular matrix components. Low c-di-GMP levels promote motility, proteolysis and the production of virulence factors. **(B)** Synthesis and degradation of c-di-GMP. C-di-GMP is synthesized from GTP by diguanylate cyclases containing GGDEF domains and is degraded into GMP by phosphodiesterase domains, containing either EAL or HD-GYP domains.

4.3 C-di-GMP effector – Protein

As a second messenger, c-di-GMP relays external stimuli (i.e. the first message) received by cell-surface receptors of sensor domains to the related effectors within the cell. Different concentrations of c-di-GMP modulate cellular functions through binding to various receptors and controlling various downstream signaling mechanisms. These c-di-GMP receptors include: 1) Effector proteins, such as PilZ

domain proteins that do not have their own enzymatic activities but regulate other proteins via domain-domain or protein-protein interaction upon binding to c-di-GMP;[88] 2) Degenerate GGDEF and EAL proteins, which have no catalytic activity.[89-91] These proteins often contain mutations in one or more of the conserved active site residues and therefore are no longer catalytically active (such as PopA from *Caulobacter crescentus*), but they still retain the ability to bind c-di-GMP.[92] 3) Proteins with enzymatic activity, and in which catalytic proficiency is increased upon binding to c-di-GMP, such as PleD of *Pseudomonas aeruginosa*;[93, 94] and 4) Transcription factors or repressors, such as FleQ of *Pseudomonas aeruginosa*, which regulate gene expression upon c-di-GMP binding[95, 96]. Collectively, these receptor proteins have affinities for c-di-GMP that range from nanomolar to lower micromolar concentrations.[88, 92, 94-101]

4.4 C-di-GMP effector – Riboswitch RNA

In addition to the proteins that bind c-di-GMP, the high frequency of GGDEF and EAL domains in bacteria implies another network of pathways involving c-di-GMP in global cellular functions. In the past few years, this hypothesis was supported by the discovery of c-di-GMP-I and II class riboswitch receptors, reported by the Breaker group (**Figure 4.2**).[102, 103] There are approximately 500 examples of class I[102] and 45 examples of class II riboswitches[103] that have been computationally identified in many different bacterial species. It is worth noting that some bacteria carry a very high number of c-di-GMP riboswitches.[102] These riboswitches are present in many pathogenic bacteria, such as *Bacillus anthracis* and *Vibrio cholerae*,

to regulate cellular processes including motility, pilus formation, biofilm formation, virulence factor expression, and flagellum biosynthesis. The broad distribution of c-di-GMP binding riboswitches throughout the bacterial domain also indicates that they are likely one of the primary targets of c-di-GMP.[102, 103]

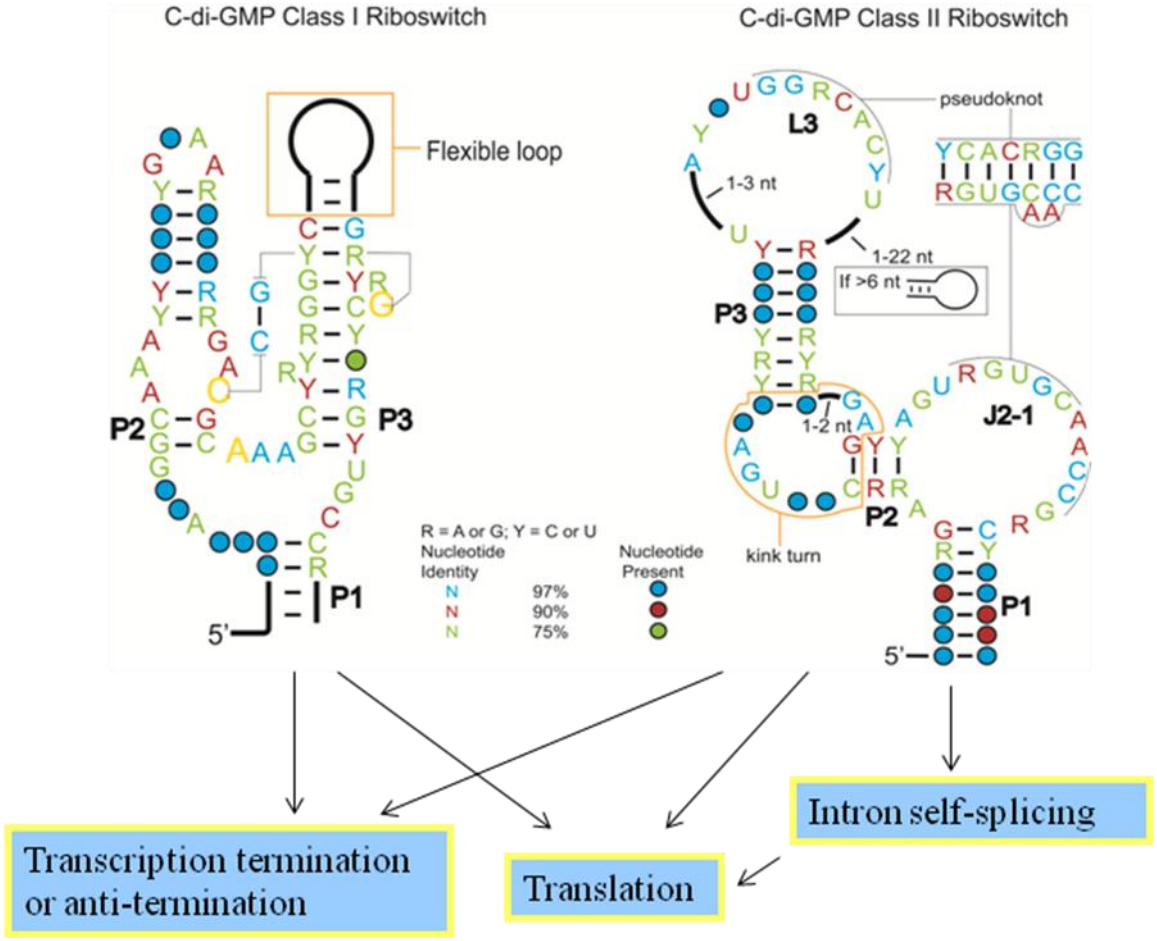
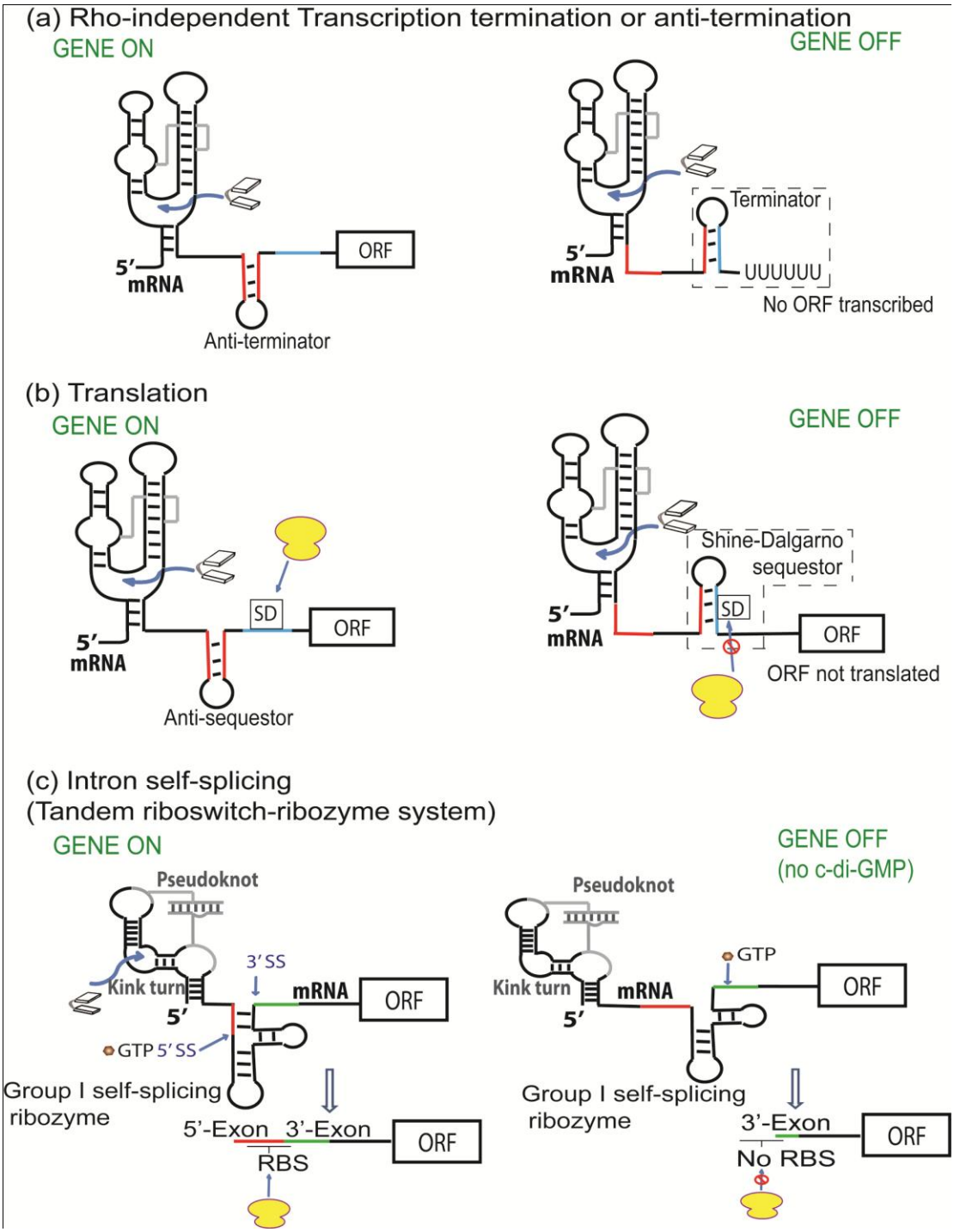


Figure 4.2. C-di-GMP class I and class II riboswitches. Figure is adapted from Lee *et al*., 2010.[103]

Although over 20 different riboswitch classes, which recognize diverse ligands, such as amino acids[104], nucleobases[105], vitamins[106], modified sugars[107], coenzymes[108], and even metals[109] or small anions[110], have been identified so far, c-di-GMP riboswitch is the first known class of RNA that directly binds a second

messenger signaling molecule. It is a section of 5' untranslated region (5'-UTR) mRNA in bacteria that contains two regions, an aptamer domain and an expression platform. The aptamer domain is responsible for recognizing and binding its target molecule – c-di-GMP, leading to a conformational change that alters the expression platform to regulate either transcription or translation (**Scheme 4.1**).[102, 103] In translational control, the Shine-Dalgarno (SD) sequence (a ribosome binding site located eight base pairs upstream of the start codon in an mRNA) is either sequestered or exposed to modulate ribosome binding. In transcriptional systems, sequestration of the terminator or anti-terminator sequence controls transcription. However, the switches between 'ON' to increase expression levels of the downstream genes and 'OFF' to decrease the expression level are independent of whether the riboswitch controls transcription or translation. One example of c-di-GMP riboswitch operating at the transcriptional control level is the Cd1 class I riboswitch from *Clostridium difficile*. The transcription of mRNA is terminated in the presence of a higher concentration of c-di-GMP, decreasing flagellum biosynthesis and repressing motility.[102] An example of c-di-GMP riboswitch using translational control is the Vc2 riboswitch in *Vibrio cholera*. In the presence of c-di-GMP, Vc2 riboswitch RNA upregulates expression of the Vc1722 gene product.[102] Additionally, there is also an example of gene expression regulation through intron self-splicing in the c-di-GMP class II riboswitch from *Clostridium difficile 630*. Its c-di-GMP aptamer resembles group I self-splicing ribozymes, in which two ligands GTP and c-di-GMP collaborate to promote self-splicing and function as a tandem riboswitch-ribozyme, initiating translation.[103] Breaker *et al.* have further proven that the tandem riboswitch-ribozyme

could check leaky splicing and serve as a backup mechanism to ensure translational inhibition when c-di-GMP is in low concentration.[111]



Scheme 4.1. Mechanisms of c-di-GMP riboswitches. Figure is adapted from Breaker *et al.*, 2012.[112]

As one class of c-di-GMP effector, c-di-GMP riboswitches have higher binding affinities than any known protein receptors, ranging from 10 pM[102, 113] for the class I riboswitch to between 200 pM and 2 nM for the class II riboswitch[103]. For the Vc2 RNA class I riboswitch from *Vibrio choleae*, it is known that the strong binding affinity (10 pM) is due to very low on-rates and off-rates ($K_{on} = 1 \times 10^6$ $M^{-1}min^{-1}$ and $K_{off} = 1 \times 10^{-5}$ min^{-1})[113]. However, such slow off-rate corresponds to a half-life of approximately 1 month, meaning that the ligand is effectively irreversibly bound in the biological timeframe. Therefore, class I riboswitches with tight binding affinity are likely controlled kinetically, in which gene regulation is determined by the competition of the on-rate with the rate of transcription. In the cell, the choice between 'ON' or 'OFF' must be made during transcription of the mRNA and binding must occur on a relevant time scale to regulate the downstream genes.[113] Except for Vc2 RNA, the kinetic properties of other class I riboswitches and class II riboswitches have not been explored.

4.5 Structure of class I and class II c-di-GMP riboswitch

To better understand the molecular basis of RNA recognition of c-di-GMP, the crystal structures of c-di-GMP bound to the aptamer domains of both class I and class II riboswitches have been determined.[113-115] Due to the three-dimensional architectures of these two classes of riboswitches, each riboswitch provides specific binding pockets for c-di-GMP and has different modes of ligand recognition (**Figure 4.2**). The class I bound form of the riboswitch consists of three helices that adopt a y-shape and c-di-GMP bound at the three helix junction[113, 115], whereas the class II

bound form riboswitch folds into a more compact conformation containing three helices, a kink turn and a pseudoknot[114].

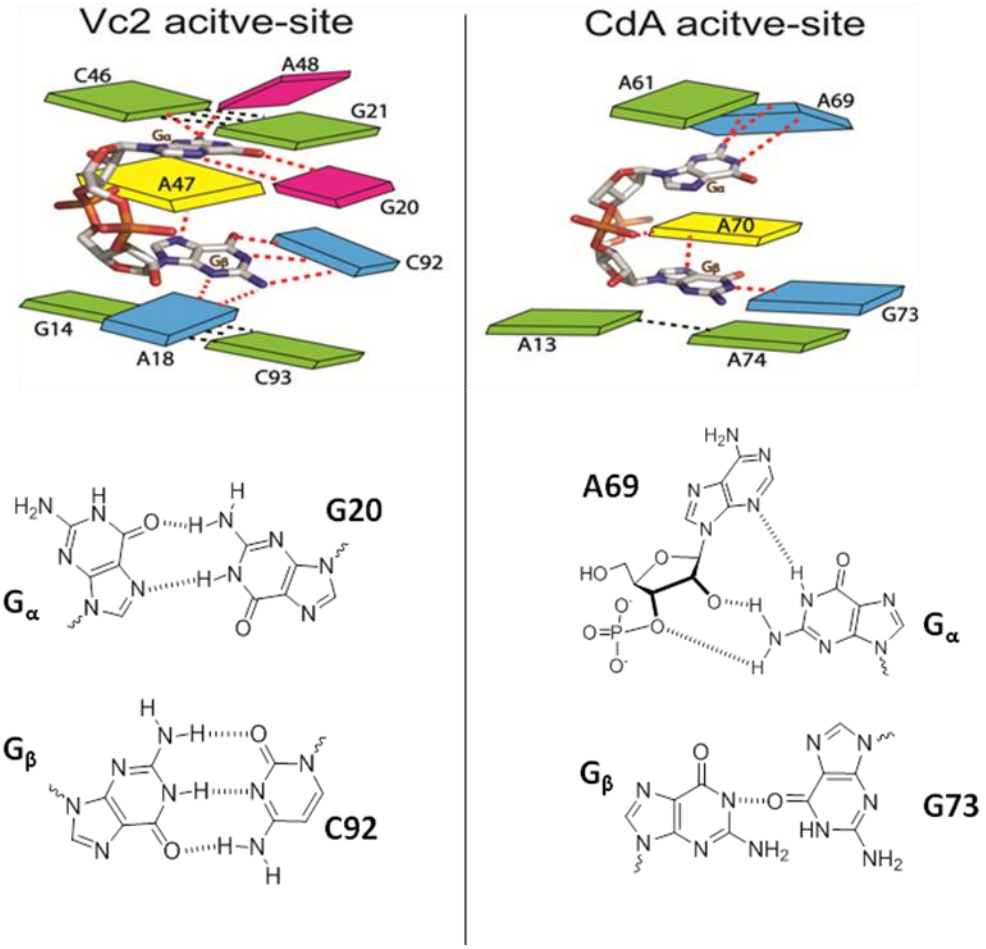


Figure 4.3. Active site architectures of class I (Vc2) and class II (CdA) riboswitches. Adenines in yellow are intercalated in both Vc2 and CdA. The blocks in green represent the stacking interactions above and below c-di-GMP. The direct interactions between nucleobases and c-di-GMP are shown in blue and magenta and the important hydrogen bonding interactions are shown in the details.

In the class I riboswitch, c-di-GMP binds at the junction of three helices, P1, P2 and P3 (**Figure 4.2**, label with residue numbers to correspond). Two critical nucleotides (G20 and C92) specifically recognize the two guanine bases (G_α and G_β) of c-di-GMP asymmetrically in canonical pairings. G20 contacts the Hoogsteen face of top base (G_α), while C92 N3 forms a canonical Watson-Crick pair with the bottom base (G_β) N1 (**Figure 4.3**).[113, 115] Besides these two hydrogen-bonding interactions,

mutational studies show that adenosine (A47) stacking between the two guanine bases is highly conserved.[116] It intercalates between the two bases of c-di-GMP to bridge the P1 and P2 helices with the ligand through extensive stacking interactions. Furthermore, recognition of the ribosyl-phosphate backbone also plays an important role in c-di-GMP binding to the class I riboswitch. The recognition between the backbone of c-di-GMP and the riboswitch is mediated by both hydrogen bonding and metal-phosphate contacts, with one phosphate on c-di-GMP more heavily recognized than the other (**Figure 4.4**).

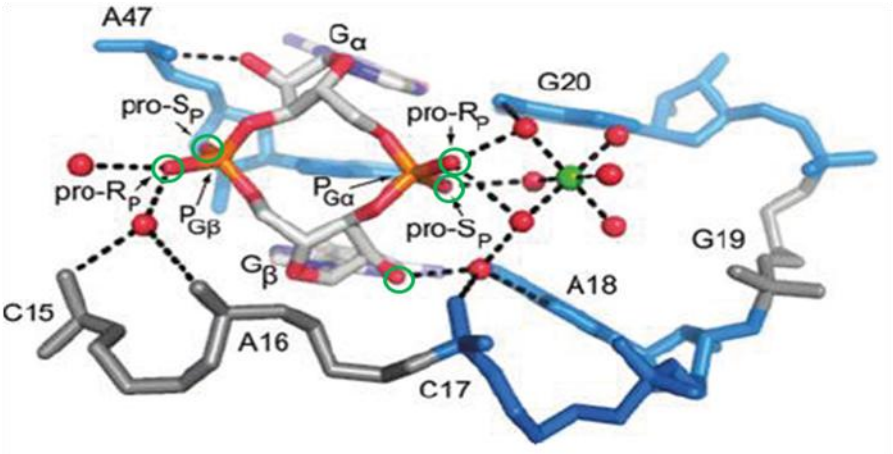


Figure 4.4. C-di-GMP ribosyl-phosphate backbone recognized by the class I riboswitch. The atoms circled in green are recognized via hydrogen-bonding and metal-phosphate contacts at the ribosyl-phosphate contacts. Figure is adapted from Shanahan *et al*., 2011.[117]

Compared to the class I c-di-GMP riboswitch, the class II riboswitch also recognizes the two guanine bases of c-di-GMP asymmetrically, but the nature of this interactions is significantly different (**Figure 4.3**).[114] Firstly, there is only one hydrogen-bonding interaction between the intercalating adenosine (A70) and a nonbridging phosphate oxygen on c-di-GMP in class I riboswitch opposite to the extensive backbone recognition of c-di-GMP by class I riboswitch. Secondly, there is no canonical base pairing observed in class II riboswitch. Instead, G_{α} of c-di-GMP is

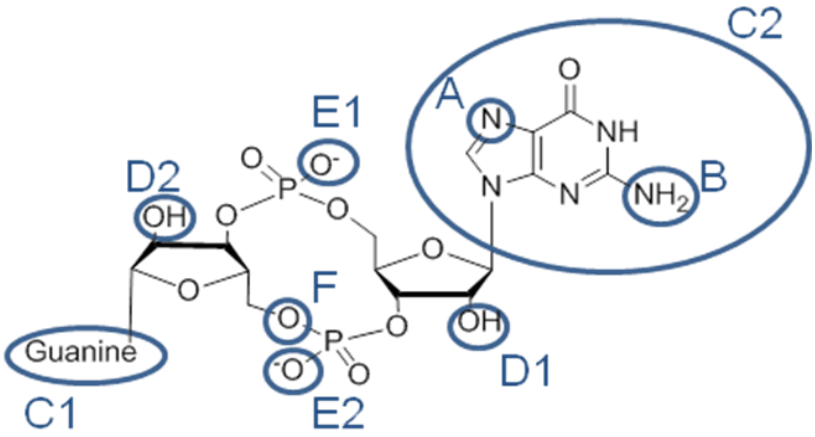
recognized as part of the a base triple with A69 and U37 of class II riboswitch, and G_{β} interacts with the A70 and G73 residues through hydrogen-bonding and hydrated magnesium. However, the only similar recognition to class I riboswitch is the stacking interactions between two guanine bases of the ligand and the RNA residues in the binding pocket. Three conserved adenosine nucleotides A61, A70 and A13 stack above, between and below the two bases of c-di-GMP respectively. This base stacking is important for ligand binding.[117] Based on these binding pocket structures of the bound forms of the class I and class II riboswitches, it is likely that the two riboswitches sense c-di-GMP via different mechanisms.

4.6 Targeting c-di-GMP binding riboswitches with different c-di-GMP analogs

The prevalence of c-di-GMP in many human pathogens, in contrast to its absence in eukaryotes, makes c-di-GMP receptors attractive as new drug targets for the development of therapeutics against bacterial infection. Recently, there has been renewed interest in developing therapeutics against RNA targets. Along these lines, we have been interested in the development of c-di-GMP – like molecules (analogs) that could specifically bind to the pathogenic bacterial riboswitch in polymicrobial systems or bind to c-di-GMP riboswitches to cause biological processes that are different from those produced by the native c-di-GMP.

To achieve the above goal, it is important to understand how the structural features of the ligand are required for specific binding and to identify which moieties of the ligand are important for binding. Herein, structure-activity relationship (SAR) studies for c-di-GMP binding to both class I and class II riboswitches using different analogs,

including base, ribosyl-phosphate modified analogs, have been investigated (**Figure 4.5** and **Table 4.1**).[117-120] In addition, binding of c-di-GMP riboswitches to pGpG – the breakdown product of c-di-GMP, and the other linear analogs GpG and GpGp, has also been studied (**Table 4.1**).[118] From these studies, it reveals that the class II riboswitch is tolerant to many c-di-GMP analogs, but the class I riboswitch is either completely discriminated against or bound with a significantly weaker binding affinity compared to the wild type ligand (**Figure 4.5** and **Table 4.1**). Then the main binding difference could be further utilized as a strategy for manipulating specific biological activities under the control of the class II riboswitch without affecting processes regulated by the class I riboswitch. For example, both the Strobel group and our group have proven that bulky modifications at the 2'-OH position of the ribose rings, e.g. OMe and the addition of biotin (see **Chapter 5**) can be used to distinguish class I and II riboswitches[114, 119]. In addition, other analogs that showed a binding preference for the class II riboswitch over the class I riboswitch include c-di-IMP, c-di-dGMP and c-(R_pS_p)-di-G_{ps} (**Table 4.1**). These observations underscore the point that the class II riboswitch is highly tolerant to modifications of the bases, ribosyl-phosphate backbone, whereas the class I riboswitch sense the ligand with all of these interactions.[114]



A. Change to CH
B. Change to H
C. Change to Adenine, Inosine
D. Change to Ome, H, F, biotin
E. Change to S
F. Change to S

Figure 4.5. Modifications of c-di-GMP at different positions.

	Analog Name	**Modification**	**Binding Affinity (K_d)**
Wild Type	C-di-GMP	-	Class I Vc2 RNA[a]: 1.4 ± 0.1 nM[117] Class II CdA RNA: 2.2 ± 0.8 nM[114]
Guanine Modification	C-c^{7}G-GMP	A1 = CH	Class I Vc2 RNA[a]: 1100 ± 80 nM[117] Class II CdA RNA: 33.0 ± 2.6 nM[117]
	C-di-c^{7}GMP	A1 = CH, A2 = CH	Class II CdA RNA: 3500 ± 630 nM[117]
	C-N^{1}mG-GMP	B1 = N-CH_3	Class I Vc2 RNA[a]: 420 ± 16 nM[117] Class II CdA RNA: 160 ± 14 nM[117]
Base Modification	C-di-IMP	C1 = inosine, C2 = inosine	Class I Vc2 RNA[a]: 1100 ±210nM[117] Class II CdA RNA: 4.1 ± 0.4 nM[117]
	C-IMP-GMP	C1 = inosine	Class I Vc2 RNA[a]: 200 ± 31 nM[117]

	C-di-AMP	C1 = adenine C2 = adenine	Class II CdA RNA: >30,000 nM[117]
	C-AMP-GMP	C1 = adenine	Class I Vc2 RNA[a]: 1600 ± 160 nM[117] Class II CdA RNA: 271.1 ± 47.1 nM[117]
Phosphate and 2'-OH Modification	C-dG-GMP	D1 = H	Class I Vc2 RNA[a]: 51 ± 5.4 nM[117]
	C-di-dGMP	D1 = H, D2 = H	Class I Vc2 RNA[a]: 2600 ± 88 nM[117] Class II CdA RNA: 11 ± 1.2 nM[117]
	C-2'OMe-G-GMP	D1 = H	Class I Vc2 RNA[a]: 200 ± 31 nM[117]
	C-di-2'OMe-GMP	D1 = OMe, D2 = OMe	Class I Vc2 RNA[a]: ~ 30,000 nM[117] Class II CdA RNA: 4.3 ± 0.7 nM[117]
	C-di-2'F-GMP	D1 = F, D2 = F	Class I Vc2 RNA[a]: 56 ±9.0nM[117] Class II CdA RNA: 5.8 ± 1.0 nM[117]
	C-(R_pS_p)di-G_{ps}	E1 = S, E2 = S	Class I Vc2 RNA[a]: 750 ±63 nM[117] Class II CdA RNA: 4.0 ± 0.8 nM[117]
	C-(R_pR_p)di-G_{ps}	E1 = S, E2 = S	Class I Vc2 RNA[a]: 150 ± 33 nM[117] Class II CdA RNA: 3.6 ± 0.6 nM[117]
Wild Type	C-di-GMP	-	Class I Vc2 RNA[b]: 15 ± 1.1 nM[118]
Linear Modification	GpG	Linear	Class I Vc2 RNA[b]: 210 ± 24 nM[118]
	pGpG	Linear	Class I Vc2 RNA[b]: 490 ± 150 nM[118]
	GpGp	Linear	Class I Vc2 RNA[b]: 22,000 ± 14,000 nM[118]

Table 4.1. Binding affinities of c-di-GMP riboswitches with different c-di-GMP analogs. [a]Class I Vc2 riboswitch used here is the G94(2AP) mutant. [b]The class I Vc2 riboswitch measured here is the C92U mutant.

As described in **Section 4.5**, the only similarity between ligand recognition by the class I riboswitch and ligand recognition by the class II riboswitch is the base stacking interactions made between the two guanine bases of the ligand and the

conserved binding pocket residues (**Figure 4.3**). Strobel *et al*. have demonstrated that binding of 7-deaza guanine modified ligands (C-c^{7}G-GMP and C-di-c^{7}GMP, **Table 4.1**) to c-di-GMP riboswitch cost more binding energy than predicted due to deceased base pairing and stacking interactions.[117] To further prove the importance of this base stacking interaction, our groups developed an endo-O modified analog endo-S-c-di-GMP (**Figure 4.5F**, modification of one of the endo-Os to S) to investigate the binding to two classes of riboswitches. Previously, we have proven that endo-S-c-di-GMP favors the 'open' or 'trans' conformation (the two guanines are on opposite sides) in solution rather than the 'closed' or 'cis' confirmation (the two guanines that are on the same side) and it is even three times less likely to populate the closed conformation than opened conformation.[121] However, the ligand in two classes of c-di-GMP riboswitches must be bound in the closed conformation in order to maintain base stacking interactions within the binding pocket nucleotides. Based on these two factors, class I and class II riboswitches bind to endo-S-c-di-GMP much more weakly than they do to the wild type c-di-GMP as we expected (see **Chapter 5**).[119]

Although there are extensive studies of different analogs binding to two classes of c-di-GMP riboswitches, no research on analogs used to specifically bind the subclass of each class of c-di-GMP riboswitch has been reported. Since there are approximately 500 examples of class I c-di-GMP riboswitches that have been computationally identified in a wide range of bacteria and most of them are found in multiple copies in pathogenic bacterial genomes,[113] we have become more interested in distinguishing the various riboswitches in class I c-di-GMP riboswitches. Herein,

our next aim is to utilize different c-di-GMP analogs to discriminate the subclass RNAs in class I riboswitch (see **Chapter 6**). The future optimized analogs are potential chemical and therapeutic tools to study the mechanism of c-di-GMP riboswitch and inhibit the c-di-GMP pathway signaling in specific bacteria in polymicrobial systems.

4.7 Dynamic conformational changes of the class I c-di-GMP riboswitch in response to Mg^{2+} and its ligand

In contrast to its regulator protein effectors, ligand binding to c-di-GMP riboswitches is directly coupled to gene regulation. Analysis of the X-ray structure of the bound form of the class I Vc2 riboswitch has prompted researchers to ask the question: How do ligands reach the deeply buried binding pocket and trigger a conformational change in the expression region to regulate the downstream gene? This interesting and important question has inspired investigation into the dynamic behavior of the riboswitches. It is known that the dynamic changes that occur in RNA structure serve a range of functions, such as regulation of gene expression and signal transduction.[111, 122, 123] Previous studies of RNAs have shown that divalent cations induce tertiary interactions to cause the RNA to fold into compact structures. For example, Mg^{2+} ions often play a particularly important role in enabling the formation and functional stabilization of compact RNA structures.[124] In addition, the high specificity and affinity of the riboswitch aptamer for its ligand is coupled to complex tertiary architectures, involving intra- and intermolecular interactions.[125]

To gain greater insight into the nature of secondary and tertiary structural rearrangement of the c-di-GMP riboswitch in the presence of Mg^{2+} and ligand, different biochemical and biophysical tools have been used.[111, 126] For example, small-angle X-ray scattering (SAXS), a powerful technique for determining global structural information of macromolecules, has been used to explore the folding of the class I Vc2 riboswitch aptamer domain induced by c-di-GMP and Mg^{2+}.[115] The angular range in SAXS contains information on the global shape of free and bound form of Vc2 riboswitch. Low-resolution molecular shapes calculated from the data suggest that in the absence of c-di-GMP, the Vc2 RNA adopts an extended conformation and then binding of c-di-GMP forms the compact conformation.[115] This result is consistent with data from in-line probing experiments, which suggest that the P1 stem is unstructured under these conditions.[102] Additionally, Wood *et al*. analyzed the interplay between Mg^{2+} ion, second-messenger ligand and tertiary interaction in the class I Vc2 riboswitch folding using single molecule fluorescence resonance energy transfer (smFRET).[127] They further confirmed the extended and compact conformations of Vc2 RNA in the absence or in the presence of Mg^{2+} and the ligand (**Figure 4.6**) and revealed that the Vc2 aptamer domain is kinetically partitioned into different folded populations: extended, intermediate and compact.[127] In the presence of Mg^{2+}, the intermediate conformation is stabilized by the interaction of GNRA tetraloop (TL) in P2 and tetraloop receptor (TLR) in P3 and base pairing of C44:G83. Then this Mg^{2+}-dependent dynamic population offers a preorganized structure that allows efficient ligand binding and conformational folding. The stabilization of the folded state with P1 stem formation is only completed upon ligand binding.[127]

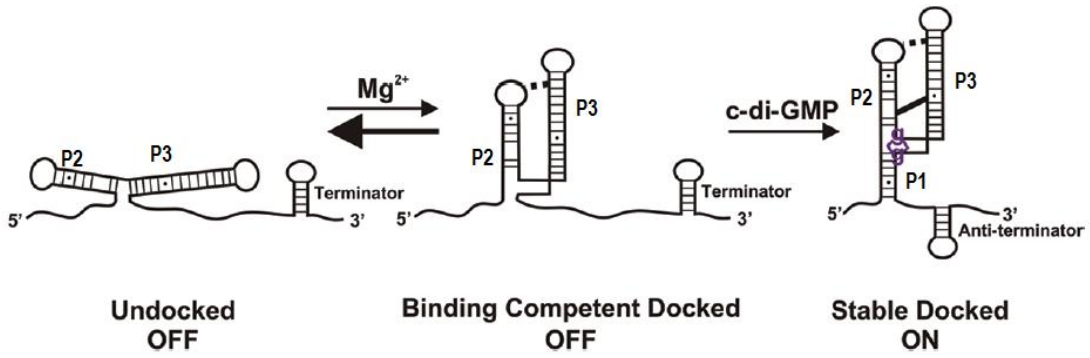


Figure 4.6. Proposed folding pathway of class I Vc2 riboswitch. Figure is adapted from Wood *et al*., 2012.[127]

Although previous smFRET and SAXS studies of the Vc2 riboswitch have suggested a stepwise global rearrangement upon Mg^{2+} ion and c-di-GMP binding,[115, 127] the degree of compaction induced by Mg^{2+} ion or ligand binding can vary even between aptamer domains of the same riboswitch class. For instance, this has been seen in the thiamine-pyrophosphate (TPP) riboswitch from *Escherichia coli* and *Arabidopsis thaliana*.[128] Thus, we are interested in investigating the conformational change and structural difference in our experimentally verified new class I riboswitch from *Clostridium tetani* E88 (Ct-E88) (see **Chapter 7**), especially given the large differences in binding affinities of c-di-GMP analogs versus the known class I Vc2 RNA (**Chapter 6**). Different complementary biochemical and biophysical tools, including selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) analysis to characterize the secondary structures, SAXS to monitor the global conformational changes, and nuclear magnetic resonance (NMR) spectroscopy to map the details of hydrogen-bonding at atomic levels, were used to investigate the details of how Mg^{2+} helps to set up the binding pocket from the extended conformation of the riboswitch and how c-di-GMP finds the binding pocket. After

such a study, we will better understand the mechanism of ligand recognition and how ligand binding drives the regulation of downsteam gene expression.

Chapter 5: C-di-GMP analogs discriminate between different c-di-GMP riboswitches[119]

5.1 Introduction

Recently, our lab and others have been interested in the synthesis of c-di-GMP analogs with the ultimate aim of using these analogs to control the biofilm phenotype via selective or differential modulation of c-di-GMP binding proteins and RNA.[121, 129-132] The majority of c-di-GMP binding proteins and RNAs have not been identified or are poorly characterized. Chemical tools, such as biotinylated c-di-GMP analogs, will therefore be useful for the identification of macromolecules that are involved in the c-di-GMP signaling cascade. It is however not known if the available 2'-biotinylated analogs of c-di-GMP are appropriate tools for the identification of c-di-GMP binding riboswitches. We believe that it is timely for this issue to be addressed. In this chapter, we investigate how 2'-modification as well as conservative phosphate modification of c-di-GMP affects binding to two c-di-GMP riboswitches. We also describe a straightforward synthesis of phosphate and biotinylated analogs of c-di-GMP (compounds **2** and **3**, **Figure 5.1**), using a simple solid phase methodology. Binding studies of c-di-GMP and analogs **2-4** (**Figure 5.1**) with class I and II c-di-GMP riboswitches reveal that the 2'-modification of c-di-GMP may affect binding to them, and hence caution has to be taken when using these capture probes for the identification of new c-di-GMP binding partners.

5.2 Materials and Methods

5.2.1 Materials

2'-O-(6-[Biotinyl]aminohexylcarbamoyl) cyclic diguanosine monophosphate (2'-[Biotin]-AHC-c-diGMP) was purchased from Axxora.

All concentrations of c-di-GMP and its analogs were determined by UV absorbance at 260 nm (ε= 26100 $M^{-1}cm^{-1}$) at room temperature.

5.2.2 Vc2 RNA preparation and purification

Vc2 RNA was prepared and purified as described.[133]

5.2.3 *C. acetobutylicum* RNA preparation and purification

CdA RNA was prepared and purified as described.[133]

5.2.4 Microdialysis assay of Vc2 / CdA RNA with c-di-GMP and analogs

The equilibrium microdialysis assay was carried out in a 2×25 µl chamber (Harvard Apparatus) separated by a 10,000 Da cut-off regenerated cellulose membrane (Harvard Apparatus). Buffer I was used for the assay: 10 mM sodium cacodylate (pH 6.8), 20 mM KCl and 5 mM $MgCl_2$. 25 µl of ligand and 25 µl of RNA (25 µM) were added to chambers A and B respectively. The microdialysis setup was agitated at 4˚C for 24 hours to reach equilibrium before measurements were taken.

5.2.5 UV_{260nm} measurement of c-di-GMP in chamber A after microdialysis

To obtain UV_{260nm} readings, the content of chamber A (25 μL) was transferred to a 1 cm path length cuvette and the absorbance at 260 nm taken, using a JASCO V-630 spectrophotometer.

5.2.6 Molecular Modeling

All dockings were performed as blind dockings (blind docking refers to the use of a grid box which is large enough to encompass the ligand in the binding pocket) using Autodock Vina 1.1.1.[134]

Autodock Vina docking was performed using an exhaustiveness value of 64. All other parameters were used as defaults. All rotatable bonds within the ligand were allowed to flexibly rotate, and the receptor was considered rigid. The bound c-di-GMP and water molecules were first removed from the binding site of the RNA and the analogs were docked onto this site. All ligand structures were created using GaussView 03W software and then brought to their energetically minimized conformers by the Gaussian 03W program utilizing optimization=modredundant hf/6-31g(d) method[135] with different fixed distance between C5'-C5' on two guanines in the ligand. Polar hydrogens were added to ligand and RNA atoms. Autodock Tools version 1.5.4[134] was used to convert the ligand and RNA receptor to the proper file formats for Autodock Vina docking. For each compound (**Figure 5.1**), only the docked structure with the largest binding affinity was assumed to be the predicted ligand bound form to the RNA receptor.

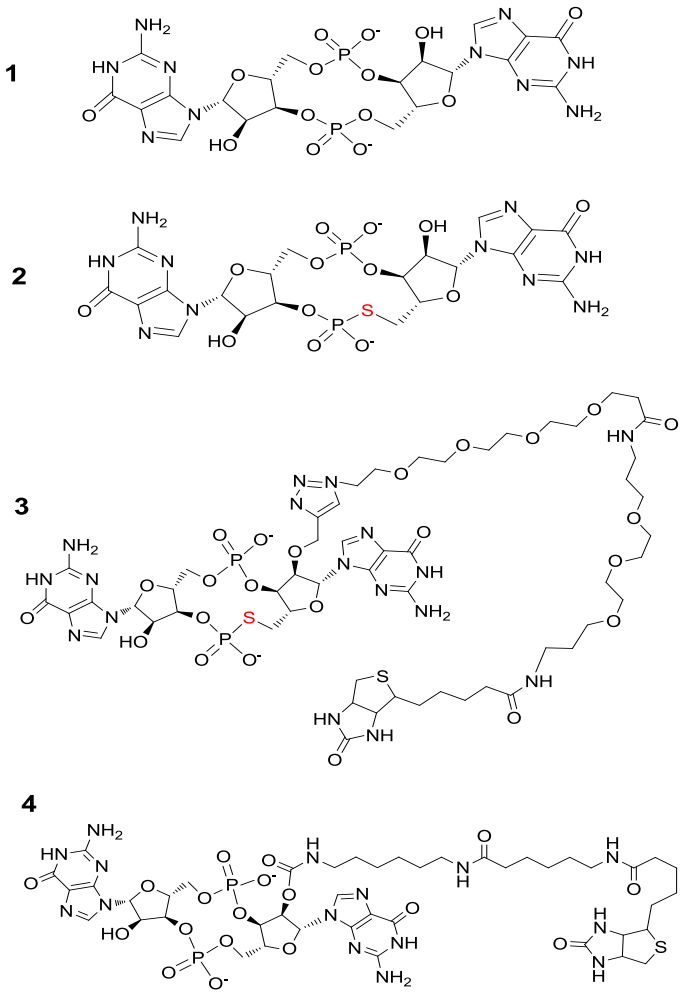


Figure 5.1. C-di-GMP and analogs. **(1)** c-di-GMP. **(2)** endo-S-c-di-GMP. **(3)** 2'-[Biotin]-endo-S-c-di-GMP. **(4)** 2'-[Biotin]-AHC-c-di-GMP (commercially available).

5.3 Results and Discussion

The attachment of a biotin moiety to biological molecules and the use of the biotinylated analogs to identify binding partners are not new in biological studies. However, a critical issue that needs to be addressed when attaching a biotin moiety to any molecule of interest is to determine which site on the molecule can be modified with the biotin fragment without abrogating the binding of the molecule of interest to its binding partners. Due to the importance of c-di-GMP in bacterial signaling, great

interest exists in generating biotinylated analogs of c-di-GMP and the use of these analogs to capture c-di-GMP proteins via affinity chromatography. So far, there is one reported synthesis of biotinylated c-di-GMP, which required eleven steps.[132] In that report, the authors appended the biotin moiety at the 2'-OH position via a click reaction of 2'-alkyne analog of c-di-GMP and biotin azide.[136] We readily prepared biotinylated c-di-GMP containing a bridging sulfur (refer to as 2'-[biotin]-endo-S-c-di-GMP (**3**), **Figure 5.1C**) following reported solid-phase methodology[137] followed by a click reaction with biotin azide (**Figure 5.2**). 2'-[Biotin]-AHC-c-di-GMP (**4**), which contains natural phosphodiester moiety, is commercially available (albeit expensive) and was purchased from Axxora, LLC for this study. We reasoned that investigations into the binding of these c-di-GMP analogs would reveal whether the 2'-position of c-di-GMP could be modified with biotin without affecting binding to c-di-GMP binding macromolecules. For this study, we chose the recently reported c-di-GMP class I & II riboswitches,[103, 113, 114] for investigations.

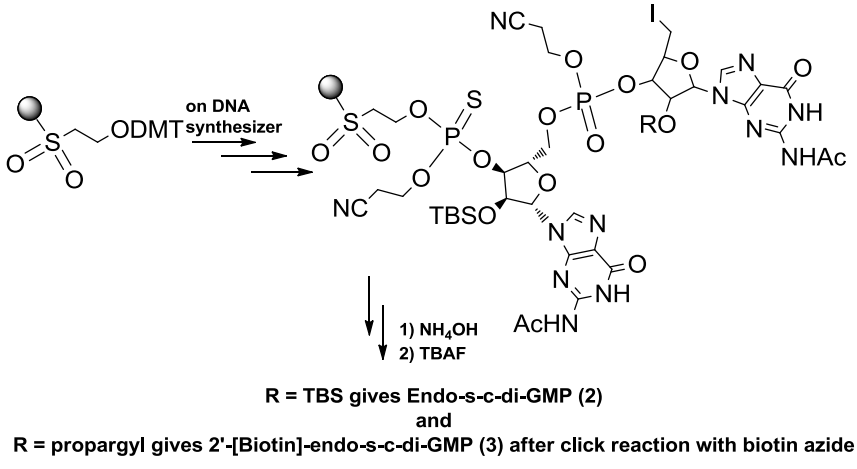


Figure 5.2. Solid-supported synthesis of endo-S-c-di-GMP analogs following procedures reported.[137]

5.3.1 Docking of c-di-GMP and analogs into RNAs

We first used computational methods to investigate the interaction of c-di-GMP and analogs (**1**-**4**) with c-di-GMP class I riboswitch Vc2 RNA (c-di-GMP binds tightly to it with Kd = 10 pM). Autodock has been reliably used to study small ligand-nucleic acid interactions[138] and we decided to use AutoDock Vina for our docking study. C-di-GMP can adopt conformations spanning *cis* and *trans* isomers. In the *cis* conformation, the two guanines are on the same side whereas in the *trans* conformation, the two guanines are on opposite sides (**Figure 5.3A**).

Using the Gaussian program, we fixed the distance between the two guanine nucleobases of c-di-GMP or analogs (C5 of the guanines) at different values (between 5.5 to 9.5 Å; **Appendix 4**) and determined their ground state conformations. To save computational time, simpler 2'-modified analogs of c-di-GMP, (**5**) and (**6**) (**Figure 5.3B**), were used to model the binding of 2'-[Biotin]-endo-S-c-di-GMP (**3**) and 2'-[Biotin]-AHC-c-di-GMP (**4**) in **Figure 5.1** to Vc2 RNA. **Figure 5.3C** shows a plot of binding energy versus distance between the two guanines of c-di-GMP, for the various analogs of c-di-GMP docked into Vc2 RNA.

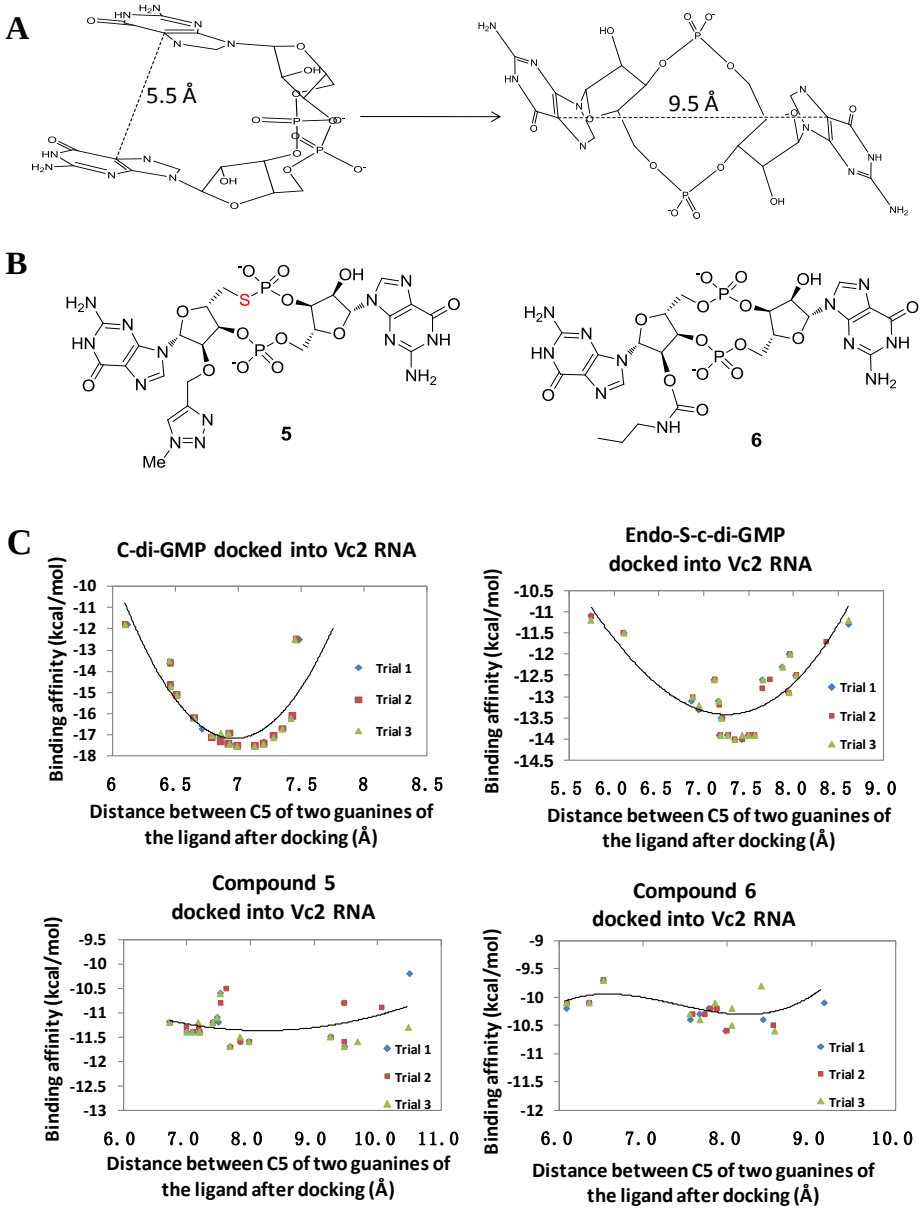


Figure 5.3. (A) The distance between C5 of the two guanines of c-di-GMP were fixed at distances ranging from 5.5 Å to 9.5 Å and the different conformers were optimized using Gaussian. (**B**) Structures of compounds **5** and **6** used as a model for biotinylated analogs **3** and **4**. (**C**) Plots of binding energies versus distance between the two guanine nucleobases of c-di-GMP, endo-S-c-di-GMP, compound **5** or compound **6** after docking into Vc2 RNA (PDB: 3IRW). Dockings were done in triplicate for each c-di-GMP or analog conformer.

Interestingly, the c-di-GMP conformer that docks best to Vc2 RNA has the two guanine bases separated by 6.99 Å, which is remarkably close to the distance found in the crystal structure of c-di-GMP (cf. 6.95 Å), when complexed with Vc2 RNA (PDB:

3IRW). Our docking study reveals that modification to the phosphate moiety of c-di-GMP affects binding to Vc2 RNA (the best docked conformer of c-di-GMP has a binding energy of -17.5 kcal/mol whereas that of endo-S-c-di-GMP is -14.0 kcal/mol). At 25 ℃, these free energies of binding translate into 368 folds tighter binding of the c-di-GMP over endo-S-c-di-GMP.

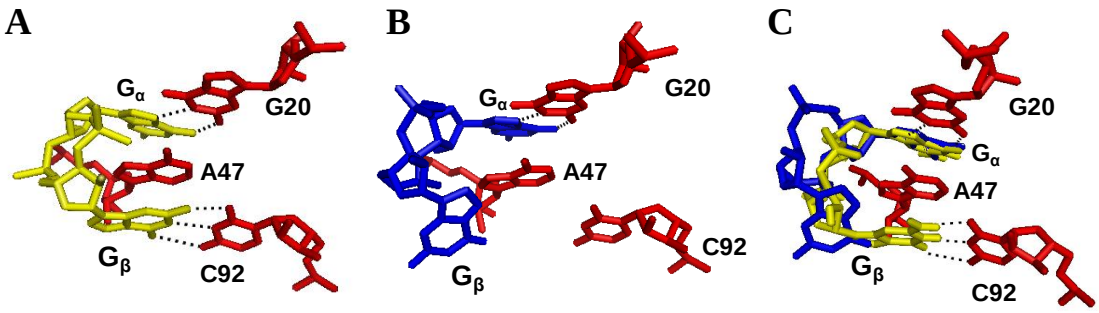


Figure 5.4. Docking of c-di-GMP and endo-S-c-di-GMP in Vc2 RNA. (A) c-di-GMP docked into Vc2 RNA (PDB ID: 3IRW). The interactions in the docked structure are identical to those found in crystal structure of c-di-GMP bound to Vc2[113, 116]. C-di-GMP is colored yellow and the RNA nucleobases are colored yellow. **(B)** Endo-S-c-di-GMP (colored blue) docked into Vc2 RNA; **(C)** Overlay of c-di-GMP and endo-S-c-di-GMP docked into Vc2 RNA. G_β of endo-S-c-di-GMP can not form Watson-Crick hydrogen bonding interactions with C92 of Vc2 RNA.

C-di-GMP makes three important interactions in the binding pocket of Vc2 RNA: Hoogsteen hydrogen bonding interactions between the G_α of c-di-GMP and G20 of the RNA aptamer; a Watson-Crick hydrogen bonding interactions beween G_β and C92 of Vc2 RNA; and a stacking interaction between A47 of the RNA and the two guanines of c-di-GMP (**Figure 5.4A**).[116] Endo-S-c-di-GMP maintains the interactions between the ligand's G_α with G20 of the RNA as well as the stacking interaction of A47 and the guanine nucleobase of the cyclic dinucleotide but not the hydrogen bonding interactions between the G_β and C92 of Vc2 RNA (**Figure 5.4B**). It appears that the substitution of one of the bridging oxygens in the phosphodiester linkages in

c-di-GMP with sulfur affects the relative orientation of the two guanine nucleobases, resulting in the inability of one of the guanine bases in endo-S-c-di-GMP to form an essential hydrogen bonding interactions with C92 of the c-di-GMP riboswitch-I aptamer (**Figure 5.4C**). Endo-S-c-di-GMP is a very close analog of c-di-GMP and can be considered as one of the most conservative modifications of c-di-GMP, therefore our docking results suggest that c-di-GMP riboswitch-I aptamer is very sensitive to modifications on c-di-GMP (**Figure 5.3**).

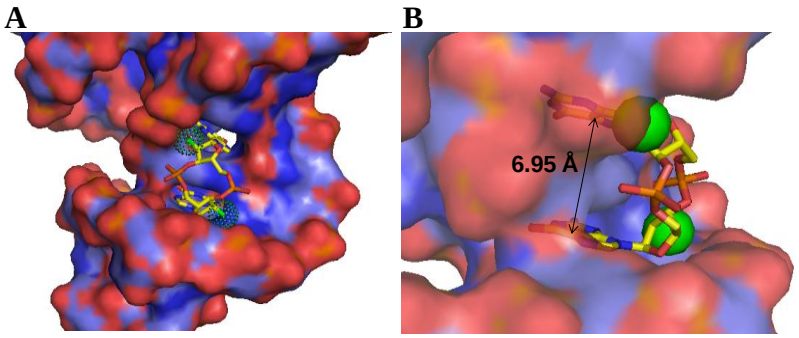


Figure 5.5. C-di-GMP in the binding pocket of Vc2 RNA (PDB: 3IRW). (A) c-di-GMP is shown in stick and colored by atom (oxygen in red, nitrogen in blue, carbon in yellow and the 2'-OH is colored green). Vc2 RNA is shown as mesh surface and colored by atom (oxygen in red, nitrogen in blue and carbon in light blue). **(B)** Zoom in of c-di-GMP in the binding pocket of Vc2 RNA. The green sphere, placed at the 2'-OH position, depicts how groups that are attached to this site in c-di-GMP would sterically clash with residues of Vc2 RNA.

Our docking results also reveal that the modification of the 2'-OH of c-di-GMP completely abrogates binding to Vc2 RNA (compare docking results of compounds (**1**) and (**2**) in **Figure 5.1** with (**5**) and (**6**) respectively in **Figure 5.3**). We rationalize that, in the docked structure of 2'-OH-modified analogs of c-di-GMP, groups that are attached to the 2'-position of c-di-GMP (shown as green sphere in **Figure 5.5B**) would have a steric clash with some of the residues of the Vc2 RNA. It also appears that the two guanine bases are buried into the binding site and could be

inappropriate sites to attach a biotin moiety. These computational studies revealed that the biotinylation of c-di-GMP could potentially render certain analogs ineffective at binding c-di-GMP class I riboswitch.

We next proceeded to dock c-di-GMP into binding receptor c-di-GMP class II riboswitch *C. acetobutylicum* RNA (CdA RNA, PDB: 3Q3Z)[114] (**Appendix 5**). For c-di-GMP class II riboswitch, only c-di-GMP (**1**) but not endo-S-c-di-GMP (**2**) or compound (**5**) or compound (**6**) displayed significant affinity for it. For c-di-GMP, our docking study reproduced experimental observations; the binding affinity of c-di-GMP is higher for binding to Vc2 RNA, compared to CdA RNA (compare **Figure 5.3C** to **Figure 5.6**). This is in agreement with previous work; which reported K_d of 10 pM and 200 pM – 2 nM of c-di-GMP binding to Vc2 and CdA respectively.[113, 114]

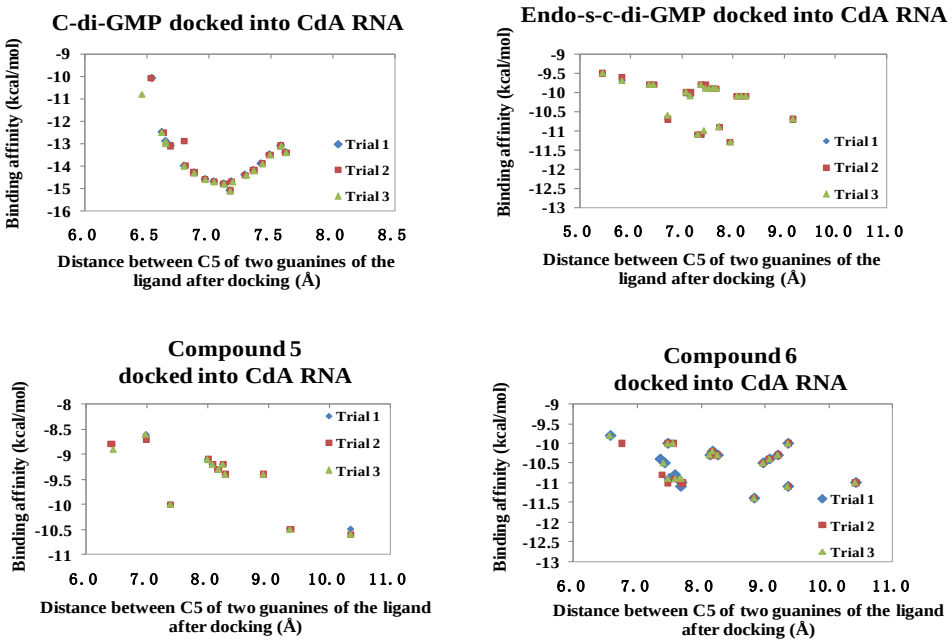


Figure 5.6. Plots of binding energies versus distance between the two guanine nucleobases (after docking) of c-di-GMP, endo-S-c-di-GMP, compound **5** or compound **6** when docked into CdA RNA (PDB: 3Q3Z). Dockings were done in triplicate for each c-di-GMP or analog conformer.

5.3.2 Experimental validation of the docking results

Having predicted, via docking studies, that modifications to the phosphate or 2'-OH positions of c-di-GMP would reduce binding to Vc2 RNA (a c-di-GMP class I riboswitch that binds monomeric c-di-GMP) and CdA RNA (a c-di-GMP class II riboswitch that binds monomeric c-di-GMP), we proceeded to test this experimentally. Equilibrium dialysis is an established method that is used for investigating binding of ligands to proteins and nucleic acids.[139] The binding of c-di-GMP or its analogs to Vc2 RNA were directly measured qualitatively by equilibrium dialysis, as shown in **Figure 5.7A**. C-di-GMP was introduced into the two chambers of a micro equilibrium dialyzer (Harvard Apparatus) that has two 25 µl chambers (A and B), separated by a membrane with a cutoff molecular weight of 10 kDa (**Figure 5.7A**). Different amounts of Vc2 RNA were then added to chamber B of the micro dialyzer and the set up was agitated at 4˚C to reach equilibrium (24 h). Then the concentration of c-di-GMP or analog in chamber A was measured by UV absorbance at 260 nm (UV_{260nm}). In the absence of c-di-GMP or analog binding to the RNA that was introduced in chamber B, the concentration of c-di-GMP in chamber A before and after the introduction of RNA into chamber B should remain constant. However, if c-di-GMP or analogs bind to the RNA that was introduced into chamber B, then the concentration of c-di-GMP in chamber A would decrease as more RNA was added to chamber B. Therefore a plot of the UV absorbance of a sample taken from chamber A at 260 nm versus the concentration of RNA that is added to chamber B (or the RNA:c-di-GMP ratio, **Figure 5.7B**) should give a qualitative information of the relative binding affinities of c-di-GMP or analogs. In complete agreement with our

docking results, c-di-GMP bound to Vc2 RNA best, followed by endo-S-c-di-GMP. Neither 2'-[Biotin]-endo-S-c-di-GMP nor 2'-[Biotin]-AHC-c-di-GMP bind to Vc2 RNA (**Figure 5.7B**).

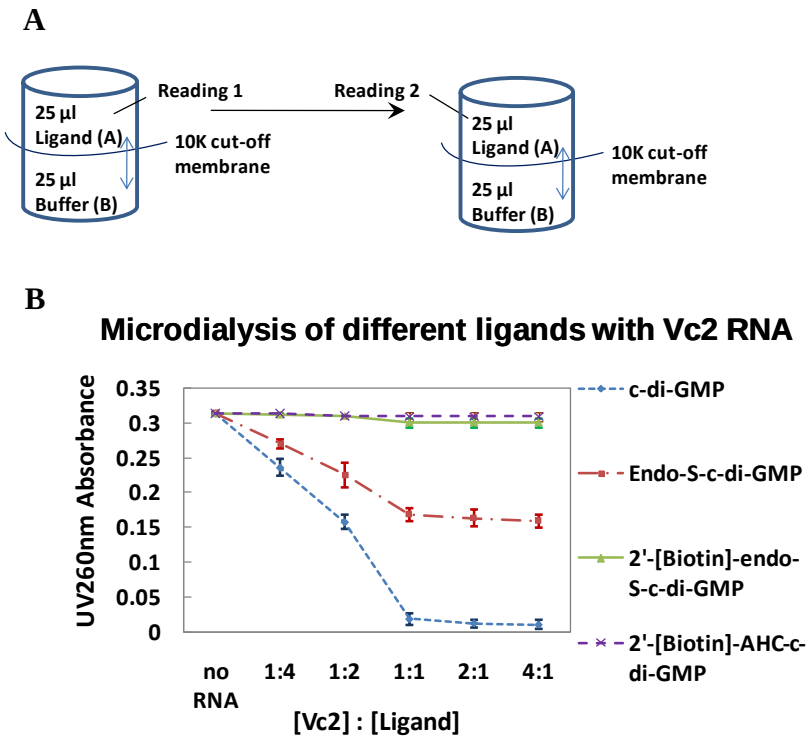


Figure 5.7. Equilibrium microdialysis of c-di-GMP and analogs with Vc2 RNA. (A) Schematic diagram for equilibrium microdialysis. Reading 1 is the UV_{260nm} measurement of ligand in chamber A when no RNA was added to chamber B and reading 2 is the UV_{260nm} measurement of ligand in chamber A when different concentrations of Vc2 RNA were added to chamber B and allowed to be reach equilibrium after 24 h incubation at 4˚C. **(B)** A plot of the concentration of c-di-GMP or analogs in chamber A (measured as UV absorption at 260 nm) versus the amount of RNA added in chamber A (denoted as RNA:ligand ratio). Error bars are generated based on standard deviation between three experiment repeats.

We also performed equilibrium microdialysis of c-di-GMP and analogs with CdA RNA, see **Figure 5.8**. Some interesting trends emerge from the binding studies: CdA RNA (a class II riboswitch) is more sensitive to phosphate modification on c-di-GMP than Vc2 RNA (a class I riboswitch) but the reverse is true for 2'-OH modification, where Vc2 RNA is more sensitive to modification than CdA RNA.[114]

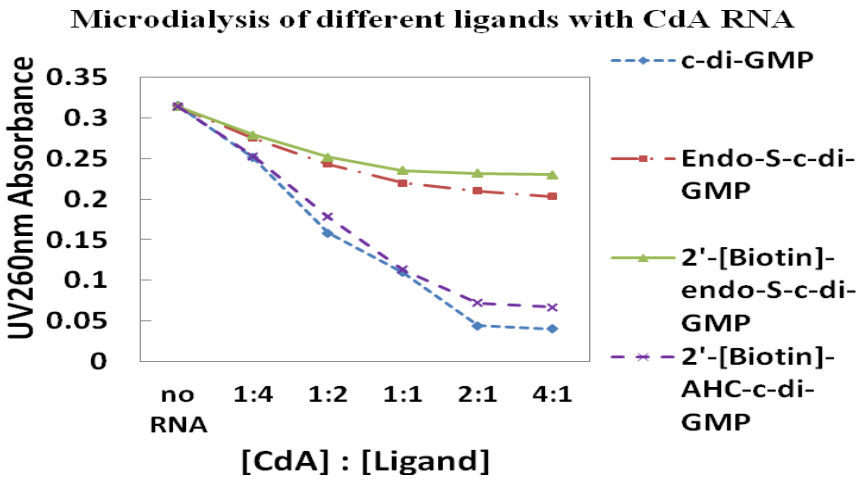


Figure 5.8. Equilibrium microdialysis of c-di-GMP and analogs with CdA RNA. Experimental conditions and analysis were same as **Figure 5.7**.

The origin of the different sensitivities of c-di-GMP riboswitch I and II towards 2'-modification has already been discussed by two excellent papers by the Strobel group.[114, 117] In riboswitch I, the two 2'-OH groups are close to RNA groups and the placement of groups at these positions would cause unfavourable steric interactions (**Figure 5.5B**). In riboswitch II, the two 2'-OH groups do not make as intimate contact with the RNA receptors, as is the case in riboswitch I, so therefore there is room for groups to be appended to this 2'-position (**Figure 5.9**). Thus, it makes sense that biotin substitution of c-di-GMP at the 2'-position does not cause as much disruption to binding to riboswitch class II as it does to class I. The differential binding of the endo-S analogue (**2)** to class I and II is however surprising because both riboswitches make similar interactions with the phosphate moiety of c-di-GMP. In class I riboswitch, A47 of the RNA contacts the exocyclic and not the endocyclic oxygen of one of the phosphate groups in c-di-GMP whereas in class II, this contact is made by A70. Based on these binding interactions, one would expect that there

would be little to no difference between the binding propensities of the endo-S analogs to both class I and class II riboswitches. We have however showed that small changes to c-di-GMP phosphodiester backbone (a thiophosphoroester modification to the bridging oxygen) can perturb the structure of the entire molecule (conformational steering).[121] Plausibly, the endocyclic oxygen-to-sulfur modification at the bridging position in the phosphodiester linkage caused a perturbation that was more detrimental to the various binding interactions of c-di-GMP and RNA in class II than class I.

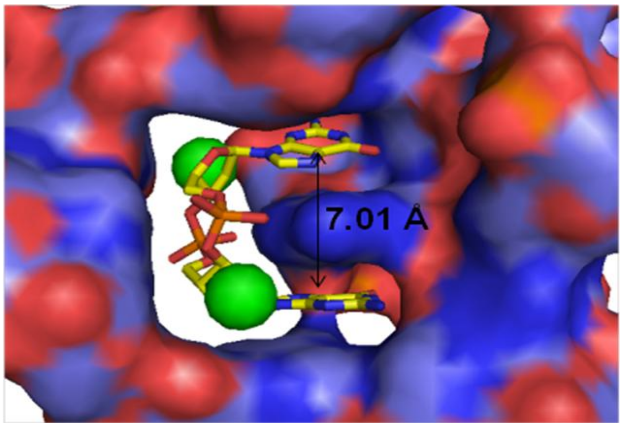


Figure 5.9. C-di-GMP in the binding pocket of CdA RNA (PDB: 3Q3Z). Where c-di-GMP is shown in stick and colored by atom (oxygen in red, nitrogen in blue, carbon in yellow and the 2'-OH is colored green); CdA RNA is shown as mesh surface and colored by atom (oxygen in red, nitrogen in blue and carbon in light blue).The green sphere, placed at the 2'-OH position, depicts how groups that are attached to this site in c-di-GMP would affect ligand recognition.

5.4 Conclusions

C-di-GMP was discovered more than two decades ago by Benziman and co-workers as a regulator of cellulose synthesis in the bacterium *Acetobacter*

xylinum.[67] It took close to twenty years after that discovery to appreciate that c-di-GMP regulates several processes that lead to biofilm formation in bacteria. Several groups have contributed to the understanding of c-di-GMP metabolism by identifying and biochemically characterizing the enzymes that make and degrade c-di-GMP (DGC and PDE respectively). In order to identify c-di-GMP riboswitches, with the ultimate aim of mapping the c-di-GMP interaction with RNAs to better understand global regulation of biofilm formation in bacteria, it is imperative that chemical tools are developed to identify c-di-GMP binding RNAs. In this research, we reveal that the site of biotin attachment on c-di-GMP can seriously affect binding to a class I riboswitch but little effect on a class II riboswitch. Endo-S-c-di-GMP prefers the *trans* conformation to the *cis* conformation, so its binding to both class I and class II riboswitches is decreased due to less stable stacking interaction of the ligand with the binding pocket nucleotides. Furthermore, this work also calls for future studies aimed at finding alternative positions to affix affinity labels on c-di-GMP for the identification of new binding partners.

Chapter 6: Selective binding of 2'-modified c-di-GMP to various class I riboswitches

6.1 Introduction

C-di-GMP functions as an important second messenger in bacteria. Recent studies have shown that many critical biological functions are controlled via c-di-GMP signalling, including exopolysaccharide synthesis, cellular motility, virulence and biofilm formation.[71] Although c-di-GMP has been studied for more than two decades,[67] it is only recently that some of the molecular details of the interactions of c-di-GMP with macromolecular targets have been uncovered. Until the recent discovery by Breaker and co-workers that c-di-GMP binds to RNA (riboswitches) to regulate gene expression and translation,[102] it had largely been assumed that c-di-GMP only binds to effector proteins, such as transcription factors, to regulate bacterial physiology.[88, 90-96, 102] The discovery of these c-di-GMP responsive riboswitches expands the role of RNA in the c-di-GMP signaling pathway. Thus, inhibiting these c-di-GMP riboswitches with small molecules could potentially lead to the inhibition of biofilm formation or virulence factor production in bacteria. Motivated by the potential promise of c-di-GMP riboswitch inhibitors, we and others have embarked on structure-activity-relationship (SAR) studies of c-di-GMP binding to riboswitches.[114, 117, 119, 120] To this end, we demonstrated that a 2'-biotinylated analog of c-di-GMP could bind to a class II c-di-GMP riboswitch, CdA RNA from *Clostridium acetobutylicum*, but not a class I riboswitch, such as Vc2 RNA from *Vibrio cholerae*.[119] Others have also shown that many other different c-di-GMP

analogs can have differential binding to class I riboswitch (Vc2) and class II riboswitch (CdA).[114, 117, 120]

Given these previous promising results and the preponderance of C-di-GMP class I and class II riboswitches, we are interested in the identification of small molecules that can discriminate between different c-di-GMP class I riboswitch. C-di-GMP class I riboswitch can be found in both symbiotic and pathogenic bacteria[113] and therefore being able to discriminate between class I riboswitches, found in symbiotic and pathogenic bacteria, will be critical for the maintenance of natural microbiota during antibacterial therapy. However, the paucity of structural data for different c-di-GMP class I riboswitches makes it difficult to design selective binders of c-di-GMP class I riboswitches *a priori.* We therefore investigated the SAR of c-di-GMP binding to different class I riboswitch (**Figure 6.1**), using c-di-GMP analogs (see **Figure 6.2** for structures of analogs) to determine if different c-di-GMP class I riboswitches use different recognition elements for binding to c-di-GMP.

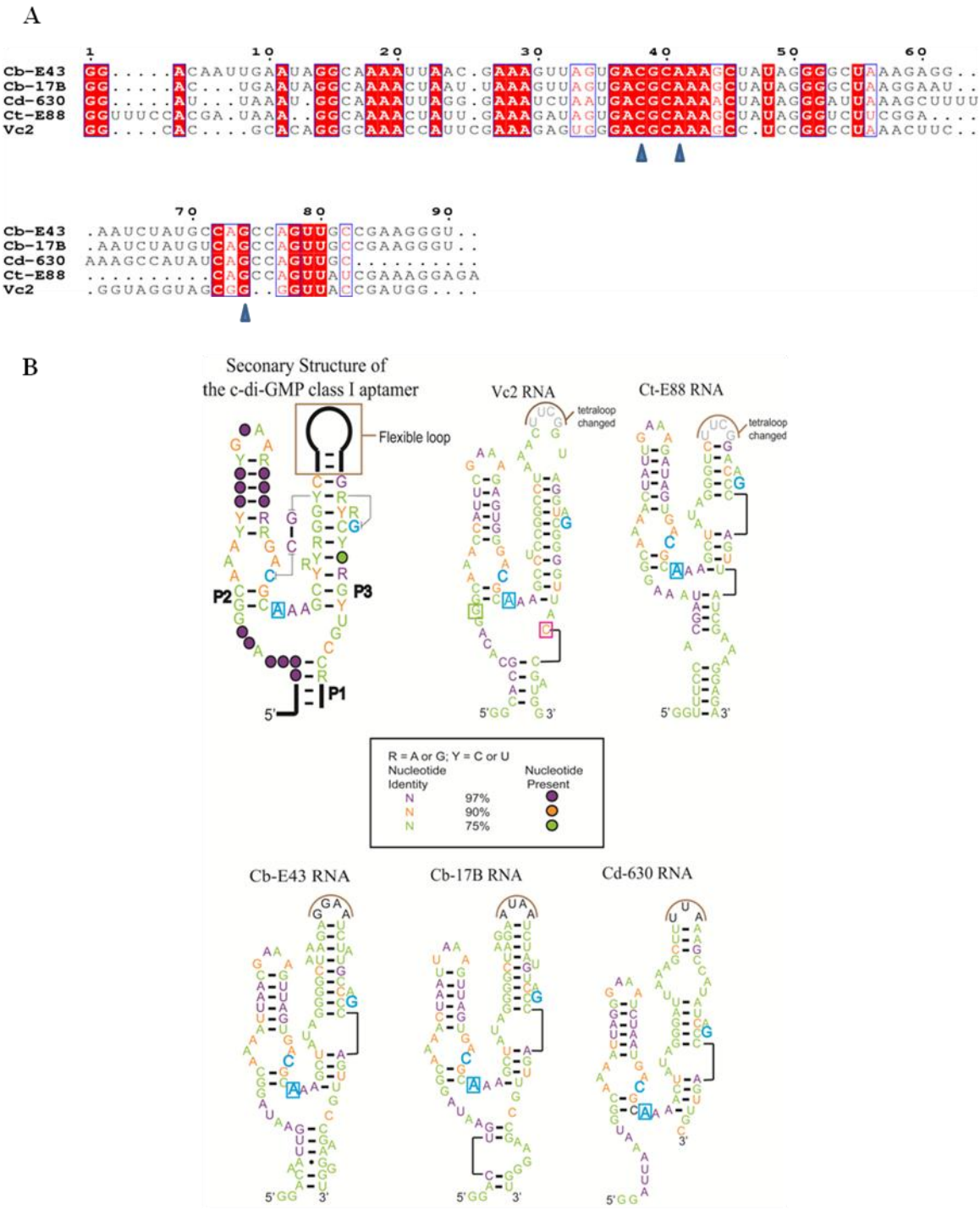


Figure 6.1. (A) Alignment of predicted c-di-GMP riboswitch sequences. The nucleotides indicated by blue arrows are the same nucleotides indicated in blue in the predicted structures. Alignment was done using Clustal W2[140] and ESPript[141]. **(B)** Predicted structures of the selected c-di-GMP-I riboswitch aptamers based on structure model of the c-di-GMP class I aptamer.[113] Cytidine on P2 stem in blue and guanine on P3 stem in blue are expected to form an interhelical Watson-Crick base pairing in the ligand bound structure and adenosine in blue box is expected to intercalate into two bases of c-di-GMP in the bound form. In Vc2 structure,

cytidine and guanosine indicated in the boxes are the two important nucleotides base pairing with the two bases of c-di-GMP.

6.2 Materials and Methods

6.2.1 Preparation of c-di-GMP class I Riboswitch Aptamer RNAs

The aptamer sequences of Vc2 RNA, Ct-E88 RNA, Cb-17B RNA, Cb-E43 RNA and Cd-630 RNA were obtained from the Rfam database.[142] For riboswitches that contained flexible loops on P3 longer than four nucleotides, this loop was changed to a UUCG tetraloop to facilitate an easy synthesis of the RNA. A pUC57 plasmid containing each target sequence, including T7 promoter sequence, was constructed by GenScript and the RNAs were prepared and purified as previously described.[119]

6.2.2 Microdialysis assay of RNA with ligand

Microdialysis was used to qualitatively gain insights into the binding between the various c-di-GMP class I riboswitches and c-di-GMP or analogs. The microdialysis assay was carried out in two 25 μL chambers (HARVARD APPARATUS) separated by a 10,000 Da cut-off regenerated cellulose membrane (HARVARD APPARATUS). 25 μL of ligand and 25 μL of RNA in the same buffer (20 mM KCl, 6 mM $MgCl_2$, and 10 mM sodium cacodylate, pH 6.8) were added to chamber A and B respectively. Equilibration was assumed to have occurred after 24 h, having left the microdialysis apparatus at room temperature (with slight shaking). UV_{260nm} readings of the 25 μL of c-di-GMP or analog in chamber A were taken after 24 h of equilibration, using JASCO V-630 spectrophotometer with 1 cm path length cuvette at 25 ℃.

6.2.3 Preparation of ^{32}P-labeled c-di-GMP

E. Coli BL21 (DE3) was used for production of WspR (D70E) from pVL791. The bacteria were grown at 30 ℃ in Luria-Bertani (LB) medium with shaking. IPTG was added at $OD_{600\ nm}$ = 0.6 and protein expression was induced for 6 h. The protein, which has a His-tag, was purified on a nickel resin and dialyzed into a 10 mM Tris, 100 mM NaCl solution. α-^{32}P-c-di-GMP was generated from purified WspR (D70E) protein as described.[94]

6.2.4 K_d measurements of c-di-GMP and analogs by gel-shift

RNA was folded by heating to 70 ℃ and slow cooling in folding buffer (20 mM KCl, 6 mM $MgCl_2$, and 10 mM sodium cacodylate, pH 6.8) in the presence of 50 pM of radiolabeled c-di-GMP. Binding reactions were incubated at room temperature (25 ℃) until equilibrium was assumed to have been reached (24 h for different class I riboswitches). The dissociation constant (K_d) of c-di-GMP for each class I riboswitch was then determined using a gel-shift assay. Unbound c-di-GMP was separated from RNA/c-di-GMP complex using 10% native PAGE (100 mM Tris/HEPES pH 7.5, 0.1 mM EDTA, 1 mM $MgCl_2$) at 4 ℃. Gels were dried and scanned using a STORM phosphorimager (Molecular Dynamics). For each ligand concentration, the bands corresponding to free and bound c-di-GMP were boxed and quantitated using ImageQuant software (Molecular Dynamics). The fraction of bound c-di-GMP was calculated using Equation (1).

$$FB = I_B / (I_F + I_B) \qquad \text{Equation (1)}$$

Where FB = fraction of c-di-GMP bound; I_B = intensity of bound c-di-GMP band; I_B = intensity of free c-di-GMP band;

Then all data was fit to the following Equation (2) to determine K_d values[143]:

$$FB = \frac{([RNA]_{tot}+[P32cdGMP]+Kd)}{2\times[P32cdGMP]} - \frac{\sqrt{([RNA]_{tot}+[P32cdGMP]+Kd)^2-4\times[RNA]_{tot}\times[P32cdGMP]}}{2\times[P32cdGMP]}$$

Equation (2)

Where FB = fraction of c-di-GMP bound; [P32cdGMP] = concentration of α-^{32}P- c-di-GMP, which is at 50 pM; K_d = dissociation constant of c-di-GMP for the riboswitch RNA; $[RNA]_{tot}$ = total concentration of riboswitch RNA.

Competition experiments to determine the K_d values of analogues were performed as described above. In this case, P-32 c-di-GMP and unlabeled competitor analog were premixed in folding buffer before adding RNA to a final concentration of 100 nM of RNA. RNA was heated to 70 ℃ for 3 min and slowly cooled in the presence of both labeled and unlabeled ligand and incubated at room temperature for 48-72 h before resolving free and bound c-di-GMP by 10% native PAGE. Then the fraction bound (FB) of labeled c-di-GMP was quantified and the Kd value of the unlabeled competitor analogue was determined from quadratic solution of the Lin and Riggs equation determined by Weeks and Crothers[144, 145], Equation (3):

$$FB = FB_\infty + \frac{FB_o}{2\times[P32cdGMP]} \times \{K_d^{cdGMP} + \frac{K_d^{cdGMP}\times[analog]}{K_d^{analog}} + [RNA]_{tot} + [P32cdGMP] - [\left(K_d^{cdGMP} + \frac{K_d^{cdGMP}\times[analog]}{K_d^{analog}} + [RNA]_{tot} + [P32cdGMP]\right)^2 - 4\times[RNA]_{tot}\times[P32cdGMP]]^{\frac{1}{2}}\}$$

Equation (3)

Where FB = fraction of c-di-GMP bound; FB_∞ = fraction bound in the absence of c-di-GMP competitor; FB_0 = fraction bound in the presence of c-di-GMP competitor; [P32cdGMP] = concentration of P-32 labeled c-di-GMP, which is at 50 pM; K_d^{cdGMP} = dissociation constant

of c-di-GMP for the riboswitch RNA; [analog] = concentration of unlabeled competitor analog; $[RNA]_{tot}$ = total concentration of riboswitch RNA; K_d^{analog} = dissociation constant of competitor analog.

The binding energy (ΔG_{bind}) of c-di-GMP or its analog was calculated from the following equation:

$$\Delta G_{bind} = RT\ln K_d \qquad \text{Equation (4)}$$

Where R = universal gas constant (1.985×10^{-3} $kcalK^{-1}mol^{-1}$) and T = temperature (295K).

The change in binding energy ($\Delta\Delta G_{bind}$) was determined according to

$$\Delta\Delta G_{bind} = \Delta G_{bind}\,(analog) - \Delta G_{bind}\,(c\text{-}di\text{-}GMP) \qquad \text{Equation (5)}$$

6.3 Results and Discussion

Although over 500 examples of c-di-GMP-I class riboswitch RNAs have been predicted computationally,[113] only two have been experimentally verified to bind to c-di-GMP. We selected four c-di-GMP-I class riboswitch aptamer candidates (Ct-E88, Cb-17B, Cb-E43 and Cd630 RNAs) from GEMM motif in Rfam database (**Figure 6.1**).[142] These putative c-di-GMP riboswitches are found in different pathogenic bacteria that cause human disease, including *Clostridium tetani*[146] (Ct-E88), *Clostridium botulinum*[147] (Cb-17B and Cb-E43) and *Clostridium difficile*[148] (Cd-630). Vc2 RNA, from *Vibrio cholerae*, which has been biochemically and structurally characterized as a c-di-GMP binding riboswitch, was used as a positive control.[102, 113, 115]

C-di-GMP-I class riboswitch aptamers are predicted to form a three-helix Y-shaped structure, with the ligand-binding site at the junction of helices P1, P2 and

P3. The crystal structure of Vc2/c-di-GMP complex reveals that 5'- and 3'- flanking nucleotides extended the putative P1 helix, to include a Waston-Crick G:C base pair between C92 residue of the RNA (boxed red in Vc2 model, **Figure. 6.1B**) and one of two guanosines of c-di-GMP (G_β). Helices P2 and P3 are parallel to each other in the presence of Mg^{2+}, while helix P1 only forms when c-di-GMP is present.[127] Residue G20 (boxed green in Vc2 model, **Figure 6.1B**) in helix P2 of Vc2 contacts the top nucleobase of c-di-GMP (G_α) along its Hoogsteen face in the bound form. The highly conserved adenosine A47 (shown in boxed blue in **Figure 6.1B**) intercalates between the two nucleobases of c-di-GMP, resulting in extensive stacking interactions that bridge the P1 and P2 helices. Several nucleotides are conserved in the selected class I aptamers (**Figure 6.1A**) and the predicted structures of the selected riboswitches are similar, so we would expect that they all bind to c-di-GMP.

6.3.1 Experimental validation of selected class-I RNA aptamer binding to c-di-GMP

To verify if the predicted class I riboswitches could bind to c-di-GMP, equilibrium microdialysis experiment was performed for each RNA (Vc2, Cd630, Cb-17B, Ct-E88 and Cb-E43) with different ligands. Equilibrium dialysis is a simple but effective assay to quickly and qualitatively determine if a ligand binds to a macromolecular target.[117, 119] One of the advantages of the equilibrium dialysis assay is that sampling is done at equilibrium so it is possible to detect even the low affinity interactions that are difficult to detect using non-equilibrium methods, such as gel shift assays. Herein, the binding of each class I aptamer to c-di-GMP was directly

measured qualitatively by this method as described previously.[119] A plot of UV260nm of ligand in chamber A versus the concentration of RNA in chamber B measures the relative binding affinities of selected c-di-GMP class I with c-di-GMP (**Figure 6.3B**). The equilibrium dialysis results showed that c-di-GMP could bind to Cb-E43, Ct-E88 and Vc2 RNAs with similar affinities, whereas Cb-17B and Cd-630 RNAs had relative low binding affinities to c-di-GMP (**Figure 6.3**). We postulate that the low binding affinity of Cd-630 or Cb-17B RNA with c-di-GMP is likely due to the lack of a P1 helix in these two RNAs (**Figure 6.1**). Our microdialysis data suggests that Cb-E43 and Ct-E88 are likely to be functional c-di-GMP riboswitches in *Clostridium tetani* (Ct-E88) and *Clostridium botulinum* (Cb-E43).

6.3.2 Recognition of 2'-position modified c-di-GMP analog

Having identified Ct-E88 and E43 as potential functional c-di-GMP riboswitches, we next sought to investigate further their binding modes. Previous reports had demonstrated that modification of the 2'-position of c-di-GMP differentially affected binding to different classes of c-di-GMP riboswitches. The X-ray crystal structure of c-di-GMP/Vc2 complex indicated that one of the 2'-OH of c-di-GMP interacts with a nonbridging phosphate oxygen of A47 of Vc2 and this might partly explain why c-di-GMP class I riboswitches are sensitive to 2'-modification of c-di-GMP.[114, 117, 119]

However, different c-di-GMP class I riboswitches have different residues in the predicted ligand binding pocket. Thus we hypothesized that discrimination between these riboswitches, which belong to the same class, could be achieved with c-di-GMP analogs that provide different puckering modes, or hydrogen bonding capabilities, or

both, at the 2'-position. To test this hypothesis, we synthesized three different c-di-GMP analogs with the 2'-position modified, including 2'-F-c-di-GMP (**2**), 2'-H-c-di-GMP (**3**) and 2'-OMe-c-di-GMP (**4**) (**Figure 6.2**) and investigated the binding affinities of these analogs for different c-di-GMP class I aptamers (Vc2, Cd630, Cb-17B, Ct-E88 and Cb-E43) using microdialysis.

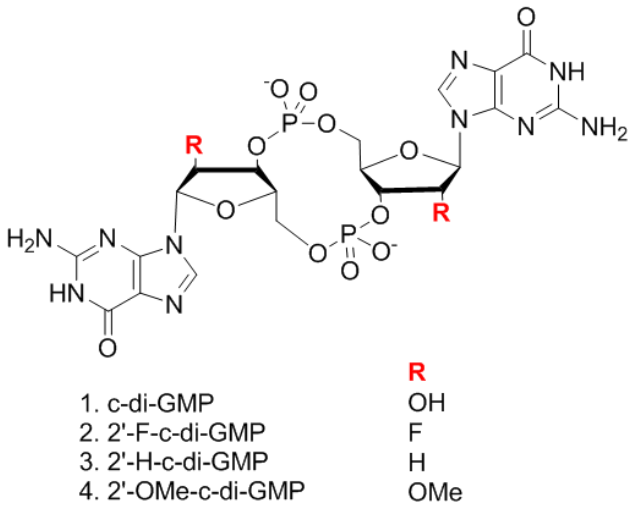


Figure 6.2. Family of c-di-GMP analogs designed and synthesized.

Our microdialysis assay (**Figure 6.3**) indicates that c-di-GMP binds to Vc2, Ct-E88 and Cb-E43 RNAs with similar affinities, but these RNAs bind to different c-di-GMP analogs with different affinities. For example, it appears that Vc2 and Cb-E43 bind to 2'-F-c-di-GMP better than Ct-E88 (**Figure 6.4A**). Cb-E43 also binds to 2'-H-c-di-GMP slightly better than the other RNAs (Vc2, Ct-E88, Cd630 and Cb-17b, see **Figure 6.4B**). All of the tested c-di-GMP class I riboswitches are sensitive to steric encumbrance at the 2'-position because none of the RNAs bind to the 2'-OMe-c-di-GMP (see **Figure 6.4C**).[119]

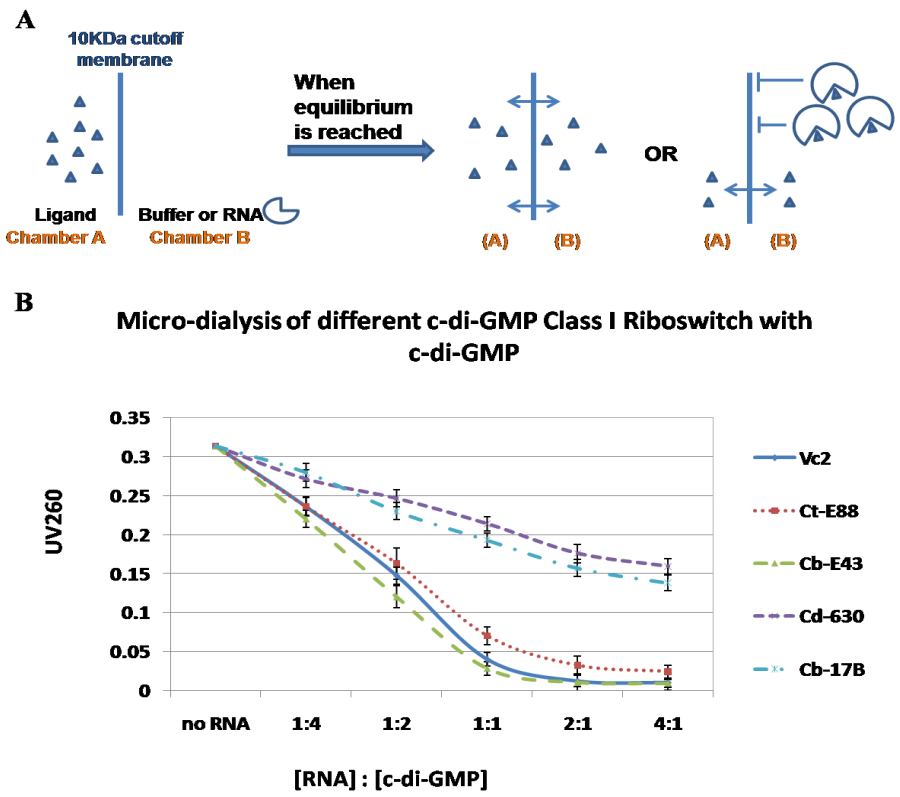


Figure 6.3. Equilibrium Microdialysis of selected c-di-GMP class I riboswitch aptamer with ligand. **(A)** Schematic diagram for equilibrium microdialysis. **(B)** A plot of the concentration of c-di-GMP in chamber A versus the amount of different c-di-GMP class I riboswitch aptamer added in chamber B. Error bars are generated based on standard deviation between three experiment repeats.

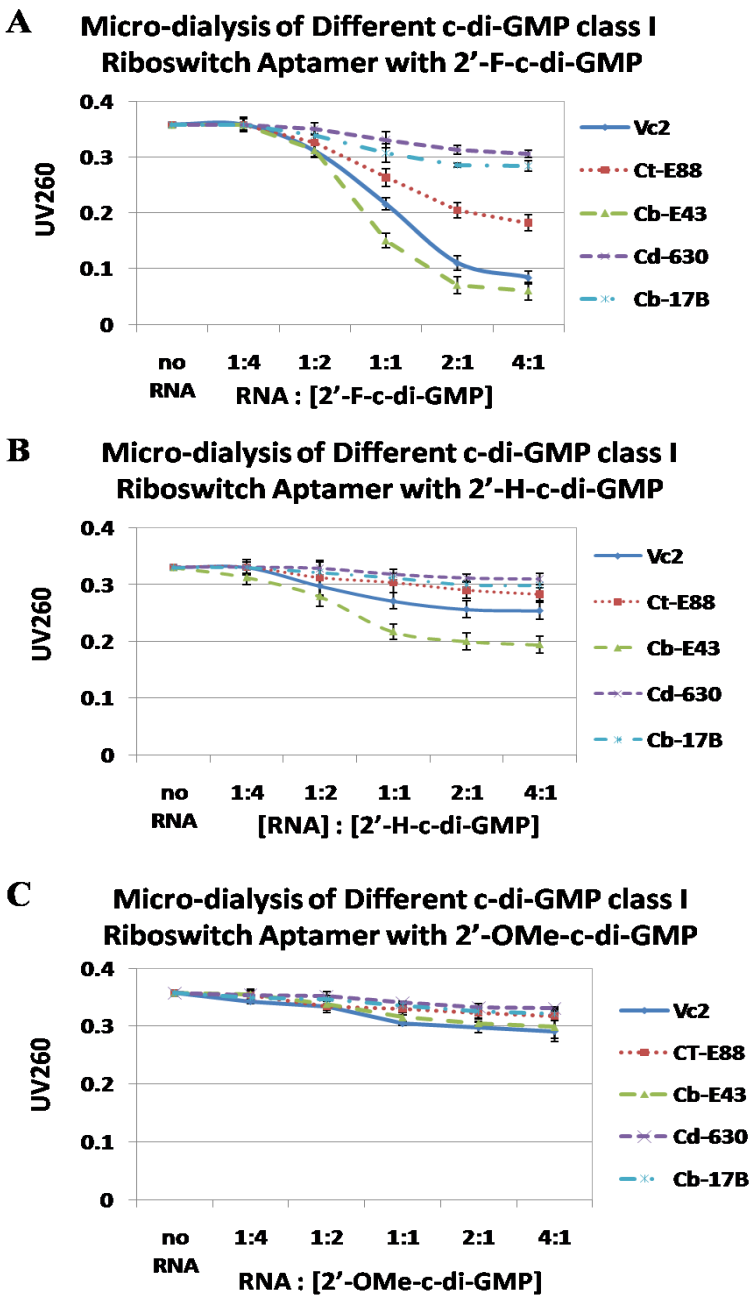


Figure 6.4. Equilibrium microdialysis of c-di-GMP analogs. **(A)** 2'-F-c-di-GMP, **(B)** 2'-H-c-di-GMP and **(C)** 2'-OMe-c-di-GMP with various class I riboswitch aptamers.

Strobel *et al*. have suggested that the deleterious effects of 2'-deoxy or 2'-methoxy modifications are probably due to their inability to maintain a favored 3'-endo conformation.[117] The 2'-fluoro modification however maintains this favoured

conformation, and this might explain why this particular analog, but not the deoxy or methoxy analogs, binds to Vc2, Ct-E88 and Cb-E43 RNAs.

Future structural studies should help delineate the factors that contribute to the differential binding of c-di-GMP and analogs to these tested RNAs. Our biochemical data suggests that although the c-di-GMP class I riboswitches use some common binding determinants (such sensitivity to steric bulk at 2'-position), there are unexpected differences in the binding modes with c-di-GMP and these differences could be utilized for the development of selective small molecules that disrupt c-di-GMP signaling in some bacteria and not others.

6.3.3 Affinity measurements of c-di-GMP and analogs for different class I aptamer

Although the microdialysis assay provided a qualitative ordering of c-di-GMP/analog binding to the tested RNAs, to obtain quantitative binding constants required electrophoretic mobility shift assay (EMSA). We therefore used gel-shift assay to determine the K_d for c-di-GMP and the various class I riboswitches. The binding affinity of various class I riboswitch aptamer to c-di-GMP was measured by titrating a radiolabeled c-di-GMP probe with increasing concentration of RNA. The fraction of radiolabeled c-di-GMP that was bound was determined from the gel shift (**Figure 6.5**) and the data was fitted to a quadratic Equation (2) (see *Materials and Methods*) to obtain the K_d (**Figure 6.5** and **Table 6.1**).

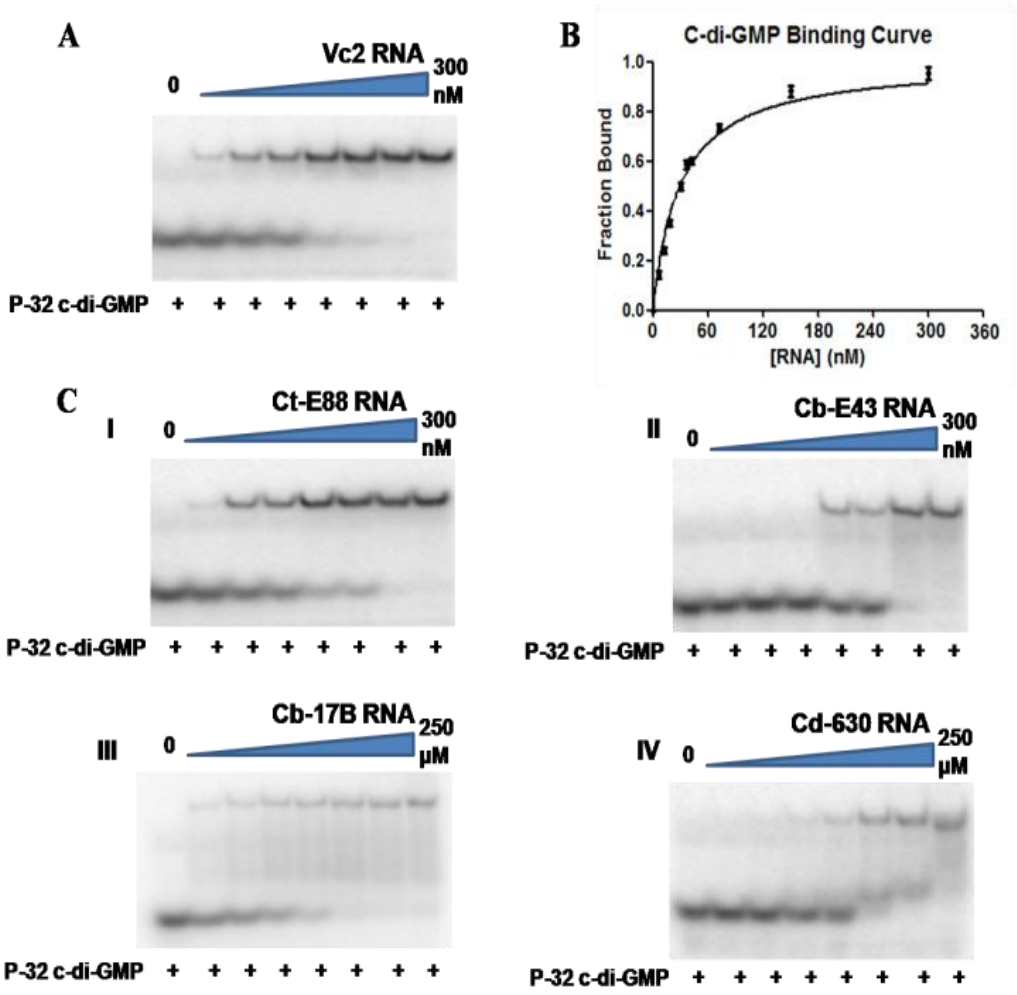


Figure 6.5. K_d measurement of c-di-GMP by gel-shift assay, using radiolabeled c-di-GMP. **(A)** Representative gel-shift experiment for measuring the K_d of c-di-GMP with Vc2 RNA by direct binding. **(B)** Representative c-di-GMP binding curve for class I aptamer RNA. **(C)** Gel-shift experiments for measuring the dissociation constants (K_d) of c-di-GMP with class I aptamers (Ct-E88, Cb-E43, Cb-17B and Cd-630 RNA). (See **Appendix 6** for each binding curve fitting data)

RNA	Ligand	K_d (nM)	$\Delta\Delta G_{bind}$ (kcal/mol)	Fold Loss
Vc2	c-di-GMP	47.2 ± 3.7		
	2'-F-c-di-GMP	176.0 ± 6.7	0.8	3.7
	2'-H-c-di-GMP	8560 ± 980	3.0	180

	2’-OMe-c-di-GMP	> 1 mM	> 5.8	> 21,000
Ct-E88	c-di-GMP	60.3 ± 5.3		
	2’-F-c-di-GMP	44700 ± 9410	3.9	741
	2’-H-c-di-GMP	52130 ± 5720	4.0	863
	2’-OMe-c-di-GMP	n.b.[a]		
Cb-E43	c-di-GMP	128.3 ± 18.0		
	2’-F-c-di-GMP	996.0 ± 10.1	1.2	7.7
	2’-H-c-di-GMP	12200 ± 6200	2.6	95.3
	2’-OMe-c-di-GMP	n.b.[a]		
Cb-17B	c-di-GMP	> 1 μM		
Cd-630	c-di-GMP	> 1 μM		

Table 6.1. Binding affinities of c-di-GMP and analogs for five class I riboswitches: Vc2, Ct-E88, Cb-E43, Cb-17B and Cd-630 RNAs. [a]No detectable binding at highest ligand concentration (600 μM) tested.

In agreement with the equilibrium microdialysis results, c-di-GMP bound to Vc2, Cb-43 and Ct-E88 RNAs tightly with nanomolar K_d values (**Table 6.1**), whereas no binding was observed for Cb-17B and Cd-630 RNAs, even at ligand concentrations of 1 μM or more. To measure the affinities of the unlabeled c-di-GMP analogs, a competition gel-shift assay with radiolabeled c-di-GMP was used since the preparation of radiolabeled c-di-GMP analogs is non-trivial. Class I aptamer and labeled c-di-GMP were incubated in the presence of increasing concentrations of the unlabeled competitor analogs until equilibrium was achieved. The fraction of free radiolabeled c-di-GMP, which was displaced from the bound riboswitch was

monitored to determine the affinity of the competitor (**Figure 6.6** and **Table 6.1**).[114, 117]

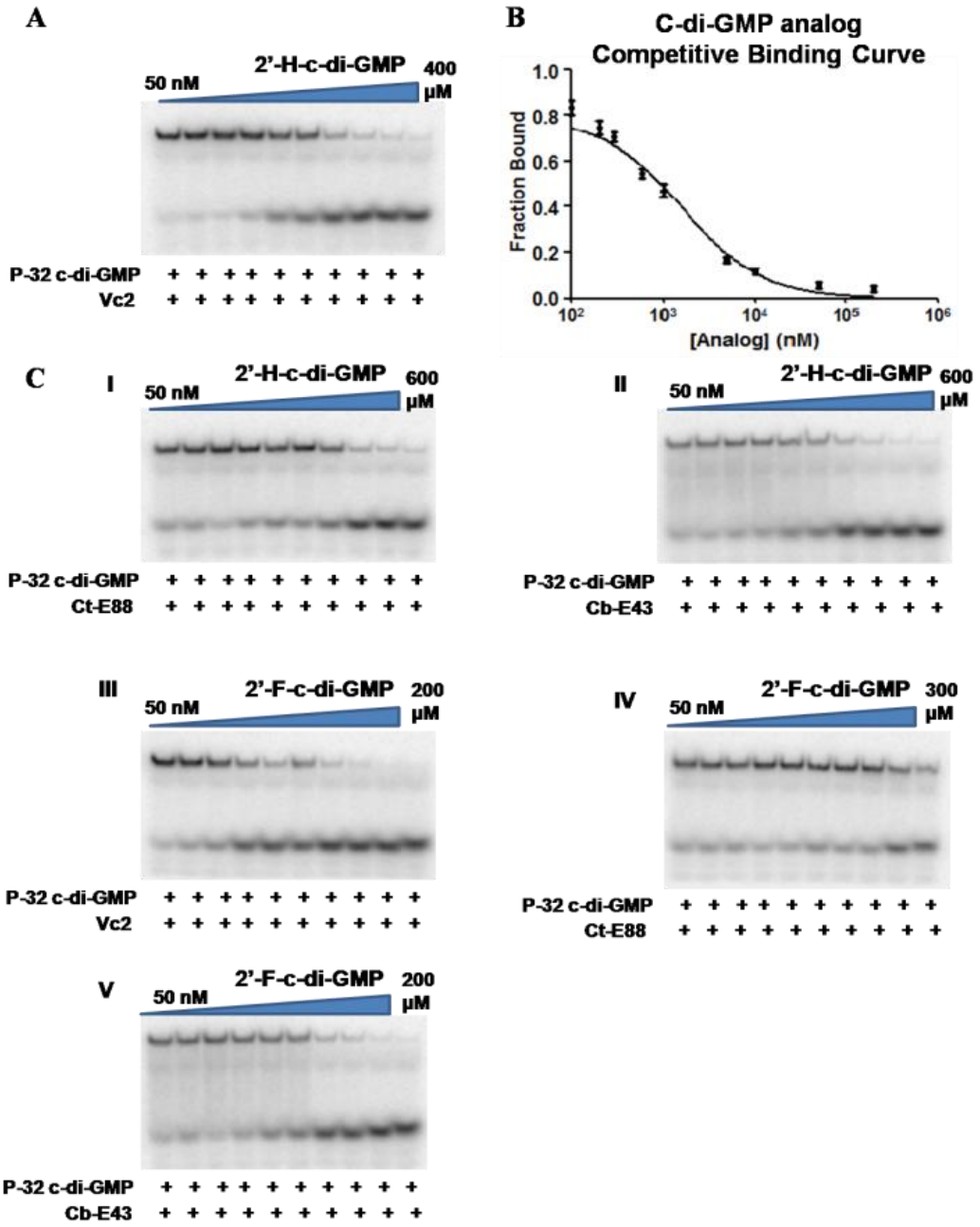


Figure 6.6. K_d measurement of c-di-GMP analogs using competition gel-shift assay with radiolabeled c-di-GMP. **(A)** Representative competition gel-shift experiment of Vc2 RNA with 2'-H-c-di-GMP. **(B)** Representative binding curve from the competition gel-shift assay. **(C)** Gel-shift experiments for measuring the Kd of 2'-H-c-di-GMP and 2'-F-c-di-GMP for each class I aptamer **(**Vc2, Ct-E88, Cb-E43 RNA) by competition assay. RNA, radiolabeled c-di-GMP, and increasing concentrations of competitor analog are incubated at room

temperature until equilibrium. Free P-32 c-di-GMP is separated from RNA-bound c-di-GMP by 10% native PAGE. (See **Appendix 6** for each binding curve fitting data)

Our data suggests that deletion of the 2'-OH group in c-di-GMP to give the 2'-H-c-di-GMP was detrimental for class I riboswitch binding. The loss in binding energy was highest for Ct-E88 RNA (4.0 kcal/mol loss in binding energy), compared to Vc2 and Cb-E43 RNA (loss of 3.0 kcal/mol and 2.6 kcal/mol binding energies respectively). The loss in binding energy seen with 2'-H-c-di-GMP could be due to an abrogation of hydrogen bonding interactions between the RNA and c-di-GMP. Alternatively, it is also possible that replacing the 2'-OH group with hydrogen affects the ribose puckering, which then affects the relative orientation of the different moieties, including the nucleobases of c-di-GMP, to offset important interactions between c-di-GMP and RNA. To differentiate the sugar pucker effect from hydrogen bonding effect in c-di-GMP binding to these RNAs, the binding affinities of 2'-F-c-di-GMP to Vc2, Ct-E88 and Cb-E43 RNAs were measured. The fluoro group is a good isostere for the OH group and it has been demonstrated that 2'-fluoro nucleotides also exist in the 3'-endo sugar puckering mode.[149] Although Fluorine is more electronegative than oxygen, when bonded to carbon, fluorine is not as good a hydrogen bond acceptor as oxygen is.[150] Secondly, the OH group can act as a donor or acceptor of hydrogen bond whereas the fluoro group can only accept a hydrogen bond. The loss in binding energy when the 2'- OH in c-di-GMP was replaced with a fluoro group was 3.9 kcal/mol for Ct-E88 RNA, resulting in 741-fold loss in binding affinity compared to c-di-GMP. Interestingly, the loss in binding affinity was only 3.7-fold and 7.7-fold for Vc2 and Cb-E43 RNA respectively. These observations

suggest that the hydrogen bonding interactions in the binding pocket of some class I riboswitch, such as Ct-E88 RNA, are very different to that of others, such as Vc2 and Cb-E43 RNAs.

6.4 Conclusions

We demonstrate that although c-di-GMP class I riboswitches have similar predicted folds, subtleties in the ligand binding pockets could allow for selective targeting of these riboswitches. The ability to selectively target one particular class I riboswitch, while not affecting others, could be used to affect c-di-GMP signalling in some pathogenic but not symbiotic bacteria. Although this dream remains to be realized, our results provide some optimism for the discovery of such selective inhibitors.

Chapter 7: Conformations of C-di-GMP class I aptamer (Ct-E88 RNA) – Preliminary data

7.1 Introduction

In **Chapter 6**, we verified that a class I c-di-GMP riboswitch aptamer from *Clostridium tetani* E88 (Ct-E88) could bind c-di-GMP with nanomolar affinity and hence is likely to be a functional c-di-GMP riboswitch. It was found that Ct-E88 RNA could bind c-di-GMP as tight as Vc2 RNA (the first c-di-GMP riboswitch to be identified). Interestingly, whereas both c-di-GMP and 2'-F-c-di-GMP could bind to Vc2 RNA with similar affinity, this was not the case with binding with Ct-E88. For Ct-E88 RNA, the binding affinity of c-di-GMP was close to three orders of magnitude tighter than 2'-F-c-di-GMP. This difference prompted us to ask the question of how Ct-E88 RNA recognizes the ligand differently from Vc2 RNA. The groups of Ferré-D-Amaré and Rueda have attempted to characterize ligand recognition and ligand-induced folding processes in the Vc2 riboswitch, using small-angle X-ray scattering (SAXS) and single-molecule fluorescence resonance energy transfer (smFRET). They have concluded that both Mg^{2+} cation and the c-di-GMP ligand are important for the proper folding of Vc2 riboswitch.[115, 127] In their experiments, it was shown that the addition of Mg^{2+} or c-di-GMP or both caused significant conformational change in the RNA. We therefore wondered if the structure of Ct-E88 riboswitch is also modulated by Mg^{2+} ion, or ligand binding, or both.[151] The conformational study of Ct-E88 RNA, in the presence of various analytes, would reveal if different class I riboswitches underwent similar or different

conformational changes that was observed for the prototypical class I c-di-GMP riboswitch Vc2.[115, 127]

Based on the conservation of many residues between Ct-E88 and Vc2 RNAs, we hypothesized that Ct-E88 RNA would have a similar global structure to Vc2 and would also adopt multiple conformations, which are modulated by ions and c-di-GMP. However, because we also observed differences in binding to different analogs of c-di-GMP, especially significant difference in affinity for 2'-F-c-di-GMP, we also hypothesized that the binding pocket of Ct-E88 is likely different from that of Vc2 RNA. To test these two hypotheses, different biochemical and biophysical tools, including native gel mobility shift assay, selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) technique, small-angle X-ray scattering (SAXS) analysis and NMR spectroscopy were used to gain some insights into how both cation and the nucleotide ligand affect the folding of Ct-E88.

7.2 Materials and Methods

7.2.1 RNA preparation

The RNAs used in this study is the UUCG-tetraloop modification of P3 helix. The RNAs were transcribed and purified as described before.[152] The crude RNA was purified using a 12% denaturing PAGE with 1×TBE buffer. The RNA pellet was dialyzed in a Biodialyzer (Nestgroup) with a 500 MWCO membrane (Nestgroup) for 24 h against 100 mM potassium phosphate buffer (pH 6.4), 0.5 M KCl, 10 mM EDTA, and then 1 and 0,1 mM EDTA, and finally against two changes of double distilled H_2O water before it was lyophilized as described by Dayie 2006.[50]

7.2.2 Electrophoretic mobility shift assay (EMSA)

To explore the conformational change between Ct-E88 RNA free and bound form, EMSA was performed.[115] RNA was folded by heating to 70 ℃ for 3 min and slow cooled in folding buffer (20 mM KCl, 6 mM $MgCl_2$, and 10 mM sodium cacodylate, pH 6.8) in the presence of saturated ligand concentration (2 × concentration of RNA). The reaction mixtures were incubated at room temperature (25 ℃) before separating on the gel. Electrophoresis was performed using native PAGE (10% acrylamide/bisacrylamide (19:1), 100 mM Tris/HEPES pH 7.5, 0.1 mM EDTA and 1 mM $MgCl_2$) at 4 ℃. The gels were stained with ethidium bromide and images were scanned using STORM phosphoImager (GE Healthcare).

7.2.3 Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE)

To identify local changes of RNA structure and characterize the general structural stabilization, SHAPE experiment was performed.

Materials

SuperScript^{TM} III First-Strand Synthesis System for RT-PCR kit was purchased from Invitrogen. T4 polynucleotide kinase was purchased from New England Biolabs. γ-[^{32}P]-ATP was ordered from Perkin Elmer. The experiment procedures were adapted with little modification from Wilkinson *et al.*[153] (**Figure 7.1**).

RNA folding and modification

The RNA construct for SHAPE experiment is as below:

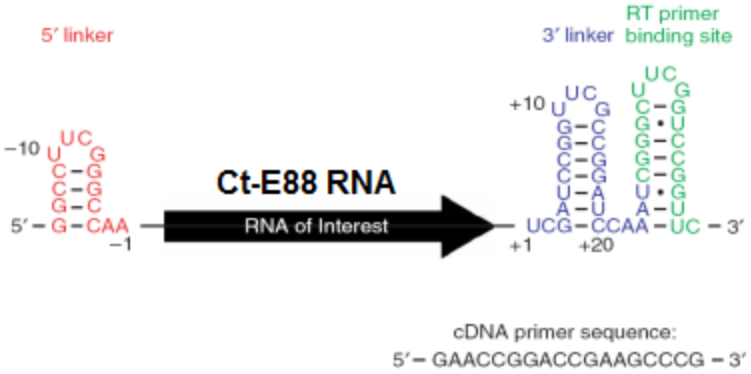


Figure 7.1. Construct of Ct-E88 used for SHAPE experiment. Figure is adapted from Wilkinson *et al.*[153]

Because SHAPE reactivities are assessed in a reverse transcription primer extension reaction, the 10-20 nucleotides adjacent to the primer binding site cannot to be read due to the presence of short cDNA fragments caused by pausing during the initiation phase of primer extension by the reverse transcriptase enzyme and the 8-10 positions at the 5'-end of the RNA also cannot to be visualized due to the presence of the intense band corresponding to the full-length extension product.[153] Therefore, the interested Ct-E88 RNA can be embedded within a cassette (**Figure 7.1**) that contains 5' (red) and 3' (blue) flanking sequences with a primer binding site (green) to allow all positions within the RNA of interest to be visualized in a sequencing gel. Both 5' and 3' linker extensions are designed to fold into stable hairpin structures that do not interfere with folding of the interested internal RNA.[154] The sequences of the 5' and 3' linker were checked by MFold to ensure that they were not prone to form stable base pairing interaction with the Ct-E88 RNA sequence.

The sequence of Ct-E88 RNA used for SHAPE:

5’- GG CCU UCG GGC CAA UUU CCA CGA UAA AGG CAA AAC UAU UGA AAG AUA GUG ACG CAA AGC UAU AGG GUC UUC GGA CAG CCA GUU AUC GAA AGG AGA UCG AUC CGG UUC GCC GGA UCC AAA UCG GGC UUC GGU CCG GUU C -3‘ (138 nt)

Where the nucleotides indicated in red are 5’-linker, nucleotides indicated in blue are 3’linker and nucleotides in green are RT primer binding site.

RNA was prepared and purified as previously described.[50]

1 pmol of RNA was folded in 9 µL of folding buffer containing 100 mM HEPES (pH 6.8) and 100 mM NaCl. To this was added either 3 mM $MgCl_2$, or ligands, or both. Each mixture was incubated overnight at room temperature before adding 1 µL of N-methylisatoic anhydride (NMIA) (65 µM) or 1 µL of pure DMSO (labeled as (-) NMIA) as negative control. Then the reaction mixtures were incubated at 37 ℃ for 45 min. After the reaction had gone to completion, the modified RNAs were recovered by ethanol precipitation with 90 µL of water, 4 µL of 5 M NaCl, 1 µL of 20 mg/mL glycogen, 2 µL of 100 mM EDTA (pH 8.0) and 350 µL of absolute ethanol and incubated at -20^{o} C overnight.

^{32}P-primer for reverse transcription

50 pmol of RT binding oligo [5’- GAA CCG GAC GAC CGA AGC CCG -3’] (2 µL of 25 pmol/µL) was mixed with 4 µL of 10 x T4 PNK buffer (New England Biolabs, cat. no. M0201S), 30 µCi of γ-[^{32}P]-ATP (6 × 10^6 Ci/mol, 10 Ci/L, Perkin Elmer cat. no. BLU502Z) and 0.5 µL of T4 polynucleotide kinase (New England Biolabs, cat. no. M0201S) in a total volume of 40 µL. The mixture was incubated at 37 ℃ for 1 h before purification by G25 column (GE Healthcare).

Primer extension and RNA sequencing

RNA samples were redissolved in 8 µL of water and a 6 µL aliquot was added to the reaction mixture containing 4 µL of ^{32}P-primer solution as described above.

Primer annealing was carried out at 65 ℃ for 5 min, at 35 ℃ for 5 min and then on ice for 1 min. The reverse transcription reaction was then started using (-) NMIA and (+) NMIA samples, 3 μL of SuperScript III enzyme mix (0.25 μL of dNTP mix, 0.25 μL of 0.1 M DTT, 1 μL of 5x Superscript buffer, 0.25 μL of superscript III enzyme and 1.25 μL of water) by adding it into 2 μL of RNA and primer mixture above. For the four sequencing samples (A/C/U/G), 2.9 μl of SuperScript III enzyme mix (same as above) was added into 1.9 μL of RNA (from DMSO control reaction) and primer mixture with 1 μL of each ddNTP respectively. All the samples were incubated at 52 ℃ for 20 min. Then 3 μl of formamide loading dye was added to each sample for a final volume of 8 μL and heated at 95 ℃ for 5 min before loading on the 8% denaturing polyacrylamide (19:1 acrylamide: bis-acrylamide, 8 M urea) gel.

7.2.4 Small angle X-ray scattering experiments and data analysis

To explore Ct-E88 RNA global conformational change, the small angle X-ray scattering (SAXS) and wide angle X-ray scattering (WAXS) data were collected at beam line 12-ID of the Advanced Photon Source (APS), Argonne National Laboratory. The wavelength, λ, of the X-ray radiation was set to 1.033 Å. The detailed protocols of SAXS measurement was followed as described.[155] In brief, the Ct-E88 RNA was dissolved in buffer A (100 mM HEPES/ 100 mM NaCl pH 8.0). The sample was heated at 90℃ for 2 min followed by cooling at room temperature for 10 min. $MgCl_2$ or c-di-GMP or both, were added to buffer A to prepare two additional buffer conditions: buffer B (100 mM HEPES/ 100 mM NaCl pH 8.0 with 3 mM $MgCl_2$) and buffer C (100 mM HEPES/ 100 mM NaCl pH 8.0 with 3 mM $MgCl_2$

and 10 μM c-di-GMP). RNA sample A was in buffer A and sample B was in buffer B. Sample C was in buffer C but also included extra c-di-GMP ([c-di-GMP]:[RNA] 2:1) and was incubated overnight at room temperature. Each RNA sample was then prepared in three concentrations (0.5, 1.0 and 3.0 mg/mL) making a total of nine sample conditions for the SAXS/WAXS data collection.

The q-range used for data analysis was between 0.002 and 0.300 $Å^{-1}$. The quality of the SAXS data was evaluated by the linearity of the Guinier plot (a plot of [log I(q)] vs. q^2, where I is intensity and q is the momentum transfer and is defined as $q = 4\pi \sin(\theta)/\lambda$, λ is the X-ray wavelength and the scattering angle is 2θ). WAXS data were used to guide the background subtraction for the corresponding SAXS data.[111, 156] For each sample condition, 25 images were taken to get good statistics and each was converted to 1D scattering profiles. Then the resulting scattering data sets were averaged before buffer background subtraction using IGOR PRO (Wavemetrics) software. The GNOM software was used to determine the radius of gyration (R_g) [111], which were in good agreement with those obtained from Guinier approximation [$\ln[I(q)] = \ln[I(0)] - R_g^2 q^2 / 3$]. Pair distance distribution function (PDDF) was computed as an indirect Fourier transform of the scattering data using this software.

7.2.5 Low resolution *ab initio* model reconstructions of Ct-E88 aptamer

The low resolution *ab initio* model reconstructions of Ct-E88 aptamer RNA under the three solution conditions described above were obtained by ATSAS 2.4 suite of programs including GNOM[157], DAMMIN[158] and DAMAVER[159] as described previously Chen *et al* 2012.[111] To avoid underestimation of the molecular dimension,

the parameter R_{max} and the upper end of distance *r* were chosen. The resulting PDDF has a near-zero-value tail at large *r*. The maximum distances (D_{max}) for each condition above were estimated from the PDDF. Then we used the program DAMMIN to obtain an approximate molecular envelope. To avoid distortion caused by possible underestimation of D_{max}, the R_{max} was set to be about 10 Å greater than D_{max}. All reconstructions for Ct-E88 aptamer under the three conditions were performed in the 'slow' mode. DAMAVER[159] was then used to average the resulting structural models and filtered dummy atom model. The normalized spatial discrepancy (NSD) values between each pair of models were also computed in this program. The model with the lowest average NSD value was chosen as the reference model and the remaining models were superimposed onto the reference model using SUPCOMB. Finally, a total of 20 independent DAMMIN runs were performed and the resulting bead models were averaged using DAMAVER.[159]

7.2.6 NMR sample preparation and NMR experiments of Ct-E88 RNA

To obtain greater insight into hydrogen-bonding interactions at Ct-E88 RNA binding pocket, NMR experiments were performed**.** The purified Ct-E88 RNA sample was lyophilized and dissolved in water. The corresponding amount of buffer stock solutions was added and the RNA sample was refolded in 30 mM HEPES-D18 / 30 mM NaCl/ 3 mM $MgCl_2$ pH 6.8. Then 10% D_2O and saturated c-di-GMP (2×concentration of RNA) were added and incubated overnight at room temperature. For the NMR experiments, the sample concentration was ~1 mM and the final concentration of c-di-GMP used was ~2 mM.

NMR experiments were performed on a Bruker Avance 600 MHz spectrometer. The spectra were recorded using the WATERGATE water suppression pulses[160] and ^{1}H-^{15}N HSQC on uniformly ^{15}N-GTP/CTP labeled Ct-E88 RNA, and 2D NOESY experiments on unlabeled Ct-E88 RNA were carried out using standard pulse sequences at 295 K.[161] 2D homonuclear NOESY spectra used to determine imino proton connectivities were recorded with a mixing time of 300 ms. All spectra were processed and analyzed using Bruker's TOPSPIN 2.1 as described in Chen *et al* 2012.[111]

7.3 Results and Discussion

7.3.1 Native gel mobility shift assay

Native PAGE is an easy and quick technique to test RNA folding and conformational changes upon ligand binding. If an alternative state of RNA after binding is present and the exchange rate between RNA in the free and bound forms was slow relative to the migration rate of the RNA through the gel matrix, then the bound RNA with different migration rates would be observed. The gel from the assay showed a fast migrating strong band for Ct-E88 RNA – c-di-GMP bound form and a weak band for RNA – endo-S-c-di-GMP bound form (**Figure 7.2A**), but there no change in migration pattern for the bands of Ct-E88 RNA binding the analogs with bulky modification at the 2'-OH position such as 2'-[Biotin]-AHC-c-di-GMP and 2'-[Biotin]-endo-S-c-di-GMP (**Figure 7.2B**). These results were consistent with the equilibrium microdialysis results of different analogs binding to Ct-E88 RNA in **Chapter 5**. The faster migrating Ct-E88 RNA – c-di-GMP band further suggested

that Ct-E88 RNA bound form adopted a compact conformation compared to the unbound RNA.

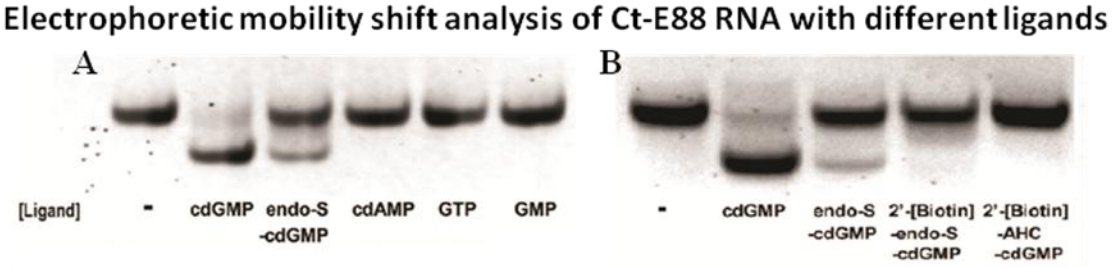


Figure 7.2. Electrophoretic mobility shift assay of Ct-E88 RNA binding to different ligands. **(A)** RNA binding to c-di-GMP, endo-S-c-di-GMP, c-di-AMP, GTP and GMP. **(B)** RNA binding to c-di-GMP, endo-S-c-di-GMP, 2'-[Biotin]-endo-S-c-di-GMP and 2'-[Biotin]-AHC-c-di-GMP.

7.3.2 Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) analysis of different conformations of Ct-E88 RNA

Next, we investigated how Ct-E88 RNA undergoes conformational change upon Mg^{2+} and ligand binding. Recently, SHAPE has become a valuable tool that allows simultaneous investigation of all nucleotides in the free and bound forms.[162, 163] It would help us understand ligand-free and bound state secondary structures and the changes in RNA dynamics that occur upon Mg^{2+} induced compaction and ligand binding. The approach employs various electrophiles such as N-methylisatoic anhydride (NMIA) as acylation reagent that selectively modifies 2'-OH groups within the regions of the RNA where the backbone is conformationally flexible (**Figure 7.3A**).[153] In contrast, nucleotides that are constrained or adopt highly structured tertiary interactions display low reactivity. Using this method, we investigated the structural changes of Ct-E88 RNA in the absence or presence of Mg^{2+} or different ligands. From our SHAPE gel, significant differences in SHAPE reactivity for the

free and bound form were observed and these regions were indicated in the different colors in **Figure 7.3B and C**. Bands intensities were quantified and absolute SHAPE reactivities from nucleotide position 36-89 within the Ct-E88 RNA in the different conditions were obtained by subtracting the (-) control intensities from the (+) NMIA intensities (**Figure 7.4A**). The predicted secondary structure based on the absolute SHAPE reactivities under each condition was drawn (**Figure 7.4B**), in which most of the base paired positions in RNA had lower intensities whereas single-stranded, bulged or mismatched nucleotide had higher intensities.

We found the reactivity of the various RNA elements (color coded yellow, light green and dark green) differed significantly in the presence and absence of Mg^{2+} **Figure 7.4A**). We think the changes in the light green and dark green regions were likely caused by the interaction between the GNRA tetraloop within helix P2 and the tetraloop receptor within helix P3. Likewise the guanosine in the yellow region is probably involved in a tertiary G:C base pair formation as seen in Vc2 RNA (**Figure 7.3C** and **Figure 7.4**). [127] These changes suggest that Mg^{2+} likely preorganizes the RNA before c-di-GMP binding.

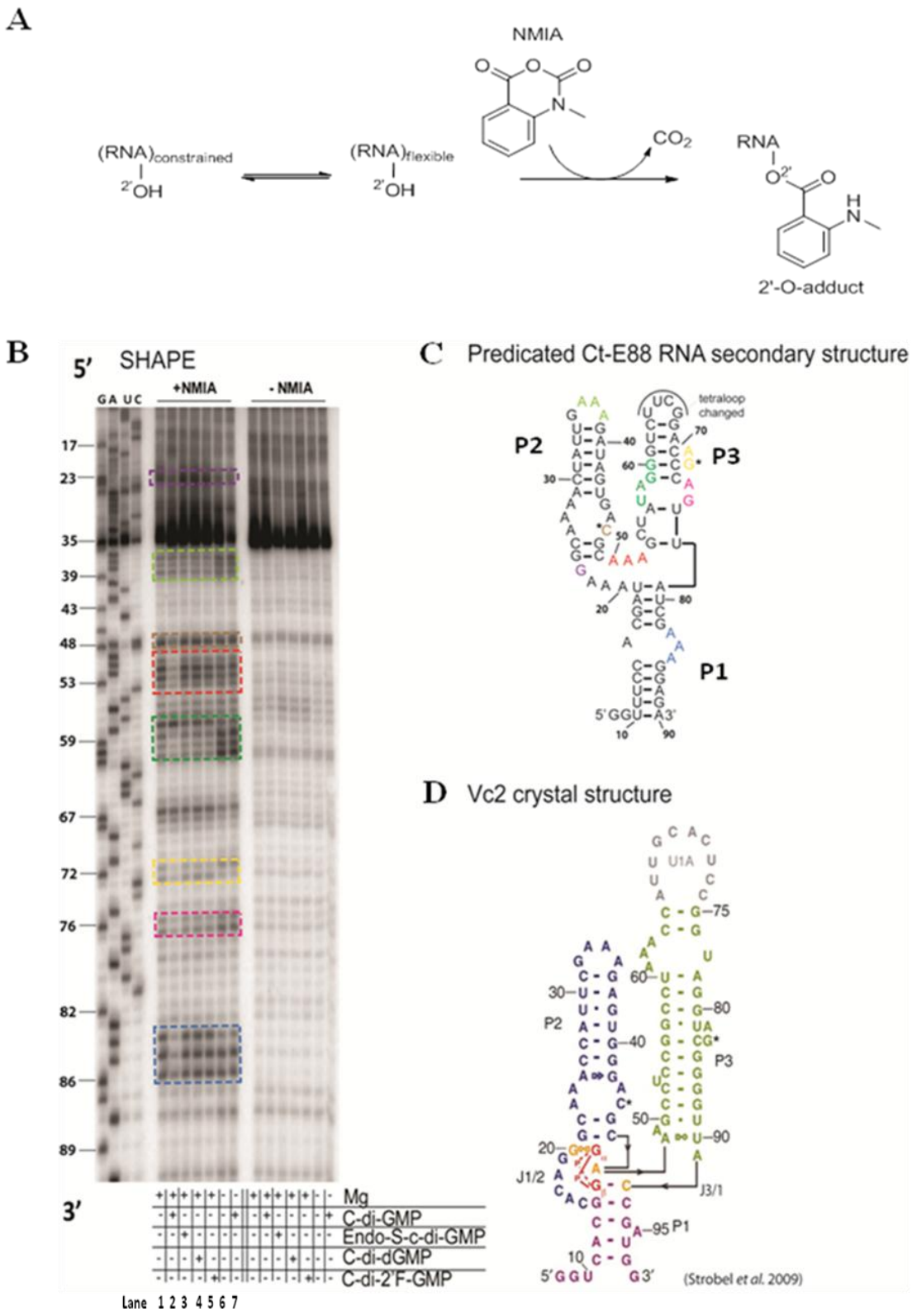
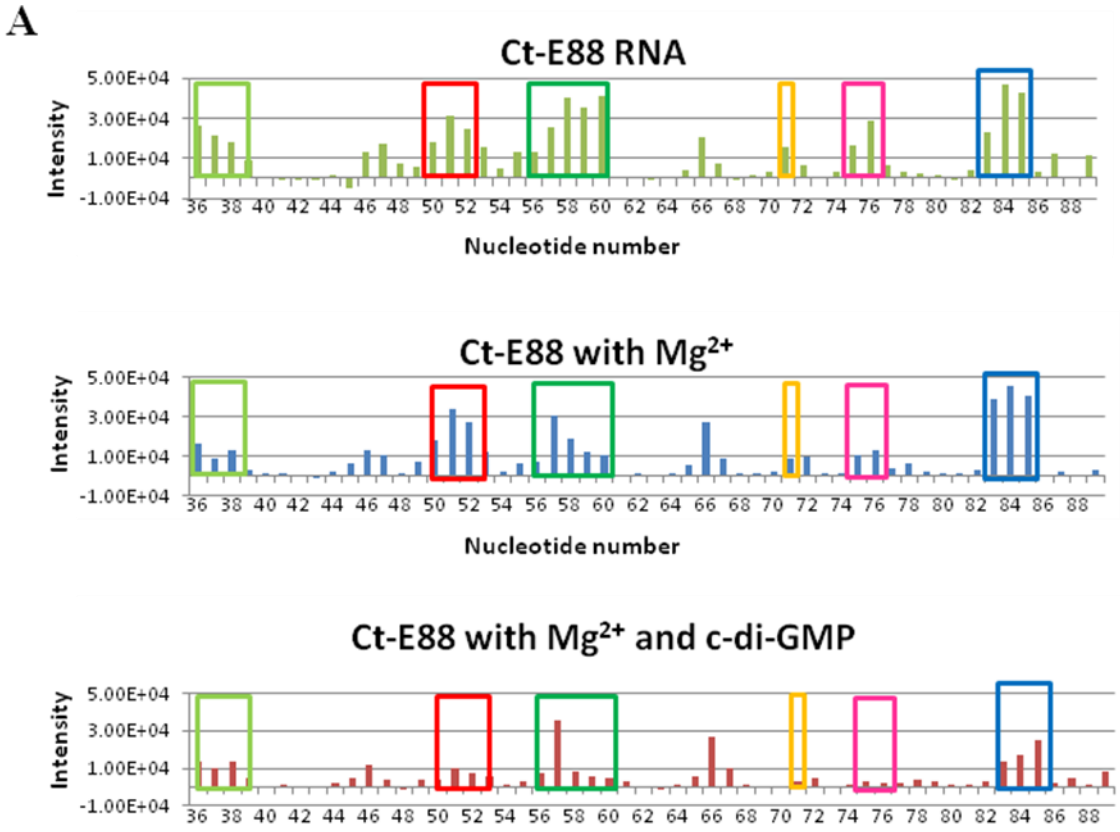


Figure 7.3. (A) Mechanism of RNA SHAPE. NMIA reacts preferentially with the 2'-hydroxyl group in flexible regions to form a stable 2'-O-ester adduct. **(B)** SHAPE analysis

for Ct-E88 RNA. G, A, U and C sequencing reactions were performed using cytidine, thymidine, adenosine and guanosine dideoxy terminating nucleotides. These marker lanes are one nucleotide longer than the corresponding NMIA lanes. (+) and (-) NMIA lanes represent reactions differing only in the presence of the reagent (See Methods). Regions that have significant changes in the presence of Mg^{2+} or ligand are boxed in different colors. **(C)** Predicted secondary structure of Ct-E88 RNA. The colored nucleotides correspond to the boxed region in the gel. **(D)** Known secondary structure of the ligand bound Vc2 RNA deduced from the crystal structure solved by Strobel *et al*. for comparison.[113]



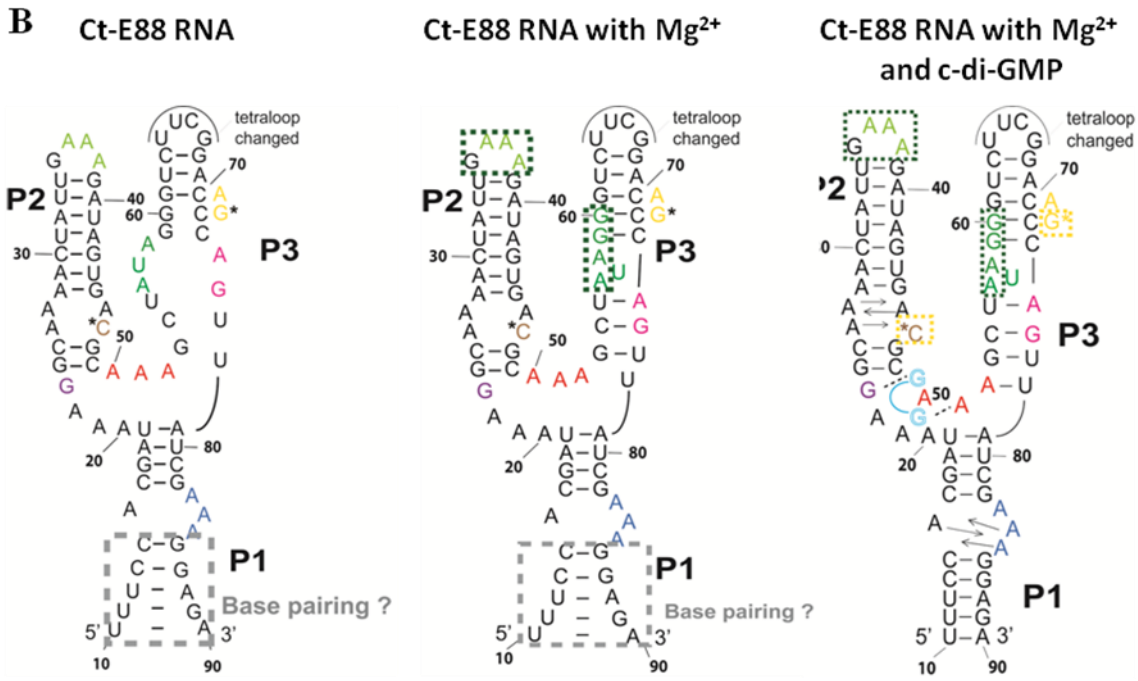


Figure 7.4. (A) Absolute SHAPE reactivities as a function of nucleotide position. The colored boxes correspond to the colored boxes in **Figure 7.3**. **(B)** The predicted secondary structures based on SHAPE reactivies in different conditions. The colored nucleotides correspond the nucleotides in the colored boxes in **(A)**.

Comparison of Ct-E88 RNA bound to Mg^{2+} in the absence or presence of c-di-GMP (**Figure 7.4A**), also showed some pronounced differences. These differences were likely from a remolding of the binding pocket. In Ct-E88 bound form (Lane 2 in **Figure 7.3B**), SHAPE reactivity from red (A50, A51 and A52), purple (G23) and blue (A83, A84 and A85) regions (**Figure 7.3B and C**) decreased dramatically, indicating that c-di-GMP induced these regions into more constrained conformations. Combined with the predicated secondary structures of Ct-E88 from sequence alignment (**Chapter 6, Figure 6.1**), we further predicted that one of A50, A51 and A52 was likely involved in intercalation between adenine and two bases of c-di-GMP (equivalent to A47 in Vc2 RNA) and G23 was likely base pairing to one of

the guanine (G_α) of c-di-GMP along the Hoogsteen face (equivalent to G20 in Vc2). The strong G:C base pair interaction from the other guanine (G_β) of c-di-GMP with cytosine (C92) on Vc2 RNA was not very obvious from our SHAPE data. It was likely that G_β of c-di-GMP formed the non-canonical G_β:A81 base pairing interaction (**Figure 7.4B**). Also, the triple adenines A83, A84 and A85 on Ct-E88 had significant difference in SHAPE reactivity in the presence of c-di-GMP. It led us to tentatively conclude that these triple adenines (A83, A84 and A85) probably formed the adenines intercalating interaction with the opposite A15 to stabilize the P1 stem in the presence of c-di-GMP.

In addition, it was also noteworthy that no difference in SHAPE reactivites was seen for Ct-E88 RNA binding c-di-GMP analogs (endo-S-c-di-GMP, c-di-dGMP and c-di-2'F-GMP) when lanes 1, 3, 4, and 5 were compared, which indicated that Ct-E88 was more discriminatory in ligand binding than Vc2. However, the details of hydrogen bonding interactions in Ct-E88 RNA binding pocket still remained to be investigated by NMR.

7.3.3 Following the conformational changes of Ct-E88 aptamer induced by c-di-GMP and Mg^{2+} ion with SAXS

We next tested the nature of the structural response elicited by Mg^{2+} and c-di-GMP using SAXS, which would provide more information on the global shape of the RNA in the free or bound forms. Solution scattering profiles were collected on Ct-E88 aptamer in the absence or presence of Mg^{2+} and c-di-GMP. The overlaid three scattering curves representing the three states (U, I and B states in **Table 7.1**)

exhibited significant differences, indicating the existence of three possible conformational states along the RNA-folding pathway (**Figure 7.5A**). Next, the SAXS data were analyzed using a Guinier plot (a plot of [log I(q)] vs. q^2, where I is intensity and q is the momentum transfer and is defined as $q = 4\pi \sin(\theta)/\lambda$, λ is the X-ray wavelength and the scattering angle is 2θ). From the scattering data, the radius of gyration (R_g) was fit at low q values using the linear, first-order Guinier approximation: $\ln[I(q)] = \ln[I(0)] - R_g^2 q^2 / 3$. The data suggested that R_g of Ct-E88 aptamer compacted in response to both Mg^{2+} and c-di-GMP (**Table 7.1**). In the absence of Mg^{2+} and c-di-GMP (U state), the RNA adopted an extended conformation with an R_g of 29.5 Å. In the presence of 3 mM Mg^{2+} (I state), R_g decreased to 28.9 Å which was correlated to the preorganized intermediate state (I). In the presence of both Mg^{2+} and c-di-GMP (B state), RNA was further compacted to the R_g at 24.3 Å. To further highlight the degree of RNA compaction, SAXS data were compared in Kratky representations (**Figure 7.5B**), where the scattering intensity is weighted by the square of the momentum transfer (q). It was seen that the scattering curve of RNA in U state exhibited a flattened plateau with scattering intensity falling slightly at large angle, suggesting that the unbound (U state) Ct-E88 RNA was locally disordered. Although it is different from an upward rising scattering intensity representative of a completely denatured conformation according to Porod's law, this U state is consistent with a partially 'folded state' with an extended conformation.[164] Compared to the U and I states, B state had a pronounced concave peak indicative of the folded structure. To provide more information on the shape and maximum intramolecular distance of the particles, the pair distance distribution PDDF (P(r) vs r)

plots were compared (**Figure 7.5C**).[165, 166] It was seen that RNA in U and I states (**Table 7.1**) had relative larger maximum intramolecular distance D_{max} than the RNA-c-di-GMP bound state (B). This observation revealed a marked ligand-induced compaction of the RNA in the presence of Mg^{2+}, likely indicative of folding of the c-di-GMP binding pocket.

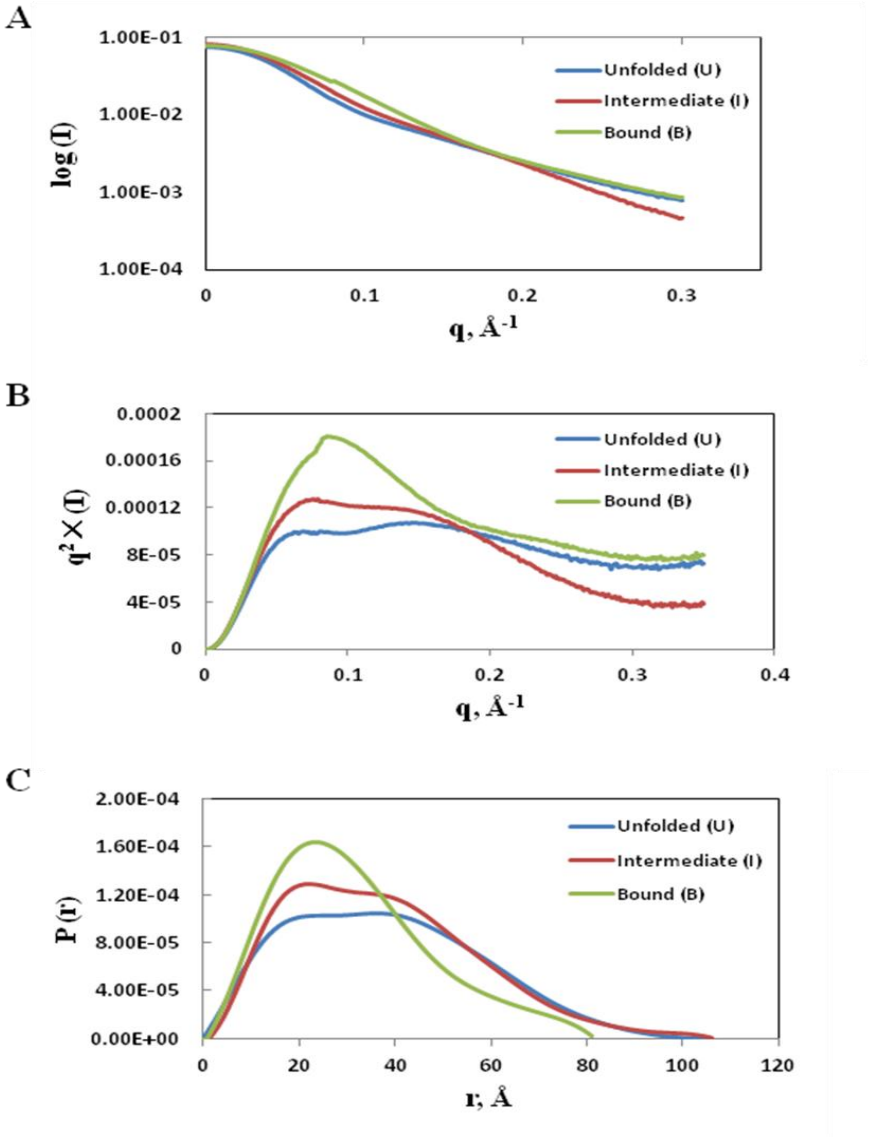


Figure 7.5. SAXS analysis of Ct-E88 RNA under three different states. **(A)** Comparison of experimental scattering profiles of Ct-E88 RNA. **(B)** Comparison of the SAXS data in Kratky representation for Ct-E88 RNA under the three states. **(C)** Comparison of P(r) plot determined from experiments using GNOM for the three different conditions.

Structural properties of Ct-E88 RNA under different conditions				
States	Conditions	Rg (Å)	D_{max} (Å)	NSD
Unfolded (U)	Buffer	29.5 ± 0.2	106	0.95 ± 0.02
Intermediate (I)	Buffer + 3 mM Mg^{2+}	28.9 ± 0.1	106	0.74 ± 0.04
Bound (B)	Buffer + Mg^{2+} + c-di-GMP	24.3 ± 0.1	81	0.60 ± 0.03

Table 7.1. Parameters and quality indicators derived from the x-ray scattering data for the Ct-E88 aptamer under different conditions. The radius of gyration (R_g) values were obtained from Guinier analysis of the low angle scattering data, the maxium intramolecular distances (D_{max}) were obtained using GNOM program and the NSD values of three low resolution reconstruction models were generated from DAMAVER program.

7.3.4 Low resolution *ab initio* modeling of Ct-E88 RNA

To gain further insight into the nature of the Mg^{2+} and c-di-GMP – induced conformational change, we constructed low resolution *ab initio* models of Ct-E88 RNA in the three states based on the SAXS data using the program DAMMIN[158] as described in Chen *et al* 2012.[111] The program package DAMAVER[159] was used to average the resulting 20 models for each sample condition. The NSD between each pair of models was computed and listed in **Table 7.1**. It provides a quantitative estimate of similarity between models, where identical models give NSD values at zero, similar models give NSD values around one and models with different shapes give NSD values greater than one.[167] The final low-resolution reconstruction obtained was the average molecular envelope of the ensemble. Here the average NSDs for the three reconstructions were 0.95, 0.74 and 0.60 (each NSD was <1) (**Table 7.1**), indicating good convergences in both individual DAMMIN fits and overall bead

model ensembles for Ct-E88 RNA under these three states. B state had the lowest average NSD value, consistent with the most rigid and folded surface as we expected.

The low resolution models corresponding to the three states of Ct-E88 RNA were presented in **Figure 7.6**, They were shown in different views of the final surface of the a*b initio* models (molecular envelope) as mesh representations with 'filtered' models inside the molecular envelopes indicated in beads. Comparison of the low resolution models of these three states indicated that the unfolded state had four arms, and then two of them were merged into one arm in the I state. Further folding into a compact conformation occurred in the presence of c-di-GMP. This observation was likely similar to the proposed folding pathway of Vc2 RNA from smFRET analysis.[127] Herein helices P2 and P3 are packed side-by-side through tetraloop and tetraloop receptor interaction in the presence of Mg^{2+}, but P1 is not yet formed. This Mg^{2+}-dependent state offers a preorganized intermediate state that would allow more efficient binding of c-di-GMP. Thus c-di-GMP binding would likely promote the folded compact state required for downstream genes regulation.

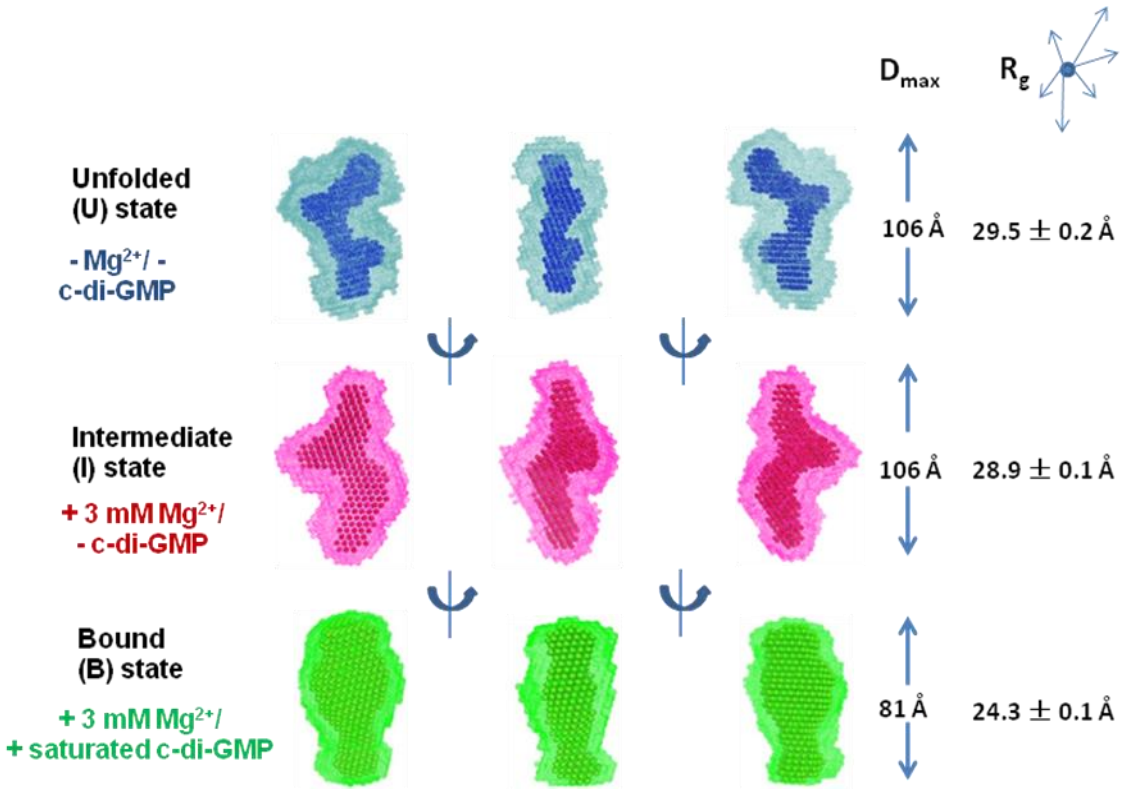


Figure 7.6. Low-resolution *ab initio* models of Ct-E88 aptamer under three different folding states (U state in blue, I state in red and B state in green). Three different views are shown for the final consensus models (molecular envelope) as mesh representatives with 'filtered' models as beads aligned inside.

7.3.5 NMR titration study of Ct-E88 RNA with c-di-GMP

Next, NMR titration experiments were performed to determine whether the Mg^{2+} ion and c-di-GMP induced structural rearrangements resulted in any changes of hydrogen-bonding interactions in Ct-E88.

First, 1D NMR chemical shift perturbation experiment is a simple method to compare the binding sites in biomolecular complexes. Before running the titration of Ct-E88 RNA with c-di-GMP experiments, the stability of Ct-E88 RNA molecules was investigated by following temperature-induced changes in the imino region of the 1D ^{1}H NMR spectra (**Figure 7.7**). It was found that at 295 K the unbound RNA

showed the most peaks in the imino region, suggesting that it is the most optimized folding temperature condition. Therefore, 295 K was used for further titration experiments. Since the chemical shifts of the imino proton resonances in RNA strongly depend on its chemical environment, including base pairing, base stacking, ring current and hydrogen bonding to solvent molecules, the characteristic imino proton and the formation and breakage of canonical or non-canonical base pairs could be monitored to define conformational changes upon interacting with ligands.[168] The titration of Ct-E88 RNA with c-di-GMP in the presence of Mg^{2+} ion from 1D ^{1}H-NMR spectra, in the imino region, revealed that there were significant chemical shift perturbations of existing imino peaks and appearance of some strong new peaks (**Figure 7.8**), where the peaks in the region of 12-15 ppm were usually indicative of canonical Watson-Crick base pairs (A:U and G:C base pairings) and upfield chemical shifting peaks at 10-12 ppm were often indicative of noncanonical base pairs (G:U, G:A and G:G base pairings).

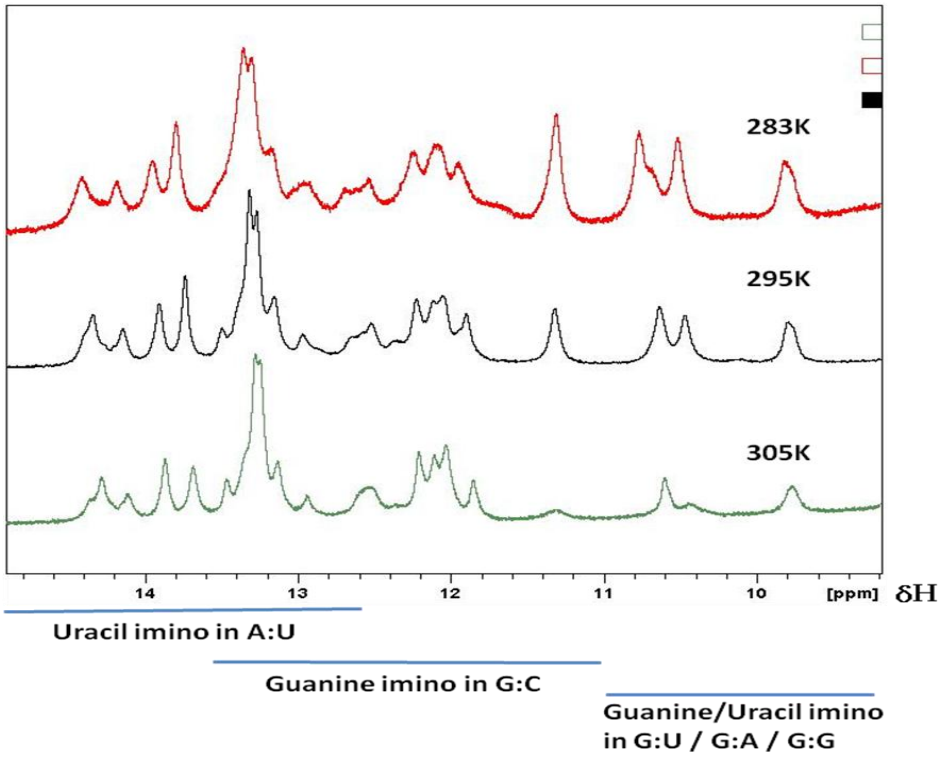


Figure 7.7. Imino region of 1D ^{1}H NMR spectra of Ct-E88 RNA with 3 mM Mg^{2+} at different temperatures.

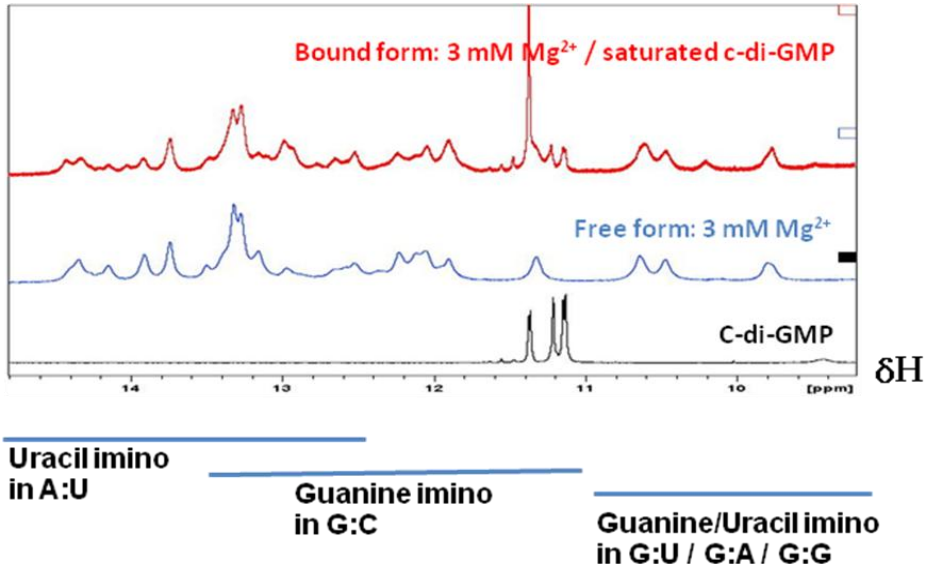
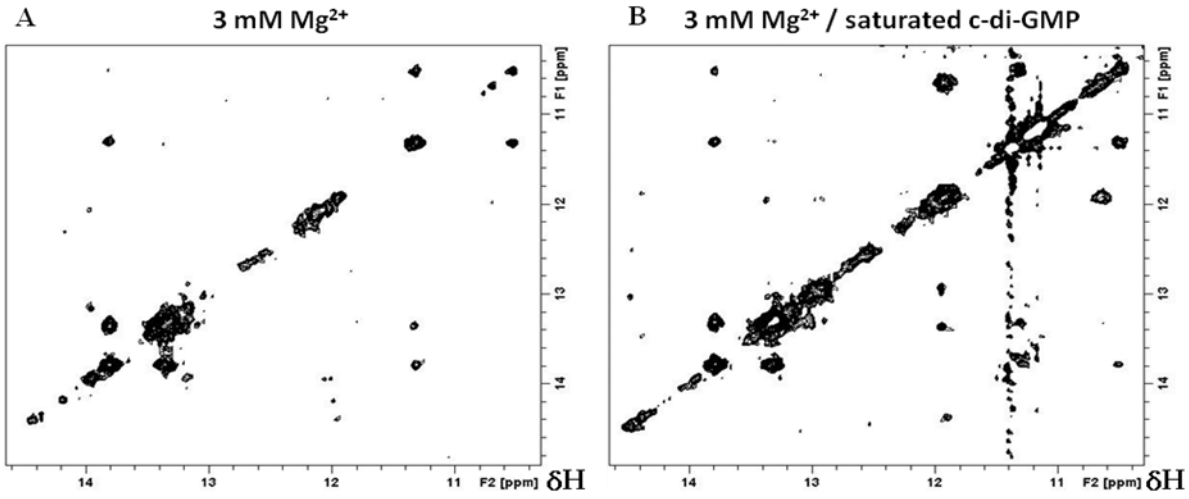


Figure 7.8. The imino region of 1D ^{1}H NMR spectra of Ct-E88 RNA with 3 mM Mg^{2+} in the presence or absence of c-di-GMP.

Second, 2D nuclear Overhauser effect spectroscopy (NOESY) experiment which correlates all pairwise protons within 5 Å would further provide us with more information on base pairing and secondary structure. From the 2D-imino region of the NOESY spectra depicting titration of CtE88-RNA with c-di-GMP, it was found that there was one non-canonical cross peak (G:U or G:A) at 10-12 ppm and some canonical cross peaks (G:C or A:U) at 12-15 ppm in the absence of c-di-GMP.[169] However, not many more cross peaks were seen when c-di-GMP was added, indicating some limited conformational change and transformation to a well-folded structure. While there were less cross peaks than we expected from SHAPE data above, it maybe caused by the slow tumbling of the large RNA structure (83 nt).



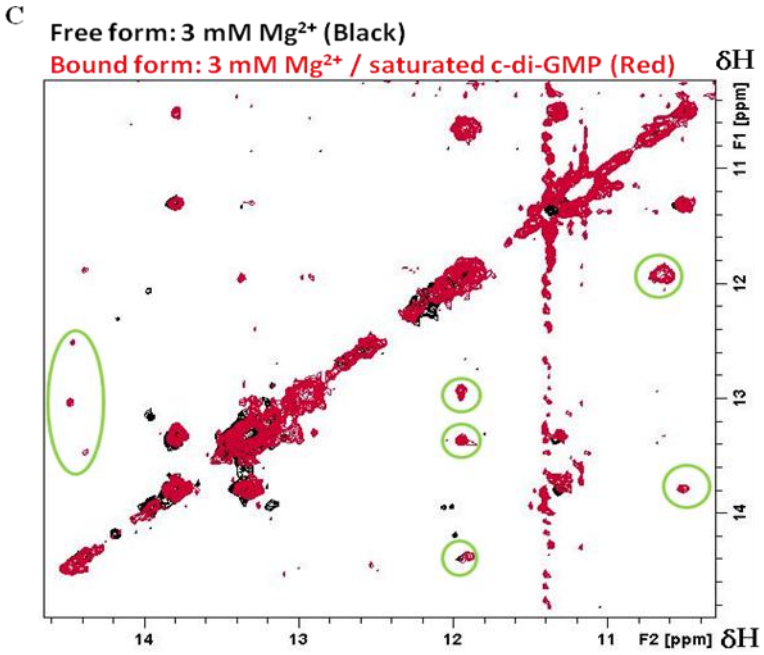


Figure 7.9. The imino region of 2D-NOESY of RNA with 3 mM Mg^{2+} **(A)** in the presence or **(B)** absence of c-di-GMP. **(C)** Overlay of **(A)** and **(B)**. Temperature is at 295K. The new peaks that appeared in the bound form are circled in green.

Third, uniformly ^{15}N – labeled GTP and CTP with unlabeled UTP and ATP Ct-E88 RNA sample was prepared and heteronuclear single quantum correlation (HSQC) experiments were recorded. In the free and c-di-GMP bound Ct-E88 RNA (**Figure 7.10**), we respectively observed 5 and 10 guanine imino resonances in the region between 12-13.5 ppm of the ^{1}H-spectrum for RNA characteristic of G:C Watson-Crick base pair formation. In the region between 9-11 ppm for RNA characteristic of noncanonical base pairing (GU, GA or GG), we found 2 and 4 guanine resonances in the free and bound Ct-E88 RNA respectively. These results suggested that Mg^{2+} helped RNA preorganize into a partial folded form and c-di-GMP binding stabilized it into a more folded structure. Some overlapped peaks between free and bound forms indicated that some base pair interactions in the free form persisted in the bound form.

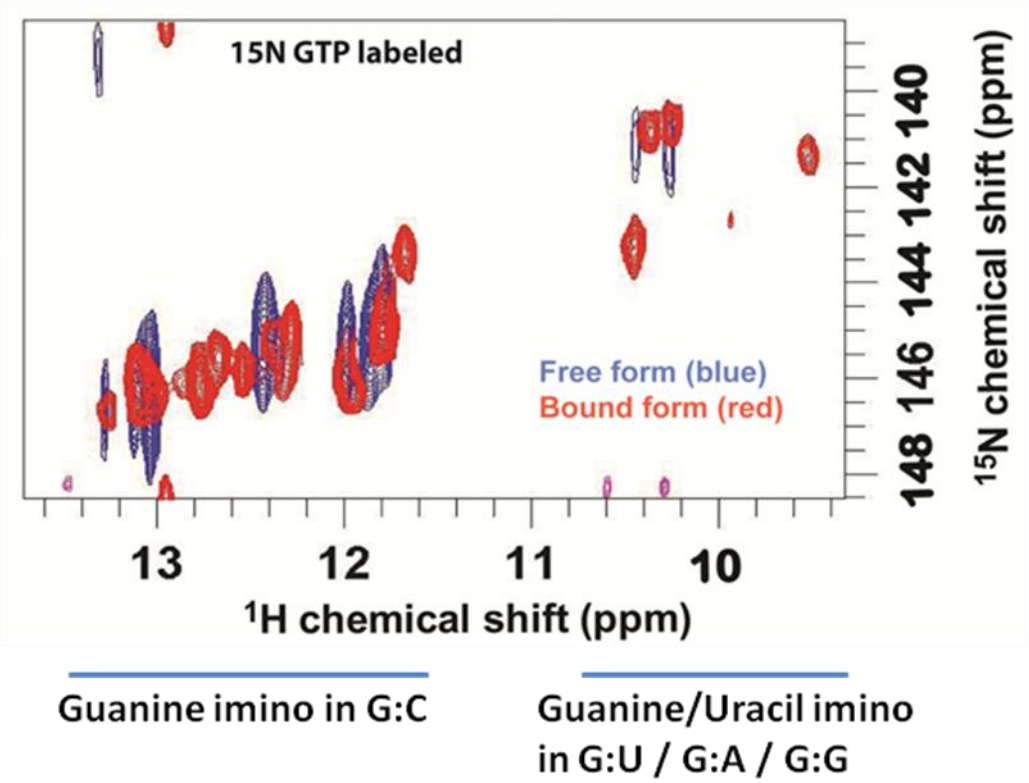


Figure 7.10. Imino HSQC NMR experiment of the ^{15}N-GTP, CTP labeled Ct-E88 RNA in the presence (red) or absence (blue) of unlabeled c-di-GMP.

Although our preliminary NMR data provided limited information on conformational change of Ct-E88 RNA recognition of ligand, details of hydrogen bonding interactions in the binding pocket are still unknown. Further heteronuclear 2D HSQC and HNN-COSY with different labeled sample will help to characterize the peaks and elucidate more details of base pairing and conformational change. Due to the large size problem of Ct-E88 RNA in NMR experiment, we will also make a new construct with two annealed pieces as the one used in smFRET for Vc2 RNA[127] and specifically label each piece using our new developed labeling techniques[170-172] to provide greater insights into the nature of secondary and tertiary structure and how c-di-GMP is recognized by Ct-E88 RNA.

7.4 Conclusions

In this study, we characterized the Ct-E88 riboswitch under different folding conditions using SHAPE, SAXS and NMR. Our preliminary data demonstrates that Ct-E88 aptamer adopts multiple conformations upon Mg^{2+} and c-di-GMP binding. Mg^{2+} ion, as an essential cofactor, facilitates RNA pre-organized folding for c-di-GMP binding. However, ligand binding further stabilizes the tertiary structure into compact fold to regulate the downstream gene. The details of hydrogen bonding interactions in the binding pocket still remain to be investigated by NMR. By combining future NMR experiments with present SHAPE and SAXS data, we will obtain greater insights into the nature of secondary and tertiary structural rearrangement of RNA in the presence of Mg^{2+} and c-di-GMP and perhaps better understand the mechanism of ligand binding via induced fit or comformational selection.

Chapter 8: Conclusions and future work – C-di-GMP riboswitch

8.1 Conclusions

The second part of this dissertation focused on the study of c-di-GMP riboswitch RNA. C-di-GMP riboswitch is the regulatory mRNA found in bacterial 3'- or 5'-UTRs (untranslated region), and it can directly sense the secondary messenger c-di-GMP using its aptamer domain to modulate downstream gene expression.[102] Due to the central role that c-di-GMP plays in bacterial biofilm formation and virulence, there is interest in inhibiting c-di-GMP signaling pathways in specific bacteria in polymicrobial systems through riboswitch regulation. We hypothesized that even within the same class, different riboswitches would use different recognition elements to sense c-di-GMP. Therefore there is an opportunity to selectively target different bacteria, which harbor different c-di-GMP riboswitches, with selective small molecules. To test this hypothesis, we designed and synthesized different c-di-GMP analogs and investigated whether they would be differentially recognized by riboswitches that belong to different classes (**Chapter 5**) or within the same class (**Chapter 6**). We also used a panel of biophysical techniques (NMR, SAX, SHAPE, gel shift) to gain some insights into the conformations of class I riboswitches (**Chapter 7**).

In **Chapter 5**, we demonstrated, using molecular docking and equilibrium microdialysis experiments, that 2'-OH modified analogs of c-di-GMP are probes to tell apart c-di-GMP class II riboswitches from class I. In **Chapter 6**, we further demonstrated that, within the same class I, the binding affinities of c-di-GMP analogs

to the RNAs could also be very different. However, more optimized analogs to selectively target one particular class or subclass of riboswitches to inhibit c-di-GMP signalling in some pathogenic but not symbiotic bacteria still remain to be developed.

Secondly, we chose one sub-class I c-di-GMP riboswitch from *Clostridium tetani* (Ct-E88), which has large differences in binding affinities of c-di-GMP analogs to the known class I Vc2 RNA, for structural study. Such a study will help us understand how the ligand finds its binding pocket and triggers the conformational change to modulate the downstream genes in the expression region of the riboswitch. To this end, we used three different biochemical and biophysical tools: selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) to characterize the secondary structure of free or bound forms, small-angle X-ray scattering (SAXS) to explore the global shape of the RNA in the free or bound forms, and nuclear magnetic resonance (NMR) to study the H-bonding interaction between the RNA and the ligand. Our results showed that the RNA is pre-organized for efficient ligand binding in the presence of Mg^{2+} ion. Subsequently ligand binding stabilizes the tertiary structure and formation of P1 stem to regulate the downstream gene. However, further details of hydrogen bonding interactions in the binding pocket are still unknown, and NMR will be needed to fill in these critical gaps in knowledge. It will not only provide details in the nature of secondary and tertiary structural arrangement of the RNA but also help us better understand the mechanism of ligand recognition.

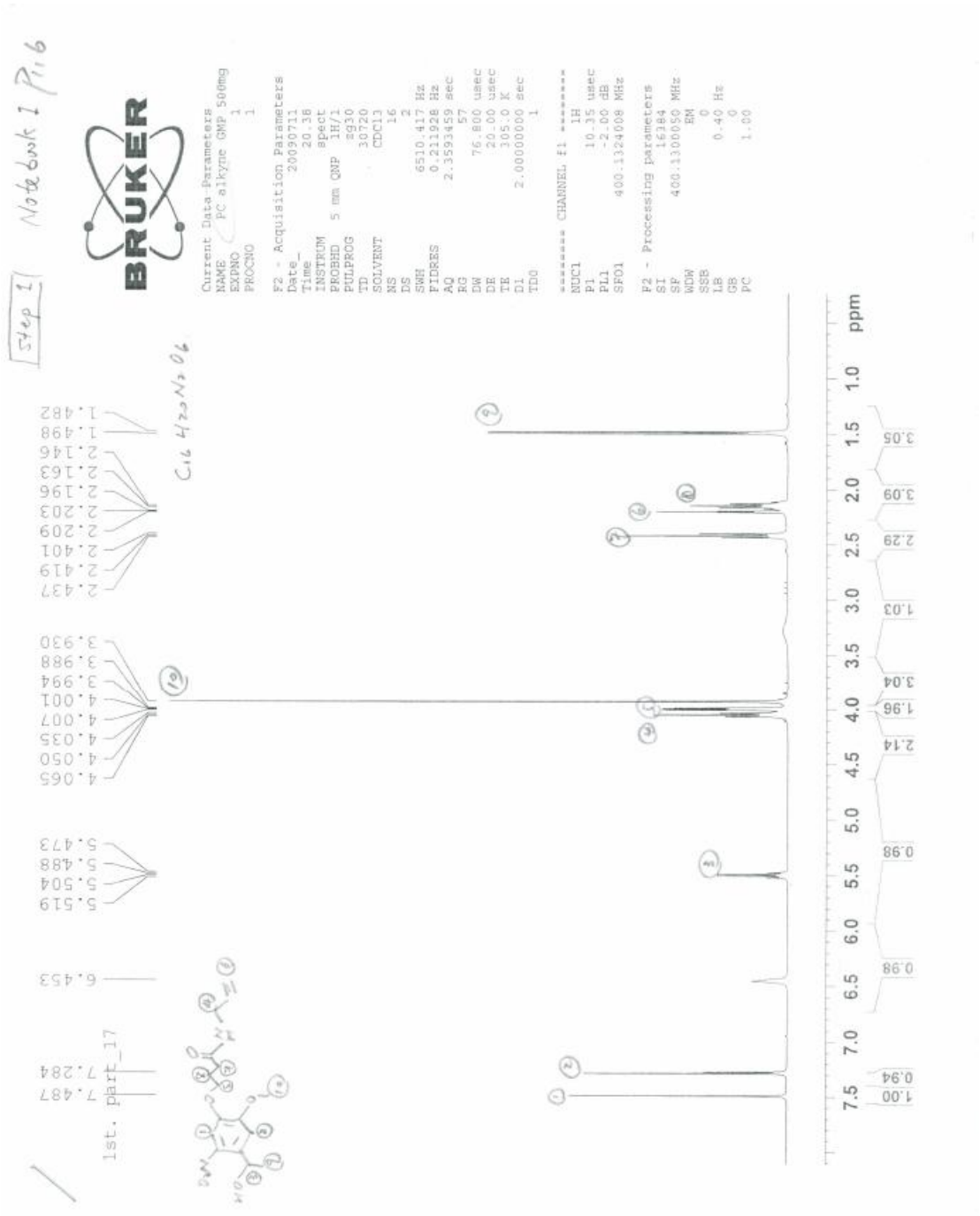
8.2 Future work

This work has demonstrated that it is indeed possible to differentiate between c-di-GMP riboswitches with small molecules. Future work could investigate how to specifically target c-di-GMP riboswitches in pathogenic, but not symbiotic, bacteria. Towards this end, two approaches could be adopted: a) modifying c-di-GMP into a pathogen-selective c-di-GMP analog and converting this analog into a cell-permeable pro-drug; b) designing non-c-di-GMP small molecule inhibitor of riboswitches, using structural data obtained from NMR and X-ray crystallography. It would therefore be fruitful to structurally characterize the new class I c-di-GMP riboswitches (Ct-E88 and Cb-E43) that were experimentally verified to bind to c-di-GMP. The structural characterization of these analogs, with and without c-di-GMP or analogs, would provide atomistic details of the binding site, which could be used to design next-generation anti-biofilm agents.

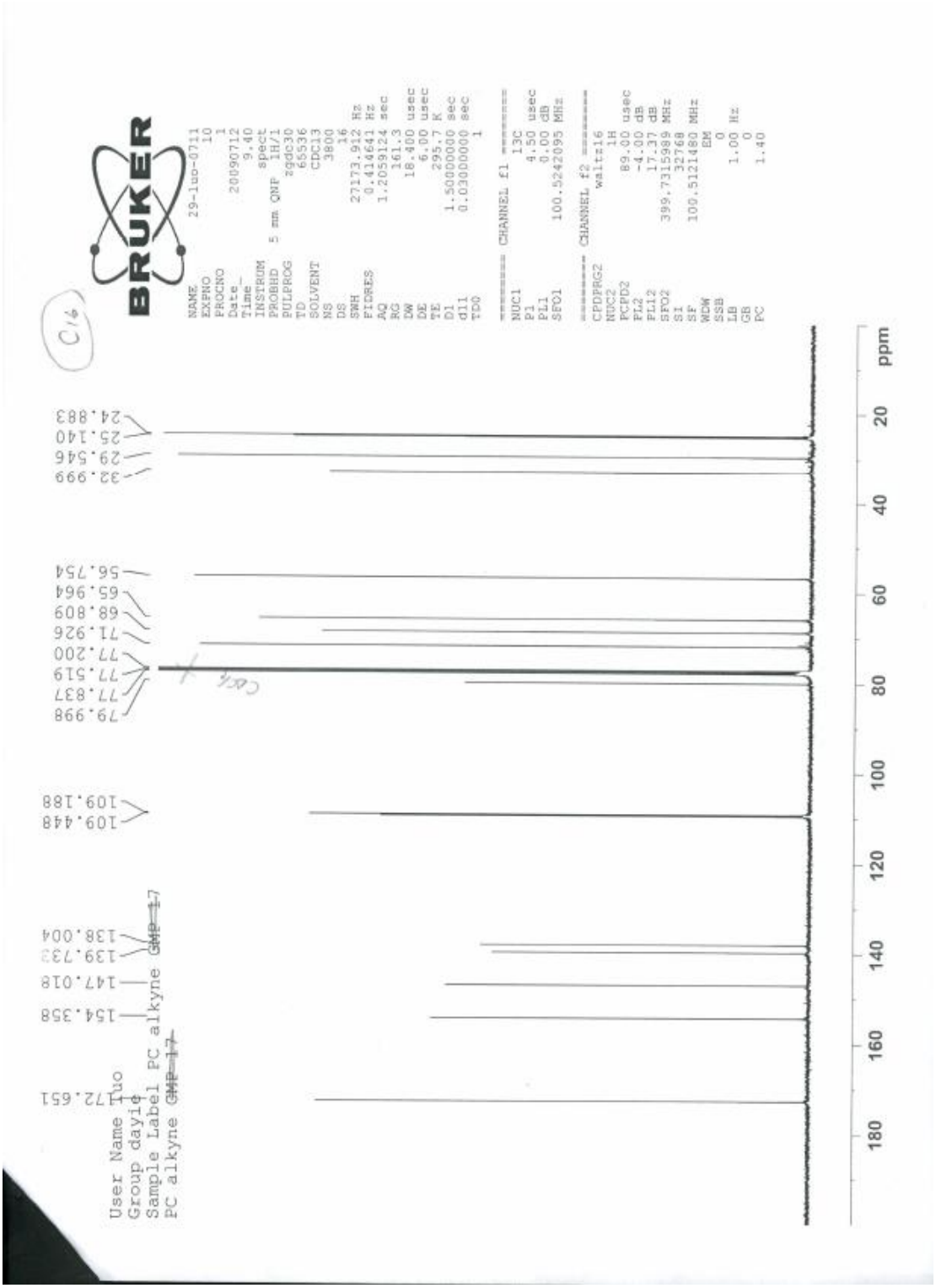
Appendices

Appendix 1: Spectra for PC-alkyne (**Chapter 2**)

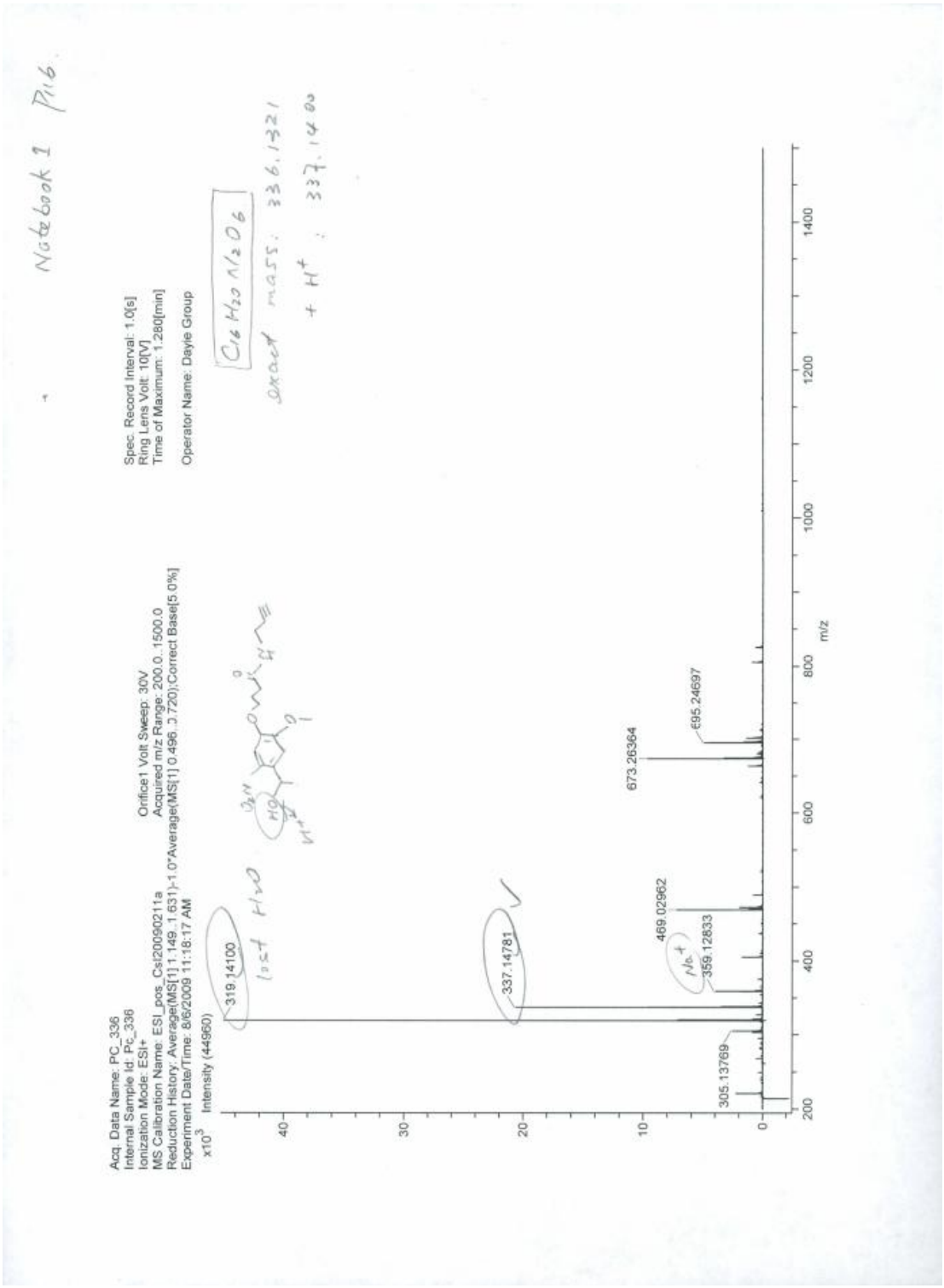
^{1}H NMR spectrum (400 MHz, $CDCl_3$) recorded for PC alkyne



^{13}C NMR (400 MHz, $CDCl_3$) recorded for PC alkyne

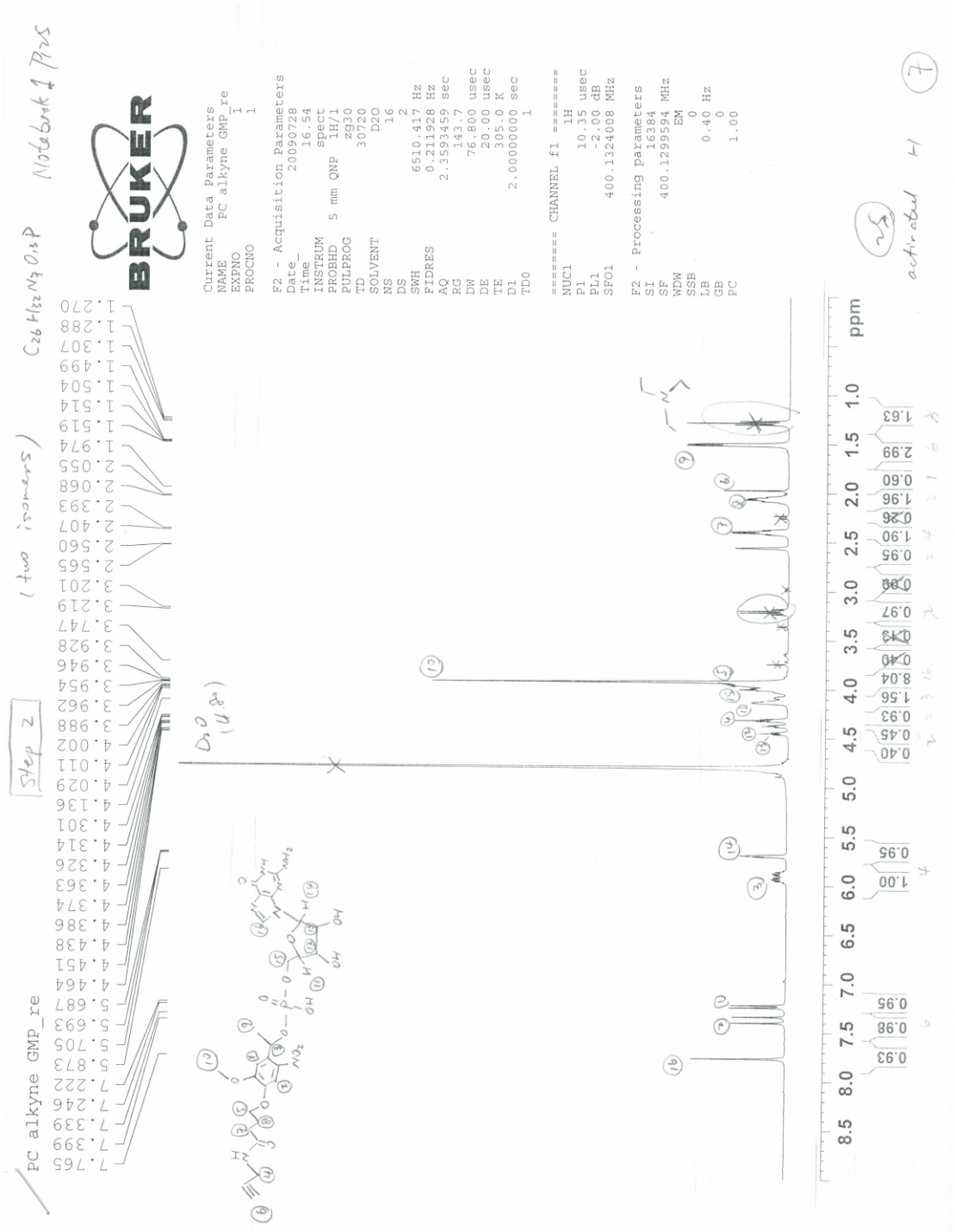


ESI-MS $[M+H]^+$ recorded for PC alkyne

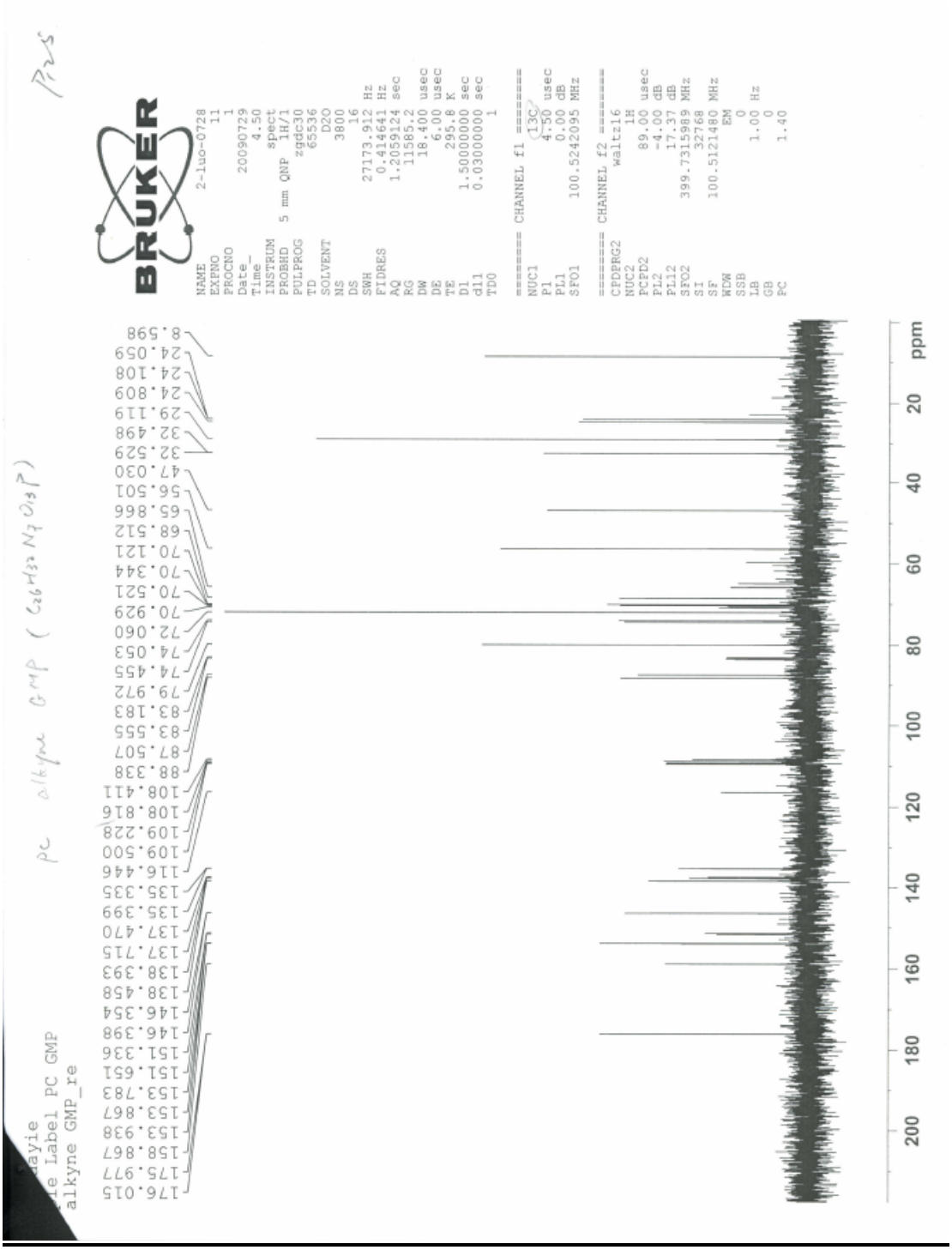


Appendix 2: Spectra for PC-alkyne GMP (**Chapter 2**)

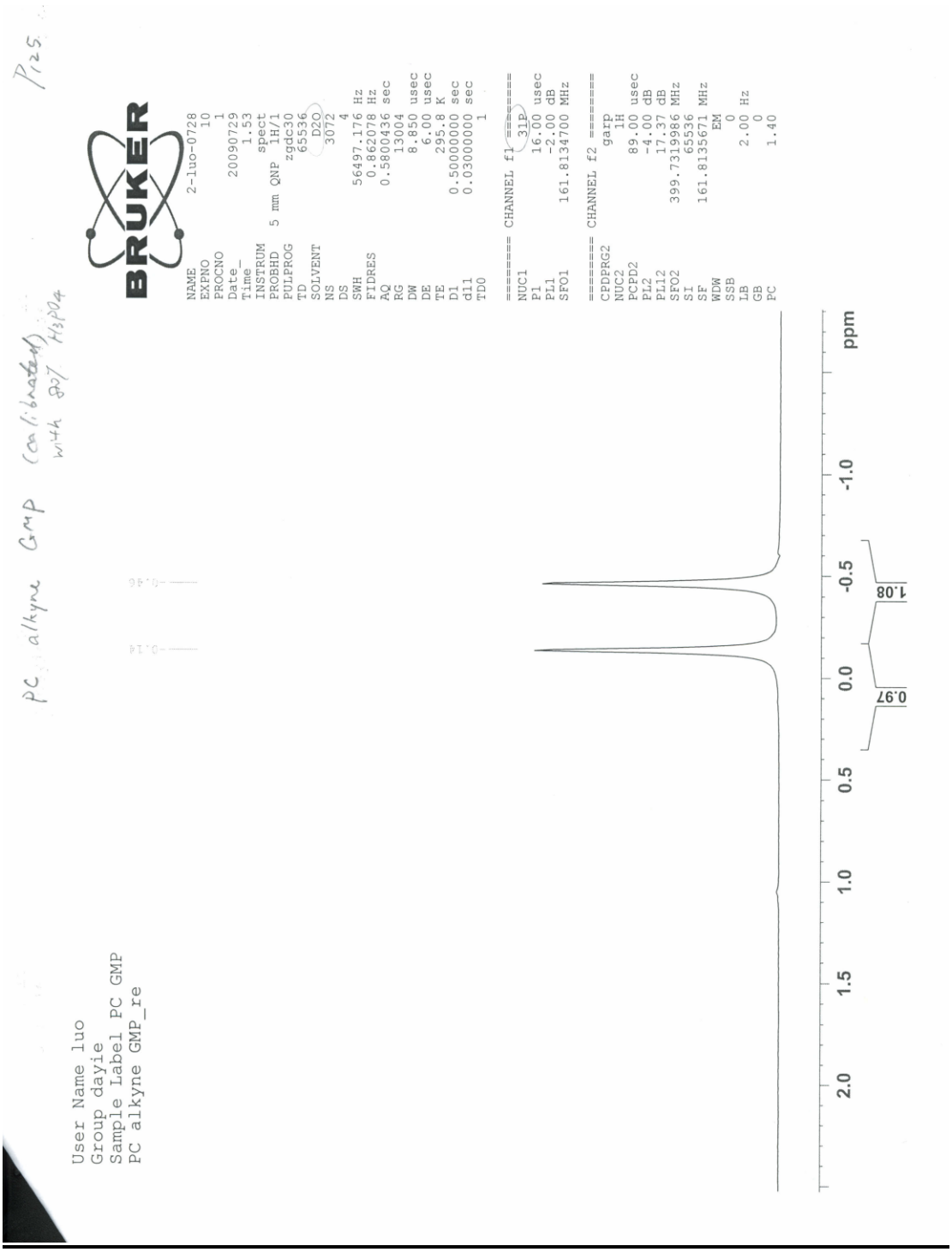
^{1}H NMR (400 MHz, D_2O) recorded for PC alkyne GMP



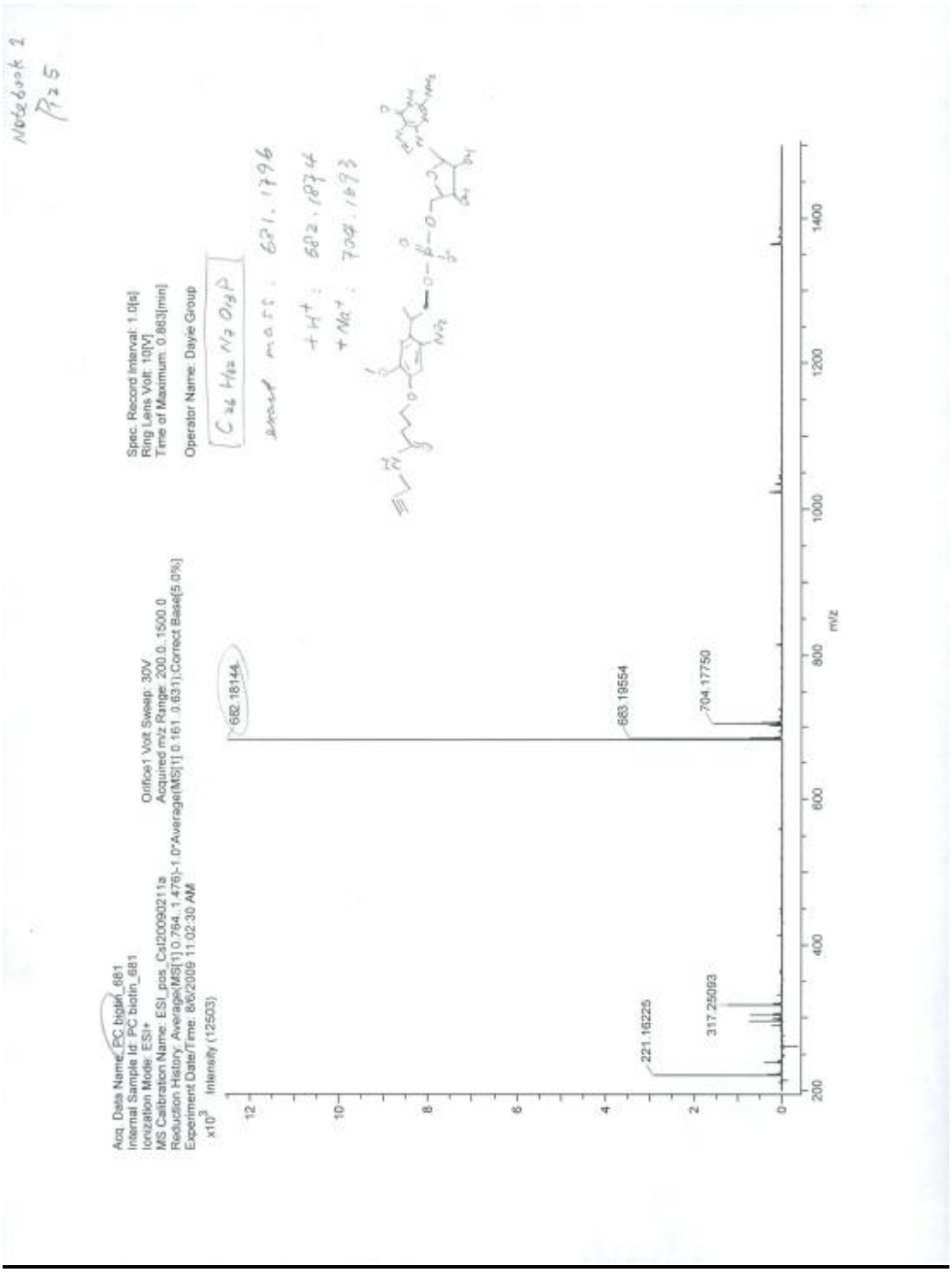
^{13}C NMR (400 MHz, D_2O) recorded for PC alkyne GMP



^{31}P NMR (400 MHz, D_2O) recorded for PC alkyne GMP

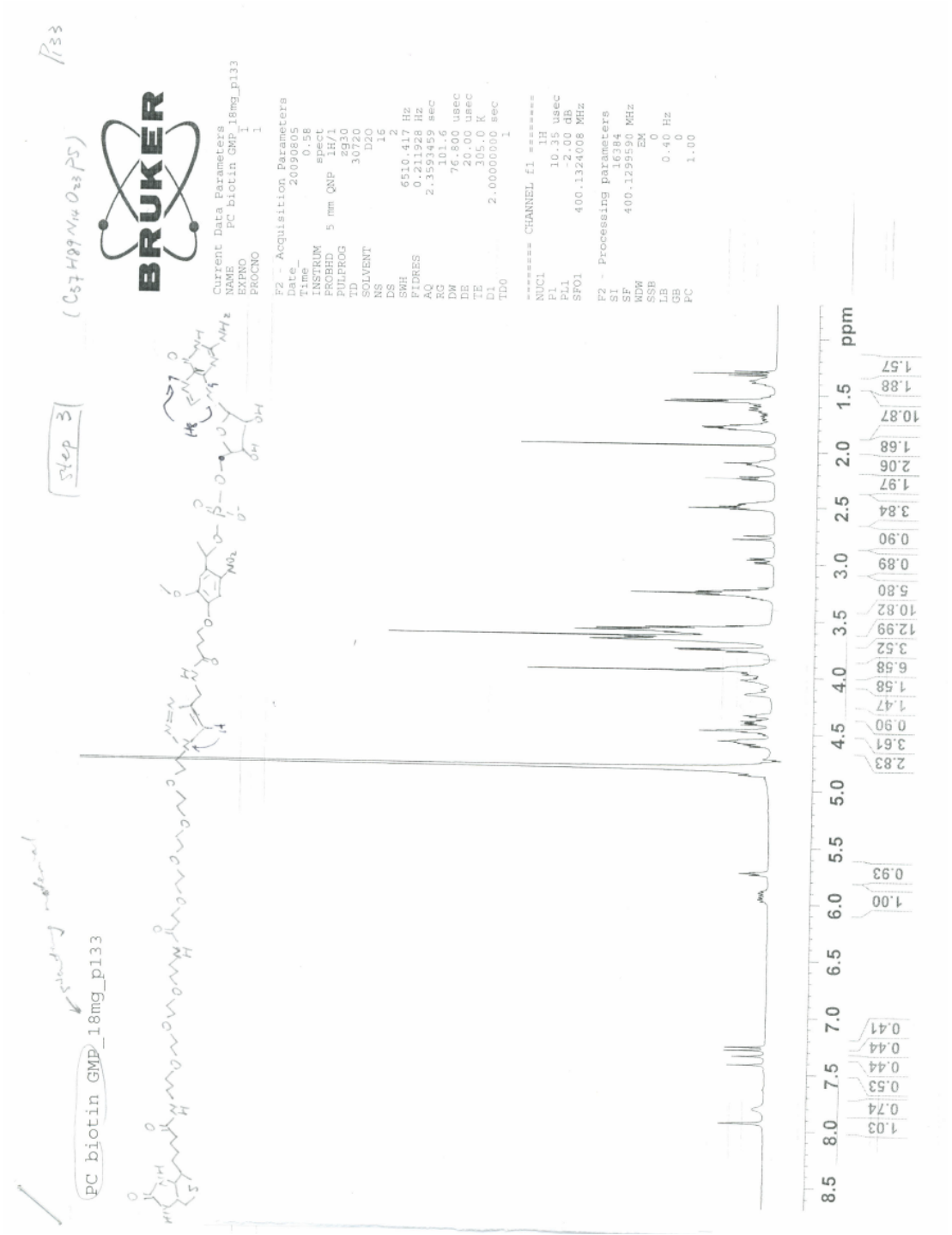


ESI-MS $[M+H]^+$ recorded for PC alkyne GMP

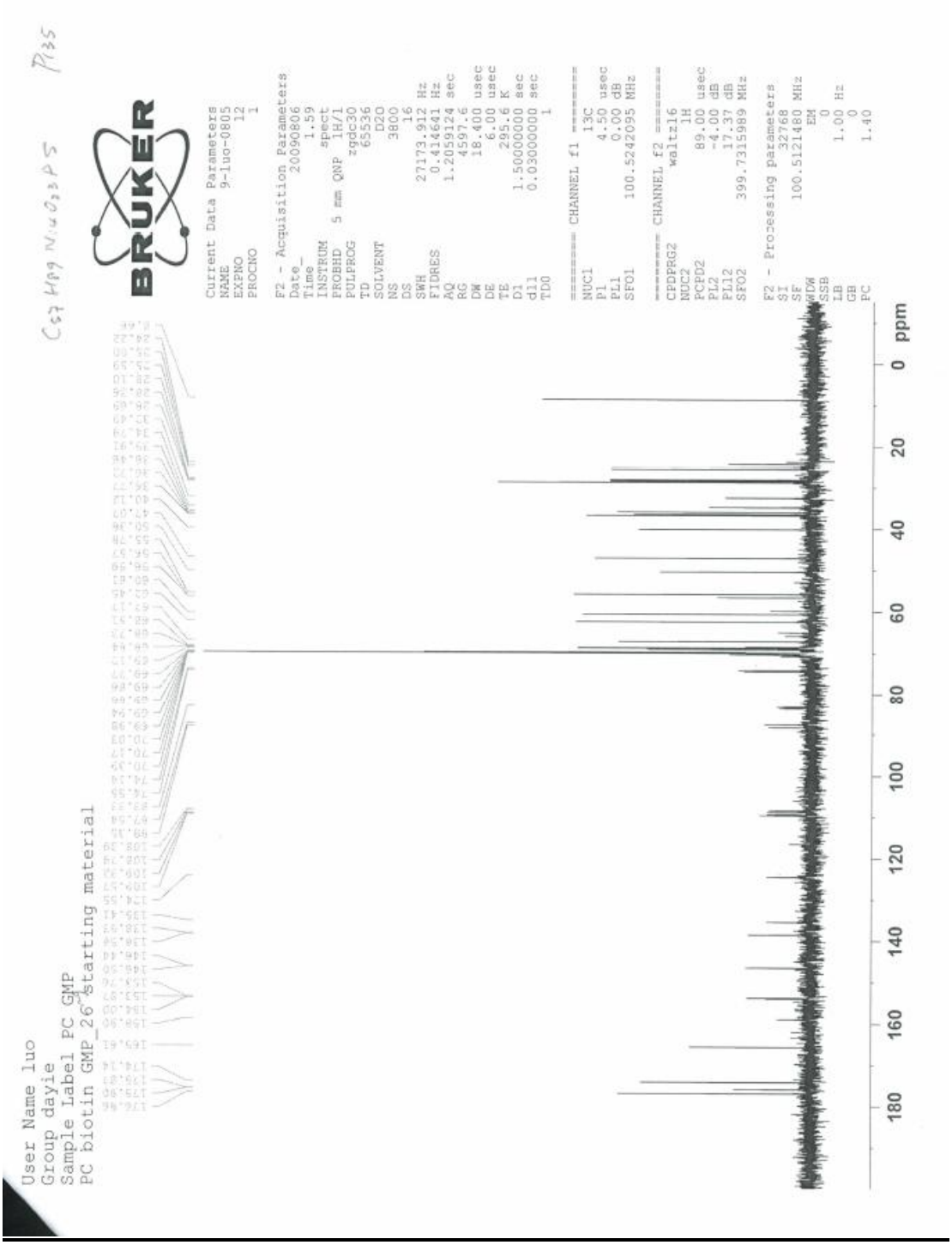


Appendix 3: Spectra for biotin-PC-GMP (**Chapter 2**)

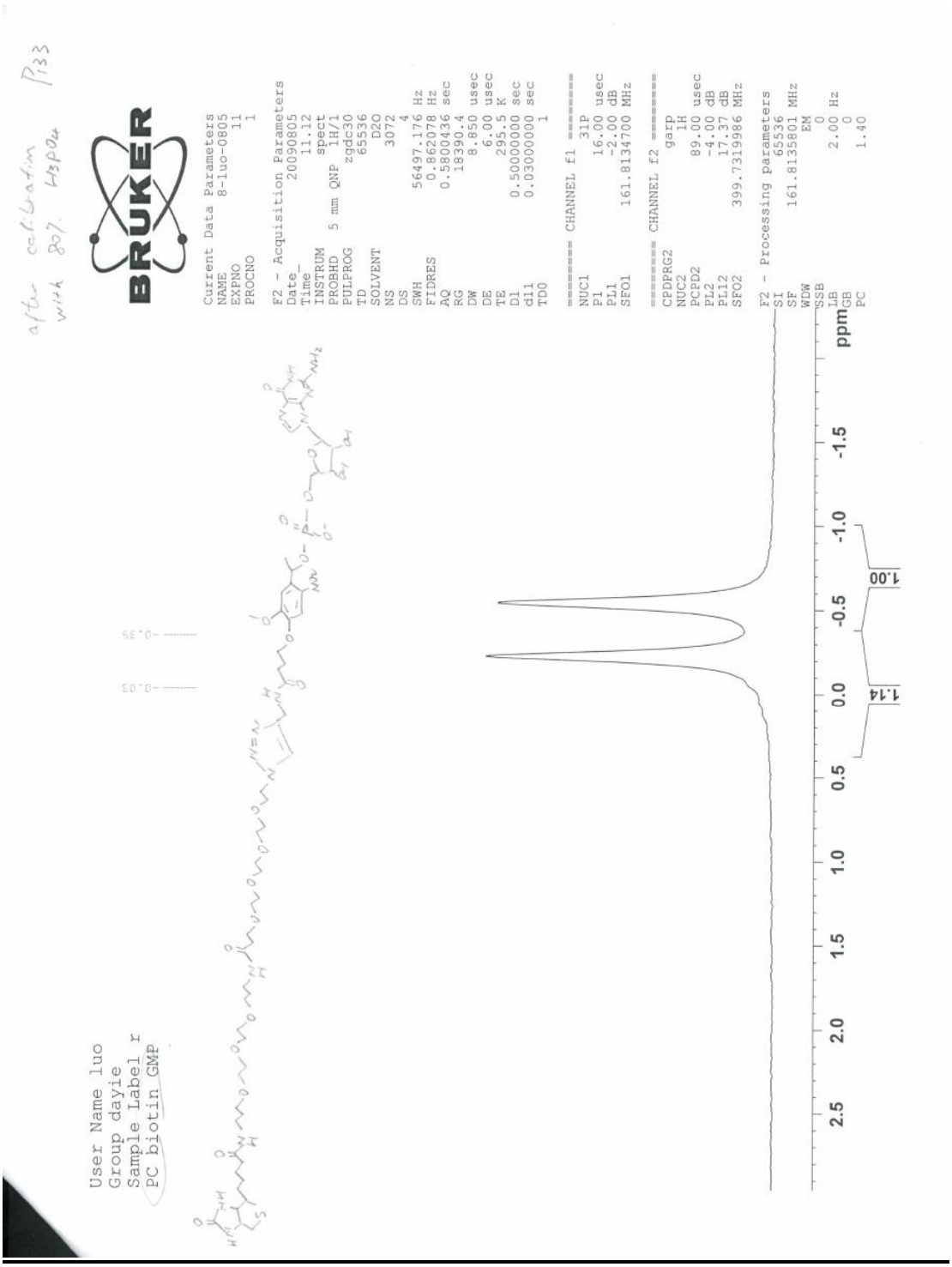
^{1}H NMR (400 MHz, D_2O) recorded for biotin-PC-GMP



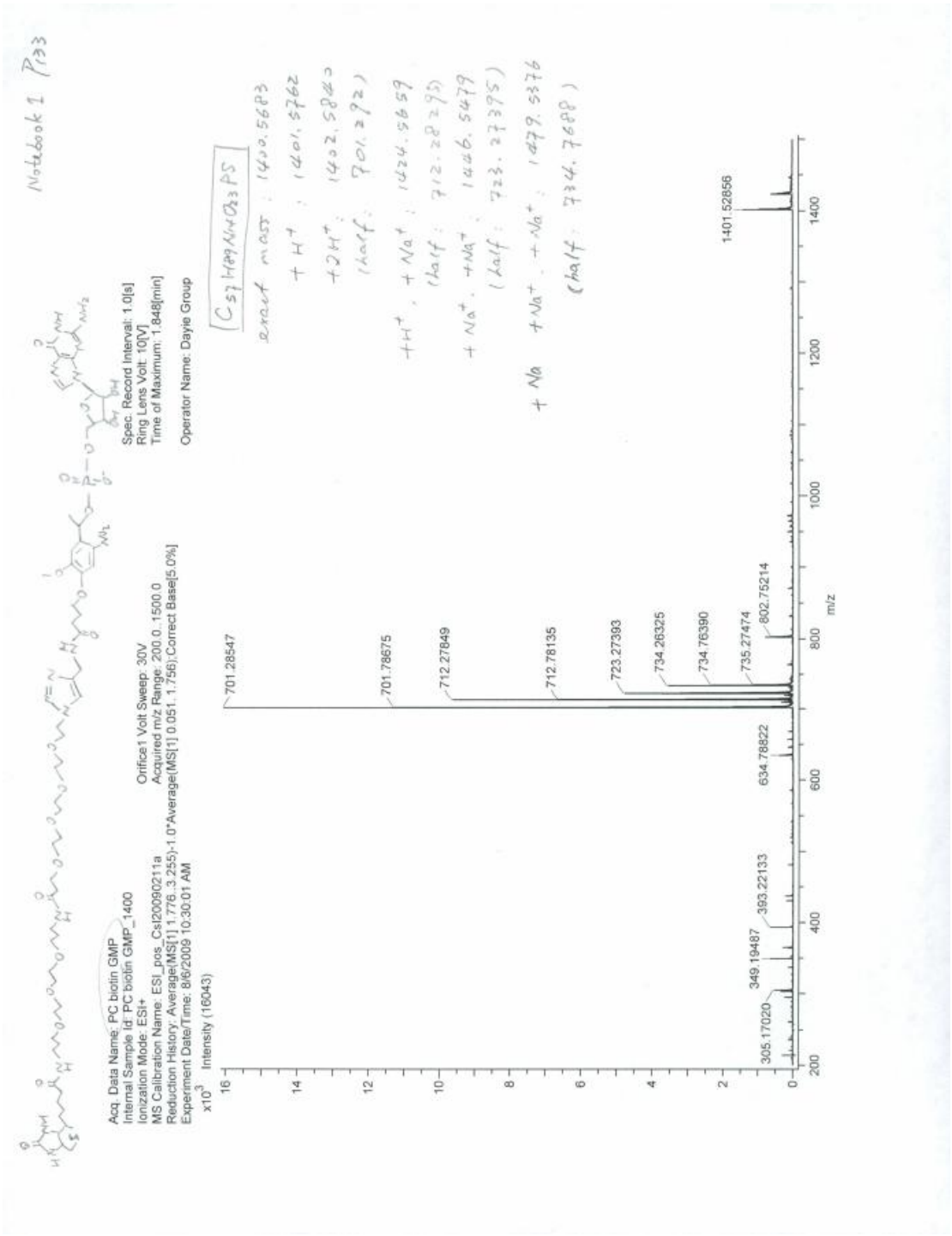
^{13}C NMR (400 MHz, D_2O) recorded for biotin-PC-GMP



^{31}P NMR (400 MHz, D_2O) recorded for biotin-PC-GMP



ESI-MS $[M+H]^+$ recorded for Biotin-PC-GMP



Appendix 4: Different ligands docking into Vc2 (**Chapter 5**)

(c-di-GMP, endo-S-c-di-GMP, compound **5)** and compound **6)** docking into Vc2 RNA (PDB: 3IRW). The different conformers of each ligand were optimized using Gaussian with fixed C5'-C5' distance of two guanines. Dockings were done in triplicate for each c-di-GMP and analog conformers.

Ligand	**Distance (Å) between C5 of two guanines in Gaussian optimized structure**	**Distance (Å) between C5 of two guanines after docking**			**Docked binding affinity (kcal/mol)**		
c-di-GMP	5.5	5.81	5.82	5.82	-12.2	-12.2	-12.2
	6.5	6.12	6.10	6.10	-11.8	-11.8	-11.8
	6.6	6.46	6.46	6.46	-14.6	-14.6	-14.7
	6.7	6.51	6.51	6.51	-15.1	-15.1	-15.1
	6.8	6.46	6.46	6.45	-13.6	-13.6	-13.5
	6.9	6.93	6.65	6.65	-11.3	-16.2	-16.2
	7.0	6.71	7.00	6.72	-16.7	-11.3	-16.7
	7.1	6.79	6.79	6.79	-17	-17.1	-17
	7.2	6.85	6.86	6.86	-17.1	-17.3	-16.9
	7.3	6.93	6.92	6.93	-17.4	-17.4	-17.4
	7.4	6.99	6.99	6.99	-17.5	-17.5	-17.5
	7.5	6.92	6.93	6.92	-16.9	-16.9	-16.9
	7.6	7.13	7.13	7.13	-17.5	-17.5	-17.5
	7.7	7.20	7.20	7.20	-17.3	-17.4	-17.4
	7.8	7.28	7.28	7.28	-17.1	-17	-17.1
	7.9	7.35	7.35	7.35	-16.7	-16.7	-16.7
	8.0	7.42	7.43	7.42	-16.2	-16.1	-16.2
	8.5	7.48	7.46	7.45	-12.5	-12.5	-12.5
	9.5	7.75	8.37	8.37	-10.4	-11.4	-11.3
endo-S-c-di-GMP	5.5	5.75	5.75	5.75	-11.1	-11.1	-11.2
	6.0	6.12	6.11	6.12	-11.5	-11.5	-11.5
	6.5	6.87	6.88	6.88	-13.1	-13	-13
	6.6	6.95	6.95	6.95	-13.3	-13.3	-13.2
	6.7	7.17	7.17	7.17	-13.1	-13.2	-13.1
	6.8	7.12	7.12	7.12	-12.6	-12.6	-12.6
	6.9	7.18	7.19	7.19	-13.9	-13.9	-13.9
	7.0	7.27	7.27	7.27	-13.9	-13.9	-13.9
	7.1	7.35	7.35	7.35	-14	-14	-14

	7.2	7.43	7.43	7.43	-14	-14	-13.9
	7.3	7.50	7.50	7.50	-13.9	-13.9	-13.9
	7.4	7.57	7.57	7.57	-13.9	-13.9	-13.9
	7.5	7.66	7.74	7.66	-12.6	-12.6	-12.6
	7.6	7.88	8.36	7.88	-12.3	-11.7	-12.3
	7.7	7.96	7.96	7.96	-12	-12	-12
	7.8	7.20	7.21	7.20	-13.5	-13.5	-13.5
	7.9	7.95	7.95	7.95	-12.9	-12.9	-12.9
	8.0	8.03	8.03	8.03	-12.5	-12.5	-12.5
	8.5	8.62	7.65	8.61	-11.3	-12.8	-11.2
	9.5	8.71	8.71	8.72	-12.1	-12.1	-12.1
Compound 5 (model for 2'-[Biotin]-endo-S-c-di-GMP)	5.5	7.50	7.19	7.18	-11.2	-11.3	-11.2
	6.5	6.73	6.73	6.73	-11.2	-11.2	-11.2
	6.8	7.00	6.99	7.00	-11.4	-11.3	-11.4
	6.9	7.09	7.09	7.09	-11.4	-11.4	-11.4
	7.0	7.20	7.20	7.20	-11.4	-11.4	-11.4
	7.1	7.20	7.20	7.20	-11.4	-11.4	-11.4
	7.2	7.41	7.41	7.41	-11.2	-11.2	-11.2
	7.3	7.53	7.53	7.53	-10.6	-10.8	-10.6
	7.4	7.48	7.62	7.48	-11.1	-10.5	-11.1
	7.5	7.68	7.68	7.68	-11.7	-11.7	-11.7
	7.6	7.84	7.84	7.84	-11.6	-11.6	-11.5
	7.7	7.98	7.98	7.98	-11.6	-11.6	-11.6
	7.8	9.26	9.26	9.26	-11.5	-11.5	-11.5
	8.0	9.48	9.48	9.48	-11.7	-11.6	-11.7
	8.5	9.47	9.47	9.69	-10.8	-10.8	-11.6
	9.5	10.50	10.07	10.49	-10.2	-10.9	-11.3
Compound 6 (model for 2'-[Biotin]-AHC-c-di-GMP	5.5	6.52	6.52	6.52	-9.7	-9.7	-9.7
	6.5	6.08	6.09	6.08	-10.2	-10.1	-10.1
	6.8	6.35	6.35	6.35	-10.1	-10.1	-10.1
	6.9	7.56	7.59	7.55	-10.4	-10.3	-10.3
	7.0	7.67	7.73	7.67	-10.3	-10.3	-10.4
	7.1	7.78	7.80	7.85	-10.2	-10.2	-10.1
	7.2	7.87	7.88	8.05	-10.2	-10.2	-10.2
	7.5	7.98	8.00	8.05	-10.6	-10.6	-10.5
	8.0	8.43	8.55	8.56	-10.4	-10.5	-10.6
	8.5	9.24	9.23	8.40	-11.7	-11.7	-9.8
	9.5	9.16	9.96	9.92	-10.1	-12.1	-11.4

Appendix 5: Different ligands docking into CdA. (**Chapter 5**)

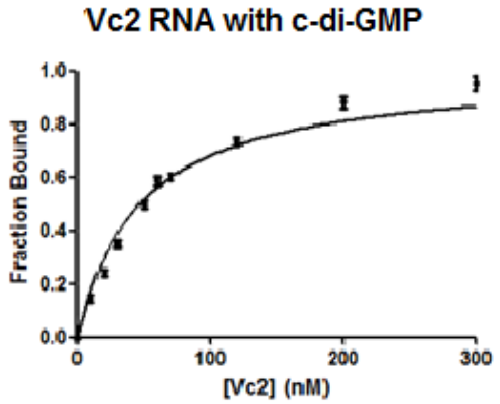
Different ligands (c-di-GMP, endo-S-c-di-GMP, compound 5) and compound 6) docking into CdA RNA (PDB: 3Q3Z). The different conformers of each ligand were optimized using Gaussian with fixed C5'-C5' distance of two guanines. Dockings were done in triplicate for each c-di-GMP and analog conformers.

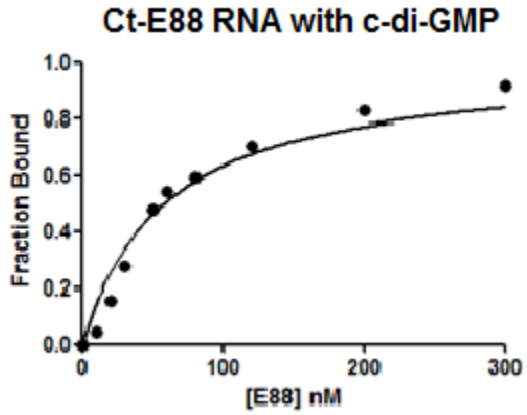
Ligand	Distance (Å) between C5 of two guanines in Gaussian optimized structure	Distance (Å) between C5 of two guanines after docking			Docked binding affinity (kcal/mol)		
c-di-GMP	5.5	5.97	5.97	5.97	-10.1	-10.1	-10.1
	6.5	6.5	6.5	6.5	-10.1	-10.1	-10.8
	6.6	6.6	6.6	6.6	-12.5	-12.5	-12.5
	6.7	6.7	6.7	6.7	-13.1	-13.1	-13
	6.8	6.7	6.8	6.7	-12.9	-12.9	-12.9
	6.9	6.8	6.8	6.8	-14	-14	-14
	7.0	6.9	6.9	6.9	-14.3	-14.3	-14.3
	7.1	7.0	7.0	7.0	-14.6	-14.6	-14.6
	7.2	7.0	7.0	7.0	-14.7	-14.7	-14.7
	7.3	7.1	7.1	7.1	-14.8	-14.8	-14.8
	7.4	7.2	7.2	7.2	-14.7	-14.7	-14.7
	7.5	7.2	7.2	7.2	-15.1	-15.1	-15.1
	7.6	7.3	7.3	7.3	-14.4	-14.4	-14.4
	7.7	7.4	7.4	7.4	-14.2	-14.2	-14.2
	7.8	7.4	7.4	7.4	-13.9	-13.9	-13.9
	7.9	7.5	7.5	7.5	-13.5	-13.5	-13.5
	8.0	7.6	7.6	7.6	-13.1	-13.1	-13.1
	8.5	7.6	7.6	7.6	-13.4	-13.4	-13.4
	9.5	8.1	8.1	8.1	-13.1	-13.1	-13.2
endo-S-c-di-GMP	5.5	5.5	5.5	5.5	-9.5	-9.5	-9.5
	6	5.8	5.8	5.8	-9.7	-9.6	-9.7
	6.5	6.4	6.4	6.4	-9.8	-9.8	-9.8
	6.6	6.4	6.5	6.4	-9.8	-9.8	-9.8
	6.7	7.4	7.4	7.4	-9.8	-9.8	-9.8
	6.8	6.7	6.7	6.7	-10.7	-10.7	-10.6
	6.9	7.5	7.5	7.5	-9.9	-9.8	-9.9
	7	7.5	7.5	7.5	-9.9	-9.9	-9.9
	7.1	7.6	7.6	7.6	-9.9	-9.9	-9.9
	7.2	7.7	7.7	7.7	-9.9	-9.9	-9.9
	7.3	7.1	7.1	7.1	-10	-10	-10
	7.4	7.2	7.2	7.2	-10.1	-10	-10.1

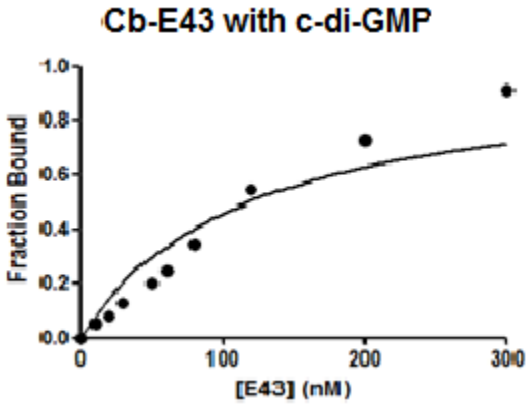
	7.5	7.7	7.7	7.7	-10.9	-10.9	-10.9
	7.6	8.1	8.1	8.1	-10.1	-10.1	-10.1
	7.7	8.2	8.2	8.2	-10.1	-10.1	-10.1
	7.8	8.2	8.2	8.3	-10.1	-10.1	-10.1
	7.9	7.3	7.3	7.3	-11.1	-11.1	-11.1
	8	7.4	7.4	7.4	-11.1	-11.1	-11
	8.5	7.9	7.9	8.0	-11.3	-11.3	-11.3
	9.5	9.2	9.2	9.2	-10.7	-10.7	-10.7
Compound 5 (model for 2'-[Biotin]-endo-S-c-di-GMP)	5.5	6.6	6.8	6.6	-9.8	-10	-9.8
	6.5	7.5	7.5	7.5	-10	-10	-10
	6.6	7.4	7.6	7.6	-10.4	-10	-10
	6.7	7.4	7.4	7.4	-10.5	-10.8	-10.5
	6.8	7.5	7.5	7.5	-10.9	-11	-10.9
	6.9	7.6	7.6	7.6	-10.8	-10.9	-10.9
	7	7.7	7.7	7.7	-11.1	-11	-10.9
	7.1	7.7	7.7	7.7	-11	-11	-10.9
	7.2	8.1	8.1	8.1	-10.3	-10.3	-10.3
	7.3	8.2	8.2	8.2	-10.2	-10.2	-10.2
	7.4	8.3	8.3	8.3	-10.3	-10.3	-10.3
	7.5	8.8	8.8	8.8	-11.4	-11.4	-11.4
	7.6	9.0	9.0	9.0	-10.5	-10.5	-10.5
	7.7	9.1	9.1	9.1	-10.4	-10.4	-10.4
	7.8	9.2	9.2	9.2	-10.3	-10.3	-10.3
	8	9.4	9.4	9.4	-10	-10	-10
	8.5	9.4	9.4	9.4	-11.1	-11.1	-11.1
	9.5	10.4	10.4	10.4	-11	-11	-11
Compound 6 (model for 2'-[Biotin] -AHC-c-di-GMP	5.5	6.4	6.4	6.5	-8.8	-8.8	-8.9
	6.5	7.4	7.4	7.4	-10	-10	-10
	6.8	7.0	7.0	7.0	-8.6	-8.7	-8.6
	6.9	8.0	8.0	8.0	-9.1	-9.1	-9.1
	7	8.1	8.1	8.1	-9.2	-9.2	-9.2
	7.1	8.2	8.2	8.2	-9.3	-9.3	-9.3
	7.2	8.3	8.3	8.3	-9.4	-9.4	-9.4
	7.5	8.3	8.3	8.3	-9.2	-9.2	-9.2
	8	8.9	8.9	8.9	-9.4	-9.4	-9.4
	8.5	9.4	9.4	9.4	-10.5	-10.5	-10.5
	9.5	10.4	10.4	10.4	-10.5	-10.6	-10.6

Appendix 6: Binding curves from gel-shift assay (**Chapter 6**)

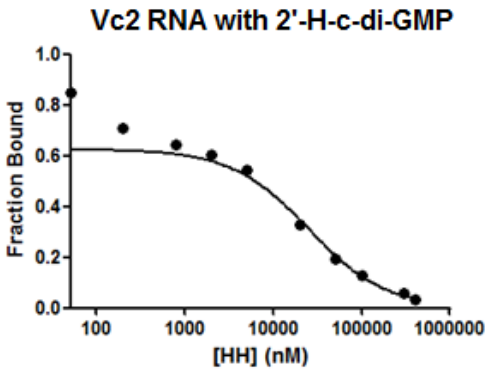
I. Binding curves of RNA with c-di-GMP

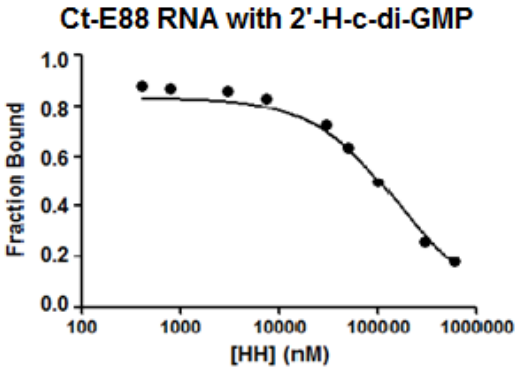


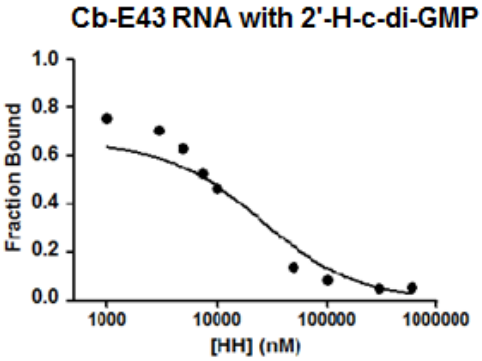




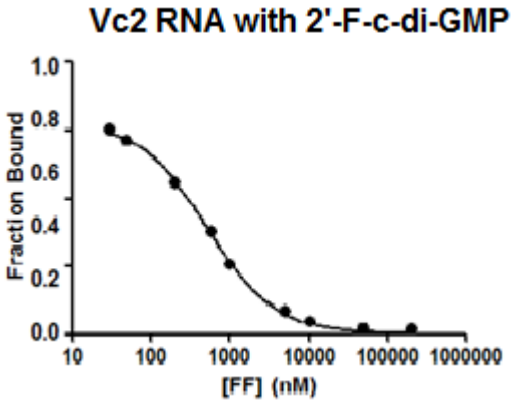
II. Competitive displacement curves of RNA with 2’-H-c-di-GMP

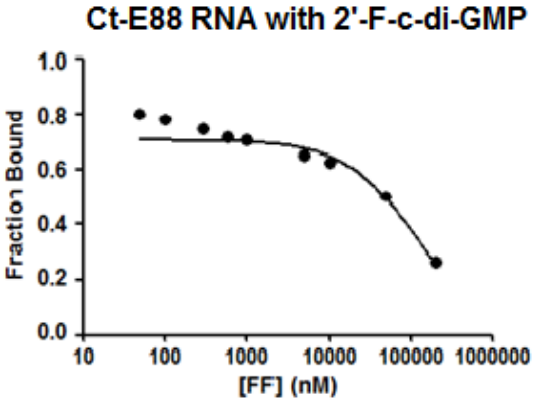


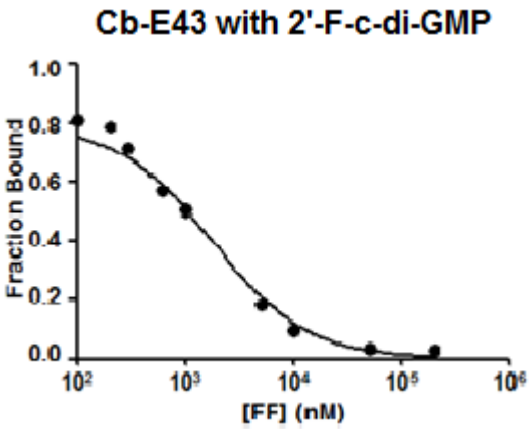




III. Competitive displacement curves of RNA with 2’-F-c-di-GMP







Bibliography

1. Breaker, R. R., Riboswitches: from ancient gene-control systems to modern drug targets. *Future Microbiol* **2009,** 4, (7), 771-3.
2. Newman, A. J.; Nagai, K., Structural studies of the spliceosome: blind men and an elephant. *Curr Opin Struct Biol* **2010,** 20, (1), 82-9.
3. Wahl, M. C.; Will, C. L.; Luhrmann, R., The spliceosome: design principles of a dynamic RNP machine. *Cell* **2009,** 136, (4), 701-18.
4. Korostelev, A.; Ermolenko, D. N.; Noller, H. F., Structural dynamics of the ribosome. *Curr Opin Chem Biol* **2008,** 12, (6), 674-83.
5. Steitz, T. A., A structural understanding of the dynamic ribosome machine. *Nat Rev Mol Cell Biol* **2008,** 9, (3), 242-53.
6. Moir, R. D.; Willis, I. M., Regulation of pol III transcription by nutrient and stress signaling pathways. *Biochim Biophys Acta* **2012**.
7. Ponting, C. P.; Oliver, P. L.; Reik, W., Evolution and functions of long noncoding RNAs. *Cell* **2009,** 136, (4), 629-41.
8. Mattick, J. S., A new paradigm for developmental biology. *J Exp Biol* **2007,** 210, (Pt 9), 1526-47.
9. Dunham, I. *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012,** 489, (7414), 57-74.
10. Djebali, S. *et al.*, Landscape of transcription in human cells. *Nature* **2012,** 489, (7414), 101-8.
11. Salazar, M.; Fedoroff, O. Y.; Miller, J. M.; Ribeiro, N. S.; Reid, B. R., The DNA strand in DNA.RNA hybrid duplexes is neither B-form nor A-form in solution. *Biochemistry* **1993,** 32, (16), 4207-15.
12. Pokrovskaya, I. D.; Gurevich, V. V., In vitro transcription: preparative RNA yields in analytical scale reactions. *Anal Biochem* **1994,** 220, (2), 420-3.
13. Milligan, J. F.; Uhlenbeck, O. C., Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol* **1989,** 180, 51-62.
14. Galburt, E. A.; Grill, S. W.; Wiedmann, A.; Lubkowska, L.; Choy, J.; Nogales, E.; Kashlev, M.; Bustamante, C., Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **2007,** 446, (7137), 820-3.
15. Yin, H.; Wang, M. D.; Svoboda, K.; Landick, R.; Block, S. M.; Gelles, J., Transcription against an applied force. *Science* **1995,** 270, (5242), 1653-7.
16. Schafer, D. A.; Gelles, J.; Sheetz, M. P.; Landick, R., Transcription by single molecules of RNA polymerase observed by light microscopy. *Nature* **1991,** 352, (6334), 444-8.
17. Paschal, B. M.; McReynolds, L. A.; Noren, C. J.; Nichols, N. M., RNA polymerases. *Curr Protoc Mol Biol* **2008,** Chapter 3, Unit3.8.
18. Helm, M.; Brule, H.; Giege, R.; Florentz, C., More mistakes by T7 RNA polymerase at the 5' ends of in vitro-transcribed RNAs. *RNA* **1999,** 5, (5), 618-21.
19. Pleiss, J. A.; Derrick, M. L.; Uhlenbeck, O. C., T7 RNA polymerase produces 5' end heterogeneity during in vitro transcription from certain templates. *RNA* **1998,** 4, (10), 1313-7.

20. Avis, J. M.; Conn, G. L.; Walker, S. C., Cis-acting ribozymes for the production of RNA in vitro transcripts with defined 5' and 3' ends. *Methods Mol Biol* **2012,** 941, 83-98.
21. Schurer, H.; Lang, K.; Schuster, J.; Morl, M., A universal method to produce in vitro transcripts with homogeneous 3' ends. *Nucleic Acids Res* **2002,** 30, (12), e56.
22. Stage-Zimmermann, T. K.; Uhlenbeck, O. C., Hammerhead ribozyme kinetics. *RNA* **1998,** 4, (8), 875-89.
23. Korennykh, A. V.; Plantinga, M. J.; Correll, C. C.; Piccirilli, J. A., Linkage between substrate recognition and catalysis during cleavage of sarcin/ricin loop RNA by restrictocin. *Biochemistry* **2007,** 46, (44), 12744-56.
24. Uhlenbeck, O. C., Keeping RNA happy. *RNA* **1995,** 1, (1), 4-6.
25. Ditzler, M. A.; Rueda, D.; Mo, J.; Hakansson, K.; Walter, N. G., A rugged free energy landscape separates multiple functional RNA folds throughout denaturation. *Nucleic Acids Res* **2008,** 36, (22), 7088-99.
26. Pereira, M. J.; Behera, V.; Walter, N. G., Nondenaturing purification of co-transcriptionally folded RNA avoids common folding heterogeneity. *PLoS One* **2010,** 5, (9), e12953.
27. Lukavsky, P. J.; Puglisi, J. D., Large-scale preparation and purification of polyacrylamide-free RNA oligonucleotides. *RNA* **2004,** 10, (5), 889-93.
28. Di Tomasso, G.; Lampron, P.; Dagenais, P.; Omichinski, J. G.; Legault, P., The ARiBo tag: a reliable tool for affinity purification of RNAs under native conditions. *Nucleic Acids Res* **2011,** 39, (3), e18.
29. Vicens, Q.; Gooding, A. R.; Duarte, L. F.; Batey, R. T., Preparation of group I introns for biochemical studies and crystallization assays by native affinity purification. *PLoS One* **2009,** 4, (8), e6740.
30. Batey, R. T.; Kieft, J. S., Improved native affinity purification of RNA. *RNA* **2007,** 13, (8), 1384-9.
31. Kieft, J. S.; Batey, R. T., A general method for rapid and nondenaturing purification of RNAs. *RNA* **2004,** 10, (6), 988-95.
32. Cheong, H. K.; Hwang, E.; Lee, C.; Choi, B. S.; Cheong, C., Rapid preparation of RNA samples for NMR spectroscopy and X-ray crystallography. *Nucleic Acids Res* **2004,** 32, (10), e84.
33. Easton, L. E.; Shibata, Y.; Lukavsky, P. J., Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA* **2010,** 16, (3), 647-53.
34. Chen, Y.; Eldho, N. V.; Dayie, T. K.; Carey, P. R., Probing adenine rings and backbone linkages using base specific isotope-edited Raman spectroscopy: application to group II intron ribozyme domain V. *Biochemistry* **2010,** 49, (16), 3427-35.
35. Kim, I.; McKenna, S. A.; Viani Puglisi, E.; Puglisi, J. D., Rapid purification of RNAs using fast performance liquid chromatography (FPLC). *RNA* **2007,** 13, (2), 289-94.
36. Ponchon, L.; Dardel, F., Large scale expression and purification of recombinant RNA in Escherichia coli. *Methods* **2011,** 54, (2), 267-73.

37. McKenna, S. A.; Kim, I.; Puglisi, E. V.; Lindhout, D. A.; Aitken, C. E.; Marshall, R. A.; Puglisi, J. D., Purification and characterization of transcribed RNAs using gel filtration chromatography. *Nat Protoc* **2007,** 2, (12), 3270-7.
38. Luo, Y.; Eldho, N. V.; Sintim, H. O.; Dayie, T. K., RNAs synthesized using photocleavable biotinylated nucleotides have dramatically improved catalytic efficiency. *Nucleic Acids Res* **2011,** 39, (19), 8559-71.
39. Diamandis, E. P.; Christopoulos, T. K., The biotin-(strept)avidin system: principles and applications in biotechnology. *Clin Chem* **1991,** 37, (5), 625-36.
40. Jones, M. L.; Kurzban, G. P., Noncooperativity of biotin binding to tetrameric streptavidin. *Biochemistry* **1995,** 34, (37), 11750-6.
41. Green, N. M.; Toms, E. J., The dissociation of avidin-biotin complexes by guanidinium chloride. *Biochem J* **1972,** 130, (3), 707-11.
42. Hofmann, K.; Wood, S. W.; Brinton, C. C.; Montibeller, J. A.; Finn, F. M., Iminobiotin affinity columns and their application to retrieval of streptavidin. *Proc Natl Acad Sci U S A* **1980,** 77, (8), 4666-8.
43. Soukup, G. A.; Cerny, R. L.; Maher, L. J., 3rd, Preparation of oligonucleotide-biotin conjugates with cleavable linkers. *Bioconjug Chem* **1995,** 6, (1), 135-8.
44. Dawson, B. A.; Herman, T.; Lough, J., Affinity isolation of transcriptionally active murine erythroleukemia cell DNA using a cleavable biotinylated nucleotide analog. *J Biol Chem* **1989,** 264, (22), 12830-7.
45. Herman, T. M.; Lefever, E.; Shimkus, M., Affinity chromatography of DNA labeled with chemically cleavable biotinylated nucleotide analogs. *Anal Biochem* **1986,** 156, (1), 48-55.
46. Shimkus, M.; Levy, J.; Herman, T., A chemically cleavable biotinylated nucleotide: usefulness in the recovery of protein-DNA complexes from avidin affinity columns. *Proc Natl Acad Sci U S A* **1985,** 82, (9), 2593-7.
47. Saran, D.; Burke, D. H., A versatile photocleavable bifunctional linker for facile synthesis of substrate-DNA conjugates for the selection of nucleic acid catalysts. *Bioconjug Chem* **2007,** 18, (1), 275-9.
48. Ide, G. J., Nucleoside 5'-[gamma-S]triphosphates will initiate transcription in isolated yeast nuclei. *Biochemistry* **1981,** 20, (9), 2633-8.
49. Sampson, J. R.; Uhlenbeck, O. C., Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed in vitro. *Proc Natl Acad Sci U S A* **1988,** 85, (4), 1033-7.
50. Gumbs, O. H.; Padgett, R. A.; Dayie, K. T., Fluorescence and solution NMR study of the active site of a 160-kDa group II intron ribozyme. *RNA* **2006,** 12, (9), 1693-707.
51. Dayie, K. T., Key labeling technologies to tackle sizeable problems in RNA structural biology. *Int J Mol Sci* **2008,** 9, (7), 1214-40.
52. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl* **2001,** 40, (11), 2004-2021.
53. El-Sagheer, A. H.; Brown, T., Click chemistry with DNA. *Chem Soc Rev* **2010,** 39, (4), 1388-405.

54. Allerson, C. R.; Verdine, G. L., Synthesis and biochemical evaluation of RNA containing an intrahelical disulfide crosslink. *Chem Biol* **1995,** 2, (10), 667-75.
55. Olejnik, J.; Krzymanska-Olejnik, E.; Rothschild, K. J., Photocleavable affinity tags for isolation and detection of biomolecules. *Methods Enzymol* **1998,** 291, 135-54.
56. Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C., Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* **1987,** 15, (21), 8783-98.
57. Huang, F.; He, J.; Zhang, Y.; Guo, Y., Synthesis of biotin-AMP conjugate for 5' biotin labeling of RNA through one-step in vitro transcription. *Nat Protoc* **2008,** 3, (12), 1848-61.
58. Fourmy, D.; Recht, M. I.; Blanchard, S. C.; Puglisi, J. D., Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **1996,** 274, (5291), 1367-71.
59. Michel, F.; Costa, M.; Westhof, E., The ribozyme core of group II introns: a structure in want of partners. *Trends Biochem Sci* **2009,** 34, (4), 189-99.
60. Kao, C.; Zheng, M.; Rudisser, S., A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **1999,** 5, (9), 1268-72.
61. Moran, S.; Ren, R. X.; Sheils, C. J.; Rumney, S. t.; Kool, E. T., Non-hydrogen bonding 'terminator' nucleosides increase the 3'-end homogeneity of enzymatic RNA and DNA synthesis. *Nucleic Acids Res* **1996,** 24, (11), 2044-52.
62. Jarrell, K. A.; Peebles, C. L.; Dietrich, R. C.; Romiti, S. L.; Perlman, P. S., Group II intron self-splicing. Alternative reaction conditions yield novel products. *J Biol Chem* **1988,** 263, (7), 3432-9.
63. Collett, J. R.; Cho, E. J.; Lee, J. F.; Levy, M.; Hood, A. J.; Wan, C.; Ellington, A. D., Functional RNA microarrays for high-throughput screening of antiprotein aptamers. *Anal Biochem* **2005,** 338, (1), 113-23.
64. Schlatterer, J. C.; Jaschke, A., Universal initiator nucleotides for the enzymatic synthesis of 5'-amino- and 5'-thiol-modified RNA. *Biochem Biophys Res Commun* **2006,** 344, (3), 887-92.
65. Fusz, S.; Srivatsan, S. G.; Ackermann, D.; Famulok, M., Photocleavable initiator nucleotide substrates for an aldolase ribozyme. *J Org Chem* **2008,** 73, (13), 5069-77.
66. Uhlenbeck, O. C., RNA biophysics has come of age. *Biopolymers* **2009,** 91, (10), 811-4.
67. Ross, P.; Weinhouse, H.; Aloni, Y.; Michaeli, D.; Weinberger-Ohana, P.; Mayer, R.; Braun, S.; de Vroom, E.; van der Marel, G. A.; van Boom, J. H.; Benziman, M., Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. *Nature* **1987,** 325, (6101), 279-81.
68. D'Argenio, D. A.; Miller, S. I., Cyclic di-GMP as a bacterial second messenger. *Microbiology* **2004,** 150, (Pt 8), 2497-502.
69. Romling, U.; Gomelsky, M.; Galperin, M. Y., C-di-GMP: the dawning of a novel bacterial signalling system. *Mol Microbiol* **2005,** 57, (3), 629-39.
70. Jenal, U.; Malone, J., Mechanisms of cyclic-di-GMP signaling in bacteria. *Annu Rev Genet* **2006,** 40, 385-407.
71. Hengge, R., Principles of c-di-GMP signalling in bacteria. *Nat Rev Microbiol* **2009,** 7, (4), 263-73.

72. Pitzer, J. E.; Sultan, S. Z.; Hayakawa, Y.; Hobbs, G.; Miller, M. R.; Motaleb, M. A., Analysis of the Borrelia burgdorferi cyclic-di-GMP-binding protein PlzA reveals a role in motility and virulence. *Infect Immun* **2011,** 79, (5), 1815-25.
73. Ahmad, I.; Lamprokostopoulou, A.; Le Guyon, S.; Streck, E.; Barthel, M.; Peters, V.; Hardt, W. D.; Romling, U., Complex c-di-GMP signaling networks mediate transition between virulence properties and biofilm formation in Salmonella enterica serovar Typhimurium. *PLoS One* **2011,** 6, (12), e28351.
74. Tamayo, R.; Pratt, J. T.; Camilli, A., Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. *Annu Rev Microbiol* **2007,** 61, 131-48.
75. Tischler, A. D.; Camilli, A., Cyclic diguanylate regulates Vibrio cholerae virulence gene expression. *Infect Immun* **2005,** 73, (9), 5873-82.
76. Karaolis, D. K.; Means, T. K.; Yang, D.; Takahashi, M.; Yoshimura, T.; Muraille, E.; Philpott, D.; Schroeder, J. T.; Hyodo, M.; Hayakawa, Y.; Talbot, B. G.; Brouillette, E.; Malouin, F., Bacterial c-di-GMP is an immunostimulatory molecule. *J Immunol* **2007,** 178, (4), 2171-81.
77. Paul, R.; Weiser, S.; Amiot, N. C.; Chan, C.; Schirmer, T.; Giese, B.; Jenal, U., Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev* **2004,** 18, (6), 715-27.
78. Schirmer, T.; Jenal, U., Structural and mechanistic determinants of c-di-GMP signalling. *Nat Rev Microbiol* **2009,** 7, (10), 724-35.
79. Schmidt, A. J.; Ryjenkov, D. A.; Gomelsky, M., The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. *J Bacteriol* **2005,** 187, (14), 4774-81.
80. Ryan, R. P.; Fouhy, Y.; Lucey, J. F.; Crossman, L. C.; Spiro, S.; He, Y. W.; Zhang, L. H.; Heeb, S.; Camara, M.; Williams, P.; Dow, J. M., Cell-cell signaling in Xanthomonas campestris involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. *Proc Natl Acad Sci U S A* **2006,** 103, (17), 6712-7.
81. Seshasayee, A. S.; Fraser, G. M.; Luscombe, N. M., Comparative genomics of cyclic-di-GMP signalling in bacteria: post-translational regulation and catalytic activity. *Nucleic Acids Res* **2010,** 38, (18), 5970-81.
82. De, N.; Navarro, M. V.; Raghavan, R. V.; Sondermann, H., Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. *J Mol Biol* **2009,** 393, (3), 619-33.
83. Hecht, G. B.; Newton, A., Identification of a novel response regulator required for the swarmer-to-stalked-cell transition in Caulobacter crescentus. *J Bacteriol* **1995,** 177, (21), 6223-9.
84. Barraud, N.; Storey, M. V.; Moore, Z. P.; Webb, J. S.; Rice, S. A.; Kjelleberg, S., Nitric oxide-mediated dispersal in single- and multi-species biofilms of clinically and industrially relevant microorganisms. *Microb Biotechnol* **2009,** 2, (3), 370-8.
85. Qi, Y.; Chuah, M. L.; Dong, X.; Xie, K.; Luo, Z.; Tang, K.; Liang, Z. X., Binding of cyclic diguanylate in the non-catalytic EAL domain of FimX induces a long-range conformational change. *J Biol Chem* **2011,** 286, (4), 2910-7.
86. Kanazawa, T.; Ren, S.; Maekawa, M.; Hasegawa, K.; Arisaka, F.; Hyodo, M.; Hayakawa, Y.; Ohta, H.; Masuda, S., Biochemical and physiological characterization of a BLUF protein-EAL protein complex involved in blue light-dependent

degradation of cyclic diguanylate in the purple bacterium Rhodopseudomonas palustris. *Biochemistry* **2010,** 49, (50), 10647-55.
87. Galperin, M. Y.; Nikolskaya, A. N.; Koonin, E. V., Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol Lett* **2001,** 203, (1), 11-21.
88. Benach, J.; Swaminathan, S. S.; Tamayo, R.; Handelman, S. K.; Folta-Stogniew, E.; Ramos, J. E.; Forouhar, F.; Neely, H.; Seetharaman, J.; Camilli, A.; Hunt, J. F., The structural basis of cyclic diguanylate signal transduction by PilZ domains. *EMBO J* **2007,** 26, (24), 5153-66.
89. Beyhan, S.; Odell, L. S.; Yildiz, F. H., Identification and characterization of cyclic diguanylate signaling systems controlling rugosity in Vibrio cholerae. *J Bacteriol* **2008,** 190, (22), 7392-405.
90. Rakshe, S.; Leff, M.; Spormann, A. M., Indirect modulation of the intracellular c-Di-GMP level in Shewanella oneidensis MR-1 by MxdA. *Appl Environ Microbiol* **2011,** 77, (6), 2196-8.
91. Petters, T.; Zhang, X.; Nesper, J.; Treuner-Lange, A.; Gomez-Santos, N.; Hoppert, M.; Jenal, U.; Sogaard-Andersen, L., The orphan histidine protein kinase SgmT is a c-di-GMP receptor and regulates composition of the extracellular matrix together with the orphan DNA binding response regulator DigR in Myxococcus xanthus. *Mol Microbiol* **2012,** 84, (1), 147-65.
92. Duerig, A.; Abel, S.; Folcher, M.; Nicollier, M.; Schwede, T.; Amiot, N.; Giese, B.; Jenal, U., Second messenger-mediated spatiotemporal control of protein degradation regulates bacterial cell cycle progression. *Genes Dev* **2009,** 23, (1), 93-104.
93. Tuckerman, J. R.; Gonzalez, G.; Gilles-Gonzalez, M. A., Cyclic di-GMP activation of polynucleotide phosphorylase signal-dependent RNA processing. *J Mol Biol* **2011,** 407, (5), 633-9.
94. Lee, V. T.; Matewish, J. M.; Kessler, J. L.; Hyodo, M.; Hayakawa, Y.; Lory, S., A cyclic-di-GMP receptor required for bacterial exopolysaccharide production. *Mol Microbiol* **2007,** 65, (6), 1474-84.
95. Hickman, J. W.; Harwood, C. S., Identification of FleQ from Pseudomonas aeruginosa as a c-di-GMP-responsive transcription factor. *Mol Microbiol* **2008,** 69, (2), 376-89.
96. Leduc, J. L.; Roberts, G. P., Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of Xanthomonas axonopodis pv. citri. *J Bacteriol* **2009,** 191, (22), 7121-2.
97. Srivastava, D.; Harris, R. C.; Waters, C. M., Integration of cyclic di-GMP and quorum sensing in the control of vpsT and aphA in Vibrio cholerae. *J Bacteriol* **2011,** 193, (22), 6331-41.
98. Krasteva, P. V.; Fong, J. C.; Shikuma, N. J.; Beyhan, S.; Navarro, M. V.; Yildiz, F. H.; Sondermann, H., Vibrio cholerae VpsT regulates matrix production and motility by directly sensing cyclic di-GMP. *Science* **2010,** 327, (5967), 866-8.
99. Newell, P. D.; Monds, R. D.; O'Toole, G. A., LapD is a bis-(3',5')-cyclic dimeric GMP-binding protein that regulates surface attachment by Pseudomonas fluorescens Pf0-1. *Proc Natl Acad Sci U S A* **2009,** 106, (9), 3461-6.

100. Pratt, J. T.; Tamayo, R.; Tischler, A. D.; Camilli, A., PilZ domain proteins bind cyclic diguanylate and regulate diverse processes in Vibrio cholerae. *J Biol Chem* **2007,** 282, (17), 12860-70.
101. Amikam, D.; Galperin, M. Y., PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics* **2006,** 22, (1), 3-6.
102. Sudarsan, N.; Lee, E. R.; Weinberg, Z.; Moy, R. H.; Kim, J. N.; Link, K. H.; Breaker, R. R., Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* **2008,** 321, (5887), 411-3.
103. Lee, E. R.; Baker, J. L.; Weinberg, Z.; Sudarsan, N.; Breaker, R. R., An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science* **2010,** 329, (5993), 845-8.
104. Garst, A. D.; Heroux, A.; Rambo, R. P.; Batey, R. T., Crystal structure of the lysine riboswitch regulatory mRNA element. *J Biol Chem* **2008,** 283, (33), 22347-51.
105. Batey, R. T.; Gilbert, S. D.; Montange, R. K., Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **2004,** 432, (7015), 411-5.
106. Nahvi, A.; Barrick, J. E.; Breaker, R. R., Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res* **2004,** 32, (1), 143-50.
107. Klein, D. J.; Ferre-D'Amare, A. R., Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* **2006,** 313, (5794), 1752-6.
108. Edwards, T. E.; Ferre-D'Amare, A. R., Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure* **2006,** 14, (9), 1459-68.
109. Ramesh, A.; Wakeman, C. A.; Winkler, W. C., Insights into metalloregulation by M-box riboswitch RNAs via structural analysis of manganese-bound complexes. *J Mol Biol* **2011,** 407, (4), 556-70.
110. Ren, A.; Rajashankar, K. R.; Patel, D. J., Fluoride ion encapsulation by Mg2+ ions and phosphates in a fluoride riboswitch. *Nature* **2012,** 486, (7401), 85-9.
111. Chen, B.; Zuo, X.; Wang, Y. X.; Dayie, T. K., Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy. *Nucleic Acids Res* **2012,** 40, (7), 3117-30.
112. Breaker, R. R., Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol* **2012,** 4, (2).
113. Smith, K. D.; Lipchock, S. V.; Ames, T. D.; Wang, J.; Breaker, R. R.; Strobel, S. A., Structural basis of ligand binding by a c-di-GMP riboswitch. *Nat Struct Mol Biol* **2009,** 16, (12), 1218-23.
114. Smith, K. D.; Shanahan, C. A.; Moore, E. L.; Simon, A. C.; Strobel, S. A., Structural basis of differential ligand recognition by two classes of bis-(3'-5')-cyclic dimeric guanosine monophosphate-binding riboswitches. *Proc Natl Acad Sci U S A* **2011,** 108, (19), 7757-62.
115. Kulshina, N.; Baird, N. J.; Ferre-D'Amare, A. R., Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch. *Nat Struct Mol Biol* **2009,** 16, (12), 1212-7.

116. Smith, K. D.; Lipchock, S. V.; Livingston, A. L.; Shanahan, C. A.; Strobel, S. A., Structural and biochemical determinants of ligand binding by the c-di-GMP riboswitch. *Biochemistry* **2010,** 49, (34), 7351-9.
117. Shanahan, C. A.; Gaffney, B. L.; Jones, R. A.; Strobel, S. A., Differential analogue binding by two classes of c-di-GMP riboswitches. *J Am Chem Soc* **2011,** 133, (39), 15578-92.
118. Smith, K. D.; Lipchock, S. V.; Strobel, S. A., Structural and biochemical characterization of linear dinucleotide analogues bound to the c-di-GMP-I aptamer. *Biochemistry* **2012,** 51, (1), 425-32.
119. Luo, Y.; Zhou, J.; Watt, S. K.; Lee, V. T.; Dayie, T. K.; Sintim, H. O., Differential binding of 2'-biotinylated analogs of c-di-GMP with c-di-GMP riboswitches and binding proteins. *Mol Biosyst* **2012,** 8, (3), 772-8.
120. Furukawa, K.; Gu, H.; Sudarsan, N.; Hayakawa, Y.; Hyodo, M.; Breaker, R. R., Identification of ligand analogues that control c-di-GMP riboswitches. *ACS Chem Biol* **2012,** 7, (8), 1436-43.
121. Wang, J.; Zhou, J.; Donaldson, G. P.; Nakayama, S.; Yan, L.; Lam, Y. F.; Lee, V. T.; Sintim, H. O., Conservative change to the phosphate moiety of cyclic diguanylic monophosphate remarkably affects its polymorphism and ability to bind DGC, PDE, and PilZ proteins. *J Am Chem Soc* **2011,** 133, (24), 9320-30.
122. Winkler, W. C.; Nahvi, A.; Sudarsan, N.; Barrick, J. E.; Breaker, R. R., An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Biol* **2003,** 10, (9), 701-7.
123. Winkler, W. C.; Breaker, R. R., Genetic control by metabolite-binding riboswitches. *Chembiochem* **2003,** 4, (10), 1024-32.
124. Draper, D. E.; Grilley, D.; Soto, A. M., Ions and RNA folding. *Annu Rev Biophys Biomol Struct* **2005,** 34, 221-43.
125. Schwalbe, H.; Buck, J.; Furtig, B.; Noeske, J.; Wohnert, J., Structures of RNA switches: insight into molecular recognition and tertiary structure. *Angew Chem Int Ed Engl* **2007,** 46, (8), 1212-9.
126. Haller, A.; Rieder, U.; Aigner, M.; Blanchard, S. C.; Micura, R., Conformational capture of the SAM-II riboswitch. *Nat Chem Biol* **2011,** 7, (6), 393-400.
127. Wood, S.; Ferre-D'Amare, A. R.; Rueda, D., Allosteric tertiary interactions preorganize the c-di-GMP riboswitch and accelerate ligand binding. *ACS Chem Biol* **2012,** 7, (5), 920-7.
128. Baird, N. J.; Kulshina, N.; Ferre-D'Amare, A. R., Riboswitch function: flipping the switch or tuning the dimmer? *RNA Biol* **2010,** 7, (3), 328-32.
129. Kiburu, I.; Shurer, A.; Yan, L.; Sintim, H. O., A simple solid-phase synthesis of the ubiquitous bacterial signaling molecule, c-di-GMP and analogues. *Mol Biosyst* **2008,** 4, (6), 518-20.
130. Ching, S. M.; Tan, W. J.; Chua, K. L.; Lam, Y., Synthesis of cyclic di-nucleotidic acids as potential inhibitors targeting diguanylate cyclase. *Bioorg Med Chem* **2010,** 18, (18), 6657-65.
131. Gaffney, B. L.; Veliath, E.; Zhao, J.; Jones, R. A., One-flask syntheses of c-di-GMP and the [Rp,Rp] and [Rp,Sp] thiophosphate analogues. *Org Lett* **2010,** 12, (14), 3269-71.

132. Grajkowski, A.; Cieslak, J.; Gapeev, A.; Schindler, C.; Beaucage, S. L., Convenient synthesis of a propargylated cyclic (3'-5') diguanylic acid and its "click" conjugation to a biotinylated azide. *Bioconjug Chem* **2010,** 21, (11), 2147-52.
133. Donaldson, G. P.; Roelofs, K. G.; Luo, Y.; Sintim, H. O.; Lee, V. T., A rapid assay for affinity and kinetics of molecular interactions with nucleic acids. *Nucleic Acids Res* **2012,** 40, (7), e48.
134. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **2010,** 31, (2), 455-61.
135. Avci, D.; Comert, H.; Atalay, Y., Ab initio Hartree-Fock calculations on linear and second-order nonlinear optical properties of new acridine-benzothiazolylamine chromophores. *J Mol Model* **2008,** 14, (2), 161-9.
136. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew Chem Int Ed Engl* **2002,** 41, (14), 2596-9.
137. Smietana, M.; Kool, E. T., Efficient and simple solid-phase synthesis of short cyclic oligodeoxynucleotides bearing a phosphorothioate linkage. *Angew Chem Int Ed Engl* **2002,** 41, (19), 3704-7; 3523.
138. Holt, P. A.; Chaires, J. B.; Trent, J. O., Molecular docking of intercalators and groove-binders to nucleic acids using Autodock and Surflex. *J Chem Inf Model* **2008,** 48, (8), 1602-15.
139. Chiou, S. H.; Kao, C. L.; Chang, Y. L.; Ku, H. H.; Tsai, Y. J.; Lin, H. T.; Yen, C. J.; Peng, C. H.; Chiu, J. H.; Tsai, T. H., Evaluation of anti-Fas ligand-induced apoptosis and neural differentiation of PC12 cells treated with nerve growth factor using small interfering RNA method and sampling by microdialysis. *Anal Biochem* **2007,** 363, (1), 46-57.
140. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G., Clustal W and Clustal X version 2.0. *Bioinformatics* **2007,** 23, (21), 2947-8.
141. Gouet, P.; Courcelle, E.; Stuart, D. I.; Metoz, F., ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **1999,** 15, (4), 305-8.
142. Gardner, P. P.; Daub, J.; Tate, J. G.; Nawrocki, E. P.; Kolbe, D. L.; Lindgreen, S.; Wilkinson, A. C.; Finn, R. D.; Griffiths-Jones, S.; Eddy, S. R.; Bateman, A., Rfam: updates to the RNA families database. *Nucleic Acids Res* **2009,** 37, (Database issue), D136-40.
143. Pollard, T. D., A guide to simple and informative binding assays. *Mol Biol Cell* **2010,** 21, (23), 4061-7.
144. Weeks, K. M.; Crothers, D. M., RNA binding assays for Tat-derived peptides: implications for specificity. *Biochemistry* **1992,** 31, (42), 10281-7.
145. Lin, S. Y.; Riggs, A. D., Lac repressor binding to non-operator DNA: detailed studies and a comparison of equilibrium and rate competition methods. *J Mol Biol* **1972,** 72, (3), 671-90.
146. Speirs, G.; Warren, R. E.; Rampling, A., Clostridium tertium septicemia in patients with neutropenia. *J Infect Dis* **1988,** 158, (6), 1336-40.

147. Hambleton, P., Clostridium botulinum toxins: a general review of involvement in disease, structure, mode of action and preparation for clinical use. *J Neurol* **1992,** 239, (1), 16-20.
148. Anand, A.; Glatt, A. E., Clostridium difficile infection associated with antineoplastic chemotherapy: a review. *Clin Infect Dis* **1993,** 17, (1), 109-13.
149. Luy, B.; Marino, J. P., Measurement and application of H-1-F-19 dipolar couplings in the structure determination of 2'-fluorolabeled RNA. *J. Biomol. NMR* **2001,** 20, (1), 39-47.
150. Kool, E. T.; Sintim, H. O., The difluorotoluene debate--a decade later. *Chem Commun (Camb)* **2006**, (35), 3665-75.
151. Baird, N. J.; Ferre-D'Amare, A. R., Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis. *RNA* **2010,** 16, (3), 598-609.
152. Dayie, K. T., Resolution enhanced homonuclear carbon decoupled triple resonance experiments for unambiguous RNA structural characterization. *J Biomol NMR* **2005,** 32, (2), 129-39.
153. Wilkinson, K. A.; Merino, E. J.; Weeks, K. M., Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **2006,** 1, (3), 1610-6.
154. Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M., RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* **2005,** 127, (12), 4223-31.
155. Wang, Y. X.; Zuo, X.; Wang, J.; Yu, P.; Butcher, S. E., Rapid global structure determination of large RNA and RNA complexes using NMR and small-angle X-ray scattering. *Methods* **2010,** 52, (2), 180-91.
156. Wang, J.; Zuo, X.; Yu, P.; Byeon, I. J.; Jung, J.; Wang, X.; Dyba, M.; Seifert, S.; Schwieters, C. D.; Qin, J.; Gronenborn, A. M.; Wang, Y. X., Determination of multicomponent protein structures in solution using global orientation and shape restraints. *J Am Chem Soc* **2009,** 131, (30), 10507-15.
157. Svergun, D. I., Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* **1992,** 25, 495-503.
158. Svergun, D. I., Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* **1999,** 76, (6), 2879-86.
159. Volkov, V. V.; Svergun, D. I., Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **2003,** 36, 860-864.
160. Piotto, M.; Saudek, V.; Sklenar, V., Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* **1992,** 2, (6), 661-5.
161. Wijmenga, S. S.; van Buuren, B. N. M., The use of NMR methods for conformational studies of nucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1998,** 32, 287-387.
162. Weeks, K. M., Advances in RNA structure analysis by chemical probing. *Curr Opin Struct Biol* **2010,** 20, (3), 295-304.

163. Mortimer, S. A.; Weeks, K. M., A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* **2007,** 129, (14), 4144-5.
164. Lipfert, J.; Doniach, S., Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu Rev Biophys Biomol Struct* **2007,** 36, 307-27.
165. Zuo, X.; Wang, J.; Foster, T. R.; Schwieters, C. D.; Tiede, D. M.; Butcher, S. E.; Wang, Y. X., Global molecular structure and interfaces: refining an RNA:RNA complex structure using solution X-ray scattering data. *J Am Chem Soc* **2008,** 130, (11), 3292-3.
166. Doniach, S., Changes in biomolecular conformation seen by small angle X-ray scattering. *Chem Rev* **2001,** 101, (6), 1763-78.
167. Kozin, M. B.; Svergun, D. I., Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography* **2001,** 34, 33-41.
168. Seetharaman, M.; Eldho, N. V.; Padgett, R. A.; Dayie, K. T., Structure of a self-splicing group II intron catalytic effector domain 5: parallels with spliceosomal U6 RNA. *RNA* **2006,** 12, (2), 235-47.
169. Furtig, B.; Richter, C.; Wohnert, J.; Schwalbe, H., NMR spectroscopy of RNA. *Chembiochem* **2003,** 4, (10), 936-62.
170. Thakur, C. S.; Luo, Y.; Chen, B.; Eldho, N. V.; Dayie, T. K., Biomass production of site selective 13C/15N nucleotides using wild type and a transketolase E. coli mutant for labeling RNA for high resolution NMR. *J Biomol NMR* **2012,** 52, (2), 103-14.
171. Thakur, C. S.; Dayie, T. K., Asymmetry of 13C labeled 3-pyruvate affords improved site specific labeling of RNA for NMR spectroscopy. *J Biomol NMR* **2012,** 52, (1), 65-77.
172. Thakur, C. S.; Sama, J. N.; Jackson, M. E.; Chen, B.; Dayie, T. K., Selective 13C labeling of nucleotides for large RNA NMR spectroscopy using an E. coli strain disabled in the TCA cycle. *J Biomol NMR* **2010,** 48, (4), 179-92.