UC Berkeley
UC Berkeley Electronic Theses and Dissertations

**Title**
Mechanisms of cyclic dinucleotide signaling: from prokaryotes to humans

**Permalink**
https://escholarship.org/uc/item/9t81b471

**Author**
Wilson, Stephen

**Publication Date**
2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

# Mechanisms of cyclic dinucleotide signaling: from prokaryotes to humans

By

Stephen C. Wilson

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jennifer Doudna, Co-Chair
Professor Russell Vance, Co-Chair
Professor Carolyn Bertozzi
Assistant Professor Danielle Tullman-Ercek

Fall 2015

Mechanisms of cyclic dinucleotide signaling: from prokaryotes to humans

© 2015

By Stephen C. Wilson

ABSTRACT

Mechanisms of cyclic dinucleotide signaling: from prokaryotes to humans

By

Stephen C. Wilson

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Jennifer Doudna, Co-Chair
Professor Russell Vance, Co-Chair

Cyclic dinucleotides (cdNs) are novel second messengers that are synthesized directly from two "molecular units of energy" in the cell, namely ATP and GTP. In prokaryotes, all three combinations of cdNs have been discovered: cyclic di-AMP, cyclic AMP-GMP, and cyclic di-GMP. Their effects on bacterial host physiology have been shown to be important and diverse, impacting processes ranging from biofilm formation to motility to sporulation. Many of their prokaryotic effectors are protein-based; however, cyclic dinucleotides have also been shown to exert genetic expression control at the RNA level by binding to mRNA regulatory elements called riboswitches. While cyclic dinucleotide signaling in prokaryotes has been appreciated for over 25 years, it was not until the last few years that cyclic dinucleotide signaling has been shown to exist and have a profound impact in mammals. cGAMP (a non-canonically linked cyclic AMP-GMP) has been discovered endogenously within humans and found to be a critical second messenger for signaling pathogen invasion to the host. cGAS, a cytosolic mammalian cyclic dinucleotide synthase, produces cGAMP upon binding pathogen DNA. Subsequently, cGAMP binds to the ER-resident transmembrane protein STING, which triggers a downstream cascade of interferon, cytokine, and chemokine production. Therefore from prokaryotes to humans, cyclic dinucleotide signaling spans billions of years in evolution, and today has led to great interest in its therapeutic targeting.

This work details the discovery of cyclic AMP-GMP and its cognate riboswitch within the bacteria *Geobacter sulfurreducens*. Interestingly, cyclic AMP-GMP signaling through its riboswitch plays a previously unappreciated role in exoelectrogenesis within *Geobacter*. The structural elucidation and signaling role of cyclic AMP-GMP (cGAMP) in humans is subsequently described, showing that humans utilize a novel regioisomer of cyclic AMP-GMP containing mixed 2′-5′ and 3′-5′ phosphodiester bonds. The ancient origin of human cGAMP signaling is illuminated by the discovery of a functional cGAS-cdN-STING pathway in *Nematostella vectensis*, an anemone species >500 million years diverged from humans. The structure of anemone STING is highly conserved compared to human STING, yet anemone

cGAS appears to produce only a 3′-5′ linked cdN. Finally, the signaling space of the cGAS-cdN-STING pathway is examined with a focus on its potential functions in lower organisms. Taken holistically, this work details the signaling mechanisms of cyclic dinucleotides in multiple organisms, while profiling the chemical evolution of a unique second messenger.

This dissertation is dedicated to
*Adam M. Ciccone*
without whom none of this would have been possible.

"Love is the mystery between two people, not the identity."
- John Fowles, *The Magus*

TABLE OF CONTENTS

1. INTRODUCTION 1

2. CYCLIC DINUCLEOTIDE SYNTHESIS AND SENSOR DEVELOPMENT

2.1 Background 4

2.2 Bioorthogonal riboswitch engineering and testing strategy 7

2.3 Synthesis of natural and unnatural cyclic dinucleotides 8

2.4 Cyclic dinucleotide fluorescent sensor development 10

2.5 Bioorthogonal riboswitch engineering and testing 12

2.6 Discussion 14

3. DISCOVERY OF CYCLIC AMP-GMP AND ITS COGNATE RIBOSWITCH IN *GEOBACTER*

3.1 Background 15

3.2 Phylogenetic analysis of the GEMM-I cyclic di-GMP riboswitch family 15

3.3 Phylogenetic GEMM-I cyclic di-GMP riboswitch – Spinach screen 20

3.4 In-line probing of cyclic AMP-GMP selective riboswitch, GM-790 24

3.5 Identification of cyclic AMP-GMP in *Geobacter sulfurreducens* cell extracts 27

3.6 GEMM-Ib class riboswitch as a genetic regulatory element 31

3.7 Discussion 33

4. A NONCANONICAL CYCLIC AMP-GMP REGIOISOMER IS A MAMMALIAN SECONDARY MESSENGER FOR INNATE IMMUNITY

4.1 Background 34

4.2 The enzymatic product of cGAS is inconsistent with canonical cyclic AMP-GMP 35

4.3 cGAS produces cyclic AMP-GMP with a single, unique 2′-5′ bond 36

4.4 Discussion 42

5. THE cGAS-cdN-STING AXIS IS AN EVOLUTIONARY CONSERVED PATHWAY

5.1 Background 44

5.2 STING cyclic dinucleotide recognition predates modern innate immunity 44

5.3 STING architecture is conserved in anemone 46

5.4 The cGAS-cdN-STING pathway is conserved in sea anemone 48

5.5 Anemone cGAS produces via a 3′,3′-linked cyclic dinucleotide 50

5.6 nvSTING specifically recognizes guanine bases in 3′,3′ second messengers 52

5.7 2′,3′ cGAMP Traps a Unique and Conserved STING Conformation 54

5.8 Discussion 56

6. ELUCIDATING THE SIGNALING SPACE OF THE cGAS-cdN-STING AXIS AND FUTURE DIRECTIONS

6.1 Background 58

6.2 cGAMP elicits a potent immune response in *Nematostella vectensis* 60

6.3 nvSTING forms puncta in response to 2′,3′ cGAMP 64

6.4 Dissection of distinct downstream signaling pathways of mammalian STINGs 66

6.5 Development of mSTING S365A and E339 truncated mouse models using cas9 72

6.6 Concluding remarks 74

MATERIALS AND METHODS

Chapter 2 76

Chapter 3 80

Chapter 4 83

Chapter 5 85

Chapter 6 89

LIST OF FIGURES

**Chapter 1**

Figure 1.1: Structure of cyclic di-GMP 1
Figure 1.2: Expanded signaling role of cyclic di-GMP in bacteria 2

**Chapter 2**

Figure 2.1: Riboswitch RNA anatomy 4
Figure 2.2: Riboswitch controlled genetic regulation 4
Figure 2.3: Riboswitches bind natural metabolites 5
Figure 2.4: Secondary structure of GEMM-I Vc2 cyclic di-GMP riboswitch bound to cyclic di-GMP 6
Figure 2.5: Binding pocket of the GEMM-I Vc2 cyclic di-GMP riboswitch 6
Figure 2.6: Rational design of cyclic CMP-GMP selective riboswitch 7
Figure 2.7: Small molecule fluorescence detection scheme using Spinach fusion 8
Figure 2.8: "One-pot" synthesis of cyclic dinucleotides 10
Figure 2.9: Design scheme for cyclic di-GMP specific fluorescent sensor 11
Figure 2.10: Spinach-based RNA fusion with cyclic di-GMP riboswitch enables cyclic di-GMP fluorescent detection 11
Figure 2.11: *In vitro* screen of mutant Vc2-Spinach sensors against cyclic di-GMP and cyclic CMP-GMP 12
Figure 2.12: *In vitro* screen of mutant Vc2-Spinach sensors against cyclic di-GMP, cyclic AMP-GMP, and cyclic di-AMP 13
Figure 2.13: G20A Vc2-Spinach fusion responds to cyclic AMP-GMP 14

**Chapter 3**

Figure 3.1: cdG program architecture for analyzing natural riboswitch phylogeny 16
Figure 3.2: Sequence diversity tree (partial) for the GEMM-I cyclic di-GMP riboswitch 19
Figure 3.3: Phylogenetic cyclic di-GMP riboswitch – Spinach screen against cyclic di-GMP, cyclic AMP-GMP, and cGAMP 20
Figure 3.4: Secondary structure model of GM-790 24
Figure 3.5: In-line probing analysis of cyclic di-AMP, cyclic AMP-GMP, cyclic di-GMP and cGAMP against GM-790 25
Figure 3.6: In-line probing $K_d$ gel of GM-790 with cyclic AMP-GMP 26
Figure 3.7: In-line probing $K_d$ gel of GM-790 with cyclic di-GMP 27
Figure 3.8: cdN cell extract analysis of *Geobacter sulfurreducens* 28
Figure 3.9: HRMS analysis of discrete fractions from *G. sulfurreducens* PCA cell extract 29

Figure 3.10: MS/MS analysis of fraction D compared to synthetic standards 30
Figure 3.11: In-line probing analysis of WT and C78U GS-814 riboswitch 31
Figure 3.12: In-line probing analysis of the GS-814 riboswitch mapped onto its secondary structure model 32

**Chapter 4**

Figure 4.1: Early model for cGAS-STING pathway activation 35
Figure 4.2: Cyclic dinucleotide activation of hSTING H232 and R232 using a luciferase-based IFN-β detection assay 36
Figure 4.3: Thin-layer chromatography analysis of cyclic dinucleotides 37
Figure 4.4: HPLC purification of cGAS enzymatic product 38
Figure 4.5: $^{1}$H-$^{31}$P HMBC of HPLC-purified cGAS product acquired at 600 MHz 39
Figure 4.6: Multiplicity-edited $^{1}$H-$^{13}$C HSQC experiment in a 900 MHz field 40
Figure 4.7: $^{1}$H-$^{1}$H COSY experiment in a 600 MHz field 41
Figure 4.8: $^{1}$H-$^{1}$H NOESY experiment in a 900 MHz field 41
Figure 4.9: Regioisomers of cyclic AMP-GMP 42
Figure 4.10: The cGAS-STING axis is a cytosolic DNA sensing pathway that elicits a potent immune response 43

**Chapter 5**

Figure 5.1: STING cyclic dinucleotide recognition predates modern innate immunity 46
Figure 5.2: Cartoon schematic of candidate N. vectensis cGAS-like enzymes and screen for cdN synthase activity using interferon-β assay 49
Figure 5.3: Screen of anemone cGAS candidate genes for synthase activity using interferon-β assay 49
Figure 5.4: Detailed sequence alignment of synthase domains from human cGAS, nv-cGAS (nv-A7SFB5.1), and related candidate synthase nv-A7S0T1.1 50
Figure 5.5: Differential activation of hSTING H232 and R232 using select cyclic dinucleotides 51
Figure 5.6: nv-cGAS induced interferon-β luciferase signal is ablated by a 3′,3′ cyclic AMP-GMP specific phosphodiesterase 52
Figure 5.7: nvSTING binds cdNs with high affinity 53
Figure 5.8: nvSTING adopts different bound conformations depending on the cdN 55

**Chapter 6**


| | | |
|---|---|---|
| Figure 6.1: | Current model for STING signaling | 58 |
| Figure 6.2: | STING is expressed during *Nematostella* embryonic development | 60 |
| Figure 6.3: | Cyclic dinucleotide overlay on 12-15 hour old *Nematostella vectensis* embryos | 61 |
| Figure 6.4: | Cyclic dinucleotide overlay on *Nematostella vectensis* primary polyps | 61 |
| Figure 6.5: | Cyclic dinucleotide microinjected into <2 hour old *Nematostella vectensis* embryos | 62 |
| Figure 6.6: | Testing nvSTING stimulated with human cGAS overexpression using a NF-κB inducible luciferase reporter | 64 |
| Figure 6.7: | Confocal microscopy images of nvSTING co-transfected with cdN synthases | 65 |
| Figure 6.8: | Interferon-β inducible luciferase reporter using indicated STING plasmids stimulated with human cGAS overexpression | 68 |
| Figure 6.9: | STAT6 inducible luciferase reporter using indicated STING plasmids stimulated with human cGAS overexpression | 69 |
| Figure 6.10: | NF-κB inducible luciferase reporter using indicated STING plasmids stimulated with human cGAS overexpression | 70 |
| Figure 6.11: | Confocal microscopy images of wild-type, S365A, and truncated mSTING co-transfected with hcGAS | 71 |
| Figure 6.12: | Strategy for generation of STING S365A knock-in mice using CRISPR/Cas9 | 72 |
| Figure 6.13: | Genotyping of cas9 modified mSTING S365A biallelic mice | 73 |

LIST OF TABLES

**Chapter 3**

| | | |
|---|---|---|
| Table 3.1: | Sample output from cdG program | 17 – 18 |
| Table 3.2: | Tested G20A GEMM-I riboswitches | 21 – 23 |

**Chapter 5**

| | | |
|---|---|---|
| Table 5.1: | nvSTING crystallographic statistics | 47 |
| Table 5.2: | ITC values for nvSTING-cdN interactions | 54 |

**Chapter 6**

| | | |
|---|---|---|
| Table 6.1: | RNA-seq of select upregulated transcripts in cGAMP treated samples compared to control | 63 |

## LIST OF ABBREVIATIONS

| | |
|---|---|
| cAG | cyclic AMP-GMP (only 3′– 5′ linked) |
| cdA, cAA | cyclic di-AMP |
| cdG, cGG | cyclic di-GMP |
| cdN, CDN | cyclic dinucleotide |
| cGAMP | cyclic AMP-GMP (mixed linkages 2′– 5′ and 3′– 5′) |
| CTT | C-terminal tail |
| DMOCP | 5,5-dimethyl-2-oxo-2-chloro-1,3,2-dioxaphosphinane |
| IFN | interferon |
| IRF3 | interferon regulatory factor 3 |
| TBK1 | TANK-binding kinase 1 |
| TBS, TBDMS | tert-butyldimethylsilyl ether |
| WT | wild type |

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge Russell and Jennifer for taking me into their labs. My time in your labs has allowed me to see science with new eyes. Through you two, I have learned how to mentor and lead with grace.

I would also like to acknowledge Philip Kranzusch for being my advocate and always "buttering me up". The skills that I have picked up from you are too numerous to detail. Most importantly, you are constant reminder for me to be kind and generous.

Lastly, I would like to acknowledge the Vance and Doudna labs. It is the people that make the science worth pursuing.

# 1. INTRODUCTION

Cyclic dinucleotides are derived directly from the energy currency of the cell. In prokaryotes, they are second messengers that regulate a diversity of physiological processes. In mammals, they act as a protective signal for the host, eliciting an immune response from the host against an evading pathogen. While their presence in nature is now known to be near ubiquitous, their impact on prokaryotic and eukaryotic physiology has not always been appreciated.

In 1987, the first cyclic dinucleotide, cyclic di-GMP, was discovered. Its biosynthesis depends on two GTPs as precursors that are enzymatically linked by a phosphodiester macrocycle, while extruding pyrophosphate (Figure 1.1). Its founder host, *Acetobacter xylinum,* produces cellulose as an extracellular product and cyclic di-GMP was found to activate this process.[1] For the next decade (1990 – 1999), subsequent studies of cyclic di-GMP largely focused on its biochemistry within *A. xylinum*, namely the enzymology associated with its anabolism, catabolism, and signaling.

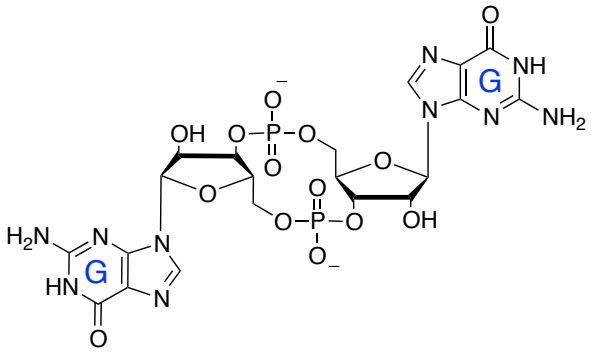


**Figure 1.1: Structure of cyclic di-GMP**

During this time (1990 – 1999), cyclic dinucleotides were also implicated as a byproduct of HIV DNA integration into the host's genome.[2] Retroviral DNA integration requires coordination of DNA cleavage and bond formation. The resulting byproduct of retroviral DNA integration was found to be a putative deoxy-cyclic dinucleotide; however, rigorously structural elucidation was never performed. Regardless, this finding coupled with its endogenous discovery in *A. xylinum* led to a renewed interest in the tractable chemical synthesis of cyclic dinucleotides, which also matured during this decade.[3,4]

The subsequent decade (2000 – 2009) led to an expansion of the signaling role of cyclic di-GMP beyond *A. xylinum* and cellulose biosynthesis (Figure 1.2). While a comprehensive profiling of all the processes involved in cyclic di-GMP signaling would be too lengthy to detail here, cyclic di-GMP has been found to regulate key signaling pathways/phenotypes including: bacterial virulence[5], biofilm formation[6-8], and motility[8,9]. By this point in time the key protein domains responsible for cyclic di-GMP synthesis (GGDEF), degradation (EAL), and signaling (PilZ) were thoroughly elucidated[10], and it was also discovered that cyclic d-GMP was a cognate ligand for metabolite-controlled mRNA regulatory elements called

riboswitches.[11]

In 2008, the first natural occurrence of cyclic di-AMP was discovered in *Bacillus subtilis*.[12] Its synthase, DisA, was crystalized, and cyclic di-AMP was found to be important for bacterial sporulation. Similar to the trajectory of cyclic di-GMP, the role of cyclic di-AMP has only expanded in the years that followed.

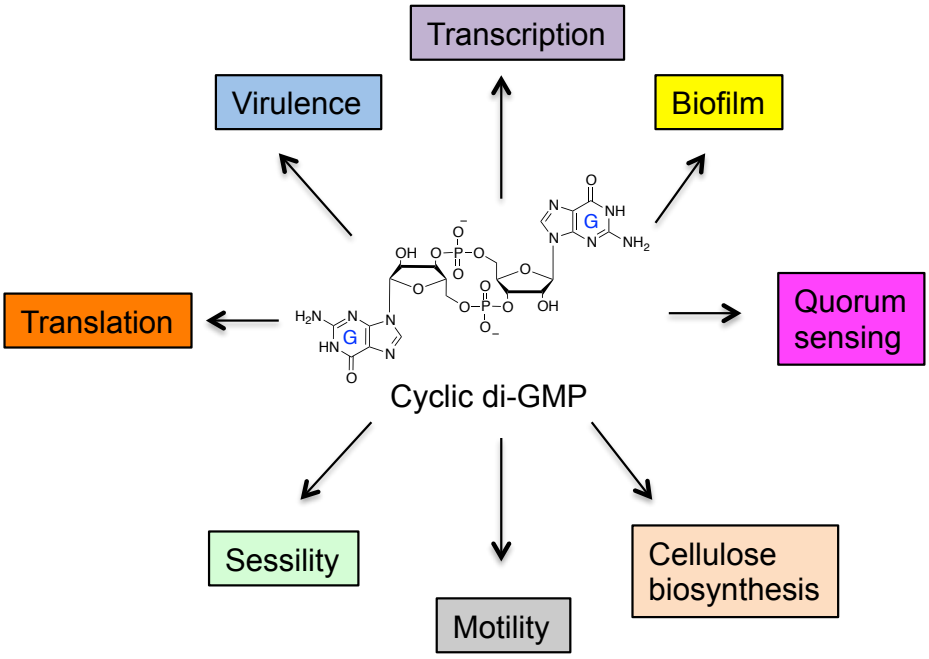



**Figure 1.2: Expanded signaling role of cyclic di-GMP in bacteria**

The last major development during this period (2000 – 2009) was that cyclic di-GMP was found to be immunostimulatory in humans. In particular, it was discovered that cyclic di-GMP co-administered with an antigen results in an enhanced IgG adaptive immune response compared to that without the cyclic dinucleotide adjuvant.[13,14] On its own, cyclic di-GMP induces type-I interferon, NF-κB, TBK1, and IRF3.[15] And finally with what has undoubtedly led to the recent interest in cyclic dinucleotides, cyclic di-GMP was found to inhibit human colon cancer proliferation.[16]

Since 2009 the cyclic dinucleotide field has grown exponentially. In particular, a base hybrid cyclic dinucleotide, cyclic AMP-GMP, was discovered in both bacteria and humans[†].[17,18] While cyclic dinucleotide signaling in prokaryotes has been appreciated for over 25 years, it was not until the last few years that cyclic dinucleotide signaling has been shown to exist and

[†] For this work, "cyclic AMP-GMP" will generally refer to the prokaryotic cdN, whereas "cGAMP" will refer to the mammalian cdN. In some cases the linkage is explicitly stated.

have a profound impact in mammals. Remarkably, the regiochemistry of a single phosphodiester linkage distinguishes the prokaryotic messenger (3′– 5′) from the mammalian messenger (2′– 5′).[19-21] Moreover, cyclic dinucleotide recognition in mammals has been shown to have tremendous impact on how immunity contends with pathogens, forcing the host to upregulate interferon (IFN), NF-κB, STAT6, and other critical cytokines/chemokines.[22-24] The implications are far-reaching and are leading to the design of novel therapeutics for vaccination and cancer immunotherapy.

This dissertation details my work with cyclic dinucleotides and focuses on cyclic AMP-GMP signaling in prokaryotes, humans, and those organisms that lie evolutionarily in-between. Background information specific to each topic is provided at the beginning of each chapter. Chapter 2 describes the synthesis of natural and unnatural cyclic dinucleotides as well as the development of RNA-based fluorescent sensors of cdNs.[25] In Chapter 3, I detail my discovery of cyclic AMP-GMP and its cognate riboswitch in *Geobacter*.[26] Chapter 4 describes my work elucidating the human cyclic dinucleotide cGAMP using NMR spectroscopy and mass spectrometry.[27] In Chapter 5, the evolutionary origins of the cGAS-cdN-STING axis is examined and traced as far back as *Nematostella vectensis* (sea anemone), where *Nematostella* STING architecture is found to be highly conserved with human STING yet its cGAS produces only a 3′– 5′ linked cdN.[28] Lastly, Chapter 6 details my efforts to elucidate and dissect the signaling space of the cGAS-cdN-STING axis in vertebrates and invertebrates. The evolution of the cGAS-cdN-STING axis is discussed and a hypothesized role in symbiosis is put forth.

## 2. CYCLIC DINUCLEOTIDE SYNTHESIS AND SENSOR DEVELOPMENT

### 2.1 Background

Riboswitches are genetic regulatory elements of mRNA that are responsive to small molecule metabolites. With the exception of the TPP riboswitch, these elements have been only discovered in prokaryotes to date. The general anatomy of riboswitch regulation is depicted in Figure 2.1 with three main components: aptamer, expression platform, and gene. The RNA aptamer recognizes and binds its small molecule cognate ligand, which results in a conformational change of the aptamer. In effect, this causes a subsequent conformational change in the expression platform, which regulates either the transcription or the translation of the gene. A generic riboswitch mechanism is shown in Figure 2.2.

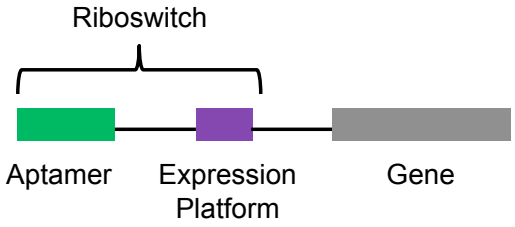



**Figure 2.1: Riboswitch RNA Anatomy**

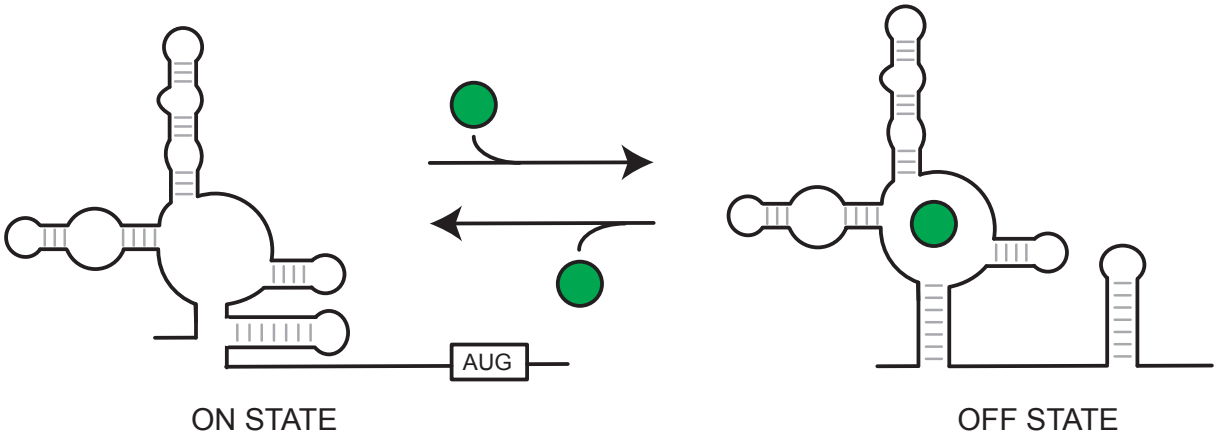



**Figure 2.2: Riboswitch controlled genetic regulation**
*Without a small molecule (green ball) effector, the gene is expressed normally (ON state). When the aptamer binds to the small molecule, the aptamer changes its conformation, which also induces a conformational change in the downstream expression platform (represented as a termination hairpin in the figure). Transcription of the gene is now effectively silenced (OFF state).*

Riboswitches have been found to recognize a diversity of small molecule metabolites including: sugars, flavin mononucleotide (FMN), fluoride, s-adenosyl methionine (SAM), and others (Figure 2.3). Binding affinities range between picomolar to micromolar depending on the ligand. However, the Vc2 cyclic di-GMP riboswitch found in *Vibrio cholerae* has the highest affinity to its cognate ligand (11 pM) of all measured riboswitches to date.[11,29]

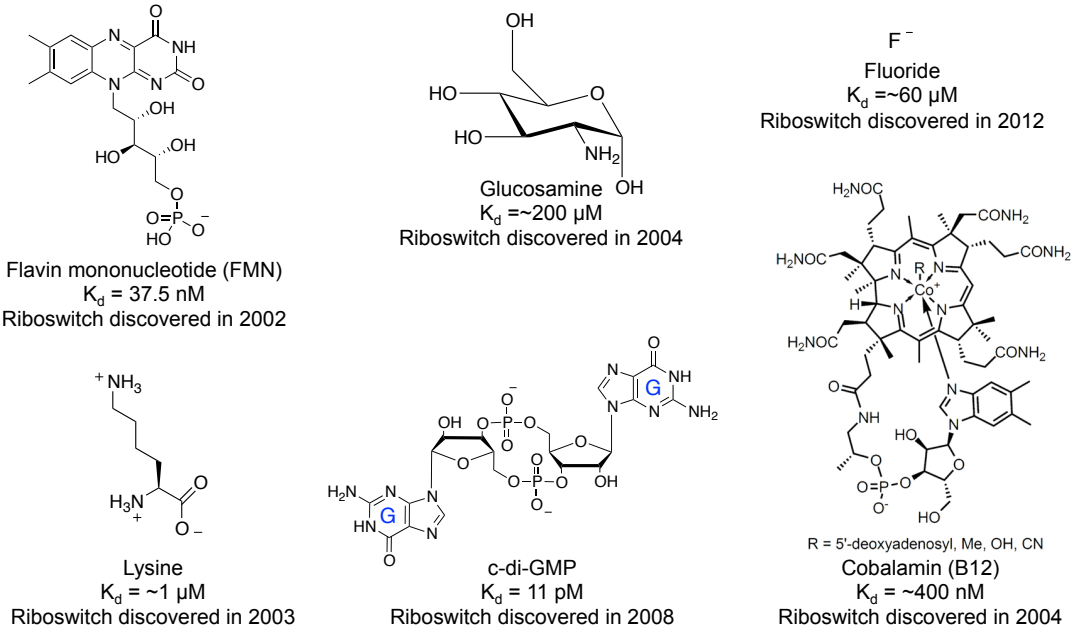



**Figure 2.3: Riboswitches bind to natural metabolites**

The specificity as well as the high affinity of the Vc2 riboswitch with cyclic di-GMP is attributed to three key interactions (Figure 2.4 and 2.5). First, the α-nucleobase of cyclic di-GMP forms a Watson-Crick base pair with a cytosine (C-92) in the riboswitch. Second, the β-nucleobase of cyclic di-GMP forms a Hoogsteen base pair with a guanine (G-20) in the riboswitch. Lastly, the riboswitch intercalates one of its adenines (A-47) in between the two bases of cyclic di-GMP, creating a pi-stacking interaction between the riboswitch and the small molecule. As such, these discrete interactions suggest that a few selective mutations to the riboswitch could confer selectivity to an unnatural cyclic dinucleotide, thereby creating a bioorthogonal riboswitch.

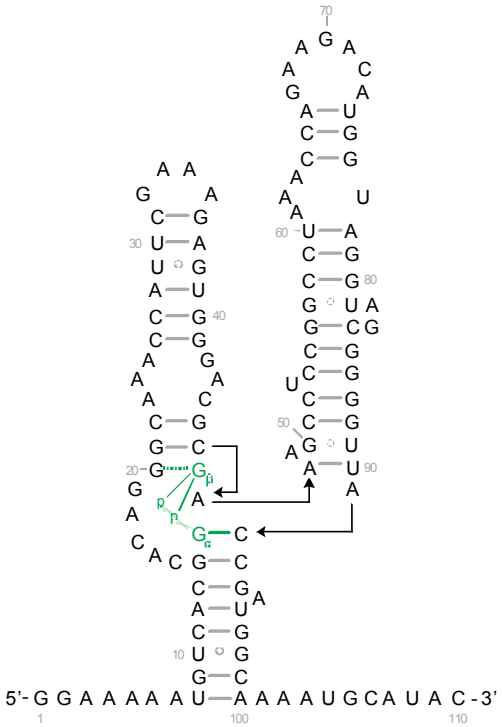



**Figure 2.4: Secondary structure of GEMM-I Vc2 cyclic di-GMP riboswitch bound to cyclic di-GMP**

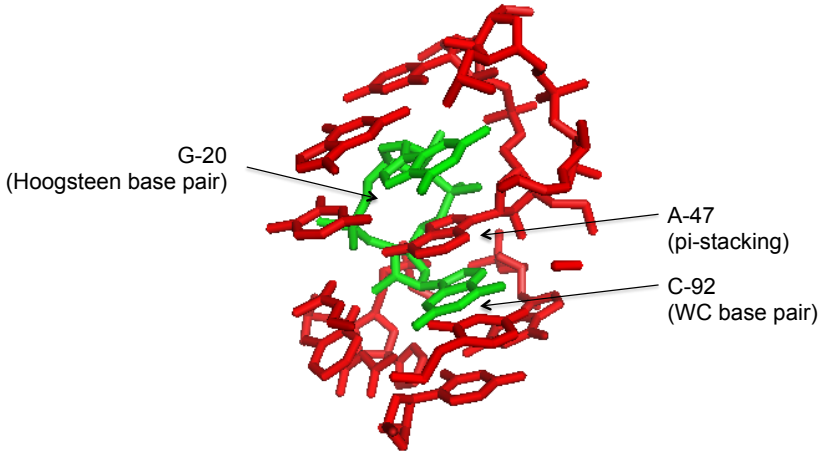



**Figure 2.5: Binding pocket of the GEMM-I Vc2 cyclic di-GMP riboswitch**
*Vc2 cyclic di-GMP riboswitch (red) binds with cyclic di-GMP (green). The key binding interactions are indicated in the figure. [PDB: 3MXH]*[29]

## 2.2 Bioorthogonal riboswitch engineering and testing strategy

Until 2012, the only natural cyclic dinucleotides confirmed to exist naturally were cyclic di-AMP and cyclic di-GMP. No base hybrid cyclic dinucleotide had been discovered up to this point in time. Therefore from an engineering standpoint, cyclic CMP-GMP and cyclic AMP-GMP were natural candidates for engineering a bioorthogonal riboswitch.

Rational design of a cyclic CMP-GMP riboswitch would require two mutations. First, mutating position 92 from a cytosine to a guanosine would preserve the natural Watson-Crick base pairing interaction (Figure 2.6). Second, an additional mutation (C15A) would also be necessary in order to prevent elongation of the P1 stem and riboswitch misfolding.

For the design of a cyclic AMP-GMP riboswitch, the Strobel lab had previously found that the G20A and C92U point mutations altered the affinities and selectivity of cyclic di-GMP and cyclic AMP-GMP for riboswitch binding.[30,31] It has been shown that the single mutant C92U binds c-AMP-GMP ($K_d$ = 19 ± 1.7 nM) and the double mutant G20A/C92U binds c-di-AMP ($K_d$ = 1,200 ± 130 nM), albeit in each case nonselectively and with much poorer affinity than the WT aptamer for c-di-GMP. The single mutant G20A also binds c-di-GMP ($K_d$ = 0.21 ± 0.07 nM), but its affinity for cyclic AMP-GMP was not determined. Therefore efforts to rationally design a cyclic AMP-GMP riboswitch focused on these two positions.

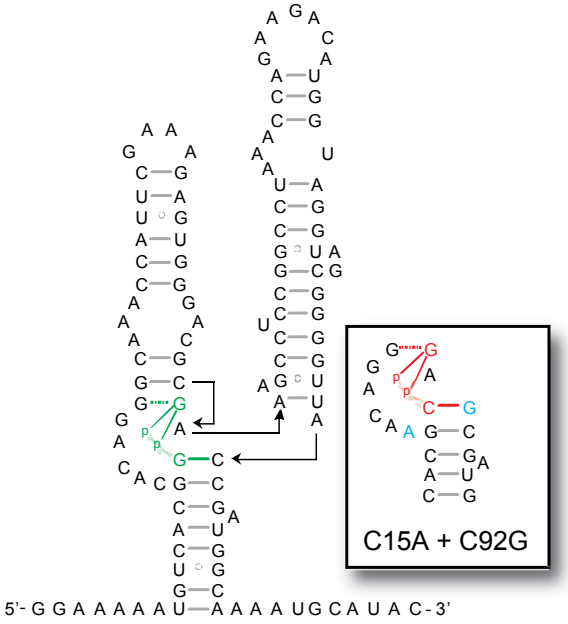



**Figure 2.6: Rational design of cyclic CMP-GMP selective riboswitch**
*The C92G mutation (blue) would preserve the Watson-Crick base pairing interaction between the riboswitch and cyclic CMP-GMP (red). However a compensatory mutation, C15A (blue), is also necessary in order to prevent P1 stem elongation.*

The classic assay to measure riboswitch-small molecule interactions is in-line probing. However, the assay is technically difficult, and it takes one week to run a single experiment. Therefore developing a faster assay to evaluate the effect of riboswitch point mutations on small molecule binding was desired. Fortunately, the Jaffrey lab developed a novel system for fluorescently sensing metabolites using riboswitches. Using SELEX, they evolved an aptamer called Spinach that is selective for a synthetic version of the GFP fluorophore (DFHBI).[32] When Spinach binds to DFHBI, DFHBI fluorescence increases. The Jaffrey lab also found that by fusing the SAM riboswitch to Spinach the fusion acted as a small molecule sensor for SAM (Figure 2.7).[33] The fluorescence emitted from the RNA fusion when exposed to ligand could then be quickly and easily measured on a plate reader. Therefore, an analogous system for the cyclic di-GMP riboswitch was developed in order to develop a bioorthogonal riboswitch (Section 2.4).

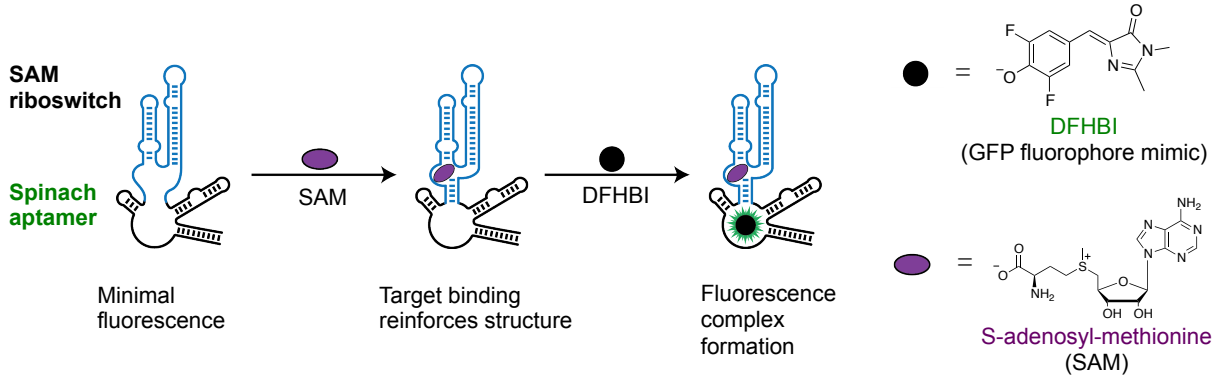



**Figure 2.7: Small molecule fluorescence detection scheme using Spinach fusion**
*Binding of both SAM and DFHBI stabilize the secondary structure of the aptamer fusion leading to increased DFHBI fluorescence.*

## 2.3 Synthesis of natural and unnatural cyclic dinucleotides

Cyclic CMP-GMP and cyclic AMP-GMP were synthesized using the "one-pot" synthesis as depicted in Figure 2.8.[34] There are multiple advantages of the "one-pot" synthesis over other published cyclic dinucleotide syntheses[4,35], which include the following:

1) Speed – the final product can be collected in two days, whereas other syntheses require at least one week.
2) Yield – the bottleneck for all cyclic dinucleotide synthesis is the cyclization step, which is the lowest yielding synthetic step. In this case, the H-phosphonate cyclization with 5,5-dimethyl-2-oxo-2-chloro-1,3,2-dioxaphosphinane (DMOCP) is more efficient than in alternative syntheses, which leads to a two fold increase in total yield compared to other strategies.
3) Purification – no column purifications need to be performed with the "one-pot" synthesis.

4) Scalability – the “one-pot” synthesis scales up remarkably well, leading to cyclic dinucleotide production on the gram-scale.

*Formation of H-phosphonate* - The commercially available phosphoramidite was prepared in a dry round bottom flask and dissolved in dry acetonitrile. Two equivalents of water were added to a stirring solution followed by 1.2 equivalents of pyridinium trifluoroacetate (pyr-TFA). The pyr-TFA acts as a nucleophilic catalyst that promotes the water to attack the phosphite diester. An excess of tetrabutylamine is then added to the solution to cause β-elimination of the cyanoethyl group, resulting in a H-phosphonate. Finally, excess DCA was used to detritylate the 5′ hydroxyl group. A small amount of water was also added to prevent retritylation during the subsequent deacidification step (addition of pyridine). The resulting mixture was evaporated multiple times to drive off any remaining water.

*Linear dinucleotide coupling* - The mixture was then redissolved in dry acetonitrile and the second phosphoramidite was added in, which reacts with the 5′ hydroxyl of the H-phosphonate immediately. Approximately 3 equivalents of t-butyl hydroperoxide were added to oxidize the dinucleotide-linking phosphorus to a phosphate. Sodium bisulfite was then used to quench any remaining peroxide. As before, excess DCA was used to detritylate the 5′ hydroxyl group. A small amount of water was also added to prevent retritylation during the subsequent deacidification step (addition of pyridine). The resulting mixture was evaporated multiple times to drive off any remaining water. The mixture was also dried on the high-vac for at least two hours prior to cyclization in order to guarantee water removal.

*Cyclization of linear dinucleotide* – The vacuum dried material from the previous step was dissolved in freshly distilled pyridine. Roughly four equivalents of DMOCP was added to the solution to cyclize the linear dinucleotide and then quenched with water and iodine, which also oxidized the H-phosphonate to a phosphate. Sodium bisulfite was then used to quench followed by workup, organic extraction, and characterization.

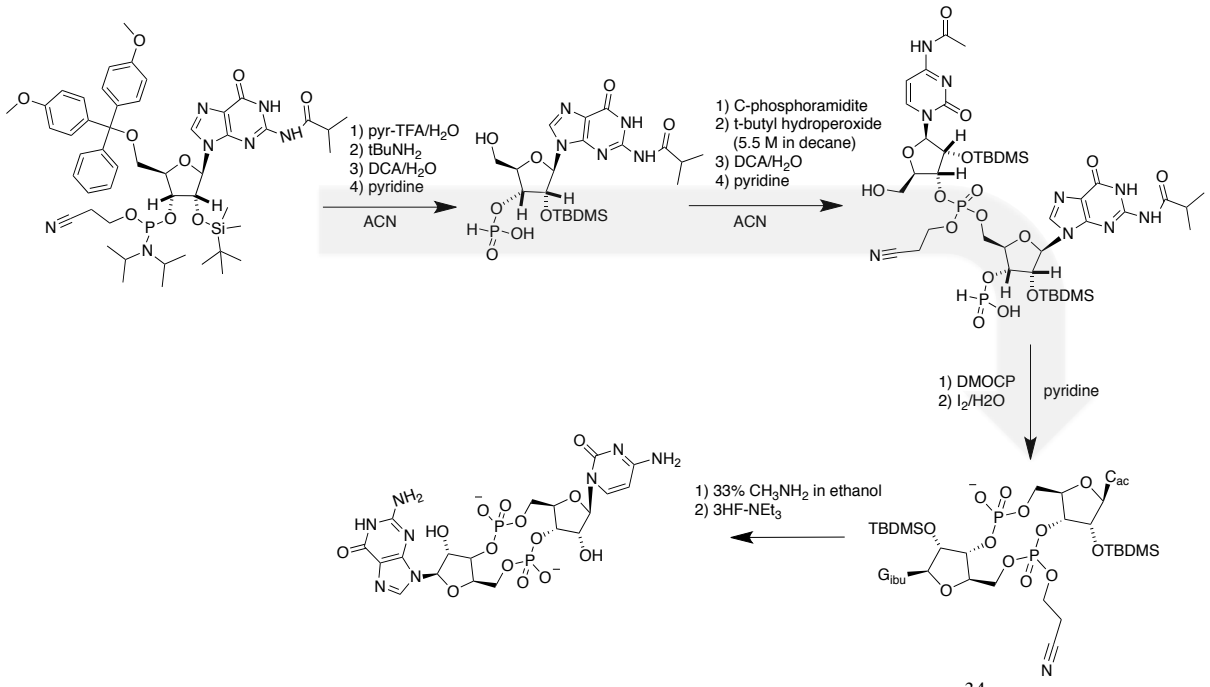



**Figure 2.8: "One-pot" synthesis of cyclic dinucleotides**[34]

*Base, phosphate, and 2′ hydroxyl deprotection* – The base and cyanoethyl protecting groups on the cyclized product were removed by dissolving the material in excess methylamine (33% in ethanol). The solution was subsequently evaporated to dryness and dissolved in excess triethylamine-hydrofluoric acid at 50 °C in order to remove the TBS protecting groups from the 2′ positions. The product was recovered by precipitating out with acetone and purifying using reversed-phase HPLC. Cyclic dinucleotides were recovered with ~30% overall yield.

## 2.4 Cyclic dinucleotide fluorescent sensor development

As a proof of principle for bioorthogonal riboswitch development and testing, the Vc2 cyclic di-GMP riboswitch was fused to Spinach with the aim of developing a cyclic di-GMP selective fluorescent sensor.[36] Binding of the fluorescent molecule DFHBI by the Spinach aptamer is dependent on formation of its second stem loop (Figure 2.9).[33] Replacing this stem loop with the aptamer domain of the Vc2 cyclc di-GMP riboswitch led to selective fluorescence activation in response to cyclic di-GMP (Figure 2.10). The related RNA fusion M1 Vc2-Spinach harbors a disruptive mutation in the riboswitch aptamer structure such that little fluorescence activation is incited in response to cyclic di-GMP (Figure 2.10a,b). In order to model physiological conditions, *in vitro* experiments were carried out at 37 °C in buffer containing 3 mM $MgCl_2$, which is in the range of the estimated free $Mg^{2+}$ in the *E. coli* cytosol.[37] Under these conditions, it was found that WT Vc2-Spinach has an apparent $K_d$ of 230 ± 50 nM for cyclic di-GMP (Figure 2.10c).

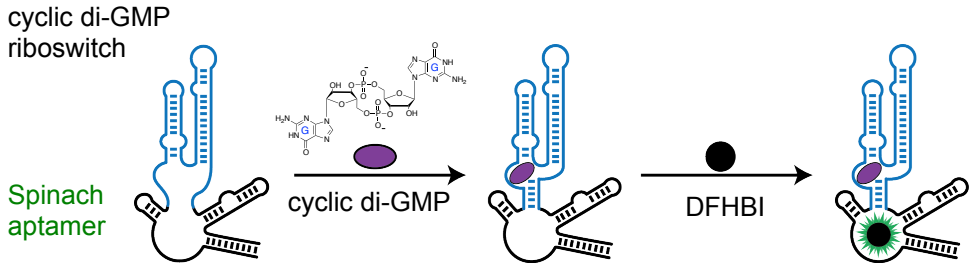



**Figure 2.9: Design scheme for cyclic di-GMP specific fluorescent sensor**

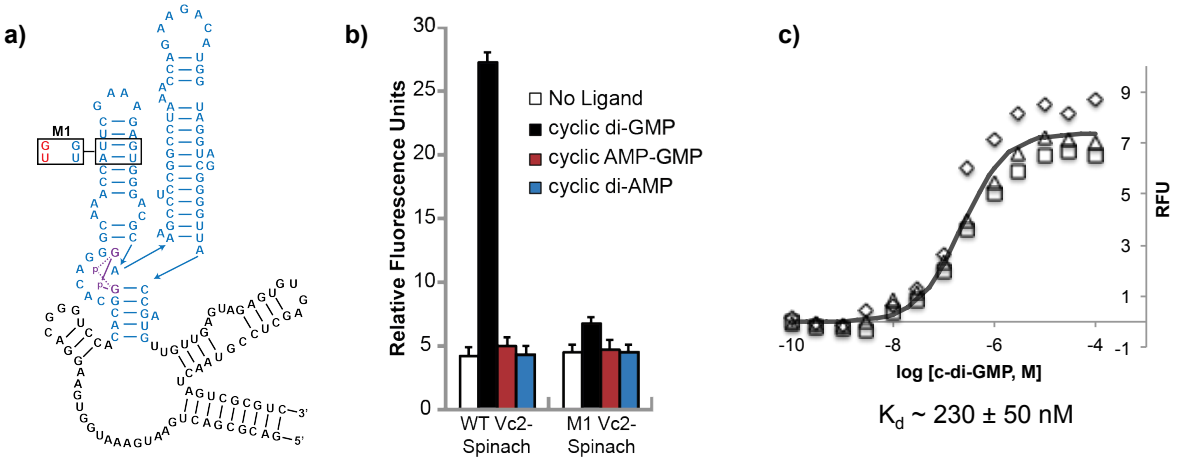



**Figure 2.10: Spinach-based RNA fusion with cyclic di-GMP riboswitch enables cyclic di-GMP fluorescent detection**
*(a) Secondary structure of Vc2-Spinach fusion. (b) In vitro screen of Vc2-Spinach sensor against cyclic di-GMP, cyclic AMP-GMP, and cyclic di-AMP. (c) In vitro analysis of WT Vc2-Spinach binding affinity for cyclic di-GMP (best fit from three independent replicates is shown). [Data acquired by Colleen Kellenberger]*

Since the cyclic di-GMP intracellular concentration in bacteria ranges from < 50 nM to micromolar[38], WT Vc2-Spinach should be capable of detecting cyclic di-GMP at biologically relevant concentrations. Furthermore, the dynamic range of the RNA-based sensor is larger than that of the previously reported protein-based sensor, which binds cyclic di-GMP cooperatively as a dimer.[39] The *in vitro* fluorescent signal of Vc2-Spinach changes from 10% to 90% between 25 and 2,000 nM cyclic di-GMP (Figure 2.10c). In comparison, the protein-based sensor exhibits similar affinity as Vc2-Spinach, but its *in vitro* FRET signal changes from 90% to 10% between 67 nM and 560 nM cyclic di-GMP, a 4 fold decrease in dynamic range.

## 2.5 Bioorthogonal riboswitch engineering and testing

Rationally-designed Vc2-Spinach mutants were constructed with the aim of creating and assaying for a riboswitch selective for an unnatural cyclic dinucleotide. Using this system, C15A and C92G Vc2-Spinach fusions (single and double mutants) were tested with cyclic d-GMP and cyclic CMP-GMP (Figure 2.11). Unfortunately, neither the single mutants nor double mutant afforded a selective fluorescence increase in response to cyclic CMP-GMP. Interestingly, WT Vc2-Spinach responded to cyclic CMP-GMP but with an attenuated signal compared to cyclic d-GMP, suggesting possible promiscuous recognition of cyclic dinucleotides. The C15A mutation was tolerated by the riboswitch and both cyclic dinucleotides elicited fluorescence turn-on from the sensor, albeit at a lower intensity. The C92G construct displayed little to no fluorescence turn-on to either cyclic dinucleotide, which could be caused either by lack of cyclic dinucleotide binding or inherent misfolding of the riboswitch.

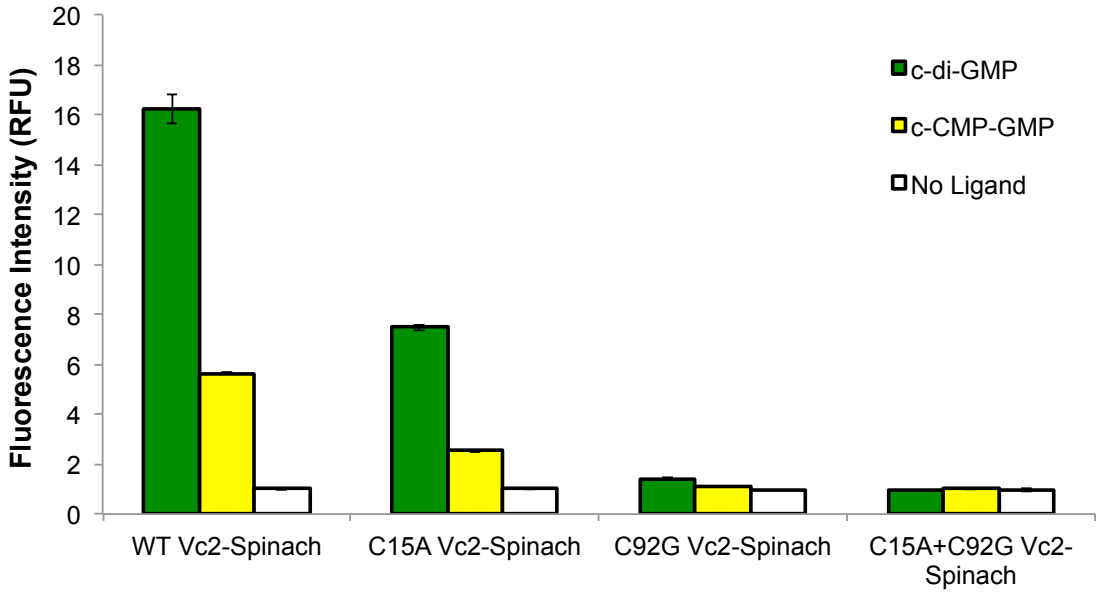



**Figure 2.11: *In vitro* screen of mutant Vc2-Spinach sensors against cyclic di-GMP and cyclic CMP-GMP**

At this point in time, cyclic AMP-GMP was still an "unnatural" cyclic dinucleotide. The Strobel lab reported that the single mutant C92U binds cyclic AMP-GMP with high affinity ($K_d$ = 19 ± 1.7 nM)[31] and the double mutant G20A/C92U binds cyclic di-AMP ($K_d$ = 1,200 ± 130 nM),[29] albeit in each case nonselectively and with poorer affinity than the WT Vc2 cyclic di-GMP aptamer. The single mutant G20A also has been shown to bind cyclic di-GMP ($K_d$ = 0.21 ± 0.07 nM),[30] but the Strobel lab never determined its affinity for the other cyclic dinucleotides. Therefore, these three mutants were assayed in the Spinach context. Remarkably, G20A Vc2-Spinach responds robustly to cyclic di-GMP and cyclic AMP-GMP

but not to related compounds (Figure 2.12), while the other variants exhibited little to no fluorescence activation by any of the cyclic dinucleotides.

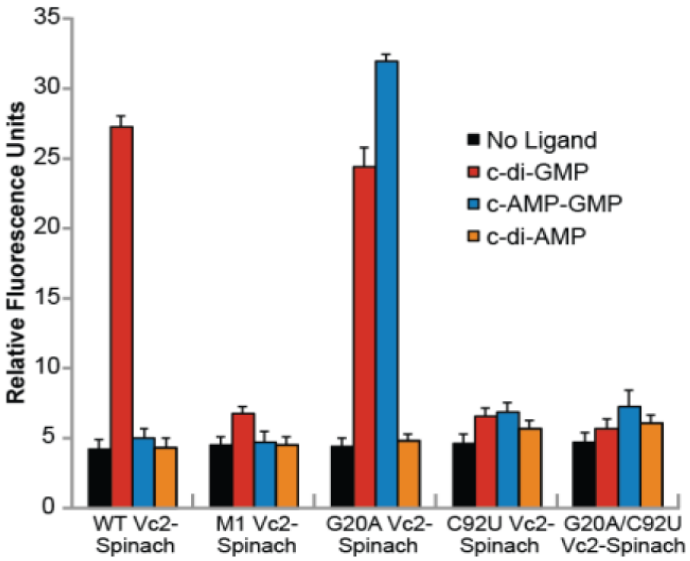



**Figure 2.12: *In vitro* screen of mutant Vc2-Spinach sensors against cyclic di-GMP, cyclic AMP-GMP, and cyclic di-AMP [Data acquired by Colleen Kellenberger]**

During this time, the Mekalanos lab discovered that cyclic AMP-GMP exists endogenous in *Vibrio cholerae*.[18] The dinucleotide synthase DncV from *V. cholerae* was shown to synthesize cyclic di-GMP, cyclic di-AMP, and cyclic AMP-GMP *in vitro*, but it preferentially makes cyclic AMP-GMP in the presence of all four nucleotide triphosphates. Mass spectrometry analysis of cell lysates also supported that cyclic AMP-GMP is produced *in vivo* by WT DncV but not by the catalytically inactive D131A/D133A DncV mutant.

In order to analyze the G20A Vc2-Spinach aptamer for use *in vivo*, its relative affinities for cyclic di-GMP and cyclic AMP-GMP were determined. The G20A Vc2-Spinach aptamer has an apparent $K_d$ of 1,000 ± 150 nM for cyclic di-GMP and 4,200 ± 320 nM for cyclic AMP-GMP at 37 °C, 3 mM $MgCl_2$ (Figure 2.13a). For *in vivo* transformation, all RNA constructs were inserted into a tRNA scaffold to improve stability upon expression in *E. coli*.[40] Cells expressing the WT Vc2-Spinach biosensor along with either DncV or the catalytically inactive DncV (D131A/D133A) exhibit similar levels of fluorescence, yet when expressed with the G20A Vc2-Spinach biosensor there is more than a 4-fold increase in fluorescence for those cells with DncV versus those with D131A/D133A DncV (Figure 2.13b,c).

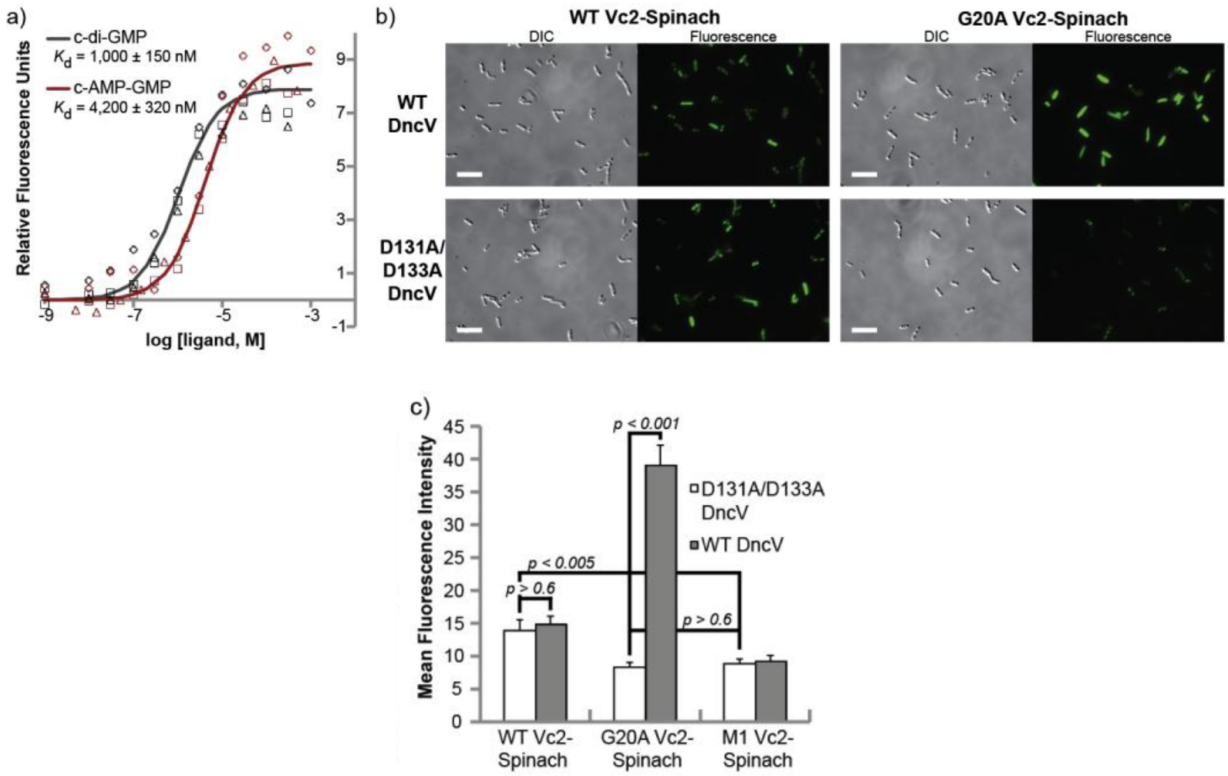



**Figure 2.13: G20A Vc2-Spinach fusion responds to cyclic AMP-GMP**
*(a) In vitro analysis of G20A Vc2-Spinach binding affinity for cyclic di-GMP and cyclic AMP-GMP (best fit from three independent replicates is shown). (b) DIC and fluorescence images of E. coli cells expressing Vc2-Spinach tRNAs and DncV enzyme variants after incubation with DFHBI. Scale bars represent 10 μm. (c) Quantitation of mean fluorescence intensity of cells. Error bars indicate SEM for at least 50 cells. P-values from student's t-test comparisons are shown. [Data acquired by Colleen Kellenberger]*

## 2.6 Discussion

The discovery of endogenous cyclic AMP-GMP in *V. cholerae* expanded the natural vocabulary of cyclic dinucleotides, leaving many questions to be answered.[18] Does cyclic AMP-GMP exist naturally in any other organisms? If so, are there riboswitches that selectively respond to cyclic AMP-GMP? How does cyclic AMP-GMP signaling differ from cyclic di-GMP or cyclic di-AMP?

An asymmetric cyclic dinucleotide encodes more information as a second messenger than a symmetric one due to the depletion of both ATP and GTP for its biogenesis. How this signal distinguishes itself from that of a symmetric cyclic dinucleotide is unclear. For organisms that signal with more than one cyclic dinucleotide, it is interesting to consider how much promiscuity and crosstalk exists between the ligands and what that means for their recognition.

# 3. DISCOVERY OF CYCLIC AMP-GMP AND ITS COGNATE RIBOSWITCH IN *GEOBACTER*

## 3.1 Background

Most riboswitches control genes responsible for its corresponding small molecule effector's metabolism. However, the GEMM-I class cyclic di-GMP riboswitch appears to control a more diverse set of genes. While many do appear to regulate putative phosphodiesterases (EAL domain) and diguanylate cyclases (GGDEF domain), many are predicted to regulate genes like chitinase, ABC transporters, RNA polymerases, type II secretion proteins, flagellar rod proteins, and glycoside hydrolases. All of these associations are predicted bioinformatically using hidden Markov models. Furthermore, the models do not mandate that cyclc di-GMP be cognate ligand for each riboswitch, simply that the general architecture of the riboswitch is preserved. It is possible that a ligand similar enough to cyclic di-GMP such as cyclic AMP-GMP could perhaps be the true cognate ligand for a subpopulation of GEMM-I class cyclic di-GMP riboswitches.

The Strobel lab previously reported the phylogenetic conservation of key residues within the Vc2 cyclic di-GMP riboswitch.[30] Intriguingly, a G20A mutation was found to exist in 23% of all bioinformatically predicted cyclic di-GMP riboswitches. Since this mutation confers increased binding affinity for cyclic AMP-GMP in the Vc2 context, it was natural to hypothesize that perhaps some riboswitch variants could be selective for cyclic AMP-GMP. As previously mentioned, cyclic AMP-GMP had only been discovered in *V. cholerae* until this time and appeared to play a role in host intestinal colonization.[18] Discovery of a natural riboswitch selective for cyclic AMP-GMP would expand the role of this poorly understood cyclic dinucleotide. Therefore, elucidating the phylogenetics and genetic context of the G20A cyclic di-GMP riboswitches was necessary.

## 3.2 Phylogenic analysis of the GEMM-I cyclic di-GMP riboswitch family

The Rfam database[41] is an open access database that contains information about families of non-coding RNAs. Within a single family of non-codings RNAs, primary sequence homology can be highly divergent yet share a common secondary structure. Therefore, Rfam often contains seed alignments (alignments of known sequences within the same family) and full alignments (alignments of potential sequences within the same family). Each multiple sequence alignment is aligned with the consensus secondary structure of the RNA, which is annotated in Stockholm format. Using the alignment and family information provided by Rfam, a Python-based program (referred to as cdG program throughout the remainder of the text) was developed to analyze the phylogenetics of the GEMM-I cyclic di-GMP riboswitch (Figure 3.1).

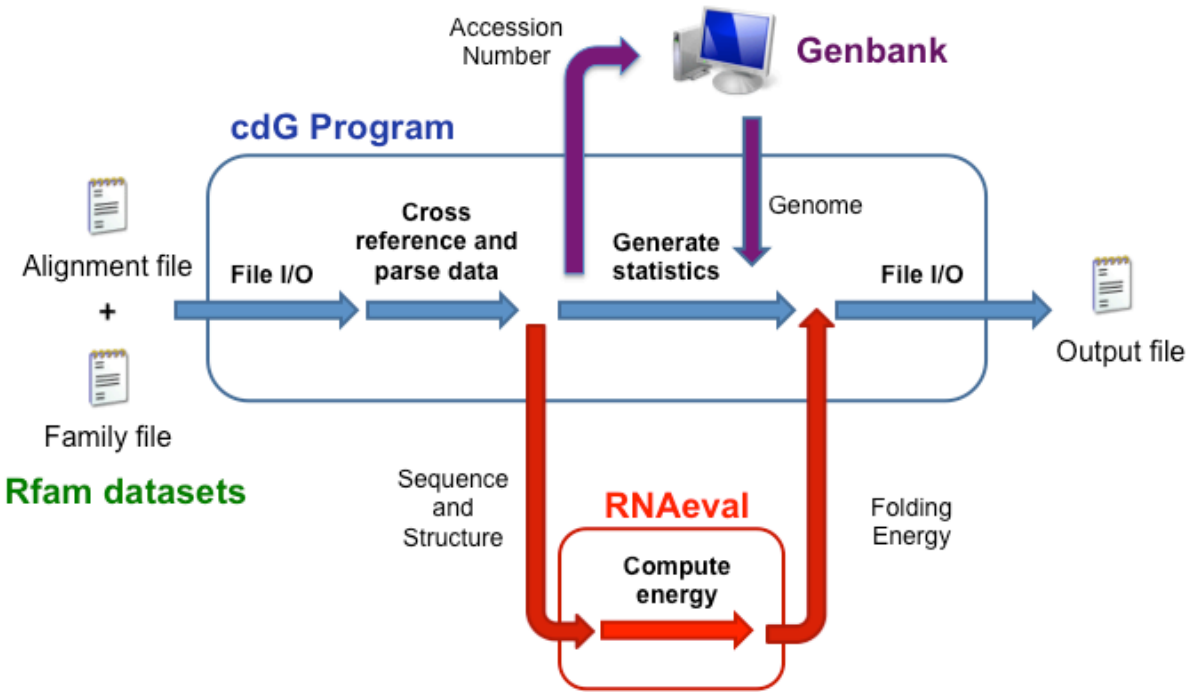



**Figure 3.1: cdG program architecture for analyzing natural riboswitch phylogeny**

The cdG program worked in multiple ways to provide relevant sequence and structure information about the phylogeny of the GEMM-I cyclic di-GMP riboswitch family. First, two Rfam files, the alignment and family file, were parsed and cross-referenced with each other to provide information about each putative riboswitch including: accession number, organism, genomic start and stop coordinates, direction in the genome, GC content, bit score, sequence, and predicted secondary structure. Nucleotide identities were also tabulated for many positions of interest within the riboswitch including position 20, which was found to be important for cyclic AMP-GMP binding. Second, once the sequence and secondary structure had been appropriately parsed, it was piped to RNAeval (part of the VIENNA software package)[42] where the RNA folding energy was calculated for each sequence. The energetics provided insight into the stability of each aptamer sequence. As expected, thermodynamic stability was strongly correlated with GC content as well as whether the bacterium was a cryo-, thermo-, or meso-philic organism. Lastly, once the genomic coordinates were established, the program docked onto Genbank (www.ncbi.nlm.nih.gov/genbank/) and retrieved the putative gene under control by each riboswitch as well its distance from the riboswitch. A riboswitch's distance from its downstream gene gives insight into how the riboswitch regulates genetic expression over that gene. A riboswitch a few nucleotides away from the gene could indicate that the riboswitch functions by obscuring the RBS (translational control), whereas a larger gap could mean that the riboswitch functions by a transcriptional control mechanism. Protein identification and sequences were also retrieved automatically from Genbank. A sample of the cdG program's output is shown in Table 3.1.

Data on exactly 1,990 sequences was extracted and annotated by the cdG program. Removal of identical sequences from the alignment file left 1,347 unique GEMM-I sequences. Those from metagenomic sources were further omitted from analysis. A phylogenetic tree was generated from the remaining sequences using the PHYLIP software package.[43] Evolutionary distances among the sequences were calculated using the F84 model, and a phylogenetic tree was constructed with the Neighbor-Joining/Unweighted Pair Group Method with Arithmetic Mean method (Figure 3.2). Sequences were then sampled throughout the phylogenetic tree and selected based on various criteria including, but not limited to: the presence of the A20 mutation, tractable P1 stems, evolutionary diversity, and interesting downstream genes.

| # | Accession | Organism | Bit Score | Energy | Length | Start | End |
|---|---|---|---|---|---|---|---|
| 1 | CP001964.1 | Cellulomonas flavigena | 46.02 | -17.4 | 74 | 2407306 | 2407232 |
| 75 | ABCF01000033.1 | Bacillus sp. | 57.56 | -21 | 94 | 2268 | 2362 |
| 221 | FN538970.1 | Clostridium difficile | 62.98 | -8.16 | 97 | 2157412 | 2157509 |
| 825 | CP000142.2 | Pelobacter carbinolicus | 61.85 | -26.8 | 81 | 2399816 | 2399735 |
| 951 | FM178379.1 | Aliivibrio salmonicida | 63.04 | -9.6 | 82 | 2124498 | 2124416 |

| # | Direction | % GC | Sequence |
|---|---|---|---|
| 1 | R | 66.67 | AGCGACAAACGGCAAACCCGCCGCAAGGUGGGGACGC AAAGCCACGGGGCCCA----CGAG---- GUCAGCCGAGCUACCGAA |
| 75 | F | 44.21 | UGUUAUAGAAGGCAAACUCAUCUGAAAAGGGAGGAC GCAAAGCCACGGGCCUACAUGCAAAAUAUUAUUUGUA UAUUGGCAGCCGGGUUACCUGU |
| 221 | F | 33.67 | AAACUAAAUCGGCAAAACUAGAGAAAUUUAGUGACG CAAAGCUAUAGGGACUAAGACUUAUAUAAAUAUCUU AUGAGUUAUGUCAGCC-AGUUGCCAAA |
| 825 | R | 56.10 | UUCCGAUAAGGGCAAAACUGGAGUAAUCCAGUGACGC AAAGCUGCGGGUGCGUGGGAAACCACGGAUAGCCGGG CUACCGAA |
| 951 | R | 50.60 | UCACGAACAGGGCAAACUGUGCGAAAGCUCAGGACGC AAAGCUUCCGGCCUACGUA--UUUA-- UAUAAGGUAGCAGGGUUGCCGAU |

**Table 3.1: Sample output from cdG program (continued on next page).**

| # | Consensus structure | Position 20 | Nearby Gene | Distance to Gene | Putative protein |
|---|---|---|---|---|---|
| 1 | .....((...((((((....)))))...))..((((.((((((...(((....)))))))..)))))).... | G | No annotation | 19 | DNA binding domain protein, excisionase family |
| 75 | .....((...(((.(((....)))))...))..((((.((((((....(((.....)))).....))))...))).. | G | flgB | 433 | flagellar basal-body rod protein B |
| 221 | .....((...((((((....)))))...))..((((.(.(((((....(((.....)))).....))))...)).. | G | cspC | 456 | cold shock protein |
| 825 | .....((...((((((....)))))...))..((((.((((((.(((((....)))).)))..)))))).... | G | csuA | 295 | sigma-fimbria adhesin, putative |
| 951 | .....((...((((((....)))))...))..((((.((((((....((((....))))...)))..)))))).... | G | tfoX | 129 | putative DNA transformation protein TfoX |

| # | Locus Tag | Protein ID | Protein sequence |
|---|---|---|---|
| 1 | Cfla_2159 | ADG75052.1 | MALHTTEAPLTQGALLTPGEVAVLFRVDPKTVT RWAQAGKLSAVRTLGGHRRFHEAEVRQLLTGV PQQRAGE |
| 75 | BSG1_18570 | EDL63937.1 | MSLFSSTIQSLENGLNYSSAKQKVISQNIANVDTP NYKAKSISFKDQLNESISTFEAKRTNPRHFEFSGR GANGYYIKSQPFQYNHNGNGVDIDKEMSDLASN QIYFNALTDRLNGKFNSLQSVIRGGK |
| 221 | CD196_1872 | CBA63577.1 | MNNGIVKWFNNEKGFGFISMEGRDDVFAHFSAI QTSGFKSLEEGQQVSFDIVKGARGPQAENITIL |
| 825 | Pcar_2060 | ABA89301.1 | MKKIVAIALAAAFTVVAGTAMAADTADLDVTA TVISSCSMTGGTLAFGNLDPTNAVEVTASSTAVT VTCTNGTAYTLSGDDGDHAVAGQKYLDNGTSTI PYSVSIPAGGTGTGAAVGVTIDGTIAANSYNTAT AGTYTDTILLSVNP |
| 951 | VSAL_I1982 | CAQ79667.1 | MDKPILKDSLRLLSSLGKITSRSMFGGFGVFIDDT MFALVVQDRLHLRASDNTINLFKDQGFEPYVYK KRGFPVVTKYFAISPECWDEPDSILIQAVVALDV AKKDKEKQKTAGPSRIKDLPNLRLATERMLKKA GITTVKELMDTGSVNAYKAIQQTHSSSVSDELL WSLEGAIKGTHWSVISTDIRNELRKQL |

**Table 3.1: Sample output from cdG program (continued from previous page).**

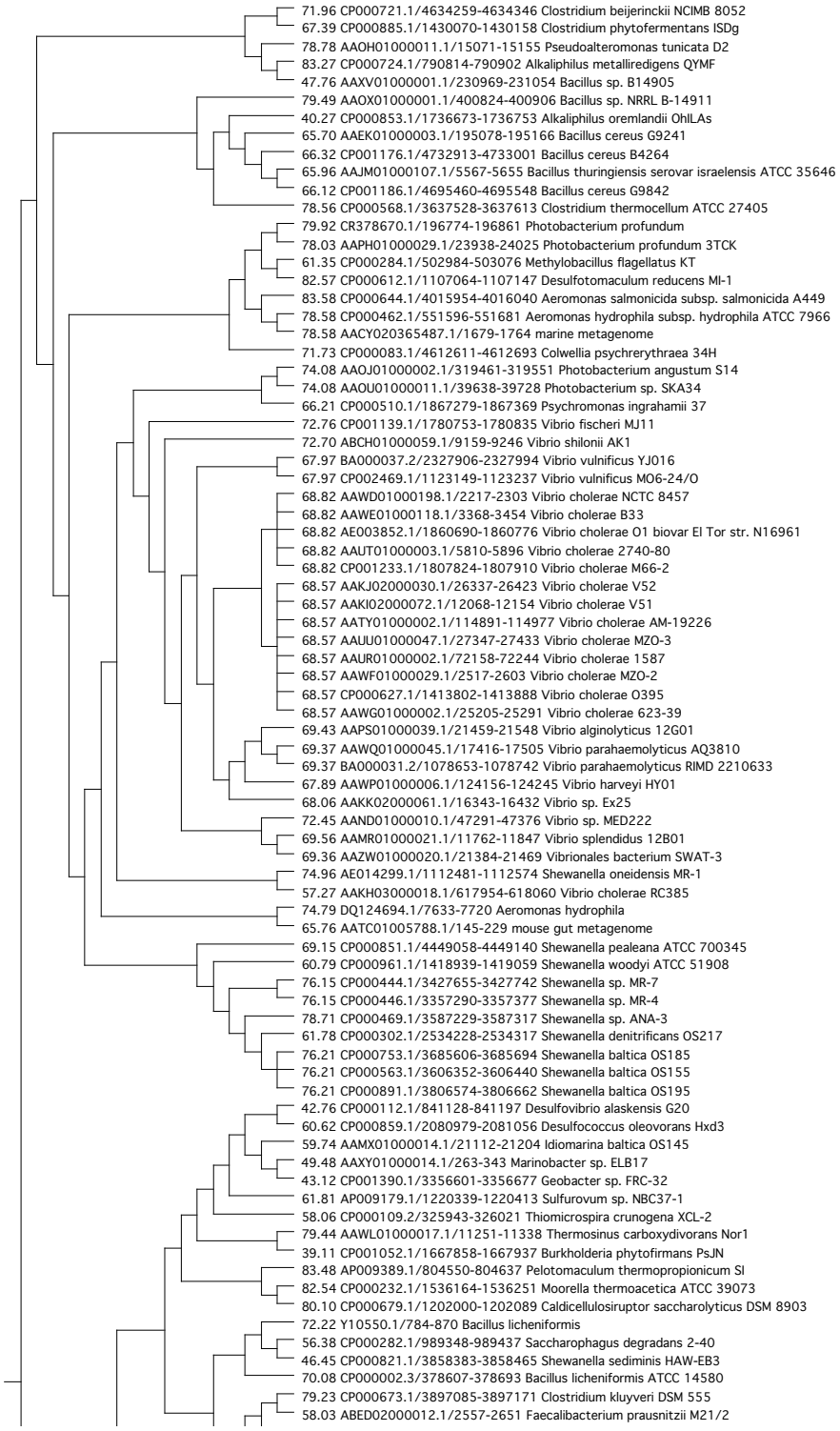



**Figure 3.2: Sequence diversity tree (partial) for the GEMM-I cyclic di-GMP riboswitch**

### 3.3 Phylogenetic GEMM-I cyclic di-GMP riboswitch – Spinach screen

The phylogenetic cyclic di-GMP riboswitch – Spinach screen consisted of 62 sequences sampled throughout the phylogenetic tree but also selectively enriched for the G20A mutation. Fusion RNAs were prepared as previously described (see Section 2.4); however, careful attention was placed on the design of the transducing (P1) stem. It was discovered that the length and constitution of the P1 stem greatly affects the intensity of fluorescence turn-on. Each construct was tested against cyclic di-GMP, cyclic AMP-GMP, and cGAMP. The results of the screen are shown in Figure 3.3 and Table 3.2.

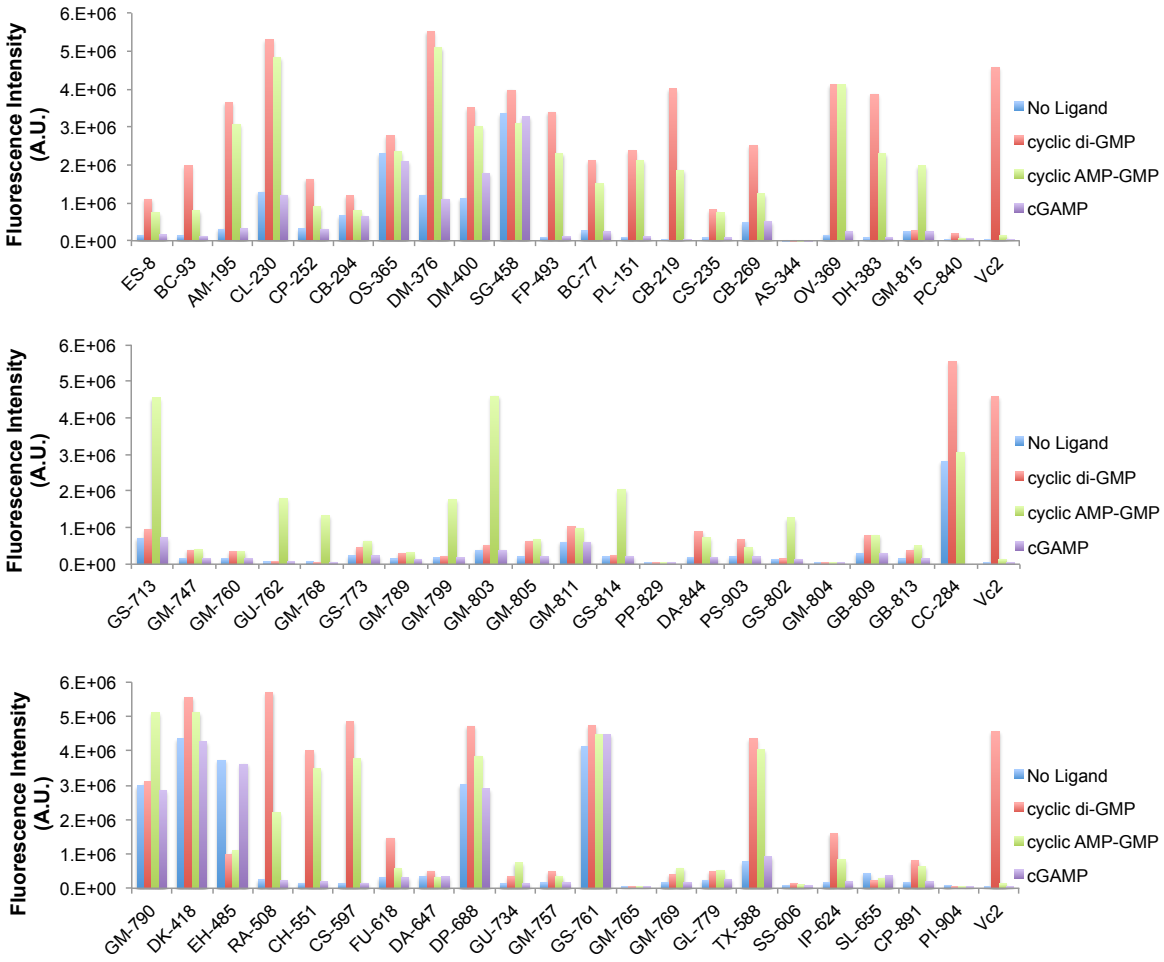



**Figure 3.3: Phylogenetic cyclic di-GMP riboswitch – Spinach screen against cyclic di-GMP, cyclic AMP-GMP, and cGAMP**
*Those constructs that show selective fluorescence turn-on in response to cyclic AMP-GMP are indicated by an *. Table 3.2 provides information about each phylogenetic cyclic di-GMP riboswitch tested.*

| Name | Accession | Organism | Start | End | Downstream Gene |
|---|---|---|---|---|---|
| **ES-8** | AP012211.1 | Eggerthella sp. | 612893 | 612976 | Hypothetical protein |
| **BC-93** | ACMP01000037.1 | Bacillus cereus | 9234 | 9324 | Collagen adhesion protein |
| **AM-195** | CP000724.1 | Alkaliphilus metalliredigens | 480071 | 480160 | methyl-accepting chemotaxis sensory transducer |
| **CL-230** | CP001666.1 | Clostridium ljungdahlii | 2698829 | 2698742 | predicted methyl-accepting transducer |
| **CP-252** | CP000885.1 | Clostridium phytofermentans | 1430070 | 1430158 | diguanylate cyclase/phosphodiesterase with PAS/PAC sensor |
| **CB-294** | ABCC02000002.1 | Clostridium bolteae | 64098 | 64013 | hypothetical protein |
| **OS-365** | AFIH01000001.1 | Oribacterium sp. | 637037 | 637112 | SCP-like protein |
| **DM-376** | AGJB01000003.1 | Desulfitobacterium metallireducens | 15355 | 15440 | metal dependent phosphohydrolase |
| **DM-400** | AGJA01000001.1 | Desulfosporosinus meridiei | 435027 | 435109 | methyl-accepting chemotaxis sensory transducer |
| **SG-458** | CP002547.1 | Syntrophobotulus glycolicus | 3225788 | 3225704 | metal dependent phosphohydrolase |
| **FP-493** | FP929045.1 | Faecalibacterium prausnitzii | 2072446 | 2072371 | pantothenate kinase, type III |
| **BC-77** | CP002394.1 | Bacillus cellulosilyticus | 4239850 | 4239934 | UDP-N-acetylglucosamine 2-epimerase |
| **PL-151** | AGIP01000013.1 | Paenibacillus lactis | 123475 | 123569 | amino acid adenylation domain protein |
| **CB-219** | CP000721.1 | Clostridium beijerinckii | 4931184 | 4931096 | MotA/TolQ/ExbB proton channel |
| **CS-235** | AFWX01000001.1 | Clostridium sp. | 1668071 | 1668159 | Not found |
| **CB-269** | ACSC01000002.1 | Clostridium botulinum | 444804 | 444895 | maltose transacetylase |
| **AS-344** | ACWB01000102.1 | Anaerostipes sp. | 21593 | 21680 | hypothetical protein |
| **OV-369** | AP012044.1 | Oscillibacter valericigenes | 81299 | 81385 | hypothetical protein |
| **DH-383** | CP001336.1 | Desulfitobacterium hafniense | 1127881 | 1127794 | metal dependent phosphohydrolase |
| **GM-815** | CP000148.1 | Geobacter metallireducens | 265065 | 264989 | hypothetical protein |
| **PC-840** | CP000142.2 | Pelobacter carbinolicus | 1987037 | 1987116 | adenylyltransferase, putative |
| **GM-790** | CP000148.1 | Geobacter metallireducens | 1079465 | 1079541 | helix-turn-helix transcriptional regulator, XRE family |
| **DK-418** | CP002770.1 | Desulfotomaculum kuznetsovii | 885656 | 885746 | Ig domain protein group 2 domain protein |
| **EH-485** | CP002400.1 | Ethanoligenens harbinense | 44570 | 44660 | peptidase C1A papain |

**Table 3.2: Tested G20A GEMM-I riboswitches (continued on next page)**

| Name | Accession | Organism | Start | End | Downstream Gene |
|---|---|---|---|---|---|
| **RA-508** | CP002403.1 | Ruminococcus albus | 1557401 | 1557482 | hypothetical protein |
| **CH-551** | CP000141.1 | Carboxydothermus hydrogenoformans | 785822 | 785906 | methyl-accepting chemotaxis protein |
| **CS-597** | ADKX01000041.1 | Coprobacillus sp. | 89095 | 89010 | hypothetical protein |
| **FU-618** | ACDH01000005.1 | Fusobacterium ulcerans | 9304 | 9229 | Not found |
| **DA-647** | CP000884.1 | Delftia acidovorans | 4023666 | 4023571 | Ig family protein |
| **DP-688** | AGJR01000009.1 | Desulfobacter postgatei | 114910 | 114826 | hypothetical protein |
| **GU-734** | CP000698.1 | Geobacter uraniireducens | 2694571 | 2694495 | Fibronectin, type III domain protein |
| **GM-757** | CP000148.1 | Geobacter metallireducens | 1212049 | 1211973 | cytochrome c |
| **GS-761** | CP002031.1 | Geobacter sulfurreducens | 2659769 | 2659689 | cytochrome c, 6 heme-binding sites |
| **GM-765** | CP000148.1 | Geobacter metallireducens | 1977793 | 1977869 | hypothetical protein |
| **GM-769** | CP000148.1 | Geobacter metallireducens | 325755 | 325831 | cytochrome c |
| **GL-779** | CP001089.1 | Geobacter lovleyi | 645064 | 645139 | Fibronectin type III domain |
| **TX-588** | CP002739.1 | Thermoanaero-bacterium xylanolyticum | 2097303 | 2097218 | diguanylate cyclase and metal dependent phosphohydrolase |
| **SS-606** | AGHO01000006.1 | Streptococcus sobrinus | 8967 | 8882 | Not found |
| **IP-624** | CP002282.1 | Ilyobacter polytropus | 410241 | 410316 | PAS/PAC sensor signal t histidine kinase |
| **SL-655** | CP001965.1 | Sideroxydans lithotrophicus | 2226794 | 2226883 | adenylate/guanylate cyclase |
| **CP-891** | CP000083.1 | Colwellia psychrerythraea | 807375 | 807459 | thermostable serine protease, subtilase family |
| **PI-904** | CP000510.1 | Psychromonas ingrahamii | 1867279 | 1867369 | hypothetical protein |
| **GS-713** | CP002479.1 | Geobacter sp. | 3835282 | 3835206 | diguanylate cyclase/phosphodiesterase with PAS/PAC sensor(s) |
| **GM-747** | CP000148.1 | Geobacter metallireducens | 1911272 | 1911348 | cytochrome c |
| **GM-760** | CP000148.1 | Geobacter metallireducens | 3913064 | 3913140 | peptidoglycan L,D-transpeptidase lipoprotein, YkuD family |
| **GU-762** | CP000698.1 | Geobacter uraniireducens | 2697411 | 2697487 | hypothetical protein |
| **GM-768** | CP000148.1 | Geobacter metallireducens | 158170 | 158246 | hypothetical protein |

**Table 3.2: Tested G20A GEMM-I riboswitches (continued on previous and next page)**

| Name | Accession | Organism | Start | End | Downstream Gene |
|---|---|---|---|---|---|
| **GS-773** | CP002479.1 | Geobacter sp. | 2018597 | 2018673 | Fibronectin type III domain protein |
| **GM-789** | CP000148.1 | Geobacter metallireducens | 2280692 | 2280768 | N-acylhomoserine lactone synthetase-related protein |
| **GM-799** | CP000148.1 | Geobacter metallireducens | 271494 | 271418 | lipoprotein, putative |
| **GM-803** | CP000148.1 | Geobacter metallireducens | 3913272 | 3913348 | peptidoglycan L,D-transpeptidase lipoprotein, YkuD family |
| **GM-805** | CP000148.1 | Geobacter metallireducens | 1324617 | 1324541 | cytochrome c |
| **GM-811** | CP000148.1 | Geobacter metallireducens | 1324478 | 1324402 | cytochrome c |
| **GS-814** | CP002031.1 | Geobacter sulfurreducens | 1889823 | 1889898 | lipoprotein cytochrome c, 3 heme-binding sites |
| **PP-829** | CP000482.1 | Pelobacter propionicus | 609688 | 609613 | ErfK/YbiS/YcfS/YnhG family protein |
| **DA-844** | CP002629.1 | Desulfobacca acetoxidans | 907011 | 907086 | EppA_BapA family protein |
| **PS-903** | BADT01000169.1 | Pseudo-alteromonas sp. | 30238 | 30326 | chitin-binding, domain 3 protein |
| **GS-802** | CP001661.1 | Geobacter sp. | 3691135 | 3691211 | Fibronectin type III domain protein |
| **GM-804** | CP000148.1 | Geobacter metallireducens | 2447619 | 2447695 | cytochrome c |
| **GB-809** | CP001124.1 | Geobacter bemidjiensis | 4525294 | 4525370 | peptidoglycan L,D-transpeptidase lipoprotein, YkuD family, SPOR domain-containing |
| **GB-813** | CP001124.1 | Geobacter bemidjiensis | 4525108 | 4525184 | peptidoglycan L,D-transpeptidase lipoprotein, YkuD family, SPOR domain-containing |
| **CC-284** | ADLJ01000004.1 | Clostridium citroniae | 176443 | 176357 | hypothetical protein |

**Table 3.2: Tested G20A GEMM-I riboswitches (continued on previous two pages)**

A screen of natural A20 GEMM-I sequences revealed that the majority display promiscuous binding to cyclic AMP-GMP and cyclic di-GMP, similar to G20A Vc2-Spinach. A few sequences retain cyclic di-GMP selectivity or have undetermined selectivities. However, one initial candidate, GM-790, demonstrated a marked selectivity for cyclic AMP-GMP (Figure 3.3 and Table 3.2). This sequence is from the bacterium *Geobacter metallireducens* GS-15, which has 17 predicted GEMM-I riboswitches annotated in its genome (GenBank accession NC_007517.1). Analysis of all GEMM-I sequences from this species revealed that 15/17 contain an A20 and are selective for cyclic AMP-GMP or have undetermined selectivity. The remaining 2/17 contain a G20; one binds both cyclic AMP-GMP and cyclic di-GMP, whereas the other has undetermined selectivity. The majority of GEMM-I sequences from *Geobacter*

species preferentially bind cAG. However since all of these aptamers were tested in the Spinach context, their binding affinities to cyclic di-GMP and cyclic AMP-GMP needed to be determined outside of the fusion.

### 3.4 In-line probing of cyclic AMP-GMP selective riboswitch, GM-790

In order to validate that selectivity is not altered through the context of the Spinach aptamer and that GM-790 is naturally selective for cyclic AMP-GMP, in-line probing analysis was performed using all four currently known naturally-occurring cyclic dinucleotides: cyclic di-AMP, cyclic AMP-GMP, cyclic di-GMP, and cGAMP (Figure 3.4 and Figure 3.5). Ligand-induced changes in the RNA fold are revealed by alteration of the in-line probing pattern, which corresponds to the extent of RNA self-cleavage at each nucleotide position. Analysis of the in-line probing pattern provides information regarding the aptamer's secondary structure and its dissociation constant. The GM-790 aptamer was found to have a dissociation constant of ~16 nM for cyclic AMP-GMP and ~26 μM for cyclic di-GMP (Figures 3.6 and 3.7), equating to a ~1,600-fold selectivity for cyclic AMP-GMP over cyclic di-GMP. High concentrations of up to 0.5 mM of cyclic di-AMP or 2′,3′ cGAMP induced very little conformational change in either aptamer, indicating that the two riboswitch aptamers do not recognize these other cyclic dinucleotides (Figure 3.5). Similar binding affinities were determined for other cyclic AMP-GMP selective riboswitches within the *Geobacter* genus. At the time of writing, these RNAs are the first specific effectors for cyclic AMP-GMP that have been identified. These sequences constitute a previously unidentified subclass of GEMM-I riboswitches, herein termed GEMM-Ib, that respond selectively to cyclic AMP-GMP.

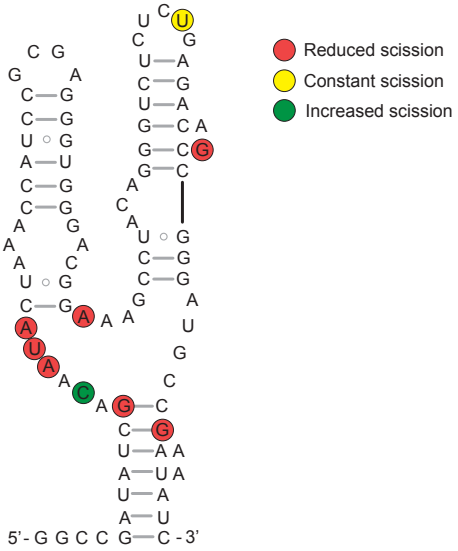



**Figure 3.4: Secondary structure model of GM-790**
*The in-line probing analysis is mapped onto its secondary structure model. The cAG bound conformation is shown.*

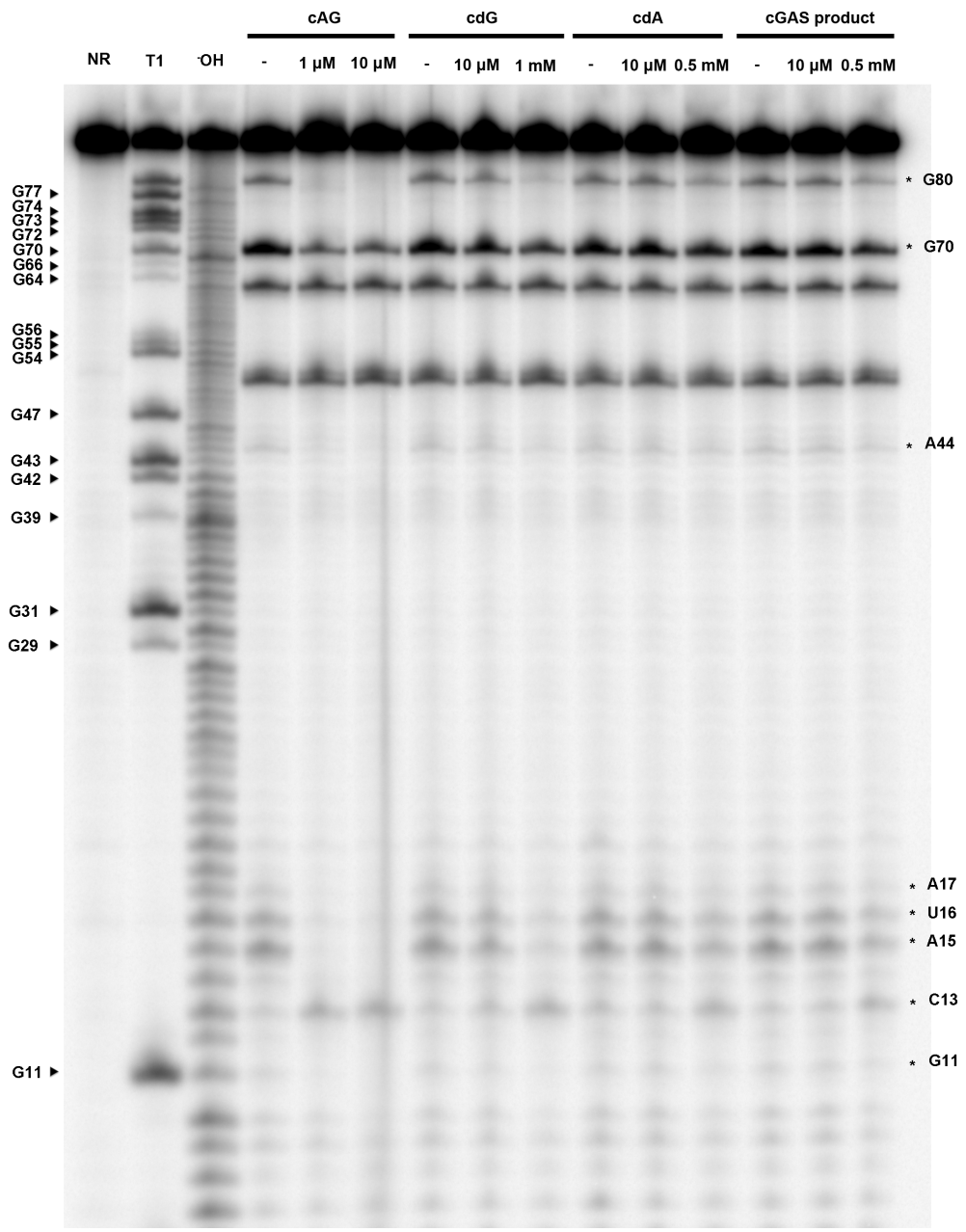



**Figure 3.5: In-line probing analysis of cyclic di-AMP, cyclic AMP-GMP, cyclic di-GMP and cGAMP against GM-790**

*Nucleotides annotated on the right side of the gel indicate positions of modulated cleavage in response to cyclic dinucleotides. Lanes NR, T1, and ⁻OH refer to "no reaction", T1 ribonuclease digest, and base hydrolysis, respectively.*

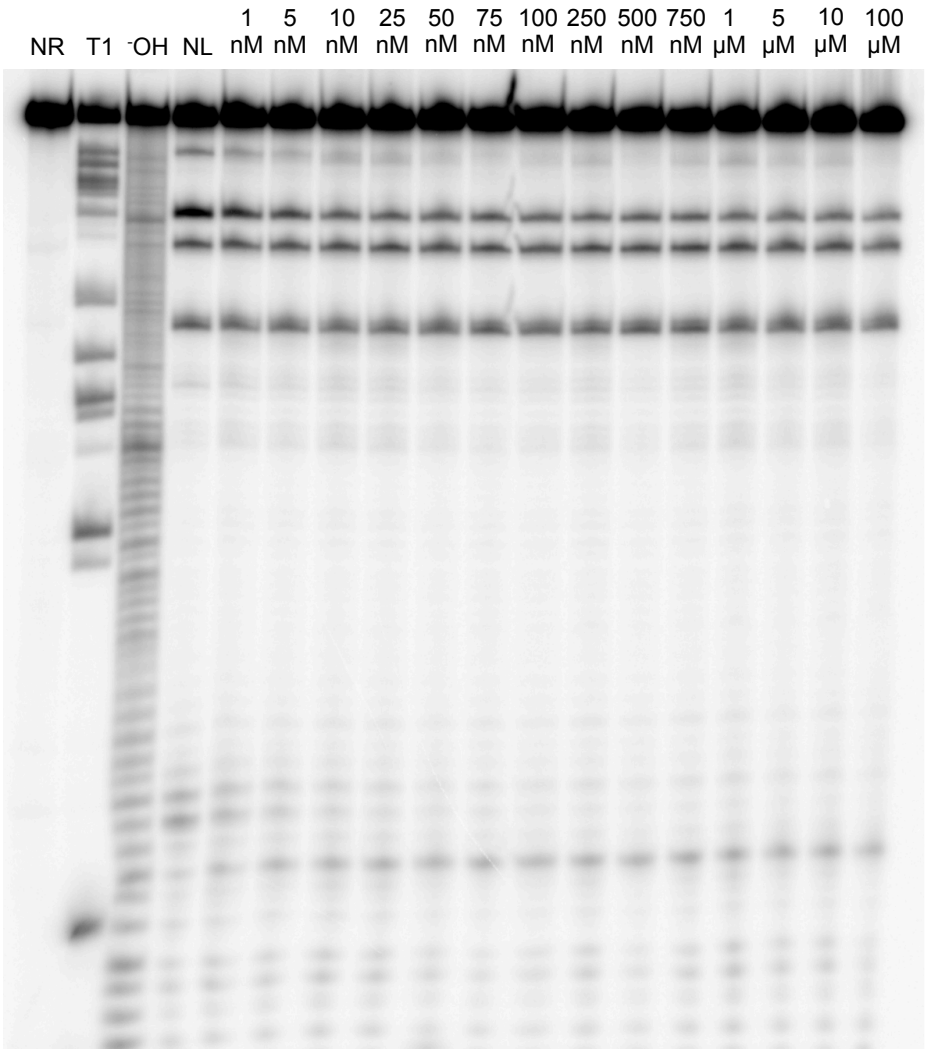



**Figure 3.6: In-line probing $K_d$ gel of GM-790 with cyclic AMP-GMP**
*Dissociation constant determined to be ~12 nM. Lanes NR, T1, and ⁻OH refer to "no reaction", T1 ribonuclease digest, and base hydrolysis, respectively.*

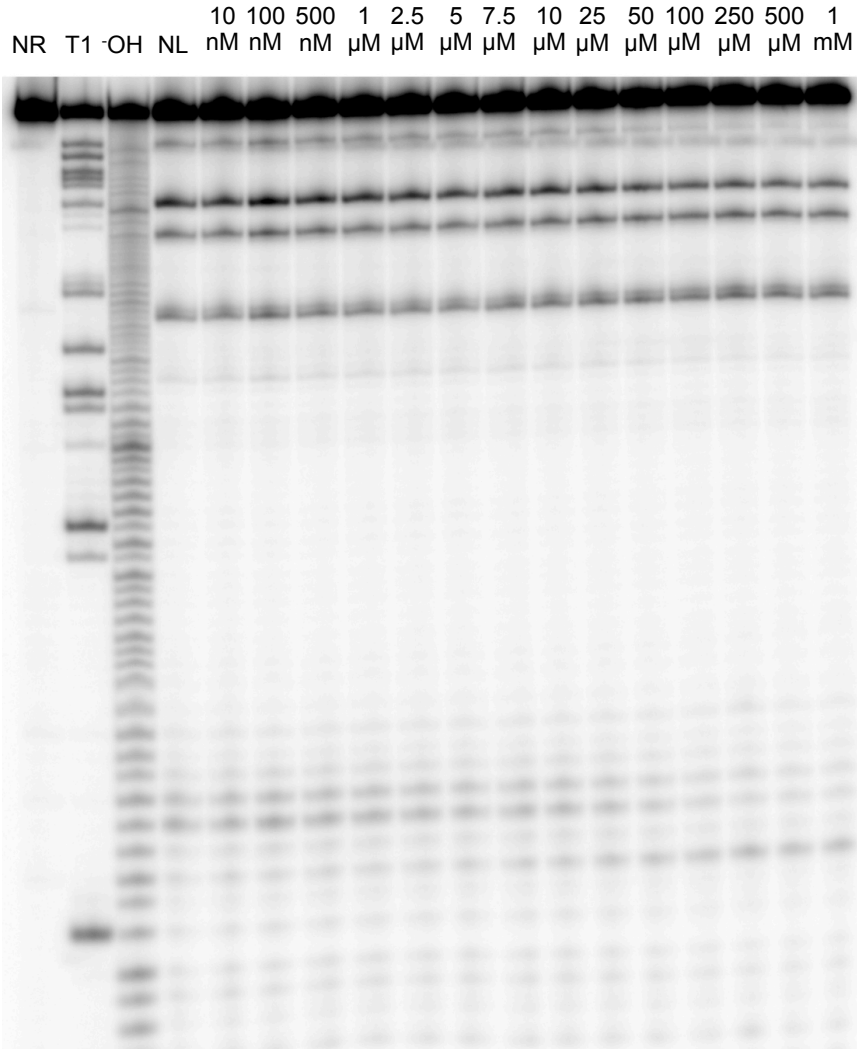



**Figure 3.7: In-line probing $K_d$ gel of GM-790 with cyclic di-GMP**
*Dissociation constant determined to be ~26 µM. Lanes NR, T1, and ⁻OH refer to "no reaction", T1 ribonuclease digest, and base hydrolysis, respectively.*

### 3.5 Identification of cyclic AMP-GMP in *Geobacter sulfurreducens* cell extract

At the time of experimentation, cyclic AMP-GMP had only been found in *Vibrio cholera*.[18] Therefore even though *Geobacter* harbored seemingly cyclic AMP-GMP specific riboswitches, it was uncertain to whether the small molecule existed within the genus. The cyclic dinucleotide content of *Geobacter* bacteria had not been analyzed previously, although bioinformatics analyses have predicted the presence of enzymes related to cyclic di-GMP and cyclic di-AMP signaling. However, no enzymes containing homology to DncV, the cyclic AMP-GMP synthase in *V. cholera*, were predicted. Thus, the cyclic dinucleotide content of

*Geobacter sulfurreducens* was analyzed.

First, an organic-aqueous extraction of cyclic dinucleotides from *G. sulfurreducens* PCA cells was performed. Then, liquid chromatography and mass spectrometry were used to analyze the extract's small molecule content. Liquid chromatography-mass spectrometry (LC-MS) and high-resolution mass spectrometry (HRMS) analysis revealed that all three bacterial cyclic dinucleotides (cyclic di-AMP, cyclic AMP-GMP, and cyclic di-GMP) are present in *G. sulfurreducens* cell extracts (Figure 3.8 and Figure 3.9). Cyclic AMP-GMP was then collected in a discrete HPLC fraction (Figure 3.9) and subjected to MS/MS fragmentation analysis. Compared against two different regioisomer standards of cAG, fragmentation of the cyclic AMP-GMP isolate from *G. sulfurreducens* is consistent with a cyclic dinucleotide that is solely 3′, 5′ linked (Figure 3.10). This finding marked the second known natural occurrence of 3′-5′, 3′-5′ cyclic AMP-GMP.

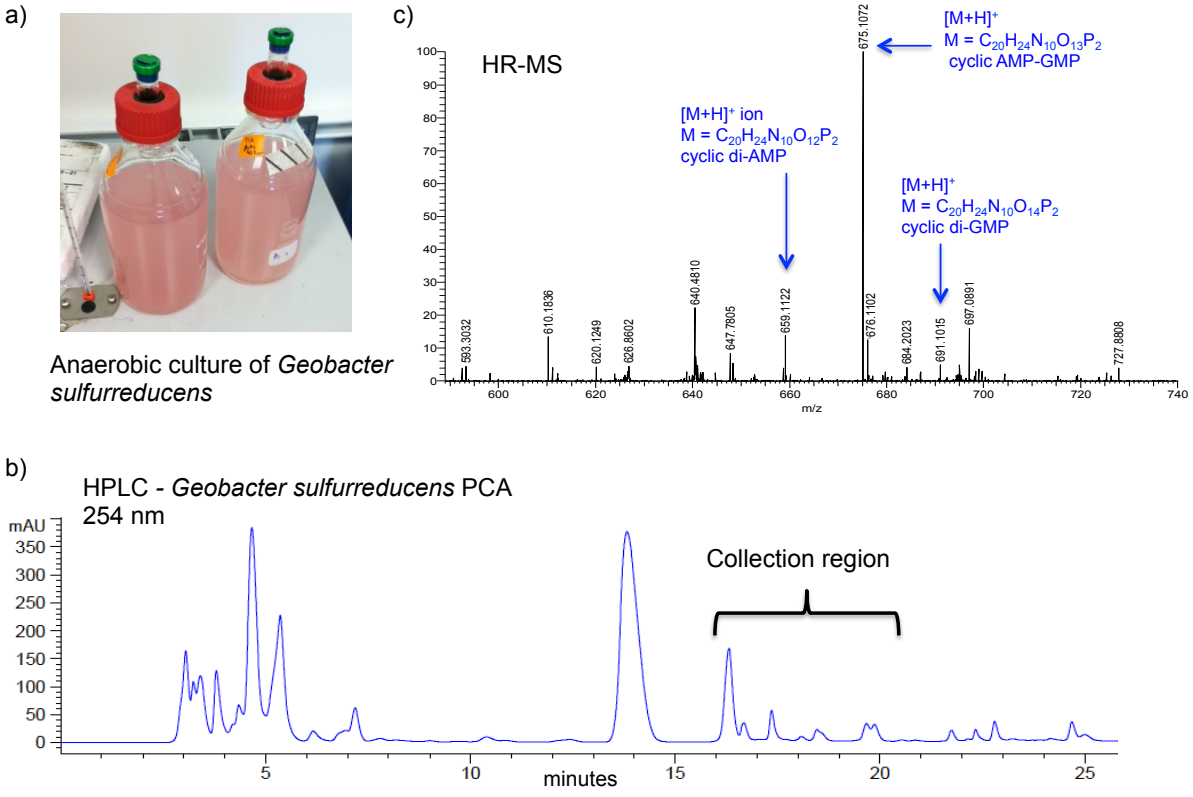



**Figure 3.8: cdN cell extract analysis of *Geobacter sulfurreducens***
*(a) Anaerobic culture of Geobacter sulfurreducens PCA. (b) HPLC chromatogram of Geobacter sulfurreducens PCA cell extract. Cyclic dinucleotide collection region indicated. (c) HR-MS of collection region indicated in (b). Cyclic di-AMP, cyclic AMP-GMP, and cyclic di-GMP are confirmed in the extract.*

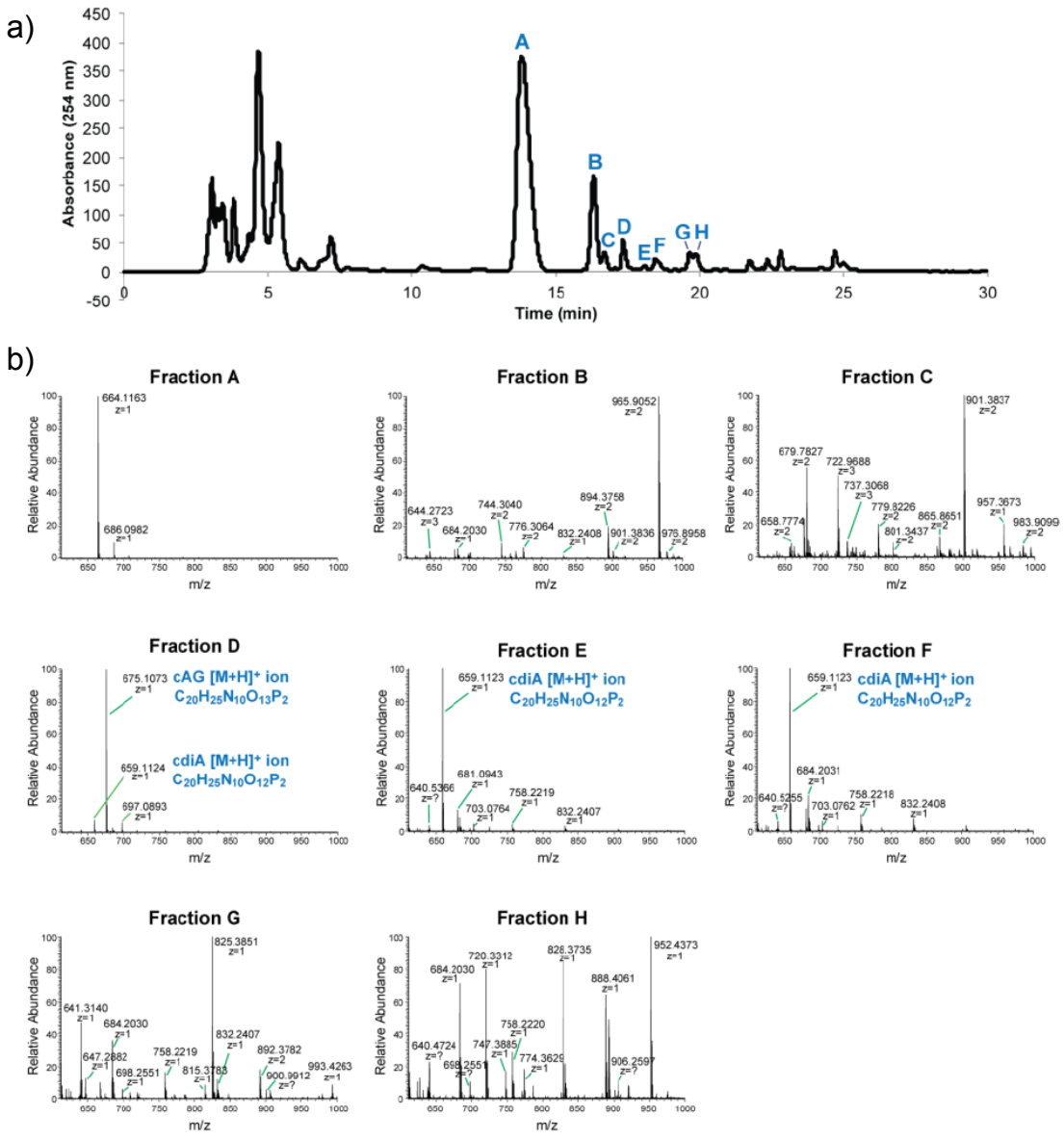



**Figure 3.9: HRMS analysis of discrete fractions from *G. sulfurreducens* PCA cell extract**
*(a) HPLC chromatogram of G. sulfurreducens PCA cell extract. (b) HRMS analysis of fractionated HPLC collections. Fraction D contains cyclic AMP-GMP (expected: 675.1078, found: 675.1073).*

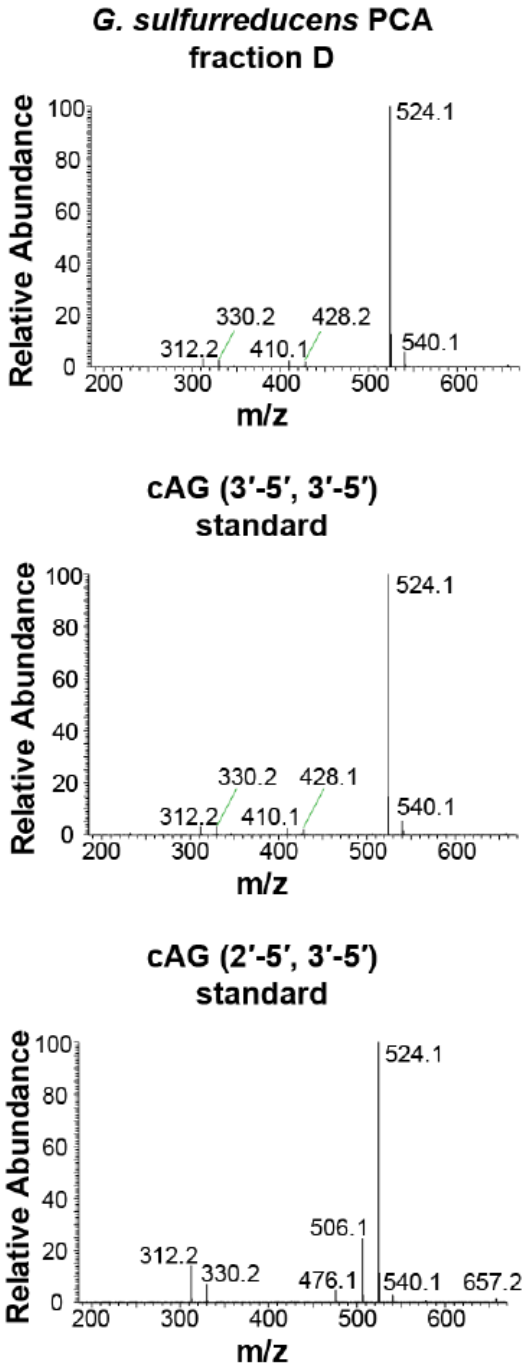



**Figure 3.10: MS/MS analysis of fraction D compared to synthetic standards**
*The fragmentation patterns of cyclic AMP-GMP in fraction D versus standard of cAG (3′- 5′, 3′- 5′) and cGAMP (2′-5′, 3′-5′) are shown. MS/MS analysis confirms that the fragmentation pattern matches cyclic AMP-GMP (3′-5′, 3′-5′).*

### 3.6 GEMM-Ib class riboswitch as a genetic regulatory element

A key metabolic feature of the *Geobacter* genus is the ability to perform extracellular electron transfer. In addition to playing an important role in natural geochemistry through the reduction of environmental metals[44], *Geobacter* have been described as "living batteries" that have the capacity to generate electricity from organic waste.[45] Significantly, the genes regulated by GEMM-Ib class riboswitches and thus activated by cAG signaling include many that are functionally linked to this process.[46] For example, the gene regulated by the GS-814 riboswitch, pgcA, encodes a periplasmic cytochrome c protein that is conserved in the *Geobacter* genus and is more highly expressed during growth on insoluble Fe(III) oxide versus Fe(III) citrate.[47,48]

Interestingly, a C-to-U transition that maps to the GS-814 riboswitch in the pgcA transcript was shown to double the rate of Fe(III) oxide reduction by *G. sulfurreducens*.[49] When the C78U mutation for the GS-814 riboswitch was analyzed by in-line probing it was found that the point mutation indeed destabilizes the terminator stem, as the observed pattern is similar to that of WT GS-814 (although not identical) bound to cAG (Figure 3.11). This suggests that the C78U mutation would constitutively turn on gene expression, which supports the ~15 fold increase of pgcA transcripts observed in the mutant strain (Figure 3.12).[49] Furthermore, the Breaker lab demonstrated that genes downstream of GM-790 are regulated in an identical manner. In the presence of cyclic AMP-GMP, transcription of the full-length downstream genetic cassette increased, demonstrating that cAG turns on genetic expression.[46]

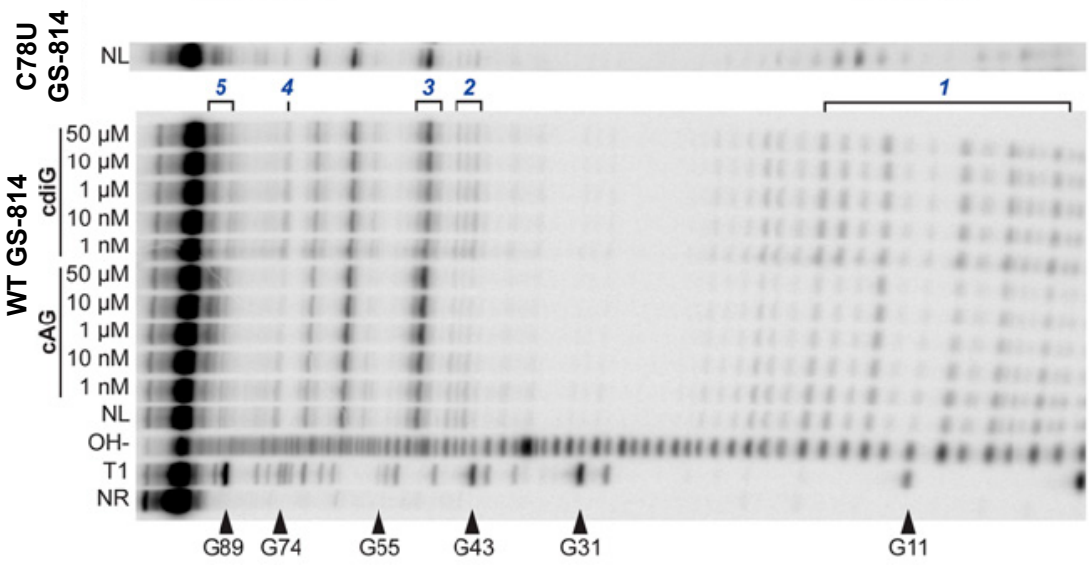



**Figure 3.11: In-line probing analysis of WT and C78U GS-814 riboswitch**
*The C78U mutation destabilizes the terminator stem, which leads to a cleavage pattern similar to that of WT GS-814 bound to cAG. [Data acquired by Scott Hickey]*

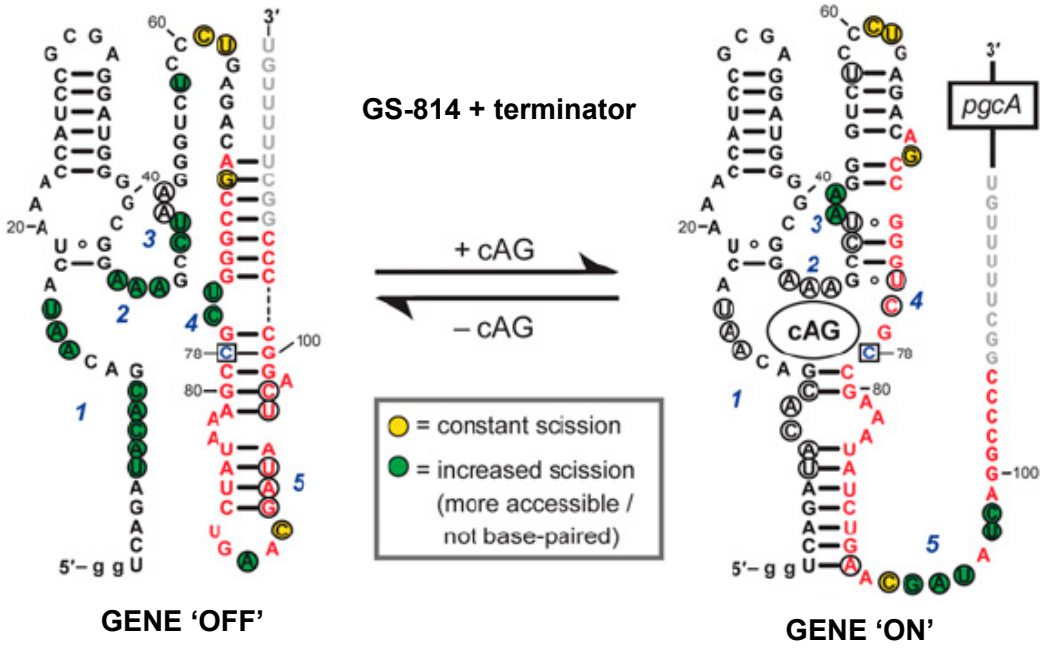



**Figure 3.12: In-line probing analysis of the GS-814 riboswitch mapped onto its secondary structure model**
*A termination hairpin is formed in the absence of cyclic AMP-GMP, which acts as an off-switch for the downstream gene. When bound, the hairpin is resolved and transcription is resumed.*

Several other cytochrome c-containing and biogenesis proteins appear to be regulated by cAG, including OmcS, which has been shown to be essential for metal oxide reduction and more highly expressed upon growth on insoluble Fe(III) or Mn(IV) oxides.[50,51] Although cytochrome genes are remarkably abundant in *Geobacter* species, they are not well conserved, making it difficult to base functional assignments of electron conduits on phylogenetic arguments.[52] The identification of a subset of cytochromes co-regulated by cAG suggests that they are part of a specific pathway and distinguishes them from others encoded in the genome.

In addition to extracellular cytochromes, *Geobacter* produce specialized type IV pili that are intimately involved in contacting insoluble metal sources, including electrodes, and there is evidence for direct electron shuttling by the pili[53], although this mechanism is still under active debate[54]. The GM-790 and GS-745 riboswitches reside within an operon conserved in *Geobacter* that contains genes encoding minor pilins and pilus assembly proteins. Thus, the genes of known function associated with GEMM-Ib appear to be involved in extracellular electron transfer and suggests a role for cAG signaling in the regulation of *Geobacter* electrophysiology. However, the majority of genes associated with GEMM-Ib are of unknown

or uncharacterized function, including two that are co-transcribed with PilMNOPQ and OmcAHG.

### 3.7 Discussion

This chapter demonstrated that cAG is produced by *G. sulfurreducens*, and that *Geobacter* species harbor riboswitches as effectors that are selective for cAG. While the cAG synthase DncV is found in *V. cholerae*, homology searches to DncV did not identify any candidate cAG synthase in *Geobacter* genomes. Therefore, it is an open question to how cAG is created in *Geobacter*. However, it is intuitive to hypothesize that since a subpopulation of the cyclic di-GMP riboswitch GEMM-I class responds to cAG, a subpopulation of GGDEF domain containing enzymes (i.e. diguanylate cyclases) could make cAG. In an energy powerhouse like *Geobacter*, cAG production could be indicative of the conditional energy status of the bacterium (i.e. ATP and GTP levels).

It is clear that much remains to be discovered about cAG signaling, and how it distinguishes itself from other cdN signaling. The most unsophisticated scenario is that the cdN signals are completely redundant (i.e. different inputs lead to the same signal output) or completely distinct (i.e. different inputs lead to different signaling outputs). Both of these cases are somewhat unlikely. Many A20 GEMM-I riboswitches bind promiscuously to both cyclic di-GMP and cyclic AMP-GMP (Figure 3.3), suggesting that a well-defined role for each cdN is not conserved. On the other hand, *Geobacter* harbors many of its GEMM riboswitches in a dual tandem fashion (data not shown), implying that both upstream riboswitches need to be modulated in order to elicit downstream genetic effects. The specificity for these riboswitches is still under investigation, but thus far some appear to be cAG specific, whereas others are promiscuous for cdG and cAG. Regardless, dual tandem riboswitches suggest that cdN signaling is not completely redundant.

Alternatively, and perhaps more likely, the signaling space can be envisioned more like a Venn diagram, where cdG and cAG have distinct signaling pathways but there is some overlap in downstream signaling. At the molecular level, this could depend on whether cdN binding is primarily governed by nucleobase or macrocycle recognition. Nucleobase recognition would lead to more distinct roles in cdN signaling, whereas macrocycle recognition would have a more unifying effect. For example, HCN4, a cyclic nucleotide-gated channel, recognizes multiple different cdNs, which can then modulate β-adrenergic stimulation.[55] In fact, this phenomenon plays itself out with other nucleotide-based messengers outside of cdNs. For example, the growth-promoting protein $p21^{Ras}$ has been shown to bind to both GTP as well as cyclic di-GMP.[56] The culmination of the resulting signals would then be a consequence of the synergy from the overall small molecule, nucleotide-based condition of the cell and how much crosstalk exists between nucleotide-based effectors. How this crosstalk is affected by symmetric versus asymmetric cyclic dinucleotides is a question that is only starting to be addressed.[57]

# 4. A NONCANONICAL CYCLIC AMP-GMP REGIOISOMER IS A MAMMALIAN SECONDARY MESSENGER FOR INNATE IMMUNITY

## 4.1 Background

Recognition of pathogen-derived nucleic acid is a major mechanism by which innate immune responses are initiated in mammals.[58] Several families of germ-line-encoded nucleic acid sensors have been described, including the Toll-like receptors and RIG-I-like receptors.[59,60] Upon binding nucleic acids, these sensors initiate signaling cascades that lead to the production of cytokines and other immune effector proteins that provide host defense.

The cytosolic presence of foreign double-stranded DNA (dsDNA) has been previously shown to trigger a potent antiviral response dominated by the production of type I interferons (IFNs).[61,62] STING (**ST**imulator of **I**F**N** **G**enes; also called TMEM173, MITA, ERIS, and MPYS), an endoplasmic-reticulum-resident host protein, has been shown to be required for dsDNA dependent IFN response.[24,63-65] Cyclic dinucleotides have also been shown to induce an IFN response[15] that depends on STING[66,67]. They are secreted or released into the cytosol by certain bacterial pathogens[68,69] and bind directly to STING[70] (Figure 4.1). A mutant allele of mouse STING (mSTING), encoding an alanine in place of arginine 231 (R231A) was found to abolish responsiveness to cdNs but did not appreciably affect the IFN response to cytosolic dsDNA.[70] Thus, although the IFN responses to both cytosolic cdNs and dsDNA require STING, the responses to these chemically distinct ligands can be genetically uncoupled.

In 2013 two seminal papers from the Chen lab were published that detailed the discovery of a cytosolic DNA sensor in mammalian cells denoted as cGAS (**cGA**MP **S**ynthase).[17,71] This protein-based DNA sensor was found to synthesize cyclic AMP-GMP from endogenous ATP and GTP in response to binding dsDNA. Cyclic AMP-GMP was then proposed to act as a secondary messenger and bind to the downstream effector STING. Consequently, STING activates and triggers a potent immunological response causing the induction of interferon-β, cytokines, and chemokines (Figure 4.1).[23,24,70,72] Up until the 2013 Chen papers, cyclic dinucleotides such cyclic di-GMP, cyclic di-AMP, and cyclic AMP-GMP were not known to exist in eukaryotes (with the exception of *Dictyostelium*[73]). Thus, these papers represented the first identification of an endogenous, mammalian cyclic dinucleotide and established cGAS as a cytosolic DNA sensor for innate immunity. However, it soon became clear that mammalian cyclic AMP-GMP (cGAMP) is distinguished from prokaryotic cyclic AMP-GMP by a single, unique noncanonical 2′-5′ phosphodiester bond. This chapter describes how cGAMP's unique regiochemistry was discovered and elucidated.

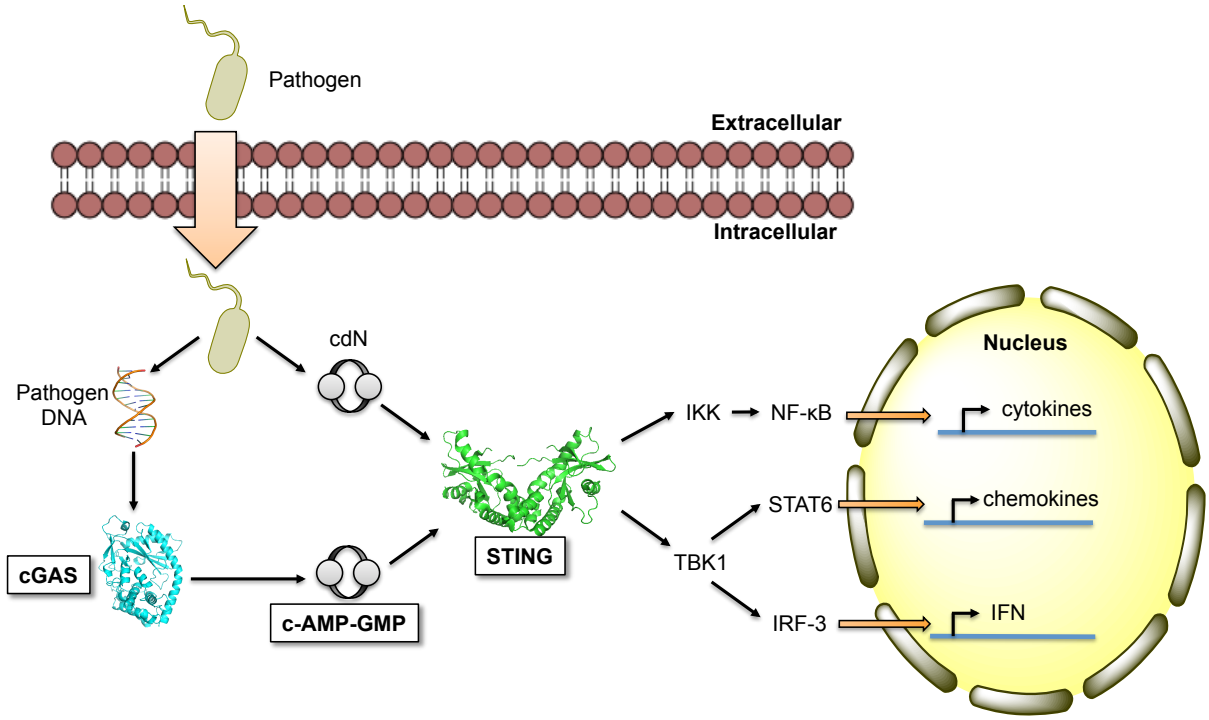



**Figure 4.1: Early model for cGAS-STING pathway activation**
*The figure indicates the product of cGAS to be canonical cyclic AMP-GMP as to highlight the prior understanding in the field before the small molecule was elucidated.*

**4.2 The enzymatic product of cGAS is inconsistent with canonical cyclic AMP-GMP**
There are four major non-synonymous variants of hSTING found in high frequencies within the human population, according to the 1,000 Genome Project Database. One such variant, R232H, occurs with an allelic frequency of 13.7% and has an attenuated IFN response to cyclic di-GMP and cyclic AMP-GMP.[74] Using a luciferase-based IFN detection assay, the activation of hSTING H232 and R232 was compared using synthetic forms of cyclic di-GMP and cyclic AMP-GMP, the same products from each respective prokaryotic synthase (i.e. WspR and DncV), and the enzymatic product produced by cGAS. As expected, the two cyclic dinucleotides from both the synthetic origin and enzymatic origin elicited a stronger IFN response with hSTING R232 than H232 (Figure 4.2). However unlike cyclic di-GMP and cyclic AMP-GMP, the cGAS enzymatic product elicited an IFN response from H232 that mimicked that from hSTING R232, suggesting that the cGAS product was not the same as the prokaryotic forms of cyclic di-GMP and cyclic AMP-GMP.

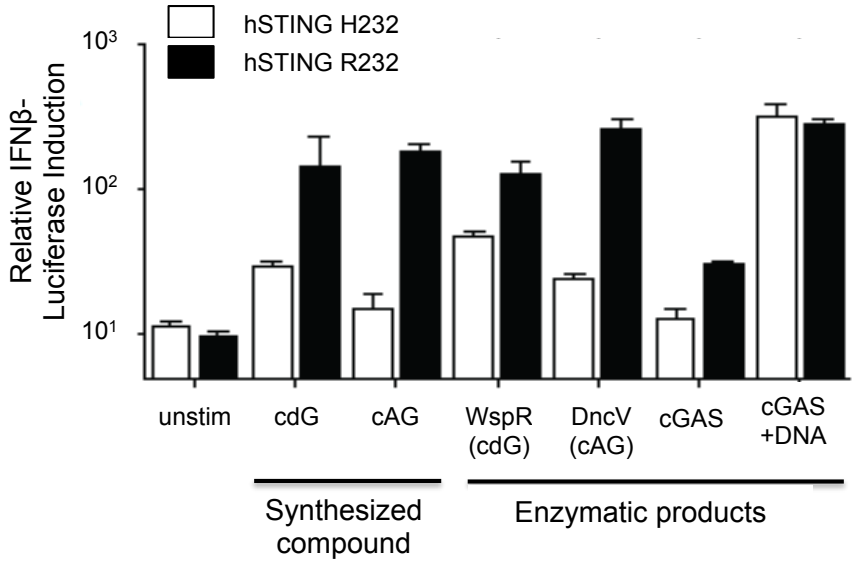



**Figure 4.2: Cyclic dinucleotide activation of hSTING H232 and R232 using a luciferase-based IFN-β detection assay**
*The cGAS product elicits a stronger IFN response from hSTING H232 compared to cdG and cAG. [Data acquired by Elie Diner and Dara Burdette]*

### 4.3 cGAS produces cyclic AMP-GMP with a single, unique 2′-5′ bond

cGAS is structurally homologous to another innate immune sensor called oligoadenylate synthase that produces a noncanonical oligoadenylate polymer containing 2′-5′ phosphodiester bonds.[71,75] Therefore, it was hypothesized that cGAS might not produce a canonical cdN as previously proposed[17,71] but, instead, might produce a cdN containing 2′-5′ phosphodiester bond(s) that could enhance stimulation of the H232 STING allele. Such a noncanonical cdN would have a mass identical to the canonical 3′-5′ phosphodiester-linked cdN, yet could produce different MS/MS patterns upon fragmentation. In fact, the previously reported MS/MS pattern of the cGAS product is not identical to that of a synthetic cyclic AMP-GMP synthetic standard.[17] This hypothesis was tested by providing radiolabelled α-$^{32}$P-GTP or α-$^{32}$P-ATP to recombinant purified cGAS or *V. cholerae* DncV and analyzing the products by thin-layer chromatography (TLC). As reported previously, DncV can produce some cyclic di-AMP (cdA) when provided with only ATP and some cdG when provided with only GTP but prefers to make cyclic AMP-GMP when provided with both ATP and GTP (Figure 4.3a).[18] cGAS requires both ATP and GTP substrates, and the resulting product migrates significantly differently than any of the canonical cdNs produced by DncV, suggesting that cGAS produces a noncanonical cdN (Figure 4.3a).

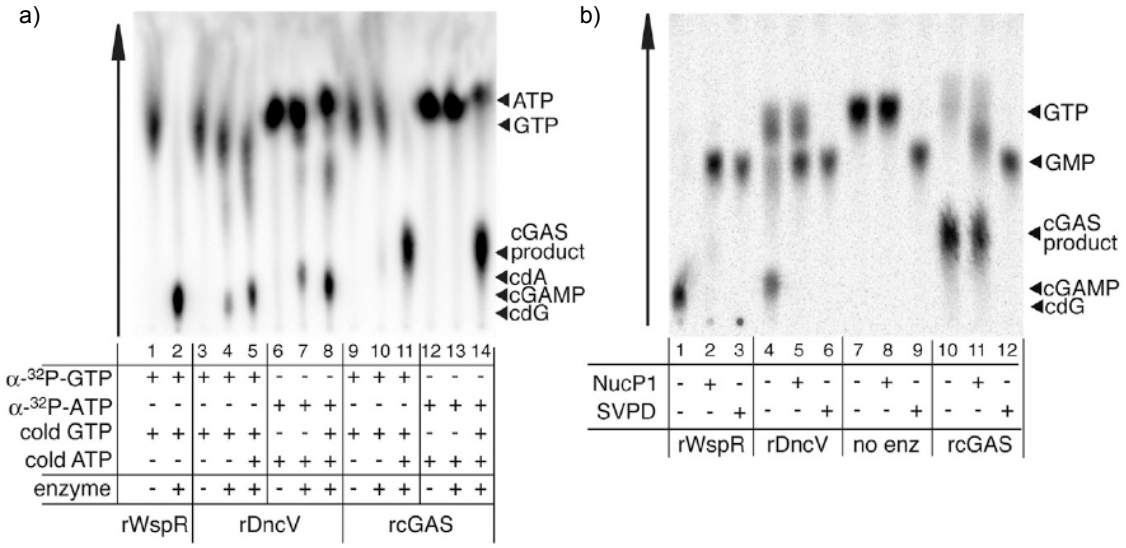



**Figure 4.3: Thin-layer chromatography analysis of cyclic dinucleotides**
*(a) Purified recombinant WspR, DncV, or cGAS were mixed with α-$^{32}$P-ATP or α-$^{32}$P-GTP and the indicated unlabeled nucleotides. Reactions were mixed with TLC running buffer, and nucleic acid species were resolved on a polyethylenimine (PEI) cellulose TLC plate. (b) WspR, DncV, or cGAS products labeled with α-$^{32}$P -GTP were digested with nuclease P1 or snake venom phosphodiesterase (SVPD) and resolved on a PEI cellulose TLC plate. [Data acquired by Elie Diner and Dara Burdette]*

The cGAS and DncV products were then analyzed by specific nuclease digestion. The cGAS product is partially cleaved by nuclease P1, which selectively digests 3′-5′ phosphodiester linkages[76], suggesting that the cGAS product contains at least one 3′-5′ phosphodiester linkage (Figure 4.3b). Nuclease P1 digestion is incomplete, given that it does not lead to the generation of GMP, in contrast to the complete digestion observed upon treatment of the DncV product with nuclease P1 (Figure 4.3b). As a control, digestion of the cGAS or DncV products with snake venom phosphodiesterase, which cleaves both 2′ -5′ and 3′ -5′ phosphodiester linkages[76], led to complete digestion (Figure 4.3b). Altogether, these results suggest that the cGAS product might also contain a 2′-5′ phosphodiester linkage. To identify the nature of the cGAS product and, in particular, ascertain the regiochemistry of the phosphodiester linkages, the cGAS product was purified by reverse-phase HPLC (Figure 4.4) and analyzed by two-dimensional nuclear magnetic resonance (NMR) spectroscopy (Figure 4.5).

Regiochemistry of the phosphodiester bonds was determined using $^{1}$H-$^{31}$P HMBC (Figure 4.5) with a low-pass J filter corresponding to 8 Hz. In this case, the phosphorous nucleus, P-11, is correlated to the 2′ ribose proton (H-12) of guanosine as well as to the 5′ ribose methylene protons (H-10). A four-bond correlation between P-11 and the 4′ ribose proton (H-

9) of adenosine is also observed. The other phosphorous nucleus (P-22) is correlated to the 3′ ribose proton (H-8) of adenosine and to the 5′ ribose methylene protons (H-21). A four-bond correlation between P-22 and the 4′ ribose proton (H-20) of guanosine is also observed. Thus, the regiochemistry of the phosphodiester linkages was determined to be cyclic[G(2′-5′)pA(3′-5′)p].

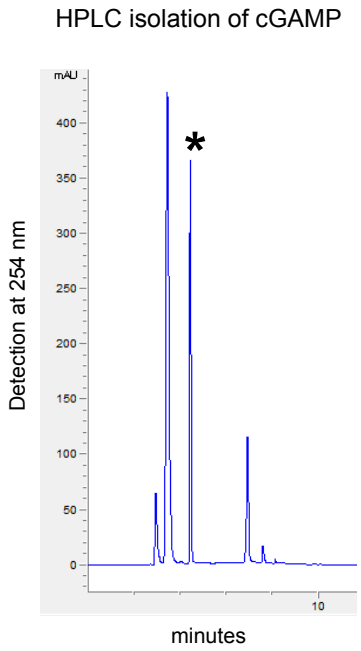



**Figure 4.4: HPLC purification of cGAS enzymatic product**
*cGAMP is annotated with an asterisk. Other peaks correspond to starting material (ex. ATP and GTP) and byproducts (ex. linear AMP-GMP).*

Resonance assignments associated with the nucleobases and ribose spin systems were made using $^{1}$H-$^{13}$C HSQC (Figure 4.6), $^{1}$H-$^{1}$H COSY (Figure 4.7), $^{1}$H-$^{1}$H NOESY (Figure 4.8), $^{1}$H-$^{1}$H ROESY, $^{1}$H-$^{13}$C HMBC, and appropriate reference spectra. $^{1}$H-$^{1}$H NOESY and $^{1}$H-$^{1}$H ROESY indicate through-space interactions between the adenine proton H-5 and the 3′ ribose proton (H-8). Weak ROEs are also detected from H-5 to H-6, H-7, and H-10. In addition, $^{1}$H-$^{1}$H NOESY and $^{1}$H-$^{1}$H ROESY depict through-space interactions between the guanine proton H-17 and the 1′ ribose proton (H-18) (Figure 4.8). The remaining protons in the corresponding ribose spin systems were identified by $^{1}$H-$^{1}$H COSY (Figure 4.7) and multiplicity edited $^{1}$H-$^{13}$C HSQC (Figure 4.6a and 4.6b), which distinguished the 5′ methylene protons (H-10 and H-21) in particular.

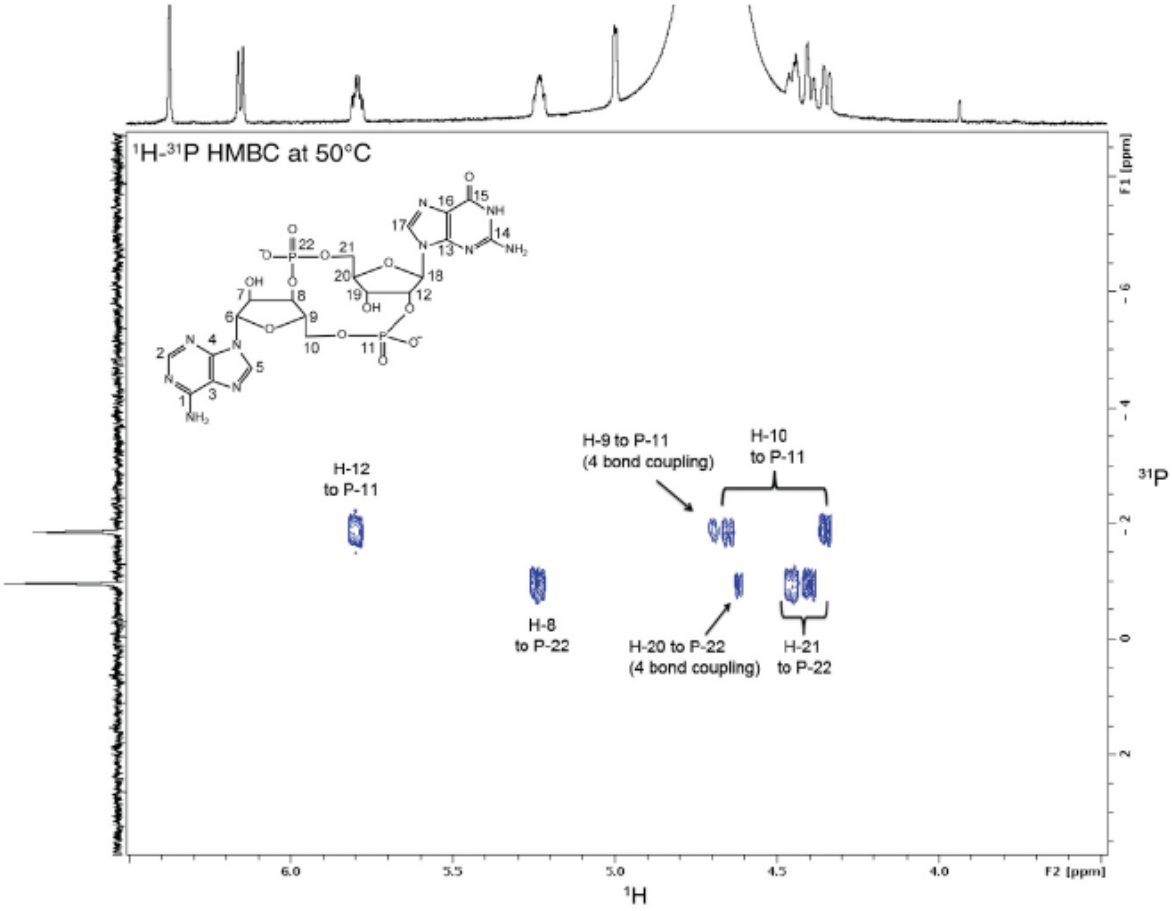



**Figure 4.5: $^{1}$H-$^{31}$P HMBC of HPLC-purified cGAS product acquired at 600 MHz**
*Critical through-bond correlations for the phosphodiester bonds are indicated. An NMR-elucidated structure of the cGAS product is also shown.*

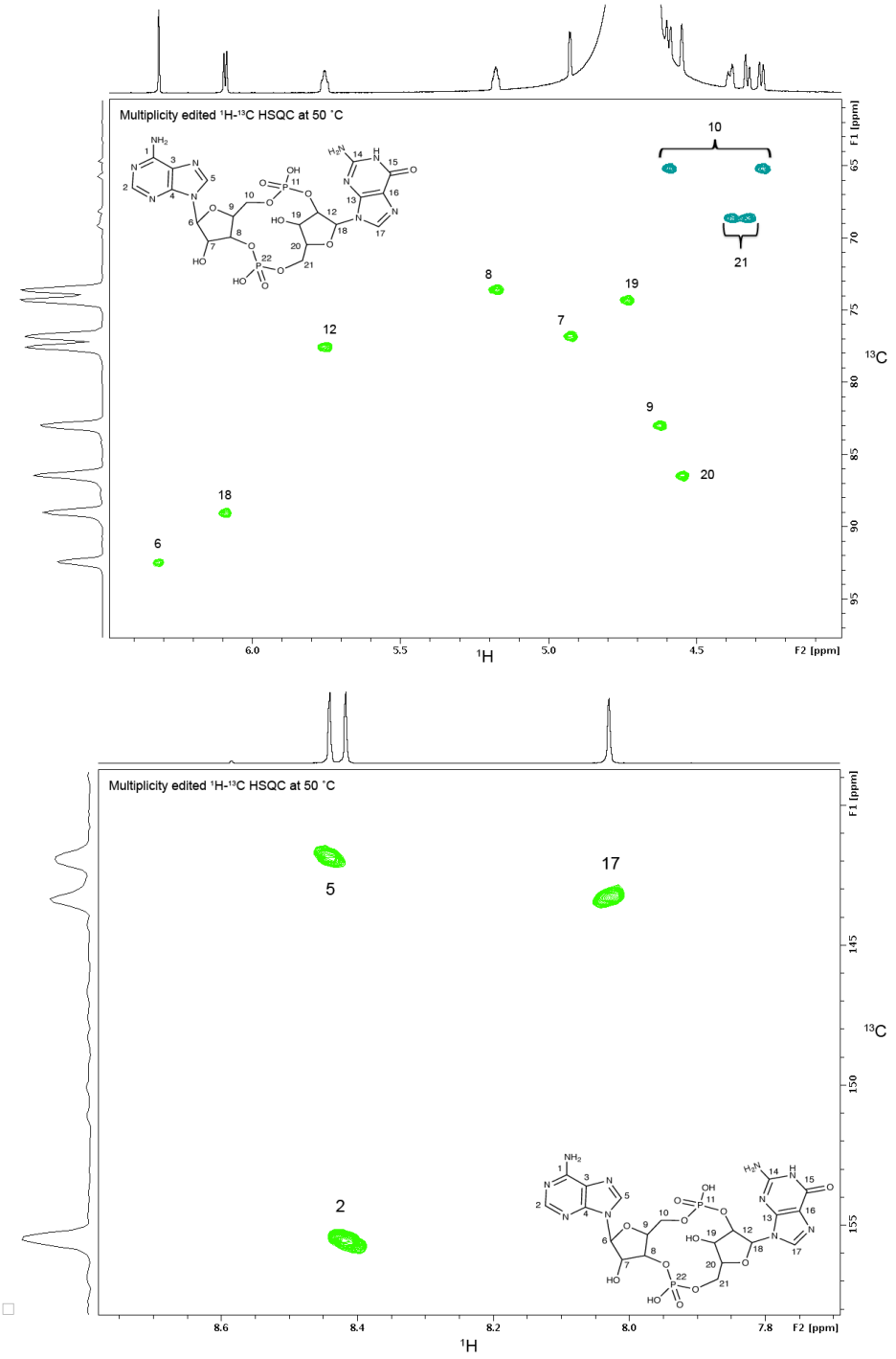



**Figure 4.6: Multiplicity-edited $^{1}$H-$^{13}$C HSQC experiment in a 900 MHz field**
*Positive phased signals corresponding to methine and methyl protons are shown in green, negative phased signals corresponding to methylene protons are shown in blue.*

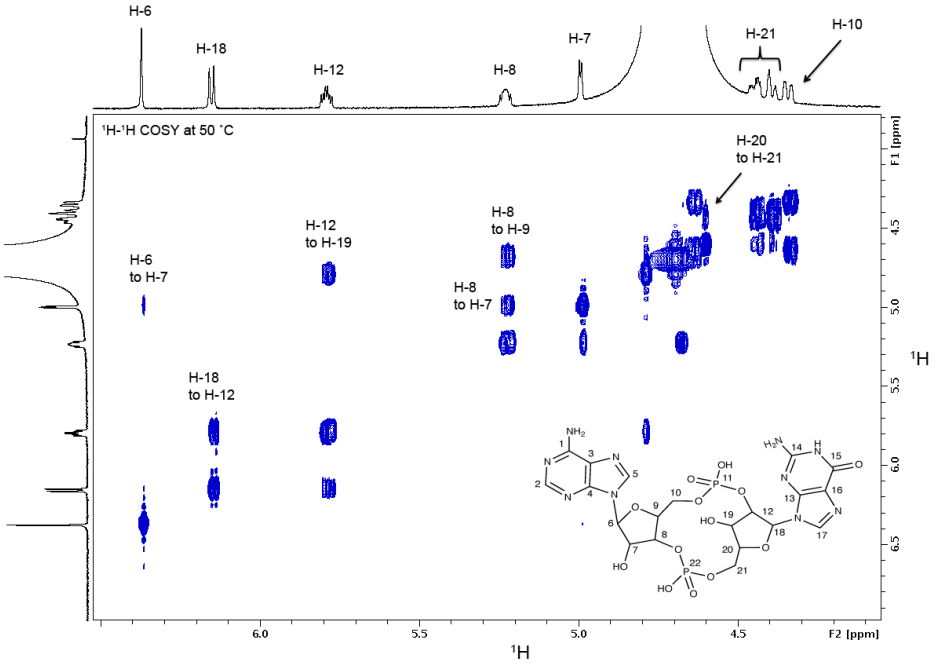



**Figure 4.7: $^{1}$H-$^{1}$H COSY experiment in a 600 MHz field**

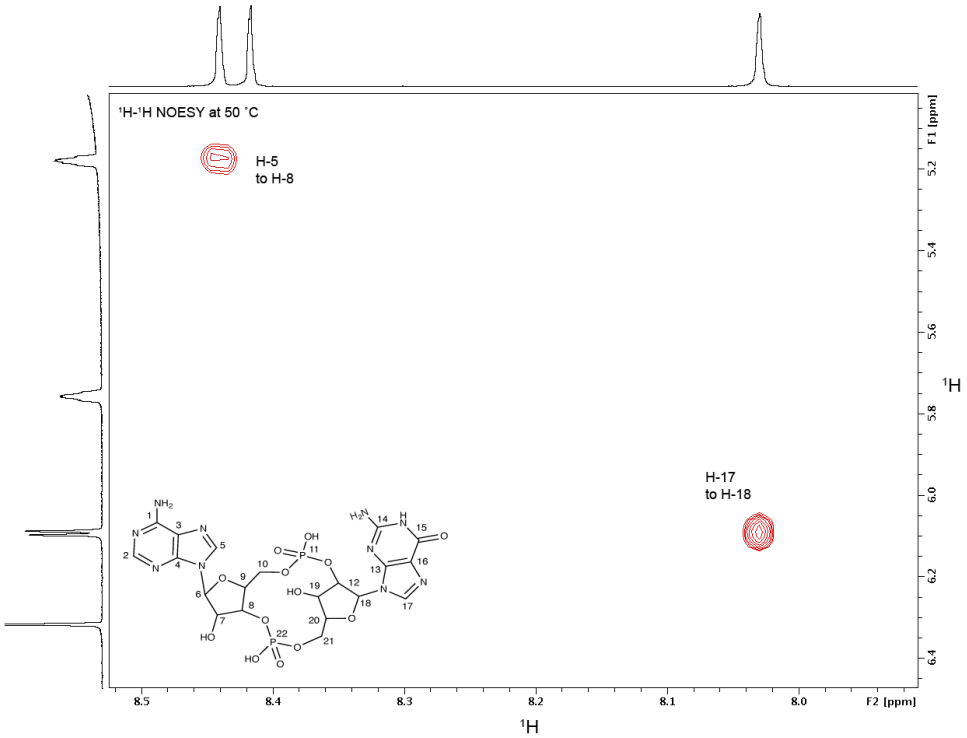



**Figure 4.8: $^{1}$H-$^{1}$H NOESY experiment in a 900 MHz field**

These results indicate that the product of cGAS is a noncanonical cdN containing a single 2′-5′ phosphodiester linkage and that its chemical structure is assigned as cyclic [G(2′-5′)pA(3′-5′)p] (Figure 4.9). This second messenger appears to be a robust activator of STING. These results demonstrate the unique ability of cyclic [G(2′-5′)pA(3′-5′)p] to stimulate diverse hSTING alleles and yet may explain why cGAS synthesizes this unusual molecule.

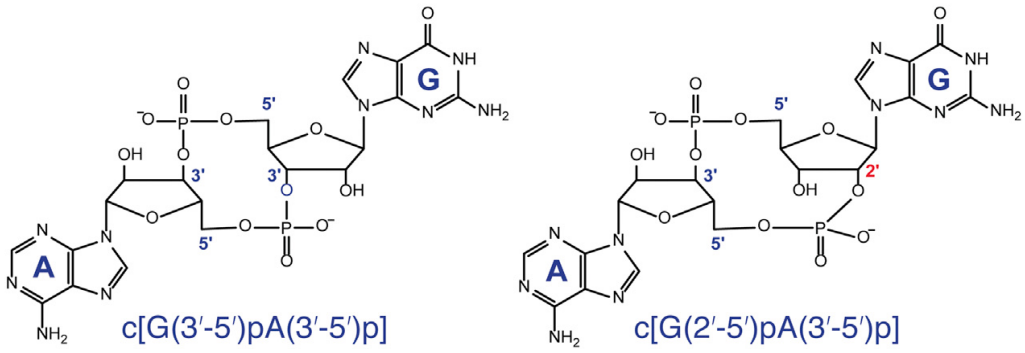



**Figure 4.9: Regioisomers of cyclic AMP-GMP**
*Chemical structures of canonical cyclic AMP-GMP, cyclic [G(3′-5′)pA(3′-5′)p] (left), and noncanonical cyclic AMP-GMP (cGAMP), cyclic [G(2′-5′)pA(3′-5′)p] (right) are shown.*

### 4.4 Discussion

These experiments have unequivocally shown that human cGAS produces a novel cyclic dinucleotide, cyclic [G(2′-5′)pA(3′-5′)p], which distinguishes itself from prokaryotic cdN messengers (Figure 4.10). However, the driving force for the modification in regiochemistry is unknown and it is still not clear why the R232H STING variant can be stimulated by the noncanonical cdN and not by the canonical cdNs. It is tempting to speculate that the R232H variant of hSTING may confer a selective advantage by reducing responses to bacterial cdNs while still retaining responsiveness to the endogenous noncanonical cdN produced by cGAS in response to viral dsDNA. Indeed, although the production of type I IFNs is essential for the control of most viruses, type I IFNs have often been found to be detrimental in the response to bacterial infections.

cdNs have been proposed to be useful as vaccine adjuvants or immunotherapeutics.[77] In addition, a synthetic STING activator, DMXAA, has been tested in human clinical trials as a chemotherapeutic agent. DMXAA was not found to be effective in humans, most likely because it is unable to stimulate hSTING.[78] In this context, these results could be significant because they suggest that noncanonical 2′-5′linked cdNs might have clinical value as potent pan-agonists of diverse STING variants, including those variants that respond poorly to canonical cdNs or DMXAA.

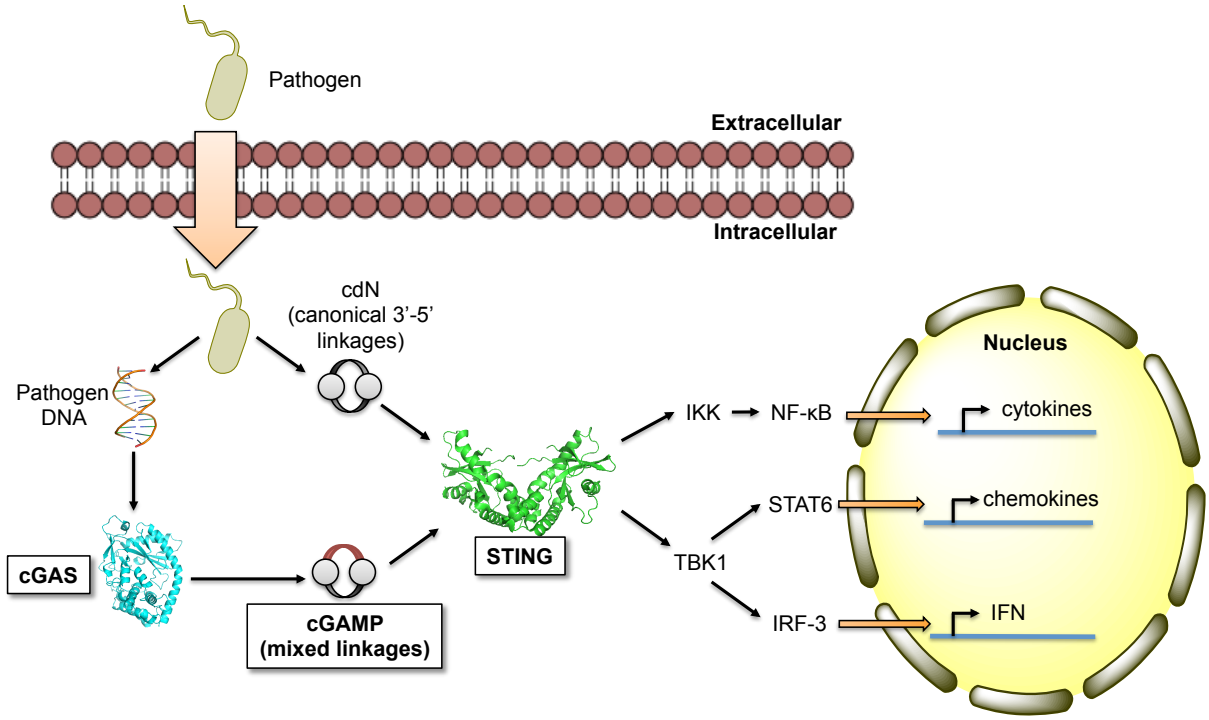



**Figure 4.10: The cGAS-STING axis is a cytosolic DNA sensing pathway that elicits a potent immune response**

*cGAS activation by dsDNA produces a noncanonical cyclic AMP-GMP (cGAMP), which binds to STING and elicits a downstream immune response.*

## 5. THE cGAS-cdN-STING AXIS IS AN EVOLUTIONARY CONSERVED PATHWAY

### 5.1 Background

From prokaryotes to humans, the idea that cyclic dinucleotide signaling evolved from solely 3′-5′ linked to a mixed linkage messenger is intriguing. Prokaryotes employ cdA, cAG, and cdG for a diverse range of processes (Figure 1.2), and humans utilize cGAMP as a means for signaling pathogen invasion. However, the necessity for this regiochemical change is not clear, nor if other metazoans utilize the same pathway for pathogen detection. If so, did metazoans always utilize a mixed linkage messenger? In order to get a better understanding of how innate immunity developed, the evolution of the cGAS-cdN-STING axis was examined.

Prior bioinformatics analyses have identified candidate STING homologs throughout diverse metazoans.[79] However, only mammalian STING proteins have been functionally characterized and shown to bind cdNs. Thus, it remains possible that ancestral STING had other functions and that the ability to bind cdNs is a recent innovation that arose in vertebrates alongside modern innate immunity. In contrast to this view, this chapter shows that binding to cdNs is a deeply evolutionarily conserved function of STING. Using a paleo-biochemical approach, the function of STING was tracked through diverse animal lineages, and functional cGAS and STING homologs were identified in the sea anemone *Nematostella vectensis*, an animal divergent from humans by >500 million years of evolution. These results reveal the surprisingly ancient evolutionary origins of cGAS-STING signaling and demonstrate that the cdN second messenger pathway in animals predates the emergence of interferon signaling and modern innate immunity.

Furthermore, results presented in this chapter demonstrate that anemone cGAS only activates the human STING R232 allele, and its cyclic dinucleotide product can only be digested by a 3′-5′ cAG phosphodiesterase, suggesting that the anemone cGAS product is the canonically linked (solely 3′-5′) cyclic AMP-GMP, not 2′-5′ containing cGAMP. A complete series of seven crystal structures of anemone STING were also solved, encompassing two unbound structures and five ligand-bound structures. These structures comprise a complete set of STING structures and reveal an unexpected nucleobase-specific recognition of cyclic AMP-GMP by anemone STING. In addition, cGAMP has an ability to trap a distinct and deeply evolutionarily conserved STING conformation not seen with 3′,3′ cyclic dinucleotides. While the function of ancient STING is still unknown, these results illuminate the evolution and ancestral origins of a critical second-messenger signaling pathway in human immunity and provide insight into the conserved functionality required for the design of broadly active STING chemotherapeutics.

### 5.2 STING cyclic dinucleotide recognition predates modern innate immunity

To probe STING function in diverse animal lineages, candidate *STING* genes were identified

by protein homology using a hidden Markov model to search published animal genomes. The C-terminal STING domains of 11 different STING orthologs were expressed, purified from *E. coli*, and tested for their ability to interact with $^{32}$P-radiolabelled cyclic dinucleotides using an electrophoretic mobility gel-shift assay. All four known naturally occurring cdNs were tested: 3′,3′ cyclic di-AMP, 3′,3′ cyclic AMP-GMP, 2′,3′ cGAMP, and 3′,3′ cyclic di-GMP (Figure 5.1a). The vertebrate STING orthologs that were analyzed readily interacted with the endogenous product cGAMP. While all mammalian STING proteins tested also bound to 3′,3′ cdNs, no 3′,3′ cdN binding with STING proteins from more diverse vertebrates including *D. rerio* (zebrafish) and *X. tropicalis* (western clawed frog) was observed. In support of preservation of a functional cGAS-cGAMP-STING pathway across vertebrate evolution, mammalian, *Xenopus*, and *Danio* genomes contain readily identifiable cGAS homologs with an intact active site and Zn-ribbon insertion critical for human cGAS function.[75,79,80]

In contrast to vertebrate STING proteins, no interaction between the insect *Drosophila* STING and any of the four naturally occurring cdNs was observed. To ascertain the generality of this result, three additional STING alleles from insect genomes (bumble bee, *B. terrestris*; wasp, *N. vitripennis*; and silkworm, *B. mori*) were tested and again there was no detectable cdN interaction other than nonspecific well-shifts. A search for cGAS homologs in insect genomes reveals possible nucleotidyltransferase superfamily-related genes but no obvious homologs of human cGAS that contain both an enzymatic active site and a Zn-ribbon domain for dsDNA-specific recognition.[75,79-81] These results suggest that insect STING proteins may have lost or decreased the ability to bind to cdNs or may only do so with additional co-factors missing from our assay.

Among invertebrates, putative STING homologs within phyla Annelida, Mollusca, and Cnidaria were characterized. Despite <30% amino acid identity with vertebrate STING proteins, the invertebrate STING homologs we tested exhibited a robust ability to bind both 3′,3′ cdNs and 2′,3′ cGAMP (Figure 5.1a). The most divergent STING homolog examined was from the starlet sea anemone (*Nematostella vectensis*), an animal with >500 million years of evolutionary divergence from humans.[82] Consistent with our biochemical findings, phylogenetic alignment of human STING (hSTING) and anemone STING (nvSTING) indicates that despite low sequence identity (29%), the key residues previously implicated in hSTING cdN recognition are conserved in ancestrally divergent STING homologs (Figure 5.1b,c).

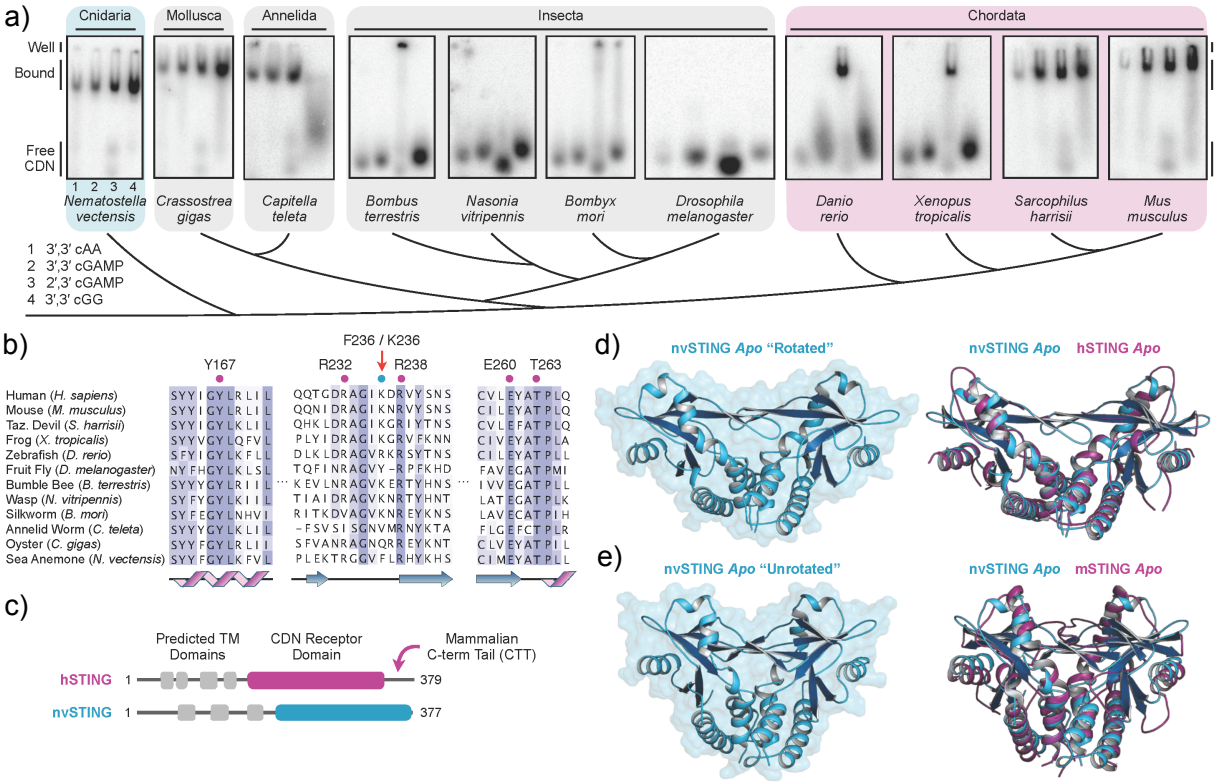



**Figure 5.1: STING cyclic dinucleotide recognition predates modern innate immunity**
*(a) Gel-shift assay measuring complex formation between recombinant STING proteins and radiolabeled cdNs. Each panel shows reactions in order: (1) 3′,3′ cyclic di-AMP, (2) 3′,3′ cyclic AMP-GMP, (3) 2′,3′ cGAMP, and (4) 3′,3′ cyclic di-GMP. (b) Phylogenetic alignment of animal STING proteins colored by amino-acid conservation. Key cdN interacting residues are indicated. (c) Cartoon schematic of human and anemone STING proteins. Predicted transmembrane domains are indicated in gray, and the crystallized cdN receptor domain is indicated in magenta or blue. Notably, the C-terminal tail required for hSTING interferon signaling is absent in nvSTING. (d and e) Crystal structure of apo nvSTING cdN receptor domain in ''rotated'' and ''unrotated'' states. Overlaid structures compare the apo nvSTING structures (blue) and apo hSTING (PDB: 4F9E) or apo mouse STING (PDB: 4KC0) (magenta) revealing conformational dynamics and an ancestrally conserved STING fold.*

## 5.3 STING architecture is conserved in anemone

To confirm that invertebrate STING homologs exhibit evolutionary conservation of key structural features, crystal structures of the C-terminal cdN receptor domain of nvSTING were solved using selenomethionine-derivatized crystals for experimental phase determination (Figure 5.1d,e and Table 5.1). *Apo* nvSTING crystallized in two distinct crystal forms, an open ''rotated'' wing conformation (2.1 Å) and a closed "unrotated" wing conformation (2.9

| | nvSTING–cGG c[G(3′–5′)pG(3′–5′)p] | nvSTING–3′,3′ cGAMP c[G(3′–5′)pA(3′–5′)p] | nvSTING–cAA c[A(3′–5′)pA(3′–5′)p] | nvSTING Apo "Rotated" | nvSTING–cGG (SeMet) c[G(3′–5′)pG(3′–5′)p] | nvSTING–cGG F276K c[G(3′–5′)p[G(3′–5′)p] | nvSTING–2′,3′ cGAMP c[G(2′–5′)pA(3′–5′)p] | nvSTING Apo "Unrotated" |
|---|---|---|---|---|---|---|---|---|
| Data Collection | | | | | | | | |
| Resolution | 48.54–1.84 Å (1.90–1.84 Å) | 40.63–1.99 Å (2.04–1.99 Å) | 40.83–2.95 Å (3.06–2.95 Å) | 41.15–2.10 Å (2.16–2.10 Å) | 48.66–2.39 Å (2.48–2.39 Å) | 40.93–2.07 Å (2.12–2.07 Å) | 42.49–2.10 Å (2.16–2.10 Å) | 43.04–2.85 Å (3.00–2.85 Å) |
| Wavelength | 1.11587 Å | 1.11587 Å | 1.11587 Å | 1.11587 Å | 0.97967 Å | 1.11588 Å | 1.11587 Å | 1.11587 Å |
| Space group | P $3_1$ | P $3_1$ | P $3_1$ | P $3_1$ | P $3_1$ | P $3_1$ | P $2_12_12_1$ | P $2_12_12_1$ |
| Unit cell: a, b, c | 81.11, 81.11, 97.07 Å | 81.27, 81.27, 99.12 Å | 81.66, 81.66, 98.76 Å | 82.30, 82.30, 100.35 Å | 80.77, 80.77, 97.33 Å | 81.39, 81.39, 100.55 Å | 40.50, 77.97, 101.35 Å | 43.04, 43.04, 170.94 Å |
| Unit cell: α, β, γ | 90.0°, 90.0°, 120.0° | 90.0°, 90.0°, 120.0° | 90.0°, 90.0°, 120.0° | 90.0°, 90.0°, 120.0° | 90.0°, 90.0°, 120.0° | 90.0°, 90.0°, 120.0° | 90.0°, 90.0°, 90.0° | 90.0°, 90.0°, 90.0° |
| Molecules per ASU | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| Total reflections: | 693,819 | 473,676 | 89,635 | 494,252 | 323,356 | 290,406 | 283,149 | 31,005 |
| Unique reflections: | 61,587 | 49,839 | 15,258 | 43,684 | 27,919 | 45,536 | 19,434 | 7,901 |
| Completeness | 98.7% (82.5%) | 99.5% (93.9%) | 98.3% (85.5%) | 98.4% (81.1%) | 99.2% (92.7%) | 99.6% (94.6%) | 99.6% (95.6%) | 99.4% (97.5%) |
| Multiplicity | 11.3 (9.3) | 9.5 (8.2) | 5.8 (5.5) | 11.3 (8.1) | 11.6 (10.9) | 6.4 (6.0) | 14.6 (6.6) | 3.9 (4.0) |
| $I/\sigma I$ | 10.3 (1.5) | 13.4 (1.8) | 6.1 (1.0) | 9.4 (1.0) | 17.1 (5.2) | 8.0 (1.1) | 16.2 (1.6) | 12.6 (3.2) |
| CC(1/2) | 99.8% (50.8%) | 99.9% (63.4%) | 95.9% (34.4%) | 99.8% (34.0%) | 99.8% (91.5%) | 99.5% (31.3%) | 99.9% (82.1%) | 99.7% (97.5%) |
| Rpim | 3.6 (69.9%) | 3.3 (73.9%) | 10.5 (89.2%) | 4.4 (88.2%) | 3.7 (16.2%) | 5.2 (90.1%) | 2.4 (54.1%) | 3.8 (23.1%) |
| No. sites | – | – | – | – | 8 | – | – | – |
| Refinement | | | | | | | | |
| Resolution | 48.54–1.84 Å | 40.63–1.99 Å | 40.83–2.95 Å | 41.15–2.10 Å | | 40.93–2.07 Å | 42.49–2.10 Å | 43.04–2.85 Å |
| Free reflections | 5% | 5% | 5% | 5% | – | 5% | 5% | 5% |
| R-factor/R-free | 17.2/20.9 | 16.9/19.0 | 22.1/24.6 | 17.2/20.2 | – | 17.2/20.6 | 24.2/25.9 | 25.7/27.8 |
| Bond distance (RMS Å) | 0.003 | 0.008 | 0.003 | 0.011 | – | 0.009 | 0.003 | 0.003 |
| Bond angles (RMS °) | 0.740 | 1.053 | 0.944 | 1.322 | – | 1.298 | 0.733 | 0.805 |
| Structure/Stereochemistry | | | | | | | | |
| Average B-factor: protein | 37.6 | 51.3 | 74.4 | 54.2 | – | 48.7 | 67.4 | 64.7 |
| Average B-factor: ligand | 24.7 | 33.5 | 85.9 | – | – | 54.4 | 31.1 | 76.5 |
| Ramachandran plot: favored | 98.4% | 97.8% | 98.6% | 97.8 | – | 98.0% | 96.1% | 94.6% |
| Ramachandran plot: allowed | 1.6% | 2.2% | 1.4% | 2.2 | – | 2.0% | 3.9% | 5.4% |
| Rotamer outliers: | 0.9% | 0.9% | 1.5% | 1.2% | – | 1.2% | 0.6% | 4.8% |
| MolProbity score | 1.06 | 1.29 | 1.49 | 1.19 | – | 1.35 | 1.72 | 2.28 |
| Protein Data Bank ID | PDB: 5CFL | PDB: 5CFM | PDB: 5CFN | PDB: 5CFO | – | PDB: 5CFP | PDB: 5CFQ | PDB: 5CFR |

**Table 5.1: nvSTING crystallographic statistics**

Å). In each case, the structures reveal a dimeric protein receptor with clear structural homology to human and mouse STING.[83-88] Two distinct ‘‘rotated’’ and ‘‘unrotated’’ *apo* nvSTING structures correlate with previously observed *apo* structures of human and mouse STING, respectively, indicating that these two forms likely represent alternatively sampled STING conformations and not species-specific differences between the human and mouse proteins (Figure 5.1d,e). Importantly, the structures of *apo* nvSTING reveal the conserved presence of a beta-strand lid domain poised to arch over a deep central ligand-binding pocket (Figure 5.1d,e). Taken together, these results provide strong evidence that the overall STING fold, and its ability to bind cdNs, is deeply conserved throughout metazoans, predating the clear emergence of interferons and “modern” innate immunity.[89]

### 5.4 The cGAS-cdN-STING pathway is conserved in sea anemone

To determine whether the entire cGAS-STING pathway is evolutionarily conserved, invertebrate genomes encoding a functional cGAS-like enzyme were analyzed. Efforts were focused on the sea anemone *N. vectensis*, as this was the most divergent species found to encode a functional STING ortholog. Twelve *N. vectensis* genes encoding putative cGAS-like proteins with predicted functional active sites were found and cloned. These enzymes were then screened for cdN synthase activity by transiently transfecting them in a hSTING R232 interferon-β luciferase reporter cell assay, which responds to all known naturally occurring cdNs.[27] Only a single anemone gene, *nv-A7SFB5.1* (Ensembl: A7SFB5.1), induced robust hSTING-dependent interferon-β signaling in human cells (Figure 5.2). Overexpression of all other candidate *N. vectensis* genes, including the closely related protein nv-A7S0T1.1, did not result in any detectable STING activation (Figure 5.3). The other candidate genes are inactive in this assay either because they are functionally diverse nucleotidyltransferases unrelated to cdN synthases or possibly because essential co-factors are missing from the cellular assay. Alignment of nv-A7SFB5.1 and human cGAS reveals a shared predicted catalytic active site (Figure 5.4), and disruption by amino-acid substitution of the predicted magnesium ion coordination motif in nv-A7SFB5.1 (E131A/D133A) abolishes all signal induction in human cells. Identification of nv-A7SFB5.1 (now re-named nv-cGAS) extends the occurrence of cdN synthases to invertebrate animals and demonstrates that cGAS and STING are ancestrally conserved signaling components in animal biology.

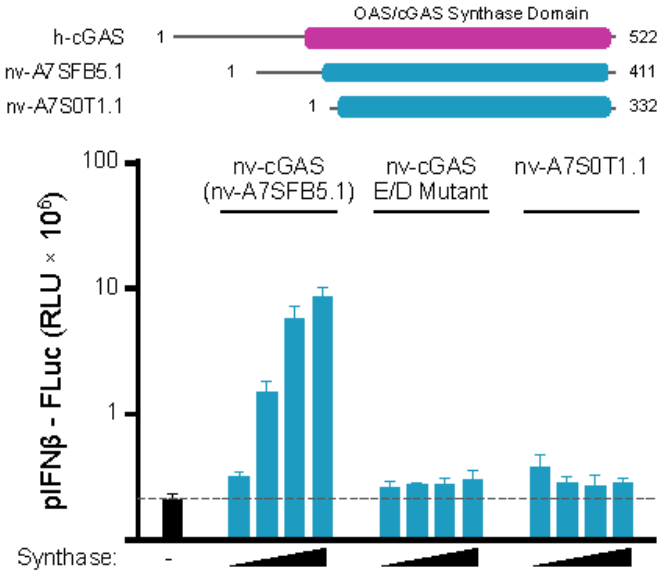



**Figure 5.2: Cartoon schematic of candidate N. vectensis cGAS-like enzymes and screen for cdN synthase activity using interferon-β assay**
*Activity is detected with gene nv-cGAS (nv-A7SFB5.1), but not in a catalytic inactive nv-cGAS control (E/D mutant) or in the closely related candidate gene nv-A7S0T1.1.*

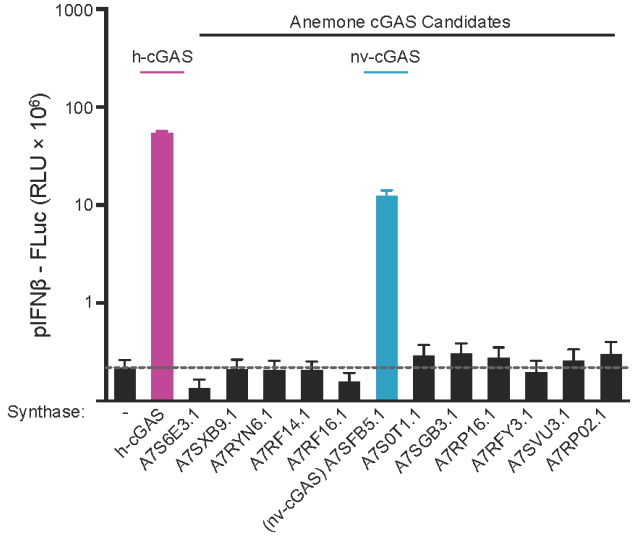



**Figure 5.3: Screen of anemone cGAS candidate genes for synthase activity using interferon-β assay**
*Only one putative cGAS homolog in Nematostella was discovered to be capable of making a cdN.*

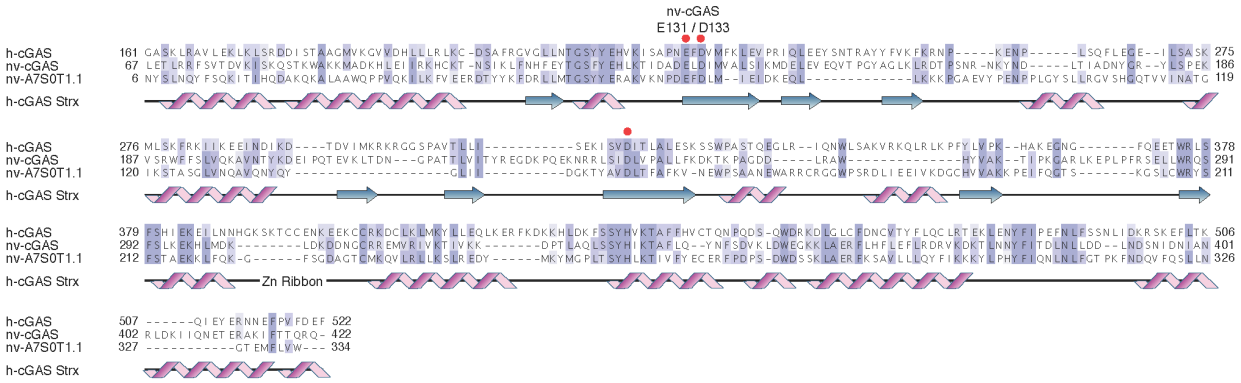



**Figure 5.4: Detailed sequence alignment of synthase domains from human cGAS, nv-cGAS (nv-A7SFB5.1), and related candidate synthase nv-A7S0T1.1**
*Alignment is colored according to phylogenetic conservation.*

## 5.5 Anemone cGAS produces via a 3′,3′-linked cyclic dinucleotide

Human cGAS produces a cyclic GMP–AMP second messenger containing a mixed 2′–5′ and 3′–5′ phosphodiester linkage (2′,3′) distinct from all previously characterized bacterial 3′,3′ cdNs. To determine the nv-cGAS product identity and linkage specificity, reconstitution of nv-cGAS enzymatic activity was attempted in vitro using purified components. In contrast to the robust in vitro activity of recombinant human cGAS, nv-cGAS activity was not detected in vitro upon stimulation in the presence of dsDNA. These results suggest that an alternative ligand present within the complex cellular environment may be responsible for controlling nv-cGAS activity.

Since the activating ligand for nv-cGAS is currently unknown, the inherent activity of nv-cGAS in transfected cells was studied. As explained in Chapter 4, the H232 and R232 alleles of STING elicit varying levels of IFN in response to cyclic dinucleotides depending on if they are canonically-linked (cyclic di-AMP, cyclic AMP-GMP, and cyclic di-GMP) or noncanonically-linked (cGAMP). Overexpression of nv-cGAS was able to activate the hSTING R232 allele but was not able to active the hSTING H232 allele (Figure 5.5), suggesting that the synthase makes only a 3′–5′ linked product.

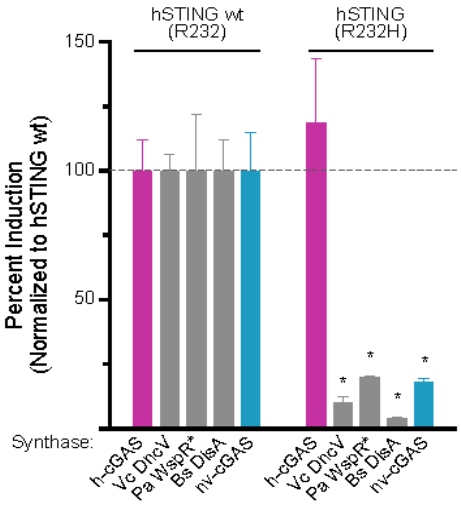



**Figure 5.5: Differential activation of hSTING H232 and R232 using select cyclic dinucleotides**

*Production of 2′,3′ cGAMP by human cGAS activates both STING alleles, while 3′,3′ cdN production by bacterial synthases and nv-cGAS only activates hSTING R232.*

While the interferon-β luciferase experiment indicates that the nv-cGAS product is 3′–5′ linked, it does not address its base identity. *Vibrio cholerae* DncV is a close structural homolog of human cGAS[90-92] and the enzyme active sites of DncV and cGAS are similar enough that a single amino acid substitution converts human cGAS into a DncV-like 3′,3′ cyclic AMP-GMP synthase[92]. Given that the nv-cGAS sequence is predicted to have similar changes to catalytic pocket amino acids important for substrate positioning (Figure 5.4), it was hypothesized that the 3′,3′ cdN nv-cGAS product is most likely to be 3′,3′ cyclic AMP-GMP. Fortunately, *Vibrio cholerae* cGAP1 efficiently hydrolyzes 3′,3′ cyclic AMP-GMP but has no activity on any other cdN product.[93] This specificity was confirmed in the luciferase-based reporter system by showing that co-expression of cGAP1 ablated DncV 3′,3′ cyclic AMP-GMP signaling but only had a minimal (<2-fold) impact on STING activation by bacterial cyclic di-GMP (*Pseudomonas* WspR) or cyclic di-AMP (*Bacillus* DisA) synthases (Figure 5.6). Importantly, cGAP1 abolished the ability of nv-cGAS to induce STING signaling (Figure 5.6), consistent with nv-cGAS producing 3′,3′ cyclic AMP-GMP. These results demonstrate that nv-cGAS produces a 3′,3′ cdN and suggest that the mixed phosphodiester linkage chemistry found in human 2′,3′ cGAMP is an evolutionarily recent immune second messenger adaptation.

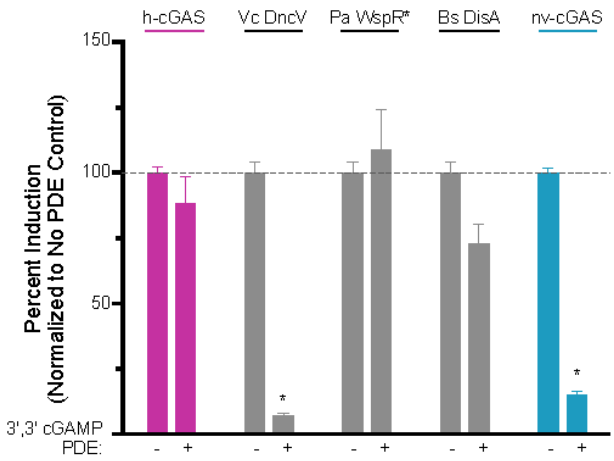



**Figure 5.6: nv-cGAS induced interferon-β luciferase signal is ablated by a 3′,3′ cyclic AMP-GMP specific phosphodiesterase**
*Cell interferon-β luciferase with supplementation of a 3′,3′ cyclic AMP-GMP specific phosphodiesterase (PDE) (Vibrio cholerae VCA0681) Vc DncV and nv-cGAS 3′,3′ cyclic AMP-GMP production is ablated by 3′,3′ cyclic AMP-GMP PDE co-expression, while other synthases are not significantly affected.*

### 5.6 nvSTING specifically recognizes guanine bases in 3′,3′ second messengers

In order to characterize the interaction of nvSTING with cdN ligands, crystal structures of nvSTING bound to cyclic AMP-GMP (2.0 Å), cyclic di-GMP (1.8 Å), and cyclic di-AMP (3.0 Å) were solved (Figure 5.7a). Interestingly, this series of structures uncovered a nucleobase-specific recognition of guanine not previously seen in mammalian STING structures. Similar to the previously reported beta-strand lid observed to form in hSTING in response to its agonist ligand, the nvSTING beta-strand lid domain undergoes an ~4 Å movement to close tightly over the top of cyclic AMP-GMP and cyclic di-GMP ligands but is notably splayed open and more loosely organized when nvSTING is bound to cyclic di-AMP (Figure 5.7b). In line with the contrasting activities of 3′,3′ versus 2′,3′ second messengers, the ability of the nvSTING beta-strand lid to specifically reorganize upon cyclic AMP-GMP and cyclic di-GMP binding is in stark contrast to previous crystal structures of hSTING bound to cyclic di-GMP, where the hSTING beta-strand lid remains disordered (Figure 5.7c).[85,88] These results suggest that recognition of the guanine nucleobase is a unique property of nvSTING.

To validate the base-specific contacts observed in the nvSTING cyclic AMP-GMP and cyclic di-GMP structures, ligand affinity was measured using isothermal titration calorimetry (ITC). nvSTING exhibited a remarkably high affinity of for cyclic AMP-GMP and cyclic di-GMP

ligands (50 and 15 nM, respectively) compared to the previously described affinity of hSTING for cyclic di-GMP (~1,000–4,400 nM) (Figure 5.7d; Table 5.1 and 5.2).[21,85,88] In further confirmation of the selective recognition of guanine nucleobases, the affinity of nvSTING for cyclic di-AMP is two orders of magnitude lower (~1,500 nM) and in closer agreement to the previously measured low-micromolar affinities of hSTING and mouse STING for 3′,3′ cdNs (Figure 5.7d; Table 5.2).

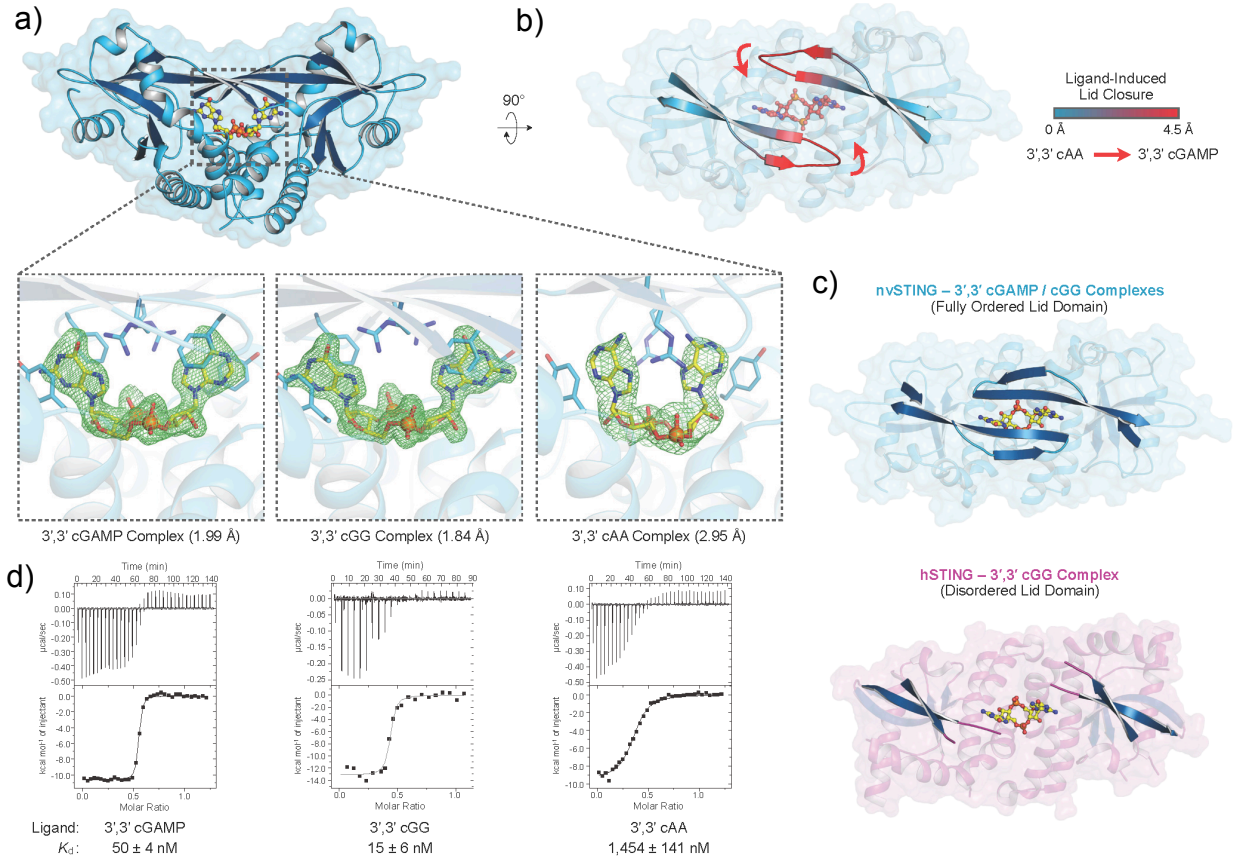



**Figure 5.7: nvSTING binds cdNs with high affinity**

*(A) Crystal structure of nvSTING in complex with 3′,3′ cGAMP (cyclic AMP-GMP), cyclic di-GMP, or cyclic di-AMP as indicated. Zoomed-in cutaway includes simulated-annealing $F_O$-$F_C$ omit maps of ligand density contoured to 5.0 σ (cyclic AMP-GMP, cyclic d-GMP) or 4.0 σ (cyclic di-AMP). (B) Top-down view of nvSTING–cyclic di-AMP crystal structure. The beta-strand lid domains are highlighted and colored according to range of movement compared to the closed nvSTING–cyclic AMP-GMP/cyclic di-GMP crystal structures. Red arrows denote the direction of beta-strand lid closure upon cyclic AMP-GMP/cyclic di-GMP binding. (C) Crystal structures of nvSTING– cyclic AMP-GMP/cyclic di-GMP (blue) and hSTING–cyclic di-GMP (hSTING R232 and H232 PDB: 45FY, 4F9G) (magenta) complexes. The beta- strand lid domain is highlighted illustrating the fully closed nvSTING lid domain*

*compared with the disordered and loosely organized hSTING lid domain. (D) ITC measurements of nvSTING affinity for specific 3′,3′ cdN ligands as indicated. ITC data are representative of at least three independent experiments. See also Table 5.1 and Table 5.2.*

| Ligand | Protein (μM) | Replicates | $K_D$ (nM) | ΔH (kcal $mol^{-1}$) | ΔS (kcal $mol^{-1}$ $K^{-1}$) | n |
|---|---|---|---|---|---|---|
| **nvSTING** | | | | | | |
| 3′,3′ cGAMP | 100 | 3 | 50 ± 4 | −11.89 ± 1.09 | −6.47 ± 3.83 | 0.57 ± 0.03 |
| 3′,3′ cGG | 15 | 5 | 15 ± 6 | −12.60 ± 0.92 | −6.36 ± 3.49 | 0.46 ± 0.09 |
| 3′,3′ cAA | 100 | 3 | 1,454 ± 141 | −9.85 ± 0.41 | −6.20 ± 1.07 | 0.29 ± 0.06 |
| 2′,3′ cGAMP | 100 | 3 | < 1 | −4.53 ± 0.09 | ND | 0.45 ± 0.03 |

**Table 5.2: ITC values for nvSTING-cdN interactions**

### 5.7 2′,3′ cGAMP Traps a Unique and Conserved STING Conformation

Collectively, the data demonstrates that the *N. vectensis* cGAS-STING signaling pathway relies on endogenous production and base-specific recognition of a 3′,3′ second messenger. However, our initial phylogenetic survey of STING alleles by gel-shift assay demonstrated that all functional ancestrally related STING alleles exhibited the ability to complex with the endogenous human second messenger 2′,3′ cGAMP (Figure 5.1). Indeed, ITC measurements with nvSTING confirmed a specific high affinity interaction with 2′,3′ cGAMP (<1 nM) (Figure 5.8a; Table 5.2). Understanding the evolutionary origins of the STING–2′,3′ cGAMP interaction might reveal the conserved structural principles underlying specific recognition of 2′,3′ cGAMP. Therefore, the crystal structure of the nvSTING–2′,3′ cGAMP complex was solved. Crystals of nvSTING–2′,3′ cGAMP grew in a unique crystal form distinct from each of the other six nvSTING structures, and the structure was determined by molecular replacement using a monomer core derived from the nvSTING–3′,3′ cyclic AMP-GMP structure as an initial search model.

Surprisingly, the structure of nvSTING–2′,3′ cGAMP (2.1 Å) reveals a partially rotated conformational state distinct from all previous nvSTING–3′,3′ cdN structures. In this partially rotated 2′,3′ cGAMP bound state, the apical wings of each STING monomer are rotated by ~15° from the *apo* positions, in comparison to the ~26° of rotation observed in the 3′,3′cdN-bound structures (Figure 5.8b–5.8d). Remarkably, the partially rotated state observed in the nvSTING–2′,3′ cGAMP structure is essentially identical to the rotation observed in the hSTING–2′,3′ cGAMP structures (Figure 5.7b). [21,94] Although nvSTING and hSTING are only 29% identical at the amino acid level, and exhibit significant structural differences when bound to 3′,3′ cdNs, the 2′,3′ cGAMP-bound structures are nearly superimposable (1.3 Å main-chain RMSD). This observation indicates that the unique chemical nature of 2′,3′ cGAMP, rather than the protein allele itself, dictates the wing pitch conformation of STING.

These results demonstrate that the unique partially rotated conformation adopted by hSTING when bound to 2′,3′ cGAMP is conserved in a distant STING ortholog. To test whether such

structural conservation is accompanied by functional conservation, we asked whether nvSTING could successfully replace hSTING in human cells. As noted above, nvSTING lacks the long unstructured carboxy-terminal tail (CTT) previously shown to be required for hSTING to recruit the critical downstream TBK1 and IRF3 signaling components.[72,95] Consistent with its lack of a CTT, WT nvSTING is unable to activate interferon-β luciferase signaling upon stimulation by human cGAS (Figure 5.8e). Remarkably however, appending the hSTING CTT (residues 341–379) to nvSTING is sufficient to permit low-level interferon-β stimulation in human cells. Together, these results demonstrate the minor changes required to adapt a STING allele to modern 2′,3′ cGAMP interferon signaling and reveal a possible path for cGAS-STING function evolution.

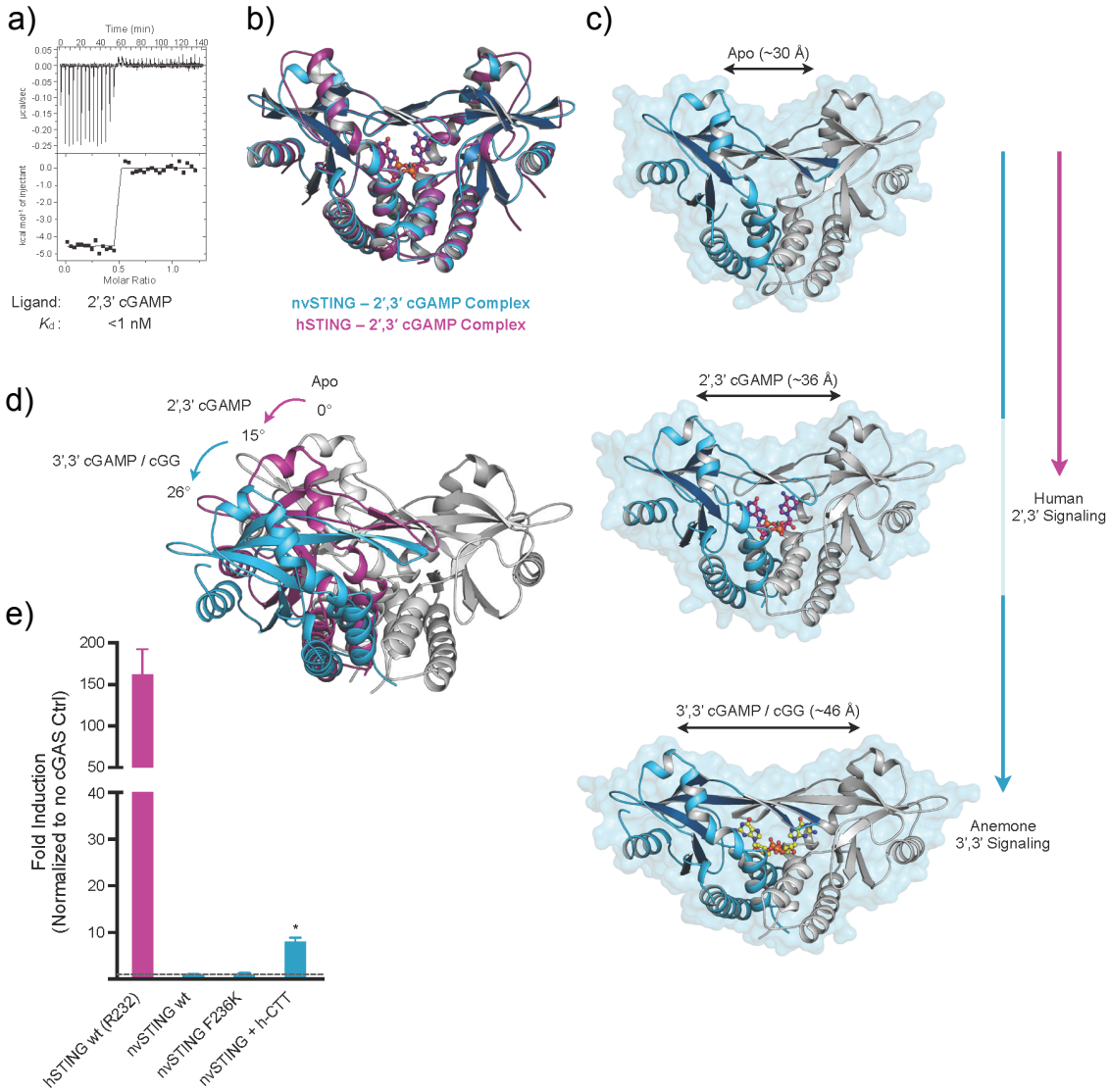


**Figure 5.8: nvSTING adopts different bound conformations depending on the cdN**
*A) ITC measurements of nvSTING affinity for 2′,3′ cGAMP. ITC data are representative of at least three independent experiments. (B) Structural overlay of nvSTING–2′,3′ cGAMP complex (blue) and hSTING–2′,3′ cGAMP complex (magenta) (PDB: 4KSY) demonstrating the monomer wing rotation and core cdN-interacting portion of nvSTING and hSTING are unchanged. (C) Structural comparison of nvSTING structures in various ligand-bound complexes. cdN ligands lock STING in alternative conformations as measured by the distance between the apical monomer wing domains (monomer 1 in blue, monomer 2 in gray). nvSTING–3′,3′ cdN interactions result in complete monomer rotation (blue vertical line), while primary hSTING–2′,3′ cGAMP signaling traps a partially rotated intermediate structure (magenta line). (D) Endogenous anemone 3′,3′ second messengers trigger an ~26° rotation in monomer wing domains from the apo state (apo gray, 3′,3′-bound in blue), while human 2′,3′ cGAMP traps an ~15° rotated structural intermediate in the nvSTING and hSTING structures (2′,3′-bound in magenta). (E) Cell interferon-β luciferase, using indicated STING plasmids stimulated with human cGAS overexpression. Error bars represent the SE of the mean of at least three independent experiments (*p < 0.002).*

## 5.8 Discussion

In vertebrates, the cGAS-STING signaling pathway enables cells to initiate interferon-mediated gene expression in response to foreign DNA from viruses, bacteria, and tumors. Although STING homologs are present in diverse animals, it has been unclear whether activation by cdNs is an evolutionarily conserved function of STING, and if so, whether STING first evolved to respond to exogenous cdNs produced by bacteria, or endogenous cdNs produced as second messengers. To determine the evolutionary origins and conserved signaling principles of cGAS-STING immunity, the pathway was investigated in divergent metazoan species. By surveying phylogenetically distant STING variants, cdN binding was found to be a deeply conserved function of STING that was likely present >500 million years ago in the common ancestor of humans and Cnidaria. This observation implies that the ability to respond to cdNs via STING may have been a conserved property of the common metazoan ancestor. Furthermore, this data suggests that most features of modern innate immunity, including the evolution of interferons, followed the evolution of the STING–cdN interaction. In addition to interferon signaling, STING–cdN recognition causes downstream activation of the NF-κB pathway[15], and ancestral NF-κB signaling[96] may explain evolutionary conservation of the cGAS-STING pathway prior to interferon-based immunity.

To investigate whether cGAS, like STING, is evolutionarily ancient, experimentation started with the most divergent species in which we observed a robust STING–cdN interaction, the sea anemone *N. vectensis*. Although *N. vectensis* contains several candidate cGAS-like genes, these enzymes all lack defining features of human cGAS, including the Zn-ribbon that is critical for binding to dsDNA. Thus, signaling assays were essential to identify a true functional ortholog of cGAS capable of activating human STING. The identification of nv-

cGAS implies that the cGAS-STING signaling pathway was also already present >500 million years ago. Recombinant purified nv-cGAS is not active in vitro and the biochemical trigger that activates nv-cGAS remains unknown. However, the ability of nv-cGAS to signal in transfected human cells suggests that, analogous to DNA-dependent activation of human cGAS, an additional cellular co-factor or activator is able to induce nv-cGAS enzymatic activity.

Anemone cGAS appears to differ from its human ortholog in the regiochemistry of its cdN product. By taking advantage of human STING variants that respond specifically to 2′,3′ cGAMP, as well as a 3′,3′ cyclic AMP-GMP-specific phosphodiesterase from *V. cholerae*, it is highly likely that nv-cGAS produces 3′,3′ cyclic AMP-GMP. This product is distinct from the cGAMP produced by vertebrate cGAS enzymes and suggests that 2′,3′ cGAMP is a recent vertebrate innovation, though it is possible that uncharacterized enzymes in anemone or other invertebrates might also make a 2′,3′-linked cdN.

This coherent series of seven nvSTING structural states now allow collective placement of available hSTING and mouse STING structures to illustrate a dynamic movement of STING monomers rotating against each other, analogous to the opening and closing of a butterfly's wings. These crystallographic studies captured two distinct *apo* forms of nvSTING. In one of these forms, *apo* nvSTING adopts an unrotated ''closed'' conformation similar to a previously published mouse STING *apo* structure.[83] The other *apo* nvSTING structure has a rotated ''open'' wing pitch and exhibits a loosely organized lid, similar to published *apo* hSTING structures.[84-88] The most parsimonious explanation of these results is that inactive *apo* STING in all species samples diverse conformations. In contrast to the existing model of hSTING activation, in which ligand binding induces a transition from a rotated-*apo* to a compact-bound conformation, this chapter supports the hypothesis that *apo* STING is structurally dynamic and that the primary function of cdN binding is to stabilize a specific STING conformation.

A unique feature of cGAMP thus appears to be its ability to stabilize an intermediate, partially rotated form that is not captured by canonical 3′,3′ cdNs (Figure 5.8). Importantly, these results demonstrate that this specific structural state, and thus the associated ability of STING to distinguish bacterial (3′,3′) from mixed-linkage (2′,3′) cdNs, is not a recent adaptation of hSTING but is instead a fundamental and deeply evolutionarily conserved feature of the STING architecture, offering an explanation for the universal signaling potential of 2′,3′ cGAMP.

# 6. ELUCIDATING THE SIGNALING SPACE OF THE cGAS-cdN-STING AXIS AND FUTURE DIRECTIONS

## 6.1 Background

As currently known, STING's molecular "signaling space" (Figure 6.1) comprises at least three transducers that are key players in the inflammatory phenotype associated with STING activation: IRF3 (leading to production of interferon-β), NF-κB, and STAT6.[15,23] This immunological profile is profound as evident from patients who suffer from a single point mutation in STING, rendering STING constitutively active. Incidents of this genetic disease was first described in 2014 and called SAVI (**S**TING-**a**ssociated **v**asculopathy with onset in **i**nfancy).[97] Those with SAVI suffer from a chronic autoinflammatory disease leading to clinical characteristics, including but not limited to: fever, fibrosis, nail dystrophy, violaceous plaques, and lesions. Patients with SAVI have a high mortality rate and tend not to live past childhood. Therefore, a molecular understanding of STING signaling transduction would be of value for treatment.

Furthermore, cyclic dinucleotides have recently gained heightened interest for their ability to ablate tumors in mouse models. Mice lacking STING are deficient in spontaneous $CD8^+$ T cell priming against tumors. Yet when STING is retained and present in a tumor environment, cyclic dinucleotides are sufficient to induce dendritic cell activation and interferon-β production, leading to tumor destruction.[98,99] However, the mechanism for innate immune recognition and activation against immunogenic tumors is not understood. Thus, a molecular understanding of STING signaling would also be useful for developing cyclic dinucleotide-based therapeutics intended for cancer immunotherapy.

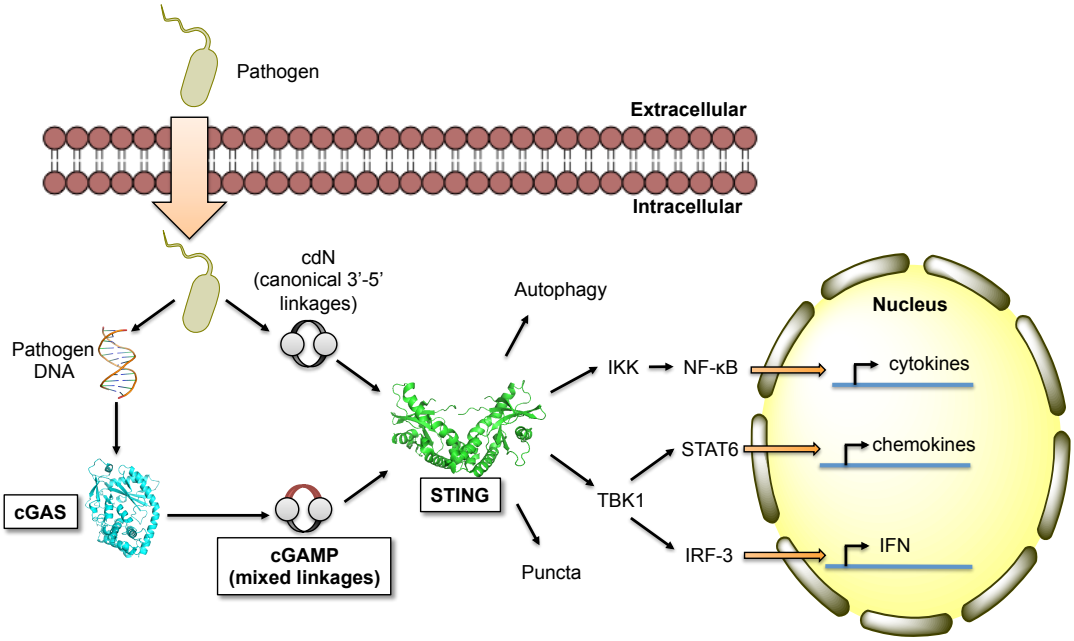



**Figure 6.1: Current model for STING signaling**

The current model for mammalian STING signaling indicates that its C-terminal domain is critical for eliciting interferon.[72,95] Once STING is activated by cdNs, STING recruits TANK-binding kinase 1 (TBK1) and interferon regulatory factor 3 (IRF3). STING, TBK1, and IRF3 are then subsequently phosphorylated; however, multiple kinases are implicated and the causative kinases for each substrate are unclear.[95,100] IRF3 phosphorylation results in auto-dimerization, which then translocates into the nucleus and elicits interferon expression. A hSTING S366A (S365A in mouse) mutation has been shown to completely ablate interferon induction by abolishing IRF3 activation yet preserves TBK1 activation.[95]

Less is known about how STING induces NF-κB and STAT6, although IKK and TBK1 have been implicated as crucial kinases. NF-κB elicits TNF-α, which is likely the leading cytokine responsible for STING-induced inflammation. STAT6 likely plays a more attenuated role but once activated, increases the expression of multiple downstream chemokines including: CCL2, CCL20, and CCL26.[23] Interestingly, it has been reported that the same hSTING S366A mutation that ablates interferon expression, preserves (and perhaps enhances) NF-κB induction.[100] The effect of the hSTING S366A mutation or CTT truncation on STAT6 induction has never been reported.

STING has also been shown to induce autophagy and form puncta when activated by cyclic dinucleotides, but even less is understood about these phenomena.[101-105] Notably, autophagic processes are activated by the cGAS-STING axis when macrophages are infected with *Mycobacterium tuberculosis*.[105] The molecular mechanism by which autophagy becomes activated are unknown, but again highlights the importance of the cGAS-cdN-STING axis in regulating host-pathogen responses.

This chapter describes efforts to examine the signaling space of the cGAS-cdN-STING axis and determine the molecular architecture necessary for diverse STING signaling. For the purpose of this chapter, STING signaling is defined as being comprised of 5 unique readouts: induction of interferon-β, NF-κB, STAT6, autophagy (i.e. formation of LC3 puncta), and formation of STING puncta. As noted in Chapter 5, many phylogenetic homologs of STING that have been demonstrated to be competent for cdN binding do not retain the C-terminal tail (CTT), which is necessary for interferon induction. In fact, many of these animals do not rely on interferon induction for immunity, suggesting that STING has a more ancient function than interferon signaling. In order to elucidate such function, cyclic dinucleotide signaling was studied in *Nematostella vectensis*. Human and mouse STING constructs containing the S366A (or S365A for mouse) mutation or CTT truncation were studied for interferon-β, NF-κB, and STAT6 induction. The same constructs were also examined for their ability to induce autophagy and form STING puncta using immunofluorescence. Lastly, CRISPR/cas9 was employed to generate STING S365A and CTT truncation mice in order to further understand the effects of STING-induced interferon signaling on host-pathogen interactions and cdN-based cancer immunotherapy.

### 6.2 cGAMP elicits a potent immune response in *Nematostella vectensis*

*Nematostella vectensis* has only one *STING* gene in its genome, which is expressed throughout embryonic development (Figure 6.2).[106] As described previously (Chapter 5; Figure 5.1c), anemone STING lacks the CTT necessary for eliciting interferon and also does not harbor any bioinformatically identifiable interferon genes. Therefore despite being expressed and capable of binding cyclic dinucleotides with high affinity (Figure 5.1a; Figure 5.7d; Figure 5.8a), nvSTING's capacity to elicit an innate immunological response when exposed to cyclic dinucleotides was unknown.

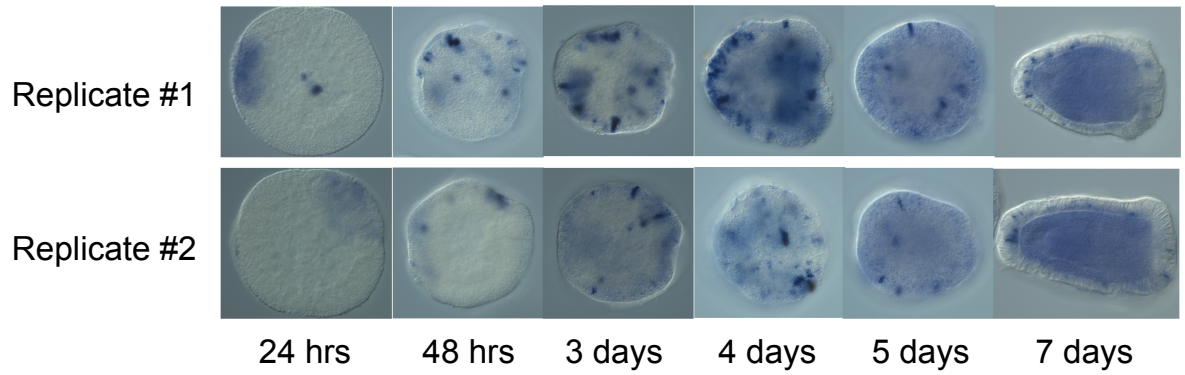



**Figure 6.2: STING is expressed during *Nematostella* embryonic development**
*In situs were performed with a nvSTING selective RNA probe. [Data acquired by Marc Servetnick]*

In order to address this question, ~12 hour old *Nematostella vectensis* embryos were incubated with cyclic di-GMP or cGAMP for a 24 hour period (Figure 6.3). Following the 24 hour overlay, embryos treated with cyclic di-GMP were slightly more gastrulated compared to controls. Their motility within the 96-well plate was also attenuated compared to the control wells. On the other hand, the wells with cGAMP had substantially more embryo debris, and the embryos exhibited significantly more rupture compared to the control and cyclic di-GMP treated wells. In addition, their motility within the 96-well plate was significantly faster compared to the control and cyclic di-GMP treated samples. Primary polyps were also exposed under the sample experimental conditions, but no visible phenotype was apparent (Figure 6.4). The remarkable differential in phenotypes of cyclic di-GMP exposed embryos versus cGAMP-exposed embryos demonstrates that a change in cdN regiochemistry, nucleobase identity, or both has a profound effect on *Nematostella* embryonic development.

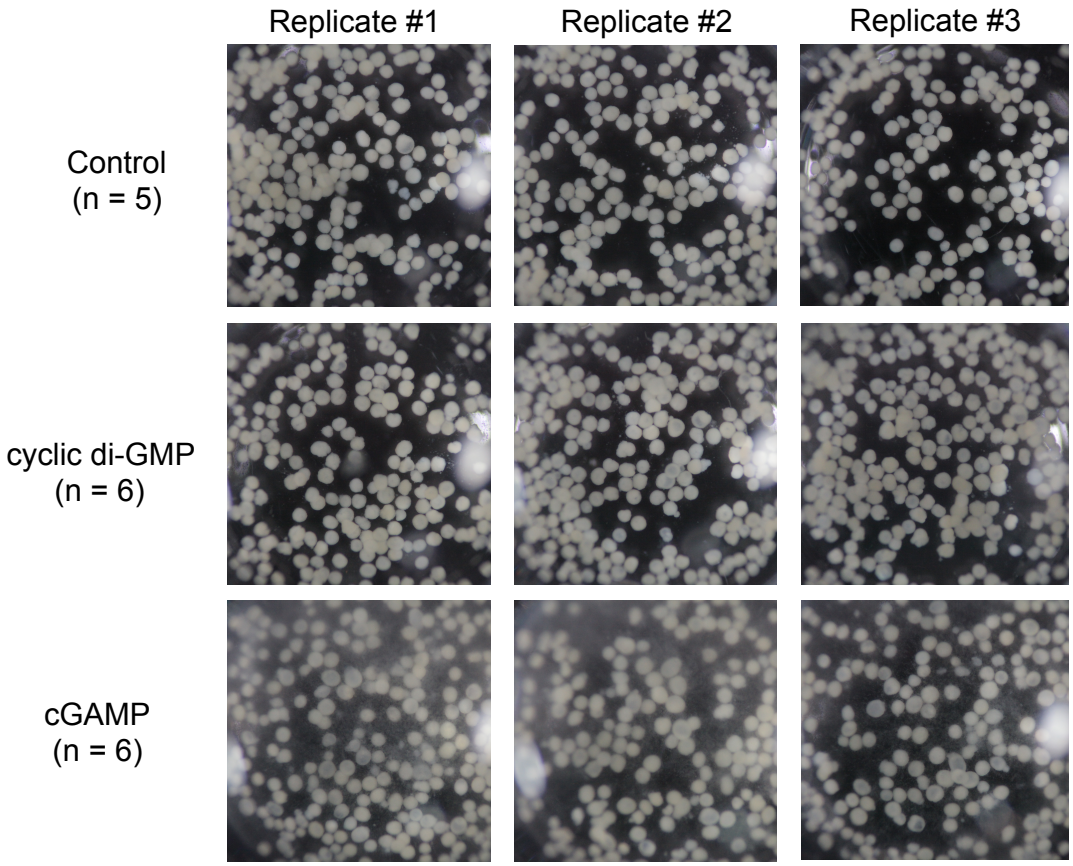



**Figure 6.3: Cyclic dinucleotide overlay on 12-15 hour old *Nematostella vectensis* embryos**
*Images were acquired ~24 hours after incubating in a 1 mM cdN salt water solution. The cloudiness in the cGAMP fields is due to debris from embryo rupture.*

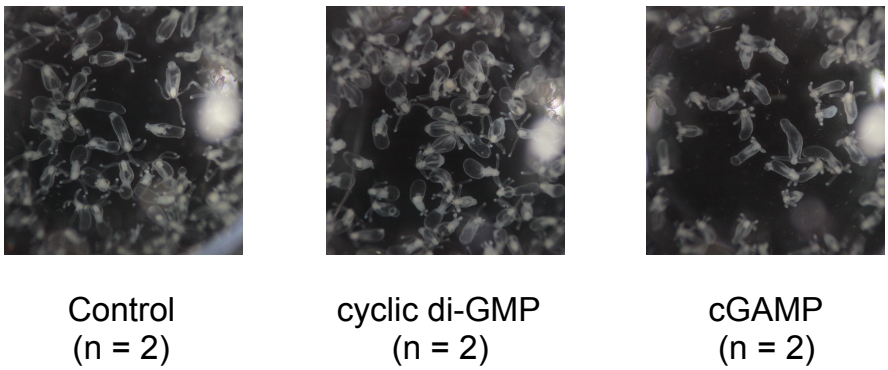



**Figure 6.4: Cyclic dinucleotide overlay on *Nematostella vectensis* primary polyps**
*Images were acquired ~24 hours after incubating in a 1 mM cdN salt water solution. No remarkable phenotype is apparent.*

While the overlay experiment implies a remarkable reaction to cyclic dinucleotides, it does not address whether cdNs are actually getting into the embryo. Therefore, cyclic di-GMP and cGAMP were microinjected into <2 hour old embryos, and their phenotypes were monitored for ~20 hours. Both the control and cyclic di-GMP injected embryos cleaved normally and proceeded thru gastrulation into the planula stage of development. However, the development of cGAMP treated embryos was arrested and exhibited no cleavage after 2-4 hours. A small number of embryos were ruptured at 4 hours; however by 19 hours, all of the embryos were completely destroyed. Taken in concert with the cdN overlay data, cGAMP has a powerful intracellular effect on *Nematostella* embryonic development that is governed by its regiochemistry and/or nucleobase identity.

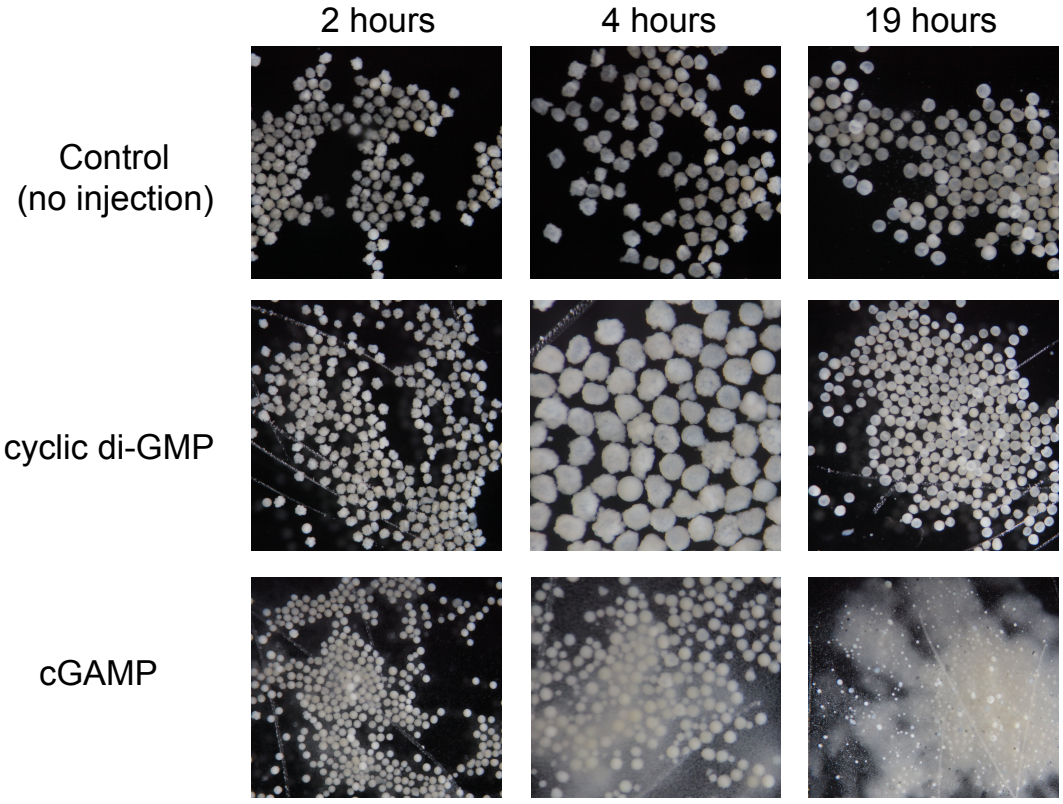
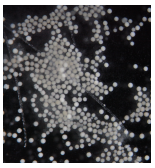



**Figure 6.5: Cyclic dinucleotide microinjected into <2 hour old *Nematostella vectensis* embryos**
*Images were acquired during a 20 hour period after exposure. Magnification is variable for each image and is meant to emphasize morphology. The cloudiness in the cGAMP fields is due to debris from embryo rupture.*

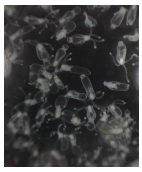
Despite the dramatic phenotype associated with the cGAMP treated embryos, the effect does not mandate that it is driven by an "innate immune" response. Consequently, the transcriptomics of the cdN treated embryo and polyp samples were analyzed with the hypothesis that there would be a significant changes in immune-related expression (Table 6.1). Both the cGAMP treated embryos and polyps exhibit a large upregulation of innate

immune associated genes. cGAMP treated embryos have larger fold changes in genetic expression compared to polyps, and their differentials in immune-related genetic expression include increases in: TNF receptor, MAVS, perforin, argonaute, and STING. It is also notable to mention that the cGAMP treated embryos exhibit a large increase in expression of chitinase and myosin, which might account for the embryo rupture and increased embryonic motility, respectively. While the changes in genetic expression are less dramatic in the polyps, there does appear to be an immune-related signature. cGAMP treated polyps demonstrate increased levels of TNF receptor, MDA-5, IRF2, argonaute, STING, MAVS, and OAS. It is important to emphasize that these are genes that were identified only through their homology to well-studied mammalian homologs. Activity and function of many of these proteins in *Nematostella* has never been shown. In addition, many genes that are upregulated in cGAMP treated embryos have no known protein homolog with an annotated function, suggesting that there may be broader and currently unknown effects of cGAMP on *Nematostella* physiology.

As expected, cyclic di-GMP treated polyps show a much different transcriptomic signature. Fewer genes were upregulated with little to no overlap with the cGAMP dataset. All of the immune related genes indicated in Table 6.1 were not appreciably changed within the cyclic di-GMP dataset. Therefore, both the phenotype and transcriptomics support a profound impact of cGAMP on *Nematostella* physiology that appears to be related in part on innate immune signaling.

Embryos 

| # | Gene | Fold change ($log_2$) |
|---|---|---|
| 1 | chitinase | 8.9 |
| 10 | TNF receptor | 6.9 |
| 11 | Ras like GTPase | 6.8 |
| 16 | MAVS | 6.3 |
| 21 | myosin | 5.8 |
| 23 | perforin | 5.8 |
| 24 | argonaute | 5.7 |
| 66 | STING | 4.3 |

Polyps 

| # | Gene | Fold change ($log_2$) |
|---|---|---|
| 2 | rhysin | 2.6 |
| 8 | chitin binding | 2.5 |
| 13 | TNF receptor | 2.5 |
| 15 | MDA-5 | 2.5 |
| 24 | IRF2 | 2.5 |
| 29 | argonaute | 2.5 |
| 34 | STING | 2.5 |
| 37 | MAVS | 2.5 |
| 43 | OAS | 2.5 |

**Table 6.1: RNA-seq of select upregulated transcripts in cGAMP treated samples compared to control**

While little is known about Cnidaria innate immunity, the data presented here suggests that cGAMP signaling is broadly conserved in lower organisms. With that said, it is curious that

*Nematostella* cGAS produces a 3′,3′ cdN (Chapter 5), yet a dramatic phenotype is only associated with a 2′,3′ cdN. It is tempting to speculate that perhaps another enzyme in *Nematostella* generates a 2′,3′ cdN. Even though all of the putative *Nematostella* cGAS homologs have been tested for cdN activity (Figure 5.3), it is possible that a necessary effector was not available within the cellular milieu for catalytic activity, rendering a possible 2′,3′ cdN synthase as a false negative. Furthermore, mammalian cGAS retains considerable homology to OAS, which imparts its 2′-5′ synthase activity to cGAS. Since *Nematostella* cGAS never gained this function, perhaps *Nematostella* OAS developed the capacity to create a 2′,3′ cdN. This hypothesis has never been tested. Further experimentation to follow up on the observations presented in this chapter include repeating the above *Nematostella* experiments using 3′,3′ cyclic AMP-GMP instead of cyclic di-GMP as well as performing a morpholino knockdown of nvSTING followed by a cGAMP overlay in order to confirm that embryo rupture is dependent on STING signaling.

### 6.3 nvSTING forms puncta in response to 2′,3′ cGAMP

Despite having a powerful impact on *Nematostella* development, the direct molecular consequence of cdNs and, in particular, its binding to STING is unclear. As previously reported in Chapter 5, nvSTING is incapable of eliciting interferon in response to cyclic dinucleotides within the HEK293T environment (Figure 5.8e). *Nematostella* STING's capacity to elicit NF-κB has also been studied under similar conditions (i.e. NF-κB inducible luciferase reporter in HEK293T cells), but no increase in NF-κB levels was observed (Figure 6.6).

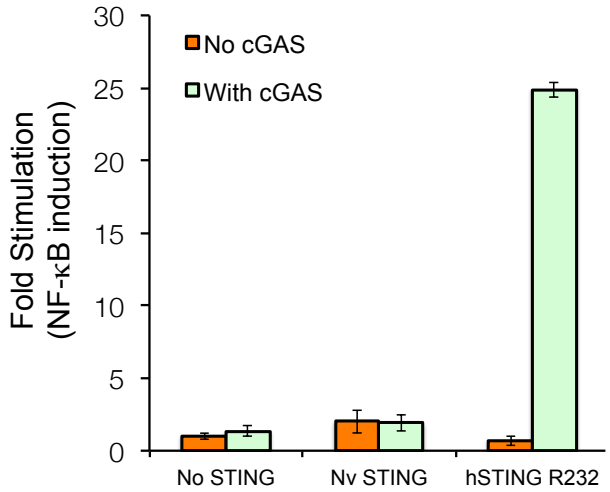



**Figure 6.6: Testing nvSTING stimulated with human cGAS overexpression using a NF-κB inducible luciferase reporter**

Using a C-terminal HA tagged version of nvSTING and a transfectable LC3-GFP construct, the effect of cyclic dinucleotide binding on the formation of STING puncta and LC3 puncta

was studied using immunofluorescence (Figure 6.7). Both nvSTING alone and nvSTING co-transfected with catalytically inactive human cGAS (E225A/D227A) do not exhibit any puncta formation, yet remarkably, nvSTING appears to still localize to the ER in a human-derived cell. In fact, co-transfection of nvSTING with human cGAS causes nvSTING to form distinct puncta. DncV was also tested with nvSTING in order to mimic a "more productive" nv-cGAS, but no STING puncta were observed. LC3 puncta were not observed under any condition.

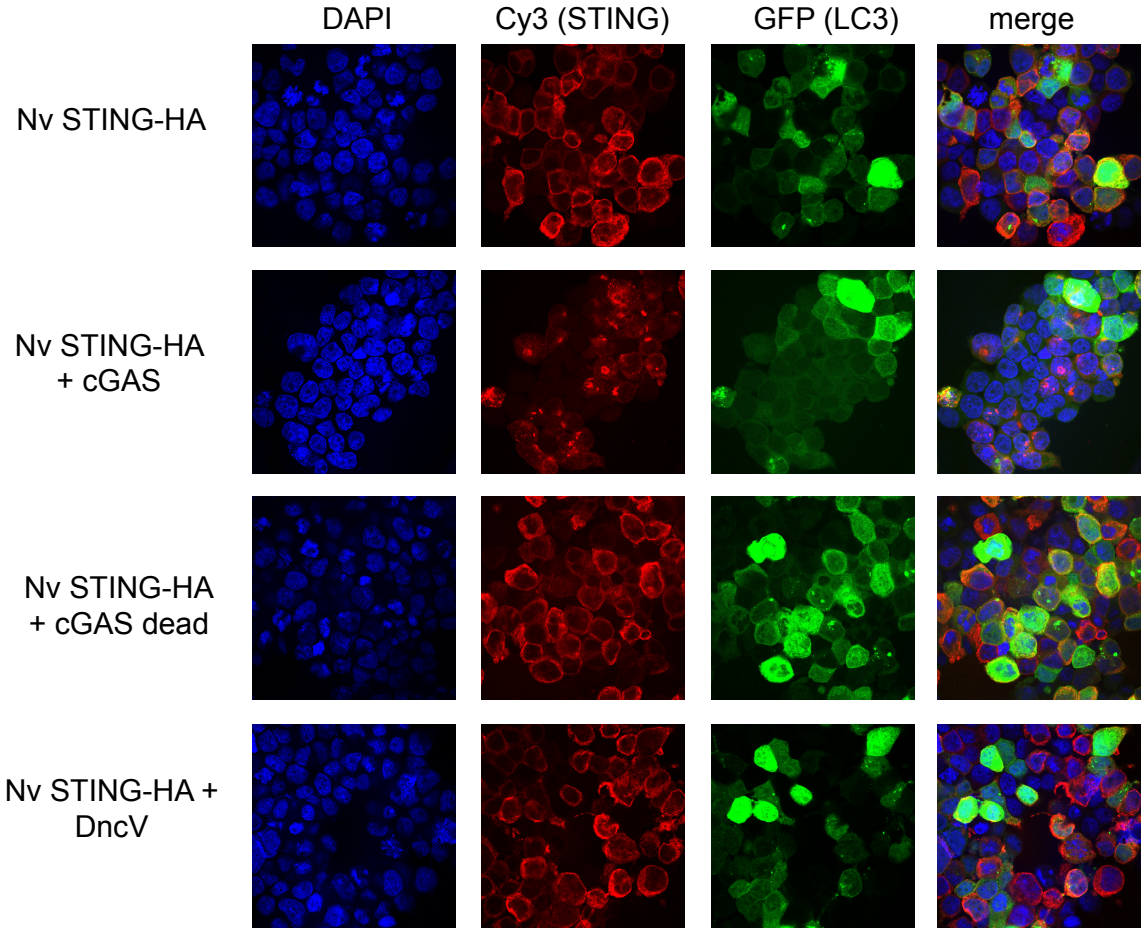



**Figure 6.7: Confocal microscopy images of nvSTING co-transfected with cdN synthases**
*Only hcGAS is capable of inducing nvSTING puncta.*

At present, induction of nvSTING is puncta by hcGAS is the only positive functional read out of nvSTING. These experiments demonstrate that STING puncta formation can be uncoupled from autophagy activation. In other words, puncta formation is not sufficient to trigger autophagy. This will be addressed in more depth in the next section, but it is certainly remarkable that puncta formation can be recapitulated using an anemone protein in a mammalian cell.

But again, it is curious to why puncta formation is only driven by the 2′,3′ cdN and not the 3′,3′ cdN, yet this observation is consistent with the results discussed previously (Chapter 6.2). One hypothesis is that 3′,3′ cdNs may signal "symbiosis" to the host, whereas 2′,3′ cdNs signal "pathogenesis". How these distinct signals arise and their downstream consequences are a mystery. However, nvSTING does adopt distinct conformations depending on whether it is bound to a 3′,3′ cdN or a 2′,3′ cdN (Chapter 5), suggesting a capacity for diverse signaling outputs and one that is witnessed in modern STINGs. Furthermore, it has been reported that *Nematostella* has closely associated bacterial consorts, many of which contain 3′,3′ cdN synthases in their proteome.[107] This would not be the first report of a modern innate immune pathway having a role in symbiosis within more primitive organisms.[108] **T**oll-**l**ike **r**eceptor (TLR) signaling has been implicated as a means for sensing and regulating bacterial colonizers in *Hydra*.[109] Gram-negative bacteria within *Drosophila* trigger the Imd (**im**mune **d**eficiency) pathway via peptidoglycan recognition proteins, which induce antimicrobial peptides, CaudaI (expression inhibitor of antimicrobial peptides), and Pirk (Imd negative regulator), thereby shaping the gut microbiota.[110-112] Lastly, the symbiotic relationship between *Vibrio fisheri* and *Euprymna scolopes* (bobtail squid) is perhaps the most extensively studied example of host-microbe interactions. Signaling through peptidoglycan and its associated proteins mediates microbe specificity and initiates the *Vibrio* colonization process, which leads to the development of the light organ in the squid.[113,114] Therefore, the role of the cGAS-STING axis within primitive organisms may also adopt a dual function.

Future experimental avenues to follow up the above work include co-transfecting nvSTING with nv-cGAS to verify that it phenocopies nvSTING + hcGAS. Testing nvSTING with WspR and DisA would also be valuable, if for no other reason to generate a complete dataset of the response of nvSTING to each known cdN. Although the localization of STING upon cdN binding is well-established, introducing ER and Golgi specific markers might also be useful. Finally, it is unlikely that nvSTING will be able to activate STAT6, but testing nvSTING using a full synthase panel along with the STAT6 luciferase reporter would complete the signaling dataset.

### 6.4 Dissection of distinct downstream signaling pathways of mammalian STINGs

Mammalian STING has multiple downstream signaling outputs, but the necessary molecular architecture to elicit each signal and how each specific signal affects host immunity is unclear. Therefore, an approach to dissect STING signaling has been developed. Signaling outputs are currently defined as interferon-β, NF-κB, STAT6, autophagy (i.e. formation of LC3 puncta), and formation of STING puncta. STING's CTT has been shown to be crucial for eliciting interferon[72,95,100], in particular human and mouse STING constructs containing the S366A (or S365A for mouse; generically referred to as S36#A below) mutation completely ablate interferon signaling. However, its effect on NF-κB, STAT6, autophagy, and STING puncta is unknown. Therefore, a CTT truncation mutant and the S36#A mutation were assayed for all 5 signaling outputs.

Using a luciferase based IFN detection assay, the CTT truncation to both mouse and human STING completely ablated interferon signaling (Figure 6.8). The S36#A mutation was sufficient to nullify interferon expression as well (data not shown for hSTING S366A). The truncation and S36#A point mutation was also sufficient to abolish STAT6 induction (Figure 6.9).

In contrast, NF-κB signaling was preserved and enhanced (Figure 6.10) with the S36#A mutation in both the human and mouse STING constructs (data not shown for hSTING S366A). This indicates that the conditions necessary for NF-κB signaling differ from that of IFN and STAT6 signaling. It is unclear why NF-κB signaling increases in with the S36#A mutation, but one report indicates that S36#A is important for protein degradation.[100] An accentuated NF-κB signal could be a consequence of STING buildup in the cell, leading to more NF-κB induction. Additionally, the truncated hSTING construct is capable of eliciting NF-κB but at a more attenuated level compared to wild-type hSTING. Interestingly, the truncated mSTING does not share the same capability even though they are truncated in identical locations. One explanation for the discrepancy is that while wildtype mouse STING still retains its ability to signal in a HEK293T cell, truncating the construct might be too much of a strain for other parts of the human cellular machinery to elicit a signal above background.

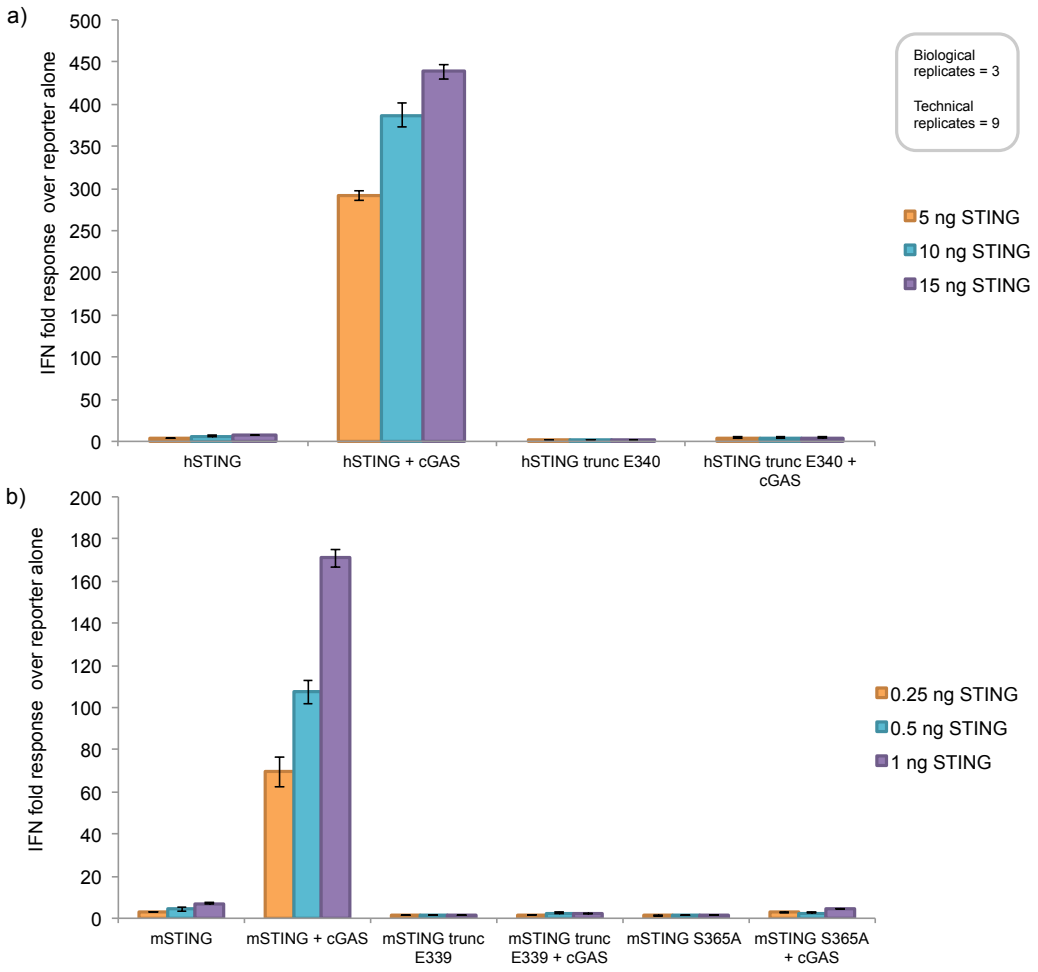



**Figure 6.8: Interferon-β inducible luciferase reporter using indicated STING plasmids stimulated with human cGAS overexpression**

*STING was transfected with the corresponding amounts of plasmid: (a) human STING (b) mouse STING.*

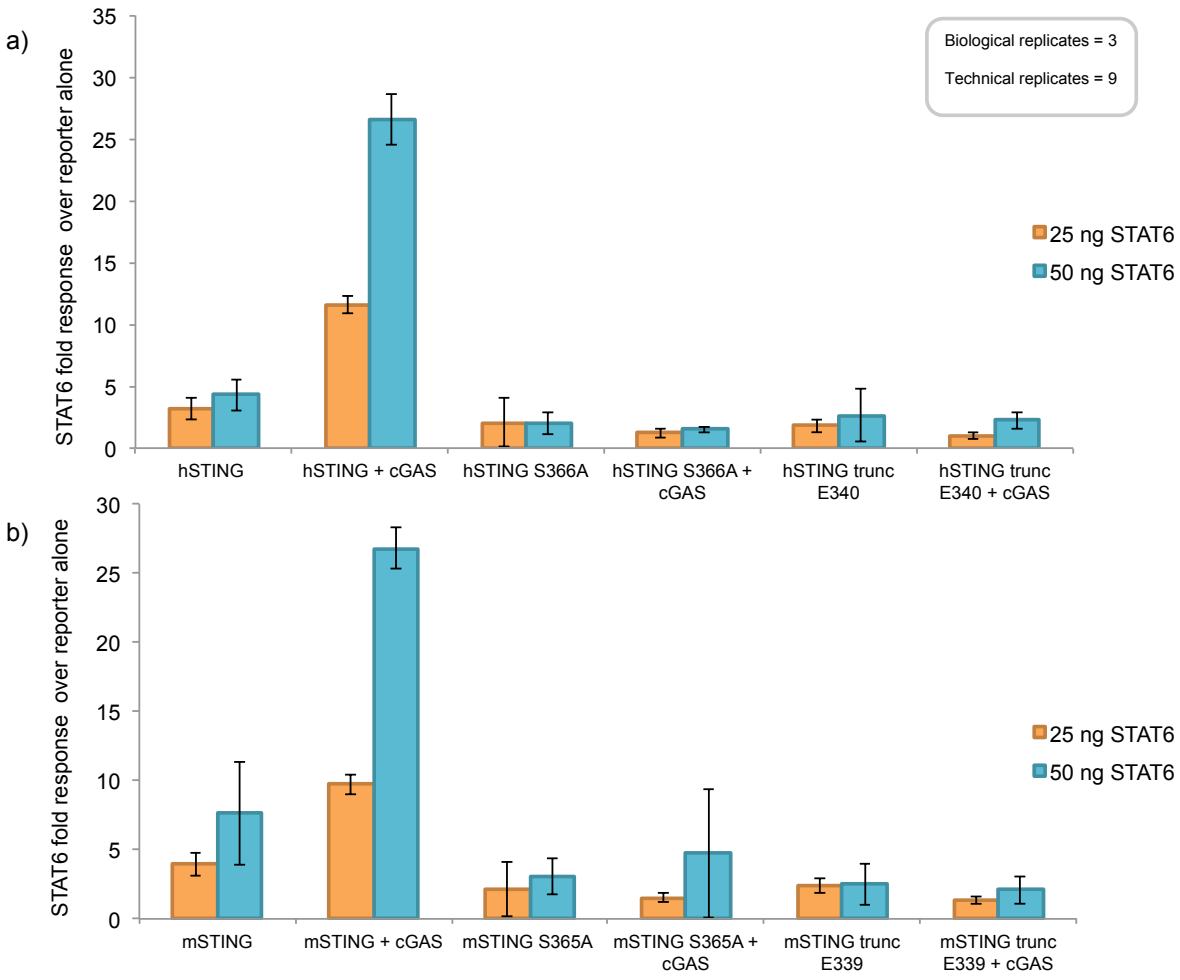



**Figure 6.9: STAT6 inducible luciferase reporter using indicated STING plasmids stimulated with human cGAS overexpression**

*STAT6 is not expressed in HEK293T cells so STAT6 was transfected with the corresponding amounts of plasmid: (a) human STING (b) mouse STING.*

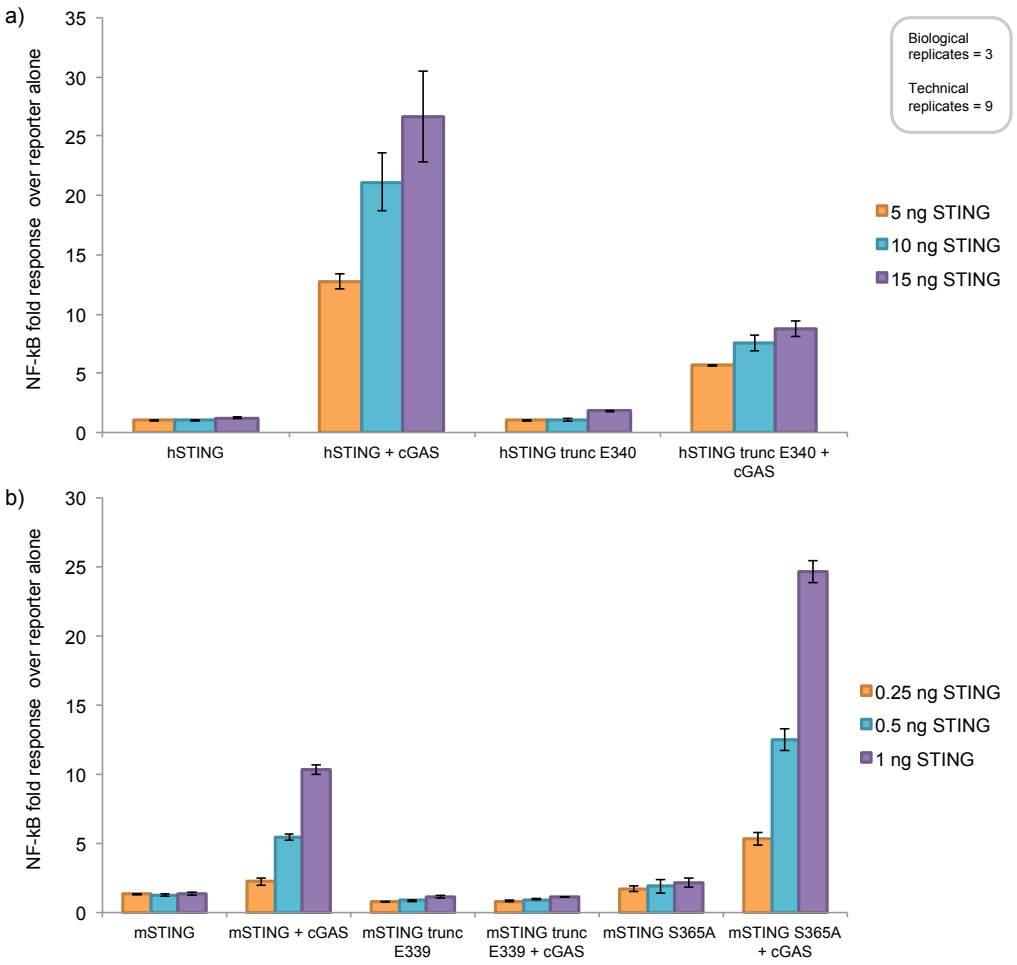



**Figure 6.10: NF-κB inducible luciferase reporter using indicated STING plasmids stimulated with human cGAS overexpression**
*STING was transfected with the corresponding amounts of plasmid: (a) human STING (b) mouse STING.*

These results indicate that NF-κB signaling can be dissected from IFN and STAT6 signaling using a S36#A point mutation and to some extent, completely truncating the CTT. In order to determine if a similar dissection was visible for STING-induced autophagy and/or the formation of STING puncta, the effect of these STING mutations were studied using immunofluorescence and confocal microscopy (Figure 6.11). Remarkably, neither the S365A point mutation nor CTT truncated mouse STING had significant impact on the formation of LC3 puncta or STING puncta, suggesting that the CTT has little control over these signaling pathways.

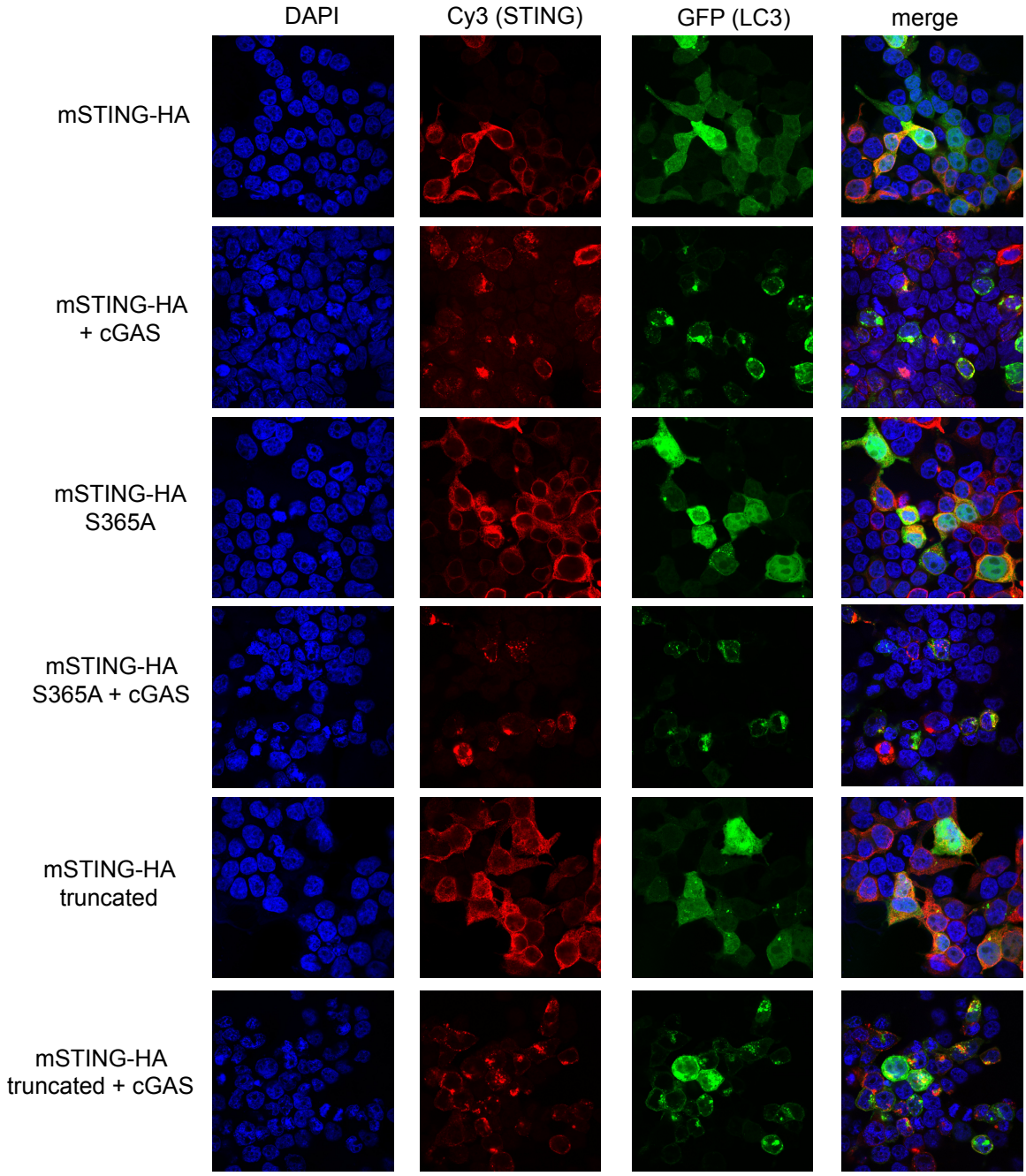



**Figure 6.11: Confocal microscopy images of wild-type, S365A, and truncated mSTING co-transfected with hcGAS**
*A transfectable LC3-GFP construct was also introduced in order to view LC3 puncta.*

Holistically, these results indicate that the S36#A mutation is sufficient to selectively dissect the IFN and STAT6 signaling pathways of STING, while leaving NF-κB signaling, autophagy, and STING puncta formation preserved. Future steps for this section include examining the effect of other cdN synthases on signaling output as well as identifying STING point mutations that selectively ablate NF-κB, STAT6, autophagy, and/or STING puncta formation while leaving interferon signaling intact. Examining TBK1 and IRF3 phosphorylation would also be useful. Once the minimum molecular architecture to elicit each signal is determined, these modifications can then be introduced into a mouse model to study how discrete STING signals affect host infection, cancer immunotherapy, and microbiota colonization.

### 6.5 Development of mSTING S365A and E339 truncated mouse models using cas9

In order to study how discrete STING signals impact host physiology, the mSTING S365A point mutation and ΔCTT were independently introduced into a mouse model using CRISPR/cas9. CRISPR/Cas9 is a highly efficient method for generating mice harboring 'knock-ins' of small point mutations.[115,116] The general strategy is illustrated in Figure 6.12, using the S365A knock-in as an example. For creating the mSTING ΔCTT mouse line, a premature stop codon was inserted before the mSTING CTT.

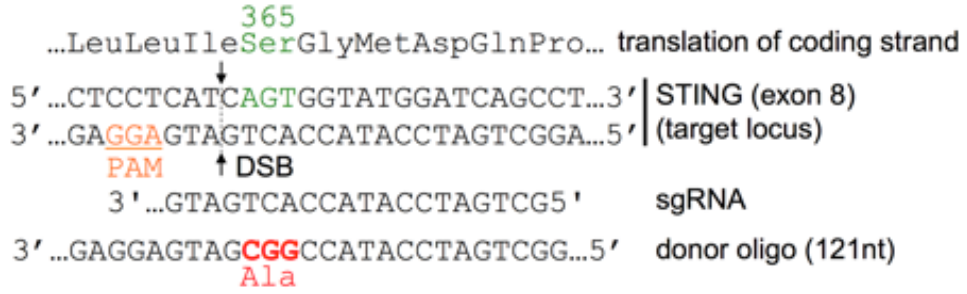



**Figure 6.12: Strategy for generation of STING S365A knock-in mice using CRISPR/Cas9**
*A double-strand break (DSB) is introduced into the STING exon 8 locus where indicated by delivery of the Cas9 nuclease into fertilized mouse eggs (zygotes) along with a sgRNA that targets Cas9 to the locus. Homology-directed repair of the DSB mediated by a co-delivered donor oligo introduces an AGT (serine) to GCC (alanine) substitution at codon 365 of STING.*

Fourteen mice were recovered from the CRISPR/Cas9 S365A knock-in injection. Of these mice, ten were wild type on both alleles, three mice had clear modifications, and one mouse was indeterminate as analyzed by sequencing. Of the three mice that had modifications, two had modifications at the desired codon, whereas one mouse had only modifications outside of

the desired codon. A single mouse was perfectly modified on both alleles with the S365A mutation and no other outstanding mutations in its proximity (Figure 6.13).

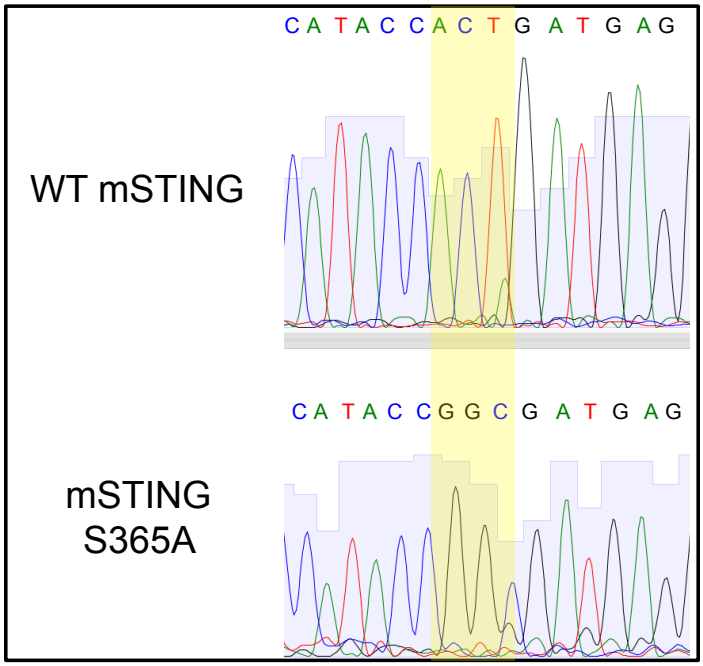



**Figure 6.13: Genotyping of cas9 modified mSTING S365A biallelic mice**
*Modified region is highlighted in yellow. mSTING is encoded on the reverse strand: TCA (serine) modified to CGG (alanine).*

Thirteen mice were recovered from the CRISPR/Cas9 ΔCTT knock-in injection. Of these mice, twelve were wild-type on both alleles and one mouse had clear modifications. Unfortunately, the one mouse with modifications also had editing outside of the desired codon. It is unclear whether the undesired mutations are on one or both alleles. Therefore, the modifications are currently being purified through the germ-line and an additional CRISPR/Cas9 ΔCTT knock-in injection has been performed. Those mice have not been analyzed yet.

Immediate future steps for this section entail purifying the desired mutations through the germline. Once there is confirmation of "clean" knock-in mice, whole-exome sequencing will be performed to ensure there are no off-target coding mutations. These mice will then be used to study how ablating STING interferon signaling affects host infection, cancer immunotherapy, and microbiota colonization.

## 6.6 Concluding remarks

In this dissertation, I have presented work that significantly contributes to the field of cyclic dinucleotide signaling, spanning from bacteria to humans. In Chapter 2, I described my work synthesizing natural and unnatural cyclic dinucleotides as well as the development of aptamer-based cyclic dinucleotide fluorescent sensors. Only purine-based cyclic dinucleotides have been discovered endogenously to date, and it is an open question to whether any pyrimidine-based cyclic dinucleotides exist naturally. CTP and UTP concentrations in *E. coli* are maintained within the same order of magnitude (low millimolar) as ATP and GTP, suggesting that endogenous concentrations are high enough for an enzyme to be able to catalyze their synthesis.[117] If pyrimidine-based cdNs do not exist naturally, it is intriguing to wonder why nature would select only purines for cdN biosynthesis.

In Chapter 3, I demonstrated how aptamer-based cdN fluorescent sensors were utilized for biological discovery. In particular, a unique subclass of the cyclic di-GMP riboswitch was discovered that is selective for cyclic AMP-GMP. The existence of cyclic AMP-GMP within *Geobacter sulfurreducens* was also confirmed by analyzing its cellular extract with MS-based analysis. Cyclic AMP-GMP's unique role in *Geobacter* physiology was also hypothesized, especially in how it distinguishes itself from cyclic di-GMP signaling. The role of cyclic AMP-GMP in *Geobacter* exoelectrogenesis is an exciting area that remains to be elucidated.

In Chapter 4, the structure of the mammalian cyclic dinucleotide, cGAMP, was elucidated and found to contain a 3′–5′ and a 2′–5′ phosphodiester linkage. This unique regiochemistry enables cGAMP to signal differently through STING compared to prokaryotic-based cdNs, which have been found to be only 3′–5′ linked to date. While the driving force for the regiochemical change is unclear, its implications for drug development relating to immunity and oncology are enormous.

Chapter 5 illuminates the evolution of the cGAS-STING axis and cyclic dinucleotide signaling in phylogentically distant organisms. Cyclic dinucleotide binding was found to be a conserved function of STING across multiple phlya, and its structure was highly conserved between humans and *Nematostella vectensis*, an anemone species >500 million years diverged from humans. Moreover, the cGAS homolog in *Nematostella* was found to create a cyclic dinucleotide more consistent with a prokaryotic messenger (i.e. only 3′–5′ linked) than a human cdN messenger (i.e. mixed linkages), suggesting that the 2′–5′ phosphodiester linkage is a recent innovation. This chapter demonstrates that the cGAS-cdN-STING axis is an evolutionary conserved pathway. However, the primitive function of the cGAS-cdN-STING axis remains unclear.

Lastly, Chapter 6 attempts to elucidate the signaling space of the cGAS-cdN-STING axis by trying to determine its primitive function in *Nematostella*, while also dissecting its many downstream signaling pathways in mice and humans. When *Nematostella* was exposed to cGAMP, it was found to elicit phenotypic and transcriptomic changes indicative of an

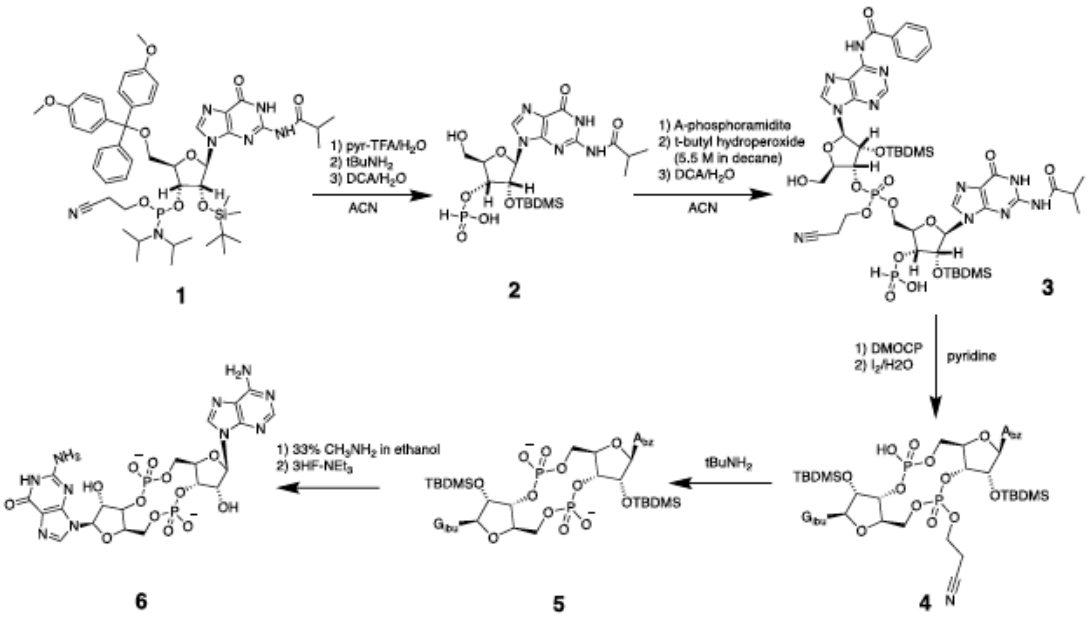
immune response. Surprisingly, 3′,3′ cdNs demonstrated little to no response, suggesting that 3′,3′ cdNs have a distinct physiological function. This chapter also examined how mouse and human STING's downstream signaling pathways can be dissected molecularly. Since STING is capable of adopting different conformations depending on the identity of its bound cdN and capable of eliciting multiple downstream signaling pathways, it is intriguing to hypothesize whether cdNs have a dual role in symbiosis as well as pathogen-response.

Genes that find themselves at the interface of host-pathogen interactions are often rapidly evolving.[118] While this work is a holistic examination of cyclic dinucleotide signaling mechanisms, it also illustrates how genes related to immunity develop remarkable functions. Antibodies, CRISPR, and endonucleases are all components of the immune system and have found utility outside of their endogenous functions for biological, therapeutic, and/or engineering purposes. It is perhaps no coincidence that those molecular tools that are the most valuable are those that are derived from conflict.

# Chapter 2

**Reagents and oligonucleotides**

DNA oligonucleotides were purchased from Elim Biopharmaceuticals (Hayward, CA) and IDT (Coralville, IA). Cyclic di-GMP, cyclic di-AMP, cyclic GMP, and pGpG were purchased from Axxora, LLC (Farmingdale, NY). DFHBI was purchased from Lucerna (New York, NY) and was prepared as a 20 mM stock solution in DMSO. Commercially available reagents were used without further purification. Chemically competent BL21 (DE3) Star cells were purchased from Life Technologies (Carlsbad, CA). *Vibrio cholerae* genomic DNA and D70E WspR DNA were gifts from the Marletta lab.




**Synthesis of cyclic AMP-GMP (6)**

Cyclic AMP-GMP was synthesized using the one-pot procedure from Gaffney *et al.*[34] with the following modification to the work-up of intermediate **5**. After deprotection of **4** with *tert*-butylamine, the crude mixture was concentrated, washed twice with acetonitrile to remove excess *tert*-butylamine, and dried. The residue was subsequently dissolved in methanol and filtered to remove solid impurities. The saved filtrate was concentrated, precipitated again with acetonitrile, and the solid material was filtered and saved. Compound **5** was purified as the di-ammonium salt using reverse-phase HPLC on a Perkin Elmer Series 200 HPLC equipped with a Perkin Elmer Prep 10-ODS C18 column (20 mm, 250 mm x 10 mm, 300 Å). Conditions used were an 85 to 45% gradient of solvent A over 30 min, flow rate 3 mL/min, where solvent A is 100 mM ammonium acetate in water and solvent B is acetonitrile. Purified

elution fractions were lyophilized overnight followed by repeated rounds of dissolution and evaporation with methanol in order to remove excess ammonia.

Characterization of **5** (isolated as di-ammonium salt). $^{1}$H NMR (600 MHz, d4-MeOD, δ): 8.84 (1H, s), 8.71 (1H, s), 8.35 (1H, s), 8.06 (2H, J = 7 Hz, d), 7.63 (1H, J = 7 Hz, t), 7.54 (2H, J = 7 Hz, t), 6.15 (1H, s), 5.96 (1H, s), 4.85 (2H, m), 4.74 (1H, J = 3 Hz, d), 4.53 (1H, J = 4 Hz, d), 4.52 (1H, d), 4.44 (1H, J = 9, d), 4.40 (1H, J = 11, d), 4.34 (1H, J = 8, d), 4.08 (1H, J = 11 Hz, d), 4.04 (1H, J = 11, d), 2.78 (1H, m), 1.21 (3H, J = 7, d), 1.19 (3H, J = 7, d), 1.04 (9H, s), 0.99 (9H, s), 0.37 (3H, s), 0.31 (3H, s), 0.28 (3H, s) 0.25 (3H, s); $^{13}$C {$^{1}$H decoupled} NMR (600 MHz, d4-MeOD, δ): (all resonances are singlets) 181.77, 168.22, 157.80, 153.05, 150.69, 149.85, 149.48, 143.27, 139.57, 134.89, 133.91, 129.76, 129.47, 125.18, 121.32, 92.16, 91.20, 81.02, 77.355, 77.292, 71.90, 71.87, 63.54, 36.79, 26.53, 19.60, 19.20, 19.09, 18.99, -3.87, -3.93, -4.54, -4.60; 31P {$^{1}$H decoupled} NMR (600 MHz, d4-MeOD, δ): (all resonances are singlets) -2.20, -2.52; HRMS (*m/z*): [M-H]$^{-}$ calculated for $C_{43}H_{62}N_{10}O_{15}P_2Si_2$, 1075.3332; found, 1075.3279. [M+Na-2H]$^{-}$ calculated for $C_{43}H_{62}N_{10}O_{15}P_2Si_2$, 1097.3151; found, 1097.3113.

Characterization of **6** (isolated as 3:2 triethylamine:**6**). $^{1}$H NMR (600 MHz, D2O, 50 °C, δ): 8.70 (1H, s), 8.50 (1H, s), 8.40 (1H, s), 6.41 (1H, br d), 6.25 (1H, br d), 5.28-5.24 (2H, br m), 5.15 (1H, br m), 5.04 (1H, br m), 4.74 (4H, m), 4.39 (2H, J = 2 Hz, t); $^{13}$C {$^{1}$H decoupled} NMR (600 MHz, D2O, 50 °C, δ): (all resonances are singlets) 157.32, 154.19, 154.04, 151.39, 149.80, 147.81, 140.10, 136.63, 118.85, 115.43, 90.57, 90.05, 80.62, 80.28, 73.71, 70.62, 62.62, 62.51; 31P {$^{1}$H decoupled} NMR (600 MHz, D2O, 50 °C, δ): (all resonances are singlets) -1.23, -1.41; HRMS (*m/z*): [M-H]$^{-}$ calculated for $C_{20}H_{24}N_{10}O_{13}P_2$, 673.0921; found, 673.0926.

***In vitro* transcription**

DNA templates were made through PCR amplification from the appropriate plasmid DNA using primers that added the T7 polymerase promoter sequence. The templates were transcribed using T7 RNA polymerase in 40 mM Tris-HCl, pH 8.0, 6 mM $MgCl_2$, 2 mM spermidine, and 10 mM DTT. RNA was purified in a denaturing (7.5 M urea) 6% polyacrylamide gel and was extracted from gel pieces using Crush Soak buffer (10 mM Tris-HCl, pH 7.5, 200 mM NaCl and 1 mM EDTA, pH 8.0). RNAs were precipitated with ethanol and resuspended in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA).

**Ligand binding assays**

To measure the fluorescence of each RNA construct in response to ligand, a solution of ligand (100 μM) and DFHBI (10 μM) was prepared in buffer containing 40 mM HEPES, pH 7.5, 125 mM KCl, and 3 mM $MgCl_2$. The RNA was renatured in buffer at 70 °C for 3 min and cooled to ambient temperature for 5 min prior to addition to the reaction solution at a final concentration of 100 nM. Duplicate 100 μL binding reactions were incubated at 37 °C in a Corning Costar (Tewksbury, MA) 96-well black plate until equilibrium was reached (see activation curves in Supplementary Figures). The fluorescence emission was measured using

a Molecular Devices SpectraMax M3 plate reader (Sunnyvale, CA) with the following instrument parameters: 460 nm excitation, 500 nm emission, 495 nm cutoff. The background fluorescence of the buffer without DFHBI was subtracted from each sample to determine the relative fluorescence units. For experiments to measure $K_d$, the ligand concentration was varied as the concentrations of RNA (30 nM unless otherwise noted) and DFHBI (10 μM) were held constant, and the fluorescence of the sample with no ligand was subtracted to determine relative fluorescence units.

**Live cell imaging of *E. coli***

For stable RNA expression *in vivo*, the human $tRNA^{Lys}_3$ scaffold was added to the 5′ and 3′ ends of WT Vc2-Spinach, as described in Paige *et al.*.[33,40] The WT Vc2-Spinach tRNA construct was amplified using primers that added a BglII restriction site and the T7 polymerase promoter sequence to the 5' end and a T7 terminator and XhoI restriction site to the 3' end. The product was then cloned into the pET31b vector for inducible expression in *E. coli*. Vc2-Spinach tRNA mutants were created by PCR amplification of the appropriate Vc2-Spinach sequence using primers that added EagI and SacII restriction sites to the 5′ and 3′ ends, respectively. Then the WT Vc2-Spinach fragment was removed from pET31b-tRNA by digestion with EagI and SacII and the appropriate Vc2-Spinach variant was cloned in its place. The WspR gene was amplified using primers that added NdeI and XhoI restriction sites to the 5′ and 3′ ends, respectively, and the product was cloned into pET24a. The DncV gene was amplified using primers that added an NheI restriction site to the 5′ end and a 6xHis tag followed by a XhoI restriction site to the 3′ end. This product was then cloned into pET24a for inducible expression in *E. coli*. Quikchange (Stratagene, La Jolla, CA) was used according to the manufacturer's protocol to produce the various mutations.

BL21 (DE3) Star *E. coli* cells were co-transformed with 60 ng each of the appropriate plasmids and cells were plated on LB/Carb/Kan plates (Carb: 50 μg/mL, Kan: 50 μg/mL). Single colonies were used to inoculate overnight LB/Carb/Kan cultures grown at 37 °C. Fresh LB/Carb/Kan cultures were inoculated with the overnight culture, and the cells were induced with 1 mM isopropyl b-D-1-thiogalactopyranoside (IPTG) once they reached an OD600 ~ 0.5. Induced cells were incubated for another 2.5 h at 37 °C, then were pelleted, washed once with M9 minimal media, pH 6.5, and resuspended in 200 μL M9 minimal media, pH 6.5 to an OD600 ~ 1. The cell suspension was pipetted onto poly D-Lysine coated coverslips and incubated at 37 °C for 45 min to allow the cells to adhere. Unattached cells were removed by washing thoroughly with M9 minimal media, and then 200 μL of 200 μM DFHBI in M9 minimal media, pH 6.5 was added to cover the cells. The cells were incubated with DFHBI on the coverslip for 1.5 h at 37 °C before placing them over microscope slides for visualization. Epifluorescence imaging experiments were performed using a Zeiss 200M AxioVert microscope (Zeiss, Jena, Germany) equipped with a mercury light source X-Cite 120 Series (Exfo Life Science Divisions, Ontario, Canada), a 63x/1.4 Plan-Apochromat oil DIC objective lens and a 1.6x tube lens. For monitoring fluorescence, a GFP filter set with an excitation 470/40 BP, FT 495 beamsplitter, and emission 525/50 BP was used (Filter Set 38 HE).

Fluorescence microscope images were normalized to the brightest sample of the entire set using ImageJ software (NIH, Bethesda, MD). This was done by automatically adjusting the brightness/contrast followed by manually setting the maximum brightness to be the same as the brightest sample.

Image analysis and fluorescence quantitation was carried out using ImageJ software. Cells were manually outlined in the differential interference contrast (DIC) image and saved as individual regions of interest (ROIs). The ROIs were overlaid on the corresponding fluorescence image to measure the mean fluorescence for each cell in the image. The average fluorescence of the black background (e.g. DFHBI solution) was determined for each image and subtracted from the value for each cell to determine the mean fluorescence intensity. At least 50 cells were analyzed to measure the average mean fluorescence intensity. Statistical significance of difference in fluorescence was assessed using a student's t test.

## Chapter 3

### General Reagents and Oligonucleotides

GEMM-I–Spinach DNA oligonucleotides were purchased as Ultramers from Integrated DNA Technologies, and other oligonucleotides were purchased from Elim Biopharmaceuticals. Cyclic dinucleotides were purchased from Axxorra. 3,5-Difluoro-4-hydroxybenzylidene imidazolinone (DFHBI) was either purchased from Lucerna or was synthesized following previously described protocols[32] and was stored as an ~20 mM stock in DMSO. *G. sulfurreducens* PCA was obtained from the J. Coates laboratory at the University of California at Berkeley. Genomic DNA from *G. sulfurreducens* was isolated using the Purelink Genomic DNA mini kit (Invitrogen).

### In Vitro Transcription

Preparation of RNAs was carried out as previously described.[25] Briefly, DNA templates for in vitro transcription were prepared via a PCR that added the T7 polymerase promoter sequence. Templates were then transcribed using T7 RNA polymerase and were purified either by a 96-well spin column (ZR 96 Clean & Concentrator, Zymo Research) for screening experiments or by denaturing (7.5 M urea) 6% (wt/vol) PAGE for characterization experiments. RNAs purified by PAGE were subsequently eluted from gel pieces in Crush Soak buffer (10 mM Tris·HCl, pH 7.5, 200 mM NaCl, and 1 mM EDTA, pH 8.0), precipitated with ethanol, and resuspended in 1 x TE buffer (10 mM Tris·HCl, pH 8.0, 1 mM EDTA). RNAs purified by a 96-well spin column were purified according to the manufacturer's protocol and were resuspended in 0.5 x TE buffer. Accurate measurement of RNA concentration was determined after performing a hydrolysis assay to eliminate the hypochromic effect due to secondary structure in these RNAs.[119]

### *In Vitro* Fluorescence Assays of GEMM-I–Spinach Variants

To screen GEMM-I riboswitch aptamer selectivity, a solution of ligand (0, 1, 50, or 100 μM) and DFHBI (33 μM) was prepared in buffer containing 40 mM Hepes, 125 mM KCl, and 10 mM $MgCl_2$ at pH 7.5. Each RNA was renatured in buffer at 70 °C for 3 min and cooled to ambient temperature for 5 min before addition to the reaction solution at a final concentration of 100 nM. Binding reactions were incubated at 30 °C in a Corning Costar 3915 96-well black plate until equilibrium was reached. The fluorescence emission was measured using a Molecular Devices SpectraMax Paradigm plate reader with the following instrument parameters: 460 nm excitation and 500 nm emission. The background fluorescence of the buffer alone (without DFHBI) was subtracted from each sample to determine the relative fluorescence units. To analyze the selectivity or binding affinity of RNA biosensor constructs, assays were run as described above, except at 37 °C with 3 mM $MgCl_2$ to mimic physiological conditions. Selectivity experiments were performed using 100 nM RNA and 100 μM ligand. Experiments to measure $K_d$ were performed with 30 nM RNA and various ligand concentrations, and the fluorescence of the sample with DFHBI and RNA in buffer without ligand was subtracted to determine relative fluorescence units.

**In-Line Probing Assays**

In vitro transcribed RNA was radiolabeled with γ-$^{32}$P ATP using T4 polynucleotide kinase (New England Biolabs) following standard procedures.[120] After PAGE purification, RNAs were passed through an Illustra MicroSpin G-25 Column (GE Healthcare) and eluted with ddH2O. Inline probing assays (1 x in-line buffer: 50 mM Tris·HCl, pH 8.3, 20 mM $MgCl_2$, 100 mM KCl) were performed as previously described[119] for the extended GS-814 riboswitch with terminator construct. The following modifications were made to the procedure to measure ligand-binding affinities below 1 nM for the GM-790 and GS-814 riboswitch aptamers. In-line reactions were prepared in 100 µL total volumes instead of the standard 10 µL. Thus, the final RNA concentration in each reaction was <<500 pM (this upper limit assumes 100% recovery from each purification step), which allowed us to determine subnanomolar dissociation constants. The reactions were quenched with 50 µL of 2 x ULB (urea loading buffer, half of the standard volume), and 25 µL of the sample was loaded per lane onto a 10% urea–PAGE gel. To achieve ~2,000–4,000 cpm per lane, the RNA had to be very efficiently radiolabeled (~240,000–480,000 cpm/pmol). The no-reaction treatment with RNase T1 (T1), and partial base hydrolysis ($^{-}$OH) ladders were prepared as 10 µL reactions and quenched with 90 µL ddH2O and 50 µL 2 x ULB before loading 25 µL of each. Dried gels were exposed on a phosphorimager screen for several days and scanned using a Typhoon laser-scanning system (GE Healthcare).

**Liquid Culture Growth of *G. sulfurreducens* PCA**

From a glycerol stock, a 10 mL starter culture of *G. sulfurreducens* PCA in fumarate–acetate media alone or supplemented with 0.1% yeast extract was grown anaerobically at 30 °C for 5–7 d without shaking in anaerobic culture tubes sealed under nitrogen with 20 mm blue butyl rubber stoppers and aluminum crimps. Intermediate 50-mL cultures were inoculated at 1:10 dilution and then used to inoculate master 1 L cultures at 1:20 dilution in the same media. Cells were grown anaerobically in fumarate–acetate media with or without yeast extract at 30 °C until an OD600 of ~0.4 was reached, which took about 2–3 d. Cells were harvested in 50 or 100 mL aliquots by centrifugation at 10,000 x g for 10 min at 4 °C, and pellets were stored at −80 °C. The fumarate–acetate media contained the following in 1 L: 4 g sodium fumarate dibasic (25 mM sodium fumarate final concentration), 1.64 g sodium acetate (20 mM final concentration), 0.49 g $NaH_2PO_4 \cdot 2H_2O$, 0.97 $Na_2HPO_4$, 0.25 g $NH_4Cl$, 0.1 g KCl, 10 mL 100 x vitamin mix, 1 mL 1,000 x NB minerals mix, and 1.0 g yeast extract (Bacto brand, if used). The media was boiled and degassed by bubbling under nitrogen gas for 30 min while cooling, then aliquoted, sealed under anaerobic conditions, and autoclaved.

**Cell Extraction from *G. sulfurreducens* PCA and *E. coli***

Cyclic dinucleotides were extracted as described previously[121] with the following modifications. A frozen cell pellet from 100 mL of liquid culture was thawed and resuspended in a 600 µL extraction buffer (40% methanol, 40% acetonitrile, 20% ddH2O). The cell solution was incubated at 4 °C for 20 min and then at ambient temperature for 20 min. After centrifugation at 4 °C for 20 min at 13,200 x g, the supernatant was carefully removed and stored on ice. The remaining pellet was extracted twice more with 300 µL of extraction

solvent as described. The combined supernatants were evaporated to dryness by rotary evaporation, and the dried material was resuspended in 250 µL ddH2O. The extract was filtered through a 10-kDa MWCO Amicon Ultra-4 Protein Concentrator (Millipore) and used immediately or stored in aliquots at −20 °C.

**HPLC Fractionation and MS Analysis (HRMS, MS/MS) of Cell Extracts**

Discrete fractions from *G. sulfurreducens* PCA cell extracts were collected using an Agilent 1260 Infinity liquid chromatograph equipped with a diode array detector and analytical-scale fraction collector. Extract samples of 100–150 µL were separated on a Polaris C18-A column (250 mm length x 4.6 mm internal diameter, 5-µm particle size, 180-Å pore size; Agilent) at 50 °C with a flow rate of 1 mL/min. Solvent A was 100 mM ammonium acetate and solvent B was methanol. The elution program consisted of 0% B for 10 min, a linear gradient to 10% B over 1 min, isocratic conditions at 10% B for 4 min, a linear gradient to 30% B over 4 min, a linear gradient to 0% B over 1 min, and isocratic conditions at 0% B for 10 min. Eluted fractions were lyophilized multiple times to remove excess salts before mass spectrometry analysis.

HRMS and tandem mass spectrometry (MS/MS) measurements of collected fractions were performed using an Agilent 1200 liquid chromatograph (LC) that was connected in-line with an LTQ-Orbitrap-XL hybrid mass spectrometer equipped with an electrospray ionization source (Thermo Fisher Scientific). This instrumentation is located in the QB3/Chemistry Mass Spectrometry Facility at the University of California at Berkeley. The LC was equipped with a C18 analytical column (Viva C18: 150 mm length x 1.0 mm inner diameter, 5-µm particles; Restek) and a 100 µL sample loop. Solvent A was H2O + 0.1% formic acid and solvent B was MeCN + 0.1% formic acid (vol/vol). The elution program consisted of isocratic conditions at 0% B for 3 min, a linear gradient to 35% B over 32 min, a linear gradient to 95% B over 0.1 min, isocratic conditions at 95% B for 4.9 min, a linear gradient to 0% B over 0.1 min, and isocratic conditions at 0% B for 19.9 min at a flow rate of 130 µL/min and 40 °C for the column compartment.

Full-scan mass spectra were acquired in the positive ion mode over the range of m/z = 100–1,000 using the Orbitrap mass analyzer, in profile format, with a resolution setting of 100,000 (at m/z = 400, measured at full width at half-maximum peak height). MS/MS spectra of selected precursor ions were acquired using collision-induced dissociation in the positive ion mode using the linear ion trap, in centroid format, with the following parameters: isolation width 3 m/z units, normalized collision energy 30%, activation time 30 ms, and activation Q 0.25. Data acquisition and analysis were performed using Xcalibur software (version 2.0.7 SP1, Thermo).

# Chapter 4

**cGAS Product Purification and Structural Characterization**

The cGAS product (prepared in vitro as described below) was purified with reverse-phase high-performance liquid chromatography (HPLC) on an Agilent 1260 Infinity HPLC equipped with an Agilent Polaris C18-A column (5 mm, 250 mm x 10 mm, 180 Å). Purification conditions included a 100% to 0% gradient of solvent A over 20 min at 50 °C and a flow rate of 5 mL/min, solvent A being 100 mM ammonium acetate in water at pH 7 and solvent B being acetonitrile. Purified elution fractions were evaporated multiple times in order to remove excess ammonia.

Resonance assignments were made with COSY, $^{1}$H-$^{13}$C HSQC, NOESY, $^{1}$H-$^{13}$C HMBC, and $^{1}$H-$^{31}$P HMBC. The characterization of cGAS product is as follows. $^{1}$H NMR (900 MHz, $D_2O$, 50 °C, δ): 8.44 (s, 1), 8.42 (s, 1), 8.03 (s, 1), 6.31 (s, 1), 6.09 (d, 1, $J$ = 8 Hz), 5.75 (m, 1), 5.18 (m, 1), 4.93 (s, 1), 4.74, 4.62, 4.59 (d, 1, $J$ = 12 Hz), 4.55 (s, 1), 4.38(m, 1), 4.33(d, 1, $J$ = 12 Hz), 4.28(d ,1, $J$ =12 Hz); $^{31}$P{$^{1}$H decoupled} NMR (600 MHz, $D_2O$, 50 °C, δ): (all resonances are singlets) -0.96, -1.86; HRMS (m/z): [M-$H^+$] calculated for $C_{20}H_{23}N_{10}O_{13}P_2$, 673.0927; found, 673.0909. [M + $Na^+$ - 2$H^+$] calculated for $C_{20}H_{22}N_{10}O_{13}P_2Na$, 695.0746; found, 695.0728.

**Luciferase Assay**

HEK293T cells were plated in TC-treated 96-well plates at 0.5 x $10^6$ cells $mL^{-1}$. The next day, the cells were transfected with indicated constructs, along with IFN-β firefly luciferase and TK-Renilla luciferase reporter constructs. After stimulation for 6 hr with the indicated ligands, the cells were lysed in passive lysis buffer (Promega) for 5 min at 25 °C. The cell lysates were incubated with firefly luciferase substrate (Biosynth) and the Renilla luciferase substrate coelenterazine (Biotium), and luminescence was measured on a SpectraMax L Luminescence Microplate Reader (Molecular Devices). The relative IFN-β expression was calculated as firefly luminescence relative to Renilla luminescence. Statistical differences were calculated with an unpaired two-tailed Student's t test with the use of Prism 5.0b software (GraphPad).

***In Vitro* Cyclic Dinucleotide Synthesis**

In vitro DncV reactions were carried out in 20 mM Tris-Cl (pH 8), 20 mM $Mg(OAc)_2$, 10% glycerol, and 1 mM dithiothreitol, 0.1 mg/ml BSA. Reactions contained 250 μM GTP and 250 μM ATP or 125 μM GTP and 125 μM ATP. In addition, 33 nM α-$^{32}$P-GTP (3,000 Ci/mmol, PerkinElmer) or 33 nM α-$^{32}$P -ATP (3,000 Ci/mmol, PerkinElmer) was included in reaction where indicated. Reactions were started by the addition of 1 μM purified DncV protein. In vitro cGAS reactions were carried out in 40 mM Tris-Cl (pH 7.5), 100 mM NaCl, and 10 mM $MgCl_2$. Cold nucleotide and α-$^{32}$P -GTP was at the same concentrations as in DncV reactions. Reactions were started by the addition of 200 nM purified cGAS. Where indicated, herring testes DNA (Sigma-Aldrich) was added to reactions at a final concentration of 0.1 mg/ml. WspR reactions were performed as described previously.[70] Reactions were incubated for 1 hr

at 37 °C, boiled for 5 min at 95 °C, and spun for 10 min at 13,000 rpm. Reactions were removed and mixed 1:5 with TLC running buffer (1:1.5 [v/v] saturated $NH_4SO_4$ and 1.5 M $KH_2PO_4$ [pH 3.6]) and spotted on polyethylenimine cellulose TLC plate (Sigma-Aldrich). Following solvent migration, the TLC plate was exposed to a phosphorimager screen and imaged with a Typhoon scanner. For in vitro product transfection into HEK293T cells, reactions were scaled up, radiolabeled nucleotide was omitted, and the concentration of ATP and GTP was increased to 2 mM.

**Protein Purifications**

WspR construct (pQE-WspR*) was a generous gift from Steve Lory (Harvard). WspR purification and cdG synthesis reactions were carried out as previously described.[70] Overexpression strains and plasmids for DncV and mutant DncV were provided by J. Mekalanos. DncV protein was overexpressed and purified as previously described.[18] Briefly, DncV protein production was induced in midlog phase for 3 hrs at 37 °C with 1 mM IPTG. Cells were lysed and DncV protein was purified under denaturing conditions. Cleared lysate was incubated with Ni-NTA and eluted in Urea Elution buffer (2 M urea, 10 mM Tris pH = 8.0, 150 mM NaCl, 250 mM imidazole). Eluted protein was dialyzed to 25 mM Tris-Cl, pH = 7.5, 300 mM NaCl, 5 mM $Mg(OAc)_2$, 10% glycerol, 2 mM DTT. His6-SUMO-mcGAS was expressed in Rosetta(DE3) pLysS cells by overnight induction with 0.5 mM IPTG at 18 °C. Cells were lysed into 50 mM Tris-Cl, pH = 8, 300 mM NaCl, 20 mM imidazole, 5 mM β-mercaptoethanol (BME) and 0.2 mM PMSF by French Press. Cleared lysate was incubated with Ni-NTA and bound protein was eluted with 20 mM Tris-Cl, pH = 7.4, 150 mM NaCl, 300 mM Imidazole. Eluant was dialyzed to 20 mM Tris-Cl, pH = 7.4, 150 mM NaCl, 5 mM BME with 10% glycerol. Protein was flash frozen and stored at -80 °C.

**Nuclease Digests**

Nuclease P1 from *Penicillium citrinum* and Snake venom phosphodiesterase I (SVPD) from *Crotalus adamanteus* were purchased from Sigma. Reactions from *in vitro* cyclic dinucleotide synthesis labeled with $\alpha$-$^{32}$P-GTP were diluted 1:5 in either P1 buffer (40 mM Tris-Cl, pH = 6, 2 mM $ZnCl_2$) or SVPD buffer (40 mM Tris-Cl, pH = 8, 10 mM $MgCl_2$) followed by digestion with 2.5 mU of nuclease P1 or SVPD, respectively. Digestions were incubated for 45 min at 37 °C and nucleotide products were resolved by TLC.

## Chapter 5

### STING Allele Identification

STING orthologs were identified by a DELTA-BLAST (http://blast.ncbi.nlm.nih.gov/) query using mouse or human STING as input. High quality hits were aligned and used to construct a profile hidden Markov model using HMMER (http://hmmer.janelia.org).[122] The model was then used to search published animal genomes.

### STING Allele Cloning and Protein Purification

Recombinant STING proteins were expressed and purified according to previously developed conditions for purification of human cGAS.[92] Codon-optimized sequence corresponding nvSTING (Genbank: XP_001627385) was cloned from synthetic DNA oligos (IDT) into a modified pcDNA vector (Invitrogen) with a Kozak start sequence using Gibson assembly according to manufacturer's protocols (NEB). The nvSTING CDN receptor domain N193–G377(C-terminus) was amplified by PCR and cloned into a modified pET vector to express an N-terminal 6 x His (KSSHHHHHHGSS)-MBP-TEV fusion protein in BL21-RIL DE3 *E. coli* cells co-transformed with a pRARE2 tRNA (Agilent) plasmid. *E. coli* was grown in 2 x YT media at 37 °C to an OD600 of ~0.5 before being cooled at 4°C for 15 min and then incubated for ~18 h at 16 °C with 0.5 mM IPTG induction. Following a PBS wash, cell pellets were lysed by sonication in lysis buffer (20 mM HEPES-KOH [pH 7.5], 400 mM NaCl, 10% glycerol, 30 mM imidazole, 1 mM TCEP) in the presence of EDTA-free Complete Protease Inhibitor (Roche). Recombinant STING protein was purified from clarified lysate by binding Ni-NTA agarose (QIAGEN). Resin was washed with lysis buffer supplemented to 1 mM NaCl, and eluted by gravity-flow chromatography at 4°C using lysis buffer supplemented to 300 mM imidazole. Recovered protein in elution buffer was diluted to 50 mM imidazole and 5% glycerol prior to concentration to ~80 mg $mL^{-1}$ and digestion with Tobacco Etch Virus protease for ~12 h at 4 °C. Homodimer STING was separated from the digested MBP tag by diluting with gel-filtration buffer (20 mM HEPES-KOH [pH 7.5], 250 mM KCl, 1 mM TCEP) and passing over a 5 ml Ni-NTA column (QIAGEN) connected in line with a 5 ml MBP-Trap column (GE Life Sciences) and further purified by size-exclusion chromatography on a Superdex 75 16/60 column. Final purified nvSTING was concentrate to ~40 mg $mL^{-1}$, used immediately for crystallography or flash frozen in liquid nitrogen for storage at −80 °C and biochemical experiments. Selenomethionine-substituted nvSTING protein was prepared as previously described.[92] All animal STING alleles were cloned and expressed using analogous conditions. However for initial EMSA experiments, recombinant STING proteins from *X. tropicalis*, *D. rerio*, *D. melanogaster*, and *N. vitripennis* were expressed and purified with an amino terminal SUMO tag rather than a MBP tag, and recombinant STING proteins from *N. vectensis*, *S. harrisii*, *C. teleta*, *C. gigas*, *B. mori*, *B. terrestris*, and *M. musculus* were expressed without any solubility tag. Recombinant Vibrio DncV, human cGAS, nv-cGAS and nv-A7S0T1.1 synthases were purified and assayed for *in vitro* CDN-synthase activity according to previously published methods.[92]

### STING–cdN Complex Gel Shift Assay

Radiolabelled cyclic dinucleotides were synthesized by incubating ~100 nM of α-$^{32}$P-ATP and/or α-$^{32}$P-GTP with 5 μM cold NTP and ~10 μM of the respective synthase: human cGAS 2′,3′ cGAMP synthase, *Vibrio cholerae* DncV 3′,3′ cyclic AMP-GMP synthase, *Pseudomonas aeruginosa* WspR* (D70E constitutively active mutant) 3′,3′ cyclic di-GMP synthase, *Bacillus subtilis* DisA 3′,3′ cyclic di-AMP synthase[12].

Reactions were performed in 40 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM $MgCl_2$ and incubated at 37 °C for 1–2 h. Human cGAS reactions were additionally supplemented with herring testes DNA. Reactions were subsequently treated with 5 U of alkaline phosphatase (NEB) for 1 hour at 37°C. Following incubation, confirmation of cyclic dinucleotide synthesis was evaluated by thin-layer chromatography. Briefly, a 0.5 μL aliquot was spotted onto a Millipore TLC PEI cellulose F plate and placed into a glass chamber with a 1.5 M $KH_2PO_4$ (pH ~3.9) mobile phase. Following the appropriate amount of capillary action, the plates were removed, dried at 80 °C for 10 minutes, and imaged in a phosphor cassette.

Reactions were then extracted with a phenol/chloroform/isoamyl alcohol mixture. The aqueous layer (top layer) was separated, mixed with 2.5 fold (v/v) ethanol, and evaporated to dryness using a Savant Speed Vac system. Radiolabelled cyclic dinucleotide stocks were typically re-suspended in 20 μL of water.

Approximately 10–100 μM of STING was used for each binding experiment. For binding experiments, purified STING proteins in ~200 mM KCl were diluted with binding buffer of 50 mM Tris-HCl (pH 7.5), 5 mM KCl, 10 mM $MgCl_2$, and 10% glycerol before adding cdN and incubating for 1 hr on ice. All STING proteins were also independently incubated with radiolabelled phosphate to serve as a negative control. A loading dye was applied to each sample and then immediately loaded onto a 6% acrylamide:bisacrylamide native gel (28:1). The gel was run in the cold room (4 °C) for 90 minutes at 12 V. The gel was then dried for 2 hours at 60 °C and imaged in a phosphor cassette.

**STING–cdN ITC Analysis**

All ITC measurements were performed on a MicroCal Auto-iTC200 (GE Healthcare). Experiments were conducted at 25 °C using a degassed buffer comprised of 20 mM HEPES-KOH (pH 7.5), 250 mM KCl, 1 mM TCEP. Protein concentrations are shown in Table 5.2. All data fitting and analysis was done using Origin.

**Crystallization and Structure Determination**

nvSTING was crystallized in *apo* form or in complex with cdN ligands at 18 °C by hanging drop vapor diffusion. Purified nvSTING cdN receptor domain was diluted on ice to 10 mg mL$^{-1}$ in the presence of chemically synthesized 3′,3′ or 2′,3′ cdN ligands (BioLog). Following a 30 min incubation, 400 nL hanging drops were set at a 1:1 ratio over 70 μL of reservoir using a Mosquito robotics platform (TTP Labtech). Three sets of optimized crystallization conditions were obtained for independent nvSTING crystal forms: 200 mM $CaCl_2$, 100 mM HEPES-KOH pH 7.5, 28% PEG-400 (nvSTING *apo* "unrotated" $P\,2_12_12_1$

crystals); 1.8 M ammonium citrate pH 7.0 (nvSTING–3′,3′ cdN and nvSTING *apo* "rotated" $P\ 3_1$ crystals); 500 mM $LiSO_4$, 2% PEG-8000 over a range of pH from ~4–7 (nvSTING–2′,3′ cGAMP $P\ 2_12_12_1$ crystals). nvSTING *apo* "unrotated" crystals grown in 28% PEG-400 conditions were harvested using reservoir solution supplemented with 15% ethylene glycol as a cryo-protectant prior to flash-freezing by submersion in liquid nitrogen. cdN bound nvSTING crystals were exquisitely sensitive and could only be harvested with oil as a cryoprotectant. Briefly, crystal drops were covered with a layer of saturated paratone-N or NVH oil (Hampton) and crystals were transferred from the drop into overlaying oil emersion using a Kozak cat whisker. The cat whisker was then used to gently clean away excess mother liquor solution prior to harvesting the oil-submerged crystal with a nylon loop and flash-freezing in liquid nitrogen. X-ray diffraction data were collected under cryogenic conditions at the Lawrence Berkeley National Laboratory Advanced Light Source (Beamline 8.3.1).

Data were processed with XDS and AIMLESS[123] using the SSRL *autoxds* script (A. Gonzalez, Stanford SSRL). Indexed 3′,3′ cdN-bound crystals belonged to the trigonal spacegroup $P\ 3_1$, and *apo* or 2′,3′ cGAMP-bound crystals belonged to two different crystal forms in the orthorhombic spacegroup $P\ 2_1\ 2_1\ 2_1$. All crystals contained two copies of nvSTING cdN receptor domain in the asymmetric unit. Experimental phase information was determined using data from a single large nvSTING–3′,3′ cyclic di-GMP crystal collected at peak selenium absorbance energy and truncated to ~2.4 Å with strong anomalous signal to ~2.5 Å. Eight sites were identified with HySS in PHENIX[124], and an initial map was calculated using SOLVE/RESOLVE[125] concurrent with phase extension to a native nvSTING–3′,3′ cyclic di-GMP data set processed to 1.84 Å. Initial maps displayed clear unbiased density for two nvSTING monomers and a central bound ligand, and model building was completed in Coot[126] prior to refinement with PHENIX. Following completion of nvSTING–3′,3′ cyclic di-GMP model building, a monomeric model of nvSTING was prepared by removing ligand, protein loops and the beta-strand lid domain and used as a search model to determine phases for the other 3′,3′ cdN datasets using molecular replacement. A molecular replacement solution of the alternative STING conformation in the nvSTING–2′,3′ cGAMP dataset required searching with a significantly truncated core nvSTING monomer domain and then manual placement and rebuilding of nvSTING monomers into the density. In all nvSTING structures, x-ray data for refinement were extended according to an $I/\sigma$ resolution cutoff of ~1.5, or extended an additional ~0.2 Å as determined by CC* correlation and $R_{pim}$ parameters where visual inspection of the resulting map warranted data extension.[127]

**Cell-Based Interferon-β Luciferase Assay**

Full-length human and bacteria cdN synthases and a 3′,3′ cGAMP phosphodiesterase were amplified from genomic DNA and cloned into pcDNA vectors (Invitrogen) or derived as following: human cGAS 2′,3′ cGAMP synthase, *Vibrio cholerae* DncV 3′,3′ cGAMP synthase, *Pseudomonas aeruginosa* WspR* 3′,3′ cyclic di-GMP synthase, *Bacillus subtilis* DisA 3′,3′ cyclic di-AMP synthase, *Vibrio cholerae* 3′,3′ cyclic AMP-GMP phosphodiesterase. Bacterial genomic DNA was a kind gift from Dan Portnoy (University of California, Berkeley) and John Mekalanos (Harvard Medical School).

Synthase assays, and the *N. vectensis* screen for cGAS-like enzymes, were conducted in 293T human kidney cells as previously described.[92] Briefly, 293T cells at ~80% confluency were transfected using Lipofectamine 2000 (Invitrogen) in a 96-well format with the following plasmids: control Renilla luciferase reporter (2 ng), interferon-β promoter driven Firefly luciferase reporter (20 ng), hSTING wt R232 (wild-type allele: R71, G230, R232, R293) or hSTING R232H plasmid (15 ng), and a cdN synthase plasmid (20-150 ng) as indicated. In Figure 5.2, synthases are titrated from 0.1–50 ng (human cGAS, Pa WspR* and Bs DisA) or 10–150 ng (Vc DncV and nv-cGAS [nv-A7SFB5.1]). In Figure 5.5, synthases are used at 20 ng (human cGAS, Pa WspR* and Bs DisA) or 150 ng (Vc DncV and nv-cGAS [nv-A7SFB5.1]). In Figure 5.6, phosphodiesterase experiments were conducted at a 3:1 (phosphodiesterase 150 ng, synthase 50 ng) ratio as previously optimized.[93] At 24 hrs post-transfection, Renilla and Firefly Dual-Luciferase values were assayed according to manufacture's protocols (Promega) using a Veritas Microplate Luminometer (Turner Biosystems). Data were combined from multiple experiments and analyzed by an unpaired, two-tailed t test in Prism.

## Chapter 6

**Immunofluorescence and confocal microscopy**
Glass coverslips were added to a 24 well plate and washed thoroughly with 70% ethanol followed by PBS. The coverslips were then incubated with FBS for 5 – 20 minutes in order to allow for better cell adhesion to the glass. Each well was then immediately seeded with 293T cells and grown to ~50% confluency.

A total of 1.25 μg of DNA was transfected into each well, using a ratio of 2.5 μL lipofectamine to 1 μg of DNA. Each well contained the following: pcDNA4-STING (10 ng), pEGFP-LC3 (5 ng) and pcDNA4 with the respective cdN synthase (1,235 ng). In cases where a synthase was not used, an empty pcDNA4 vector was transfected instead. Cells were transfected for 24 – 48 hours.

Following transfection, cells were fixed and permeabilized as follows. Cells were washed 1 x PBS, and then fixed for 15 minutes in a fresh 4% solution of paraformaldehyde. Following another 1 x PBS wash, cells were permeabilized for 5 minutes in a fresh 0.5% saponin/PBS solution. They were then washed 1 x PBS, and treated with 0.1% sodium borohydride/0.1% saponin/PBS for 5 – 10 minutes in order to consume any remaining paraformaldehyde. Finally, cells were washed 3 x PBS before staining.

Cells were stained as follows. Each well was incubated with 1% BSA/0.1% saponin/PBS for 45 minutes as an initial blocking step. Cells were then treated with 1% BSA/0.1% saponin/PBS + HA antibody (1:200 dilution, Sigma 11867423001 rat IgG from Roche) for one hour. They were then washed 3 x PBS. Each well was then incubated with 0.1% saponin/PBS + secondary antibody (1:500 dilution, Jackson ImmunoResearch, Cy3 affinipure donkey anti-rat IgG, 712-165-153) for 45 minutes. Finally, cells were washed 3 x PBS, mounted using VectaShield with DAPI, and dried overnight.

All microscopy images were acquired using a Zeiss LSM 780 NLO AxioExaminer.

***Nematostella vectensis* cyclic dinucleotide overlay and microinjection**
*Nematostella* embryos and primary polyps were cultivated as previously described.[128] Briefly, adult *Nematostella* were sexed by their ability to lay eggs and maintained in 1/3x sea water. Once an anemone had been observed to release eggs, it was transferred to a temporary bath to dilute out any sperm and then kept in a bowl designated for females. For embryo cultivation, spawning females were first separated from the males. The day prior to fertilization, males and females were fed small pieces of oyster and then incubated at 16 °C for a couple of hours. At the end of the day, they were put into a room temperature incubator with a fluorescent lamp, where they were exposed to bright light for 8 – 15 hrs. The following day *Nematostella* were monitored every 15 – 20 minutes for the release of eggs or sperm. Eggs were then cultivated and incubated with sperm not more than 2 hours after release from the females. The eggs were incubated with sperm for 15 minutes, and then incubated with 4% cysteine for 15

minutes, which acts as a dejellying solution.

Embryos were overlaid with cyclic dinucleotides approximately 28 hours post-fertilization. Approximately 250 *Nematostella* embryos were pipetted into each well of a 96-well plate and incubated with a 1 mM cyclic dinucleotide solution (in 1/3x sea water) for 24 hours at 16 °C. Phenotypic changes of the embryos were recorded. The embryos were then homogenized and stored in Triazol® and stored at -80 °C for subsequent RNA isolation.[129] Month-old primary polyps were treated in a similar manner. Between 75 – 100 polyps were used for each condition (i.e. number of polyps were well).

Cyclic dinucleotide microinjection into *Nematostella* embryos was performed using a 100 μM cdN solution. Embryos were injected < 2 hours post-fertilization as previously described.[130]

**Cell-Based Interferon-β/NF-κB/STAT6 Luciferase Assay**
Interferon-β, NF-κB, and STAT6 luciferase-based detection assays were conducted in HEK293T human kidney cells in a similar manner to that described in Chapter 4 and 5. Briefly, 293T cells at 50 - 80% confluency were transfected using Lipofectamine 2000 (Invitrogen) in a 96-well format with the following plasmids: control *Renilla* luciferase reporter (2 ng), interferon-β, NF-κB, or STAT6 promoter driven firefly luciferase reporter (20 ng), mouse (0.25 – 1 ng) or human (5 – 15 ng) STING constructs in pcDNA4 and cdN synthase plasmids (variable amount) in pcDNA4. The total amount of plasmid DNA transfected per well was fixed at 125 ng. A STAT6 expression plasmid was also transfected at the indicated amounts for all STAT6-based detection experiments.

At 20 - 24 hrs post-transfection, *Renilla* and firefly luciferase values were assayed. The supernatant in each well was aspirated, and incubated in 50 μL of passive lysis buffer for 20 minutes. Twenty microliter aliquots were then pipetted into two white opaque 96-well plates. *Renilla* or firefly luminescence was then assayed for each respective plate. The respective substrates were injected into each well during reading.

**CRISPR/cas9 modified mice**
DNA oligos were obtained from Integrated DNA Technologies as ultramers (Coralville, IA), which were cloned into a linearized T7 gRNA vector (System Biosciences, Mountain view, CA USA) and sequence verified using DNA sequencing. The sgRNA templates for *in vitro* transcription were prepared by PCR amplification using Phusion high fidelity DNA polymerase followed by PCR cleanup (Qiagen, Hilden, Germany). sgRNA transcripts were then generated by an *in vitro* transcription synthesis kit (System Biosciences, Mountain View, CA).

C57BL/6J female mice (3–4 weeks olds) were obtained from JAX-WEST (The Jackson Laboratory, Sacramento, CA), and were super ovulated by intraperitoneal injection using pregnant mare's serum gonadotropin (PMSG) followed by human chorionic gonadotropin (hCG) after 48 hours. The females were then mated overnight with C57BL/6J male mice (>7

weeks old). On next day, zygotes were harvested from the ampullae of the super ovulated females and were be treated with hyaluronidase (Sigma, St Louis, MO) to remove the cumulus cells. Zygotes were placed in KSOM medium (Millipore, Billerica, MA) and incubated in a $CO_2$ incubator (37 °C, 5% $CO_2$) before microinjection.

The microinjection mixture of Cas9 mRNA (Life Technologies, Grand Island, NY) and sgRNA for a simple indel mutation or point mutation (DNA donor oligo included) were mixed to an appropriate concentration. The mixture was filtered using a Spin-X® column (Corning Inc., Corning, NY) by centrifugation. The top 2/3 portion of the mixture was then collected as the final microinjection mixture. Microinjection was performed in M2 medium using a micromanipulator (Narishige, East Meadow, NY) and an inverted microscope (Nikon, Melville, NY). The microinjection mixture (prepared from above) in continuous flow was injected into both the pronuclei and the cytoplasm of zygotes. After injection the zygotes were cultured in a $CO_2$ incubator (37 °C, 5% $CO_2$) overnight. The next day, 2-cell stage embryos were transferred to CD1 (Charles River, Wilmington, MA) recipients (at 0.5 dpc) via oviduct transfer. After the pups were born, they were weaned after 3 weeks and genotyped.

References


1. Ross, P. *et al.* Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. *Nature* **325,** 279–281 (1987).
2. Engelman, A., Mizuuchi, K. & Craigie, R. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* **67,** 1211–1221 (1991).
3. Ross, P. *et al.* The cyclic diguanylic acid regulatory system of cellulose synthesis in Acetobacter xylinum. Chemical synthesis and biological activity of cyclic nucleotide dimer, trimer, and phosphothioate derivatives. *J. Biol. Chem.* **265,** 18933–18943 (1990).
4. Zeng, F. & Jones, R. A. Synthesis of cyclic dinucleotides by an H-phosphonate method in solution. *Nucleosides* (1996). doi:10.1080/07328319608002723
5. Ryan, R. P. *et al.* Cyclic di-GMP signalling in the virulence and environmental adaptation of Xanthomonas campestris. *Molecular Microbiology* **63,** 429–442 (2006).
6. Karaolis, D. K. R. *et al.* c-di-GMP (3‘-5’-Cyclic Diguanylic Acid) Inhibits Staphylococcus aureus Cell-Cell Interactions and Biofilm Formation. *Antimicrobial Agents and Chemotherapy* **49,** 1029–1038 (2005).
7. Waters, C. M., Lu, W., Rabinowitz, J. D. & Bassler, B. L. Quorum Sensing Controls Biofilm Formation in Vibrio cholerae through Modulation of Cyclic Di-GMP Levels and Repression of vpsT. *Journal of Bacteriology* **190,** 2527–2536 (2008).
8. Krasteva, P. V. *et al.* Vibrio cholerae VpsT regulates matrix production and motility by directly sensing cyclic di-GMP. *Science* **327,** 866–868 (2010).
9. Simm, R., Morr, M., Kader, A., Nimtz, M. & Römling, U. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. *Molecular Microbiology* **53,** 1123–1134 (2004).
10. Kalia, D. *et al.* Nucleotide, c-di-GMP, c-di-AMP, cGMP, cAMP, (p)ppGpp signaling in bacteria and implications in pathogenesis. *Chemical Society Reviews* **42,** 305 (2012).
11. Sudarsan, N. *et al.* Riboswitches in Eubacteria Sense the Second Messenger Cyclic Di-GMP. *Science* **321,** 411–413 (2008).
12. Witte, G., Hartung, S., Büttner, K. & Hopfner, K.-P. Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Molecular cell* **30,** 167–178 (2008).
13. Karaolis, D. K. R. *et al.* Cyclic Di-GMP Stimulates Protective Innate Immunity in Bacterial Pneumonia. *Infection and Immunity* **75,** 4942–4950 (2007).
14. Karaolis, D. K. R. *et al.* Bacterial c-di-GMP is an immunostimulatory molecule. *J. Immunol.* **178,** 2171–2181 (2007).
15. Mcwhirter, S. M. *et al.* A host type I interferon response is induced by cytosolic sensing of the bacterial second messenger cyclic-di-GMP. *Journal of Experimental Medicine* **206,** 1899–1911 (2009).
16. Karaolis, D. K. R. *et al.* 3′,5′-Cyclic diguanylic acid (c-di-GMP) inhibits basal and growth factor-stimulated human colon cancer cell proliferation. *Biochemical and*

*Biophysical Research Communications* **329,** 40–45 (2005).
17. Wu, J. *et al.* Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339,** 826–830 (2013).
18. Davies, B. W., Bogard, R. W., Young, T. S. & Mekalanos, J. J. Coordinated Regulation of Accessory Genetic Elements Produces Cyclic Di-Nucleotides for V. cholerae Virulence. *Cell* **149,** 358–370 (2012).
19. Diner, E. J. *et al.* The Innate Immune DNA Sensor cGAS Produces a Noncanonical Cyclic Dinucleotide that Activates Human STING. *CellReports* **3,** 1355–1361 (2013).
20. Gao, P. *et al.* Cyclic [G(2',5')pA(3‘,5’)p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* **153,** 1094–1107 (2013).
21. Zhang, X. *et al.* Cyclic GMP-AMP Containing Mixed Phosphodiester Linkages Is An Endogenous High-Affinity Ligand for STING. *Molecular cell* **51,** 226–235 (2013).
22. Mcwhirter, S. M. *et al.* A host type I interferon response is induced by cytosolic sensing of the bacterial second messenger cyclic-di-GMP. *Journal of Experimental Medicine* **206,** 1899–1911 (2009).
23. Chen, H. *et al.* Activation of STAT6 by STING Is Critical for Antiviral Innate Immunity. *Cell* **147,** 436–446 (2011).
24. Ishikawa, H. & Barber, G. N. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature* **455,** 674–678 (2008).
25. Kellenberger, C. A., Wilson, S. C., Sales-Lee, J. & Hammond, M. C. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messengers Cyclic di-GMP and Cyclic AMP-GMP. *J. Am. Chem. Soc* **135,** 4906–4909 (2013).
26. Kellenberger, C. A. *et al.* GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. *Proceedings of the National Academy of Sciences* **112,** 5383–5388 (2015).
27. Diner, E. J. *et al.* The Innate Immune DNA Sensor cGAS Produces a Noncanonical Cyclic Dinucleotide that Activates Human STING. *CellReports* **3,** 1355–1361 (2013).
28. Kranzusch, P. J. *et al.* Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2‘,3’ cGAMP Signaling. *Molecular cell* **59,** 891–903 (2015).
29. Smith, K. D. *et al.* Structural basis of ligand binding by a c-di-GMP riboswitch. *Nature Structural & Molecular Biology* **16,** 1218–1223 (2009).
30. 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* **49,** 7351–7359 (2010).
31. 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* **133,** 15578–15592 (2011).
32. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA Mimics of Green Fluorescent Protein. *Science* **333,** 642–646 (2011).
33. Paige, J. S., Nguyen-Duc, T., Song, W. & Jaffrey, S. R. Fluorescence Imaging of Cellular Metabolites with RNA. *Science* **335,** 1194–1194 (2012).
34. 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.* 1–3 (2010). doi:10.1021/ol101236b

35. Ching, S. M., Tan, W. J., Chua, K. L. & Lam, Y. Synthesis of cyclic di-nucleotidic acids as potential inhibitors targeting diguanylate cyclase. *Bioorganic & Medicinal Chemistry* **18,** 6657–6665 (2010).
36. Kellenberger, C. A., Wilson, S. C., Sales-Lee, J. & Hammond, M. C. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messengers Cyclic di-GMP and Cyclic AMP-GMP. *J. Am. Chem. Soc* **135,** 4906–4909 (2013).
37. Kung, F. C., Raymond, J. & Glaser, D. A. Metal ion content of Escherichia coli versus cell age. *Journal of Bacteriology* (1976).
38. Hengge, R. Principles of c-di-GMP signalling in bacteria. *Nat Rev Micro* **7,** 263–273 (2009).
39. Christen, M., Kulasekara, H. D., Christen, B. & Kulasekara, B. R. Asymmetrical distribution of the second messenger c-di-GMP upon bacterial cell division. *Science* (2010).
40. Ponchon, L. & Dardel, F. Recombinant RNA technology: the tRNA scaffold. *Nat Meth* **4,** 571–576 (2007).
41. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. Rfam: an RNA family database. *Nucleic Acids Research* **31,** 439–441 (2003).
42. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Research* **31,** 3429–3431 (2003).
43. Felsenstein, J. *PHYLIP–Phylogeny Inference Package (Version 3.2) Cladistics. 1989; 5: 164–166*. (… the CCAL).
44. Lovley, D. R. & Coates, J. D. Bioremediation of metal contamination. *Curr. Opin. Biotechnol.* **8,** 285–289 (1997).
45. Lovley, D. R. The microbe electric: conversion of organic matter to electricity. *Curr. Opin. Biotechnol.* **19,** 564–571 (2008).
46. Nelson, J. W. *et al.* Control of bacterial exoelectrogenesis by c-AMP-GMP. *Proceedings of the National Academy of Sciences* **112,** 5389–5394 (2015).
47. Ding, Y.-H. R. *et al.* The proteome of dissimilatory metal-reducing microorganism Geobacter sulfurreducens under various growth conditions. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1764,** 1198–1206 (2006).
48. Ding, Y.-H. R. *et al.* Proteome of Geobacter sulfurreducens grown with Fe(III) oxide or Fe(III) citrate as the electron acceptor. *BBA - Proteins and Proteomics* **1784,** 1935–1941 (2008).
49. Tremblay, P.-L. *et al.* A c-type cytochrome and a transcriptional regulator responsible for enhanced extracellular electron transfer in Geobacter sulfurreducens revealed by adaptive evolution. *Environmental Microbiology* **13,** 13–23 (2011).
50. Mehta, T., Coppi, M. V., Childers, S. E. & Lovley, D. R. Outer Membrane c-Type Cytochromes Required for Fe(III) and Mn(IV) Oxide Reduction in Geobacter sulfurreducens. *Applied and Environmental Microbiology* **71,** 8634–8641 (2005).
51. Aklujkar, M. *et al.* Proteins involved in electron transfer to Fe(III) and Mn(IV) oxides by Geobacter sulfurreducens and Geobacter uraniireducens. *Microbiology* **159,** 515–535 (2013).
52. Butler, J. E., Young, N. D. & Lovley, D. R. Evolution of electron transfer out of the cell: comparative genomics of six Geobacter genomes. *BMC Genomics* **11,** 40 (2010).

53. Vargas, M. *et al.* Aromatic Amino Acids Required for Pili Conductivity and Long-Range Extracellular Electron Transport in Geobacter sulfurreducens. *mBio* **4,** e00105–13–e00105–13 (2013).
54. Bonanni, P. S., Bradley, D. F., Schrott, G. D. & Busalmen, J. P. Limitations for Current Production in Geobacter sulfurreducensBiofilms. *ChemSusChem* **6,** 711–720 (2013).
55. Lolicato, M. *et al.* Cyclic dinucleotides bind the C-linker of HCN4 to control channel cAMP responsiveness. *Nature Chemical Biology* **10,** 457–462 (2014).
56. Steinberger, O., Lapidot, Z., Ben-Ishai, Z. & Amikam, D. Elevated expression of the CD4 receptor and cell cycle arrest are induced in Jurkat cells by treatment with the novel cyclic dinucleotide 3‘,5’-cyclic diguanylic acid. *FEBS Letters* **444,** 125–129 (1999).
57. Shi, H., Wu, J., Chen, Z. J. & Chen, C. Molecular basis for the specific recognition of the metazoan cyclic GMP-AMP by the innate immune adaptor protein STING. *Proceedings of the National Academy of Sciences* **112,** 8947–8952 (2015).
58. Barbalat, R., Ewald, S. E., Mouchess, M. L. & Barton, G. M. Nucleic Acid Recognition by the Innate Immune System. *Annu. Rev. Immunol.* **29,** 185–214 (2011).
59. Palm, N. W. & Medzhitov, R. Pattern recognition receptors and control of adaptive immunity. *Immunol. Rev.* **227,** 221–233 (2009).
60. Takeuchi, O. & Akira, S. Pattern Recognition Receptors and Inflammation. *Cell* **140,** 805–820 (2010).
61. Ishii, K. J. *et al.* A Toll-like receptor–independent antiviral response induced by double-stranded B-form DNA. *Nat Immunol* **7,** 40–48 (2005).
62. Stetson, D. B. & Medzhitov, R. Recognition of Cytosolic DNA Activates an IRF3-Dependent Innate Immune Response. *Immunity* **24,** 93–103 (2006).
63. Zhong, B. *et al.* The Adaptor Protein MITA Links Virus-Sensing Receptors to IRF3 Transcription Factor Activation. *Immunity* **29,** 538–550 (2008).
64. Ishikawa, H., Ma, Z. & Barber, G. N. STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature* **461,** 788–792 (2009).
65. Sun, W., Li, Y., Chen, L., Chen, H. & You, F. ERIS, an endoplasmic reticulum IFN stimulator, activates innate immune signaling through dimerization. in (2009).
66. Jin, L. *et al.* MPYS, a Novel Membrane Tetraspanner, Is Associated with Major Histocompatibility Complex Class II and Mediates Transduction of Apoptotic Signals. *Molecular and Cellular Biology* **28,** 5014–5026 (2008).
67. Sauer, J. D. *et al.* The N-Ethyl-N-Nitrosourea-Induced Goldenticket Mouse Mutant Reveals an Essential Function of Sting in the In Vivo Interferon Response to Listeria monocytogenes and Cyclic Dinucleotides. *Infection and Immunity* **79,** 688–694 (2011).
68. Woodward, J. J., Iavarone, A. T. & Portnoy, D. A. c-di-AMP Secreted by Intracellular Listeria monocytogenes Activates a Host Type I Interferon Response. *Science* **328,** 1703–1705 (2010).
69. Barker, J. R. *et al.* STING-Dependent Recognition of Cyclic di-AMP Mediates Type I Interferon Responses during Chlamydia trachomatis Infection. *mBio* **4,** e00018–13–e00018–13 (2013).

70. Burdette, D. L. *et al.* STING is a direct innate immune sensor of cyclic di-GMP. *Nature* **478,** 515–518 (2011).
71. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339,** 786–791 (2013).
72. Tanaka, Y. & Chen, Z. J. STING Specifies IRF3 Phosphorylation by TBK1 in the Cytosolic DNA Signaling Pathway. *Science Signaling* **5,** ra20–ra20 (2012).
73. Chen, Z.-H. & Schaap, P. The prokaryote messenger c-di-GMP triggers stalk cell differentiation in Dictyostelium. *Nature* **488,** 680–683 (2012).
74. Yi, G. *et al.* Single Nucleotide Polymorphisms of Human STING Can Affect Innate Immune Response to Cyclic Dinucleotides. *PLoS ONE* **8,** e77846 (2013).
75. Kranzusch, P. J., Lee, A. S.-Y., Berger, J. M. & Doudna, J. A. Structure of Human cGAS Reveals a Conserved Family of Second-Messenger Enzymes in Innate Immunity. *CellReports* **3,** 1362–1368 (2013).
76. Pino, S., Ciciriello, F., Costanzo, G. & Di Mauro, E. Nonenzymatic RNA ligation in water. *J. Biol. Chem.* **283,** 36494–36503 (2008).
77. Chen, W., Kuolee, R. & Yan, H. The potential of 3‘,5’-cyclic diguanylic acid (c-di-GMP) as an effective vaccine adjuvant. *Vaccine* **28,** 3080–3085 (2010).
78. Conlon, J. *et al.* Mouse, but not human STING, binds and signals in response to the vascular disrupting agent 5,6-dimethylxanthenone-4-acetic acid. *The Journal of Immunology* **190,** 5216–5225 (2013).
79. Wu, X. *et al.* Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic Acids Research* **42,** 8243–8257 (2014).
80. Civril, F. *et al.* Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498,** 332–337 (2013).
81. Hancks, D. C., Hartley, M. K., Hagan, C., Clark, N. L. & Elde, N. C. Overlapping Patterns of Rapid Evolution in the Nucleic Acid Sensors cGAS and OAS1 Suggest a Common Mechanism of Pathogen Antagonism and Escape. *PLoS Genet* **11,** e1005203 (2015).
82. Putnam, N. H. *et al.* Sea Anemone Genome Reveals Ancestral Eumetazoan Gene Repertoire and Genomic Organization. *Science* **317,** 86–94 (2007).
83. Chin, K.-H. *et al.* Novel c-di-GMP recognition modes of the mouse innate immune adaptor protein STING. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 352–366 (2013).
84. Huang, Y.-H., Liu, X.-Y., Du, X.-X., Jiang, Z.-F. & Su, X.-D. The structural basis for the sensing and binding of cyclic di-GMP by STING. *Nat Struct Mol Biol* **19,** 728–730 (2012).
85. Ouyang, S. *et al.* Structural Analysis of the STING Adaptor Protein Reveals a Hydrophobic Dimer Interface and Mode of Cyclic di-GMP Binding. *Immunity* 1–14 (2012). doi:10.1016/j.immuni.2012.03.019
86. Shang, G. *et al.* Crystal structures of STING protein reveal basis for recognition of cyclic di-GMP. *Nat Struct Mol Biol* **19,** 725–727 (2012).
87. Shu, C., Yi, G., Watts, T., Kao, C. C. & Li, P. Structure of STING bound to cyclic di-GMP reveals the mechanism of cyclic dinucleotide recognition by the immune system. *Nat Struct Mol Biol* **19,** 722–724 (2012).

88. Yin, Q. *et al.* Cyclic di-GMP Sensing via the Innate Immune Signaling Protein STING. *Molecular cell* **46,** 735–745 (2012).
89. Zhang, L. *et al.* Massive expansion and functional divergence of innate immune genes in a protostome. *Sci. Rep.* **5,** 8693 (2015).
90. Kato, K., Ishii, R., Hirano, S., Ishitani, R. & Nureki, O. Structural Basis for the Catalytic Mechanism of DncV, Bacterial Homolog of Cyclic GMP-AMP Synthase. *Structure/Folding and Design* **23,** 843–850 (2015).
91. Zhu, D. *et al.* Structural biochemistry of a Vibrio cholerae dinucleotide cyclase reveals cyclase activity regulation by folates. *Molecular cell* **55,** 931–937 (2014).
92. Kranzusch, P. J. *et al.* Structure-Guided Reprogramming of Human cGAS Dinucleotide Linkage Specificity. *Cell* **158,** 1011–1021 (2014).
93. Gao, J. *et al.* Identification and characterization of phosphodiesterases that specifically degrade 3‘3’-cyclic GMP-AMP. *Cell Res.* **25,** 539–550 (2015).
94. Gao, P. *et al.* Structure-Function Analysis of STING Activation by c[G(2. *Cell* **154,** 748–762 (2013).
95. Liu, S. *et al.* Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation. *Science* (2015). doi:10.1126/science.aaa2630
96. Wolenski, F. S. *et al.* Characterization of the core elements of the NF-κB signaling pathway of the sea anemone Nematostella vectensis. *Molecular and Cellular Biology* **31,** 1076–1087 (2011).
97. Liu, Y. *et al.* Activated STING in a Vascular and Pulmonary Syndrome. *N Engl J Med* **371,** 507–518 (2014).
98. Woo, S.-R. *et al.* STING-Dependent Cytosolic DNA Sensing Mediates Innate Immune Recognition of Immunogenic Tumors. *Immunity* **41,** 830–842 (2014).
99. Corrales, L. *et al.* Direct Activation of STING in the Tumor Microenvironment Leads to Potent and Systemic Tumor Regression and Immunity. *CellReports* **11,** 1018–1030 (2015).
100. Konno, H., Konno, K. & Barber, G. N. Cyclic Dinucleotides Trigger ULK1 (ATG1) Phosphorylation of STING to Prevent Sustained Innate Immune Signaling. *Cell* **155,** 688–698 (2013).
101. Watson, R. O., Manzanillo, P. S. & Cox, J. S. Extracellular M. tuberculosis DNATargets Bacteria for Autophagy by Activating the Host DNA-Sensing Pathway. *Cell* **150,** 803–815 (2012).
102. Dobbs, N. *et al.* STING Activation by Translocation from the ER Is Associated with Infection and Autoinflammatory Disease. *Cell Host and Microbe* 1–13 (2015). doi:10.1016/j.chom.2015.07.001
103. Saitoh, T. *et al.* Atg9a controls dsDNA-driven dynamic translocation of STING and the innate immune response. *Proceedings of the National Academy of Sciences* **106,** 20842–20846 (2009).
104. Liang, Q. *et al.* Crosstalk between the cGAS DNA Sensor and Beclin-1 Autophagy Protein Shapes Innate Antimicrobial Immune Responses. *Cell Host and Microbe* **15,** 228–238 (2014).
105. Watson, R. O. *et al.* The Cytosolic Sensor cGAS Detects Mycobacterium tuberculosis DNA to Induce Type I Interferons and Activate Autophagy. *Cell Host and Microbe*

**17,** 811–819 (2015).
106. Helm, R. R., Siebert, S., Tulin, S., Smith, J. & Dunn, C. W. Characterization of differential transcript abundance through time during Nematostellavectensis development. *BMC Genomics* **14,** 1–1 (2013).
107. Artamonova, I. I. & Mushegian, A. R. Genome Sequence Analysis Indicates that the Model Eukaryote Nematostella vectensis Harbors Bacterial Consorts. *Applied and Environmental Microbiology* **79,** 6868–6873 (2013).
108. Chu, H. & Mazmanian, S. K. Innate immune recognition of the microbiota promotes host-microbial symbiosis. *Nat Immunol* **14,** 668–675 (2013).
109. Franzenburg, S. *et al.* MyD88-deficient Hydra reveal an ancient function of TLR signaling in sensing bacterial colonizers. *Proceedings of the National Academy of Sciences* **109,** 19374–19379 (2012).
110. Gottar, M., Gobert, V., Michel, T., Belvin, M. & Duyk, G. The Drosophila immune response against Gram-negative bacteria is mediated by a peptidoglycan recognition protein. *Nature* (2002).
111. Ryu, J. H. *et al.* The Homeobox Gene Caudal Regulates Constitutive Local Expression of Antimicrobial Peptide Genes in Drosophila Epithelia. *Molecular and Cellular Biology* **24,** 172–185 (2003).
112. Kleino, A. *et al.* Pirk Is a Negative Regulator of the Drosophila Imd Pathway. *The Journal of Immunology* **180,** 5413–5422 (2008).
113. Foster, J. S., Apicella, M. A. & McFall-Ngai, M. J. Vibrio fischeri Lipopolysaccharide Induces Developmental Apoptosis, but Not Complete Morphogenesis, of the Euprymna scolopes Symbiotic Light Organ. *Developmental Biology* **226,** 242–254 (2000).
114. Koropatnick, T. A., Engle, J. T., Apicella, M. A. & Stabb, E. V. Microbial factor-mediated development in a host-bacterial mutualism. *Science* (2004).
115. Yang, H. *et al.* One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell* **154,** 1370–1379 (2013).
116. Wang, H. *et al.* One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* **153,** 910–918 (2013).
117. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat Meth* **5,** 593–599 (2009).
118. George, R. D. *et al.* Trans genomic capture and sequencing of primate exomes reveals new targets of positive selection. *Genome Research* **21,** 1686–1694 (2011).
119. Wilson, S. C., Cohen, D. T., Wang, X. C. & Hammond, M. C. A neutral pH thermal hydrolysis method for quantification of structured RNAs. *RNA* **20,** 1153–1160 (2014).
120. Regulski, E. E. & Breaker, R. R. In-line probing analysis of riboswitches. *Methods Mol. Biol* **419,** 53–67 (2008).
121. Spangler, C., Böhm, A., Jenal, U., Seifert, R. & Kaever, V. A liquid chromatography-coupled tandem mass spectrometry method for quantitation of cyclic di-guanosine monophosphate. *Journal of Microbiological Methods* **81,** 226–231 (2010).
122. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence

similarity searching. *Nucleic Acids Research* **39,** W29–37 (2011).
123. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132 (2010).
124. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).
125. Terwilliger, T. C. Reciprocal-space solvent flattening. *Acta Crystallographica Section D: Biological ...* (1999). doi:10.1107/S0907444999010033
126. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132 (2004).
127. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* (2012). doi:10.1126/science.1216980
128. Stefanik, D. J., Friedman, L. E. & Finnerty, J. R. Collecting, rearing, spawning and inducing regeneration of the starlet sea anemone, Nematostella vectensis. *Nat Protoc* **8,** 916–923 (2013).
129. Stefanik, D. J., Wolenski, F. S., Friedman, L. E., Gilmore, T. D. & Finnerty, J. R. Isolation of DNA, RNA and protein from the starlet sea anemone Nematostella vectensis. *Nat Protoc* **8,** 892–899 (2013).
130. Layden, M. J., Röttinger, E., Wolenski, F. S., Gilmore, T. D. & Martindale, M. Q. Microinjection of mRNA or morpholinos for reverse genetic analysis in the starlet sea anemone, Nematostella vectensis. *Nat Protoc* **8,** 924–934 (2013).