GEMM-I riboswitch-based fluorescent biosensors for live cell analysis of cyclic dinucleotide signaling

by

Xin Wang

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

Doctor of Philosophy

in

Molecular and Cell Biology

at the

University of California, Berkeley

Committee in charge:
Professor Ming C. Hammond, *Chair*
Professor Dave Savage
Professor Jamie Cate
Professor Ke Xu

Spring 2016

ABSTRACT

GEMM-I riboswitch-based fluorescent biosensors for live cell analysis of cyclic dinucleotide signaling

by

Xin Wang

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Ming C. Hammond, Chair

The bacterial second messenger cyclic di-GMP (c-di-GMP) holds a prominent position within the repertoire of bacterial signaling molecules. With its presence established in over 75% of sequence bacteria, c-di-GMP responds to primary environmental signals by affecting the bacterial lifestyle transition between the motile and sessile states, regulating biofilm formation, host colonization, and bacterial virulence. Over the past 30 years, many aspects of the c-di-GMP signaling pathway have become well characterized in no small part due to the plethora of tools that can quickly and conveniently detect c-di-GMP. Despite these triumphs, we have only begun to contend with the enormous scope of c-di-GMP signaling, an endeavor that would be greatly aided by new tools that are harder, better, faster, and stronger.

The four fluorescent biosensors introduced herein bring us towards that goal. These second-generation RNA biosensors were designed based on a natural c-di-GMP riboswitch aptamer fused to the Spinach dye-binding aptamer, producing a fluorescent signal upon c-di-GMP binding. Their speed, sensitivity, and selectivity secure their place as a valuable tool for studying c-di-GMP signaling, with demonstrated efficacy in monitoring c-di-GMP *in vivo* in *E. coli* in a variety of conditions, including anaerobic and zinc-exposed environments. Furthermore, these sensors were adapted towards studying the related cyclic dinucleotide cyclic GMP-AMP (cGAMP), a recently discovered second messenger known for roles in bacterial intestinal colonization and surface sensing. Both c-di-G and cGAMP-specific biosensors were used to uncover components of the cGAMP signaling pathway in organisms not previously known to have cGAMP signaling. It is envisioned that these biosensors can be used to further understand c-di-GMP and cGAMP signaling in a variety of organisms, *in vivo*, in real time.

To Nigel, the dreamer of dreams

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

When I started graduate school in 2011, I had no idea if I would actually make it through the entire program without being enticed away by any of my other disparate interests. Needless to say, I'm grateful that I did make it, and am especially grateful that Berkeley gave me so many opportunities to grow as a scientist, a teammate, a leader, a musician, an artist, and much more. None of this would have been possible without the support and guidance of my advisor, Dr. Ming Hammond, who not only trained me to tackle scientific challenges with a keen and discerning eye, but also allowed me to thrive in an environment that prioritized personal sustainability. Furthermore, thanks go to all my fellow graduate students of the Hammond Lab, past and present – Scott, Colleen, Steve, Tania, Zach, Yichi, Todd, Johnny, Andrew, Rebekah – the only ones who might actually read this (good luck on your theses; 5 cents in the RBF, please) and all of whom have run at least one of my gels. Importantly, Colleen Kellenberger laid the groundwork for many of the experiments done herein; I'm sorry that my sensors have now outdated yours.

Many collaborators, both formal and informal, contributed their expertise and insight into this work. The tRNA scaffold screen was built on work by Wanda Thi and Cindy Lam, two talented undergraduates that I had the pleasure of mentoring. The anaerobic experiments were only possible with the help of downstairs expert Joe Gallagher. The zinc-responsive time course was meticulously measured by Dr. Jongchan Yeo. All of our *in vivo* experiments have been aided by Dr. Chen Chen's offhand suggestion of autoinduction media. Our collaborators at Memorial Sloan-Kettering, Aiming Ren and Dr. Dinshaw Patel, crystallized our cGAMP-responsive riboswitch and provided precise commentary. Zach Hallberg and the Hypr GGDEF team – Todd Wright, Jongchan Yeo, Omer Ad, Beiyan Nan – crafted a groundbreaking story that has shifted a major paradigm in c-di-GMP signaling.

There is always an amazing team of administrators and support staff behind the scenes that keeps the world running as it should. Thanks to our lab administrator Susan, and all our lab managers of the years: Malathy, Ting, Wanda, CindyL, Bao, Jian; to Tim, Manik, Jason, and the team behind our oft-finicky Attune; to Sarah, Berta, Eric, Tanya, and others in our incredibly attentive MCB Graduate Affairs Office.

So much love goes to my roommates, Sam and Divya, who have become my de facto Berkeley family, and the multitude of friends – especially those musical and artistic partners in crime – too numerous to list here, who never fail to support and inspire me.

My interests in science were planted early, perhaps subconsciously as I spent my childhood wandering around my parents' Ph.D. labs, but definitely consciously in Ms. Wallmuth's Scientific Inquiries class at the Illinois Mathematics and Science Academy. To this day, IMSA remains the most intellectually fulfilling experience in which I have ever partaken, for which I am eternally grateful.

Finally, I owe everything to the unwavering support and humor of my parents, who instilled in me a healthy balance of skepticism and confidence. I would be nowhere without you.

Chapter One

Methods for c-di-GMP detection

A key facet of living organisms is the ability to respond and adapt to the ever-changing environment. As humans, our neurons translate sensory input – for instance, heat – into electrical signals that travel to our brains, informing us that the coffee is still too hot to drink comfortably, that we should wait for it to cool before the next sip. Likewise, bacteria depend on environmental signals to inform their lifestyle decisions. Surface receptors sense these primary external signals and translate them into internal signals by producing molecules known as second messengers. These second messengers then bind to cellular targets, altering protein states and gene expression patterns to produce a noticeable downstream effect.

The second messenger cyclic di-GMP (c-di-GMP) plays a central role in the bacterial environmental response. Present in over 75% of bacteria, c-di-GMP responds to temperature, light, and other ambient signals to affect bacterial movement and biofilm formation. The stay-or-go decisions informed by internal c-di-GMP levels have implications in key issues related to human health, including antibiotic resistance, as antibiotics are often impenetrable to biofilm-bound bacteria. In addition, c-di-GMP's role in mobility helps determine the composition of organisms within the various bacterial microbiomes in our bodies, as they maintain and promote our health.

With its widespread influence, it is no wonder that the overwhelming interest in understanding the c-di-GMP signaling has elicited a plethora of established methods to detect its presence. These methods seek to identify and characterize the components of the signaling pathway, from the upstream production synthases and hydrolases, to the downstream effector genes and proteins. Despite these advances, many gaps remain unfilled, particularly in the dynamics of how bacteria quickly adapt to changing environmental conditions. Furthermore, save for a few organisms, little is yet known about how c-di-GMP localization affects downstream activity.

Efficiently tackling these and other remaining challenges depends on establishing a convenient, high-throughput detection method that is sensitive, selective, and visualizable. Below, I discuss current methods and their quest towards realizing all of these characteristics.

Introduction to c-di-GMP signaling

The bacterial second messenger cyclic di-GMP (c-di-GMP) was first discovered in 1987 as Benziman and colleagues searched for a cofactor that activated cellulose synthase from Gluconacetobacter xylinus (Ross et al. 1987). After rounds of characterization via mass spectroscopy, nuclear magnetic resonance, and comparisons to synthetic standards, they finally isolated the "unusual cyclic nucleotide activator:" bis-(3' 5')-cyclic diguanylic acid, or c-di-GMP (Figure 1.1). In the years since, cyclic-di-GMP has entered the scientific limelight as a universal signaling molecule, mediating numerous bacterial processes including motility, cellular adhesion, and biofilm formation, with wide-ranging implications in microbial ecology and human health.

In the cell, cyclic di-GMP is synthesized from two GTP molecules by diguanylate cyclases (DGCs), a class of enzymes denoted by their conserved GGDEF domains (Simm et al. 2004; Ausmees et al. 2001). On the other side, phosphodiesterases (PDEs) hydrolyze c-di-GMP into its

constituent nucleotides: HD-GYP domain PDEs directly hydrolyze c-di-GMP into two GMP molecules (Galperin et al. 1999), while EAL domain PDEs undo the cyclization by creating a pGpG molecule (Ross et al. 1987; Simm et al. 2004) that is further hydrolyzed by oligoribonucleases (Orr et al. 2015) (Figure 1.1). Many of these DGCs and PDEs contain N-terminal sensory input domains that respond to primary environmental signals such as oxygen, light, or extracellular molecules, triggering DGC activation. Downstream, c-di-GMP levels are detected by a variety of effectors that include proteins, transcription factors, and riboswitches, whose responses center around mediating the transition between the bacterial motile and sessile states: low levels of c-di-GMP lead to pili formation and flagellar rotation, while high levels lead to the secretion of cellulose and other exopolysaccharides for biofilm production (reviewed in Ross et al. 1987; Römling et al. 2013) (Figure 1.2). In organisms that undergo asymmetric division, including Caulobacter crescentus, Pseudomonas areuginosa, and Dictyostelium discoideum, c-di-GMP has been implicated in determining daughter cell identity (Waters 2010; Christen et al. 2010; Z.-H. Chen & Schaap 2012).

GGDEF domain proteins have been identified in over 75% of all sequenced bacteria, securing c-di-GMP's prominence as a signaling molecule in the bacterial kingdom. The recent discovery that the mammalian protein STING (stimulator of interferon genes) also binds c-di-GMP to activate an innate immune response (Burdette et al. 2011) expanded c-di-GMP's influence into the eukaryotic realm, raising the possibility that the second messenger plays a large part beyond bacteria. Meanwhile, as our understanding of c-di-GMP signaling has grown, so has the scope of methods used to detect c-di-GMP and identify components in its signaling pathways. In this chapter, I explore many of these techniques, from chromatography to fluorescent sensors, chronicling how they have become more fast, sensitive, and high-throughput over the years.

Chromatography-based methods for *in vitro* analysis

Incidentally, the methods that Benziman and colleagues first used to characterize c-di-GMP and its synthesis pathway gave rise to the first methods for detecting c-di-GMP (Ross et al. 1987; Waters 2010). One method, thin-layer chromatography (TLC), uses a mobile liquid phase to separate components of a sample spotted on a solid phase plate. The different mobilities determined by the chemical properties of each substituent cause them to migrate to different locations on the plate. Meanwhile, radiolabeling the substrate of choice allows its location to be detected through exposure to X-ray film. Benziman et al. incubated the diguanylate cyclase with GTP radiolabeled at the α vs. γ-phosphate position to discover that the product retained α-^{32}P but not γ-^{32}P, providing clues towards its structure (Wolfe & Berg 1989; Ross et al. 1987). Using TLC to detect products from α-^{32}P labeled nucleotide substrates provides a sensitive strategy for the *in vitro* detection and characterization of different cyclic dinucleotide (CDN) products, and has been used to survey cyclase and phosphodiesterase activity (Hallberg et al. 2016; Kranzusch et al. 2013). Higher-resolution separations within more complex mixtures can be achieved via two-dimensional TLC (2D-TLC), in which extracts are sequentially exposed to two different solvents traveling along perpendicular axes (Waters 2010).

Another technique Benziman et al. employed was high-performance liquid chromatography (HPLC) followed by mass spectroscopy (MS), which allowed them to determine the identity of c-di-GMP based on its mass. The initial chromatography step involves the loading the sample onto a solid column and slowly eluting the components over high pressure through a liquid phase, whose changing composition separates the eluted products. The products are then analyzed via MS, which ionizes the samples and determines their mass by evaluating their movements through an electric field. Today, HPLC-MS remains a gold standard for determining synthase and phosphodiesterase activity due to its ability to simultaneously detect substrates, intermediates, and products in cell extracts (Antoniani et al. 2010; Spangler et al. 2010; Simm et al. 2009). Furthermore, its ability to determine CDN levels down to the low nM range makes it one of the most sensitive detection methods to date (Waters 2010; Simm et al. 2009)

These two *in vitro* chromatography methods are still widely used in c-di-GMP research, and further advancements have rendered them even more sensitive and quantitative. Preparing samples for TLC or LC-MS, however, is often a time-consuming process that oftentimes involves cell extract preparation or protein purification. The rapid growth of the c-di-GMP made it clear that more high-throughput methods would be required to meet the growing demand.

Indirect methods based on c-di-GMP phenotypes

Initially considered a niche molecule, c-di-GMP managed to stay relatively shrouded from the scientific limelight until the sequence identification of the GGDEF domains that characterize c-di-GMP-synthesizing diguanylate cyclases (DGCs) and the EAL and HD-GYP domains that characterize c-di-GMP-hydrolyzing phosphodiesterases (PDEs). The eventual availability of whole genome sequences then prompted bioinformatics analyses that found c-di-GMP regulatory proteins in a majority (>75%) of sequenced bacterial genomes (Ausmees et al. 2001). Interestingly, most genomes contain multiple DGCs and PDEs, often adjacent to each other on the same operon, suggesting multiple layers of regulation to fine-tune c-di-GMP levels inside each cell (Römling et al. 2013).

From this, c-di-GMP entered the spotlight as a near-universal signaling molecule, prompting the development of more high-throughput techniques that could quickly and accurately assay DGCs and PDEs by measuring relative c-di-GMP levels. In particular, the discovery of c-di-GMP's role in affecting motility and biofilm formation inspired methods using those phenotypes as proxies for c-di-GMP concentration. To determine motility, researchers spot a normalized amount of liquid bacterial culture on soft agar, measuring the extent of bacterial growth after a fixed amount of time (Liaw et al. 1990; Wolfe & Berg 1989). Meanwhile, biofilm formation can be measured colorimetrically via crystal violet or congo red dyes. The crystal violet assay, also called the microtiter plate assay, involves growing cells in a small quantity of liquid culture, washing the cultures, and using the crystal violet nucleic acid stain to quantify the cells that remained adhered to the surface (for review, see O'Toole et al. 1999). The congo red assay involves using low levels of congo red in the media, which then binds amyloid fibers and exopolysaccharides that indicate biofilm formation (Antoniani et al. 2010). These indirect methods continue to be used as a

relatively fast way to screen many enzymes for diguanylate cyclase and phosphodiesterase activity (Antoniani et al. 2010; Simm et al. 2005; S. Gao et al. 2014).

These phenotype-based methods have two main advantages: first, they allow the enzymes to be assayed in host systems without the time-consuming processes of cell lysis or protein purification, allowing them to be more high-throughput than any previously developed assays. Second, they measure downstream effector activity, rendering any positive signals biologically relevant. However, as with any indirect methods, these assays are not quantitative, and any positive hits require further direct confirmation. These disadvantages highlight the need for c-di-GMP detection methods that are quantitative, direct, and high-throughput.

Direct methods based on G-quadruplex formation

Early studies recognized that c-di-GMP could form spectroscopically distinct aggregates (Nakayama, Roelofs, et al. 2012; Liaw et al. 1990). At high (>10 μM) concentrations, c-di-GMP multimerizes to form highly stable G-quadruplex structures whose presence can be detected through circular dichroism (Roembke et al. 2014; Stelitano et al. 2013) and synthetic dyes, many of which are described below. The presence of cations such as potassium and magnesium, as well as planar intercalators such as acriflavin and proflavine, can also stimulate multimerization (Nakayama, Kelsey, Wang & Sintim 2011), allowing c-di-GMP detection at lower concentrations.

Dye-based assays for G-quadruplex detection offer a simple and rapid measurement of c-di-GMP concentrations *in vitro*. One method utilizes the nucleic acid dye thiazole orange, which becomes brightly fluorescent when intercalated between adjacent bases. Adding this planar aromatic dye to c-di-GMP does the double duty of stimulating aggregation and unleashing fluorescence once intercalated in the G-quadruplex structure (Nakayama, Kelsey, Wang, Roelofs, et al. 2011). This method can be used to detect c-di-GMP in cell extracts, as extracts from cells expressing a constitutively active DGC displayed higher fluorescence signals than those from WT cells (Nakayama, Kelsey, Wang, Roelofs, et al. 2011).

Another dye-based approach involves a secondary peroxidase reaction that results in a colorimetric signal. G-quadruplex structures complexed to hemin, an iron-containing porphyrin, can act as peroxidases (Travascio et al. 1999). Adding these G-quadruplex-hemin complexes to the colorless 2,20-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) oxidizes the compound into the green $ABTS^{\bullet+}$, which absorbs light at 415 nm. The addition of proflavine to stimulate c-di-GMP aggregation allowed Nakayama and colleagues to use this peroxidase assay to detect sub-micromolar levels of c-di-GMP *in vitro* (Nakayama, Roelofs, et al. 2012).

A synthetic analog containing a GMP linked to the fluorescent synthetic base 2-aminopurine (2AP) (3',3'-cG(d2AP)MP) has also been used to monitor c-di-GMP concentrations. Unlike c-di-GMP, these analogs cannot self-aggregate, but they can dimerize with c-di-GMP. This dimerization then quenches 2AP fluorescence, and the loss of fluorescence signal indicates the amount of c-di-GMP present (Roembke et al. 2014).

These G-quadruplex-based strategies can rapidly assay c-di-GMP concentrations without requiring radiolabeled ligands or time-consuming chromatography. However, they still only offer approximations of c-di-GMP concentrations, as G-quadruplex formation is very sensitive to buffer conditions including ions, temperature, and intercalators (Nakayama, Kelsey, Wang & Sintim 2011; Nakayama, Kelsey, Wang, Roelofs, et al. 2011; Liaw et al. 1990). Furthermore, the method is not sensitive enough to be used in native environments, as cellular c-di-GMP concentrations are not usually high enough for spontaneous G-quadruplex formation (Donaldson et al. 2012; Nakayama, Kelsey, Wang & Sintim 2011).

In vitro methods based on effector binding

Inside the cell, effector proteins and riboswitches bind and detect c-di-GMP levels to activate downstream signals. These natural binding interactions can hence also be leveraged for c-di-GMP detection. Identifying and characterizing these binding interactions, however, is a non-trivial process. Bioinformatic analyses can predict known binding motifs such as protein PilZ domains, the largest class of c-di-GMP-binding proteins, characterized by a RxxxR and D/NxSxxG motif (Amikam & Galperin 2006). To identify novel interactions, cell lysates can be incubated with c-di-GMP analogs functionalized with biotin or magnetic beads, which are then pulled down to reveal bound constituents (Luo et al. 2012; Düvel et al. 2012). These affinity-tagged methods, however, only identify effectors that can tolerate the functionalized ligand.

Using radiolabeled c-di-GMP provides an alternate strategy that does not require interfering with ligand structure. Although radiolabeled c-di-GMP lack functionalized handles to pull down bound effectors, their presence can be detected with high sensitivity, allowing for the quantitative characterization of binding affinities and other native interactions. One high-throughput screening method is Differential Radial Capillary Action of Ligand Assay (DRaCALA), in which radiolabeled ligand is equilibrated with putative protein effectors and spotted onto a sheet of dry nitrocellulose (Roelofs et al. 2011). Proteins are immobile on nitrocellulose, leaving them concentrated at the center of the spot along with any radioactive signal from bound ligand. Unbound ligand, meanwhile, travels on the membrane via capillary action, leading to a diffuse signal. The radius of radioactivity at the center of each spot hence provides a measurement of the amount of ligand bound, and comparing spots with varying effectors and concentrations, as well as competitors, can provide quantitative information about binding affinities.

In DRaCALA, the radioactive nature of the signal output requires that each sample contain a constant amount of radiolabeled ligand. The assay can be modified to detect or measure unknown ligand concentrations through a competitive displacement assay, in which unlabeled c-di-GMP compete with labeled c-di-GMP for effector binding (Roelofs et al. 2011). This assay, however, requires a protein effector with a known binding affinity, and is much less sensitive than direct binding assays. That said, the competitive displacement assay has tremendous value as high-throughput screen to rapidly identify inhibitors from a large compound library (Koh 2002; Lieberman et al. 2014), opening the door for applications in pharmaceutical discovery.

DRaCALA's effectiveness lies in the unique capillary action produced upon protein-nitrocellulose interaction. Yet proteins only comprise one subset of c-di-GMP effectors; RNA riboswitches also bind c-di-GMP to regulate gene expression (Sudarsan et al. 2008). Adapting DRaCALA towards assaying RNA-ligand interactions requires functionalizing the RNA with a protein, typically a biotin-streptavidin tag, to mimic protein-nitrocellulose interactions (Donaldson et al. 2012). This functionalization, however, also precludes the assay from making native measurements, as the large tag could also interfere with ligand binding.

The first class of RNA effectors for c-di-GMP was inferred bioinformatically from a broad screen for structured RNAs. These aptamers, called GEMM-I (Genes for the Environment for Membranes and for Motility), shared a two-hairpin structure and were found upstream of genes annotated as DGCs, PDEs, and c-di-GMP effectors in a large number of both gram-positive and gram-negative bacteria (Weinberg et al. 2007). To confirm that GEMM-I riboswitches bound c-di-GMP, Breaker and colleagues used in-line probing, a technique that exploits the tendency of unstructured single-stranded RNA regions to degrade. Comparing the locations of single-stranded regions in the Vibrio cholerae Vc2 riboswitch with and without c-di-GMP indicated that ligand binding prompted a conformational rearrangement in the P1 stem at the base of the two hairpins, forming the basis for riboswitch activity (Sudarsan et al. 2008). This conformational rearrangement further allowed the development of an electrophoretic mobility shift assay for riboswitch-ligand binding (Kellenberger, Sales-Lee, et al. 2015) as well as a rapid high-throughput microfluidic adaptation (Pan et al. 2014). Another riboswitch-based assay involved tethering the Vc2 aptamer to a hammerhead ribozyme such that ligand binding would induce ribozyme activity (Furukawa et al. 2012).

In cells, riboswitches toggle gene expression by occluding or exposing key downstream sequences in response to ligand binding. Engineered riboswitch reporters place riboswitch sequences upstream of reporter genes such as ß-galactosidase and fluorescent proteins, providing another method to indirectly assay ligand presence (Sudarsan et al. 2008; Fowler et al. 2010; Topp et al. 2010). Similar to the aforementioned motility and biofilm formation assays, riboswitch reporters can be very sensitive since they propagate a downstream signal, but they are not quantitative and the signal cannot be reversed upon changes in ligand concentration. Furthermore, as with all the effector binding-based methods, they cannot be used to evaluate c-di-GMP levels in real time.

Live cell imaging techniques

One of the first characterized DGCs was PleD, a response regulator from the organism Caulobacter crescentus. Caulobacter follow an asymmetric cell division, yielding a motile swarmer cell with a single flagellum and type IV pili, and a sessile stalked cell with cellular attachment mechanisms at its surface. This asymmetric division is mediated by the polar localization of various enzymes and organelles, including PleD, which localizes to the pole of the immotile stalked cell. Deletion of PleD results in symmetric division, with both cells becoming swarmer cells (Skerker & Laub 2004; Duerig et al. 2009).

PleD's characterization as a GGDEF prompted questions of whether c-di-GMP was also asymmetrically distributed during cell division. Answering these questions, however, called for c-di-GMP tracking mechanisms that could monitor spatial localization in real time. Fluorescent biosensors offer one potential solution, as their ability to directly detect and monitor ions, small molecules, pH, and voltage potential have made them invaluable tools in molecular biology research (Crone et al. 2013). In particular, genetically encoded sensors, such as those derived from green fluorescent protein (GFP) or related proteins, can be expressed and detected in the native cellular environment to study cellular response and signaling *in vivo* (Enterina et al. 2015).

These sensors commonly function by fusing fluorescent domains to a ligand-binding domain such that ligand binding triggers an internal conformational change, modulating fluorescence output. For instance, Förster Resonance Energy Transfer (FRET) sensors involve two fluorescent proteins with overlapping excitation and emission maxima such that the emission energy released from the donor protein can excite the adjacent acceptor protein (Mitra et al. 1996). Because the state of energy transfer depends on the spatial distance between the fluorophores, and a ligand-sensing domain inserted between the fluorescent proteins changes their distance change upon ligand binding. One major advantage of FRET sensors is that the output signal is a ratio between donor and acceptor fluorescence, hence giving a semi-quantitative readout of ligand concentration.

A FRET-based sensor for c-di-GMP was developed by inserting the c-di-GMP-binding PilZ domain from the *S. enterica* effector YcgR between yellow and cyan fluorescent proteins (Christen et al. 2010). This sensor was used to evaluate local changes in c-di-GMP concentration during asymmetric cell division in *Caulobacter crescentus, Pseudomonas aeruginosa*, and their respective DGC knockout strains. Despite the advantages afforded as a genetically encoded fluorescent biosensor, this c-di-GMP FRET sensor had two major limitations that prevented its immediate adoption. First, ligand binding reduceed FRET signal by only 60%, a low signal change exacerbated by the high background from the turn-off nature of the signal. Second, the sensor had a Hill coefficient of 2.07, indicating that it binds two c-di-GMP molecules in a cooperative manner, narrowing the dynamic range and convoluting the quantitative nature of the signal. Two additional FRET-based sensors for c-di-GMP were developed with different ligand-binding PilZ domains (Ho et al. 2013). These sensors, however, were also turn-off sensors with even lower signal change.

These FRET sensors exemplified the major challenge in protein-based biosensor design of finding ligand-binding proteins and constructing linker regions that propagate adequate conformational change. Furthermore, engineering and refining protein-ligand interactions *de novo* is often a tedious and time-consuming process that requires prior structural knowledge (Koh 2002). Beyond difficulties in engineering, fluorescent proteins also have the inherent disadvantage of requiring oxygen for chromophores maturation (Reid & Flynn 1997; Drepper et al. 2007; Craggs 2009; Chapman et al. 2008). This oxygen requirement makes them ill-suited for anaerobic applications, where they suffer from reduced or variable brightness and slow turn-on kinetics (Kumagai et al. 2013; Ransom et al. 2015). Meanwhile, c-di-GMP-related enzymes and effectors are found in both obligate and facultative anaerobes, and its roles in biofilm formation and host colonization affect the composition of microbes found anaerobic environments such as

the human digestive tract (Lozupone et al. 2012; Nicholson et al. 2012). C-di-GMP synthases and phosphodiesterases have also been linked to oxygen-sensing PAS domains, suggesting that oxygen can directly regulate c-di-GMP levels (Tuckerman et al. 2009; Lacey et al. 2010; An et al. 2010). These factors call for the development of biosensors for live cell imaging that do not need oxygen to function.

In the protein realm, anaerobic fluorescent reporters exist in the form of proteins that bind external chromophores, such as the flavin-binding iLOV (Drepper et al. 2007; Chapman et al. 2008) and the bilirubin-binding UnaG (Kumagai et al. 2013). These natural metabolites remain non-fluorescent until protein binding restricts their rotational movement, thereby increasing quantum yield. Although these proteins have been validated *in vivo* under anaerobic conditions, their fluorescence is dependent on the presence of endogenous metabolites whose varying concentrations can generate inconsistent signals. Furthermore, the aforementioned challenges in protein engineering have limited the scope of biosensors derived from these protein scaffolds (Crone et al. 2013; Buckley et al. 2015; Erapaneedi et al. 2015; Liem et al. 2015).

RNA-based fluorescent tools for live imaging

On the RNA side, analogous RNA aptamer-dye pairs hold promise for anaerobic applications, although none have yet been validated as anaerobic sensors. In contrast to the naturally occuring proteins and metabolites that comprise iLOV and UnaG, RNA-dye pairs typically feature synthetic chromophores with engineered scaffolds, a combination made possible by SELEX (systematic evolution of ligands by exponential enrichment) and its numerous variations (Tuerk & Gold 1990; Paige et al. 2011). In SELEX, a large pool of random aptamers is exposed to the target ligand, the binders isolated and re-amplified, and the entire process is repeated over several rounds until a consensus sequence is isolated. The first aptamer-dye pair developed this way involved the triphenylmethane dye malachite green (MG) (Babendure et al. 2003), and the resulting MG aptamer could be engineered as a biosensor by appending a ligand-binding recognition domain and a communication module to transduce binding events to the dye-binding reporting domain (Stojanovic & Kolpashchikov 2004). By generating biosensors for theophylline and flavin mononucleotide, this method demonstrated the engineering potential of this modular construction. However, malachite green has toxic effects in animal cells (Srivastava et al. 2004), discouraging the use of this system *in vivo*.

Additional RNA aptamer-dye pairs have been introduced in recent years, including Spinach (Paige et al. 2011), which binds 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), a synthetic analog of the GFP chromophore, and Mango (Dolgosheina et al. 2014), which binds thiazole orange, a nucleic acid dye. The Spinach aptamer system has also been adapted as a small molecule biosensor by inserting another ligand-binding aptamer into the Spinach sequence. Initial attempts at biosensor engineering included biosensors for S-adenosylmethionine (Paige et al. 2012), streptavidin (Song et al. 2013), GMP (Paige et al. 2012), and even c-di-GMP via a Vc2-Spinach fusion and an artifically designed transducer domain (Nakayama, Luo, et al. 2012). These early biosensors succeeded in validating the proof of concept, but from a practical

perspective, they suffered from poor sensitivity, poor ligand selectivity, and slow kinetics, and many of them could not be seen *in vivo*.

In our lab, we adapted a similar strategy towards developing a Spinach-based biosensor for c-di-GMP by appending the Vc2 riboswitch aptamer and using the natural Vc2 P1 stem as the transducer domain (Kellenberger et al. 2013) (Figure 1.3). This biosensor, termed Vc2-Spinach, exhibited a 5-fold fluorescence turn-on *in vitro* with a sub-μM binding affinity (K_D = 230 nM), with a dynamic range across the cellular c-di-GMP concentration range of 0-10 μM (Weinhouse et al. 1997; Simm et al. 2009; Spangler et al. 2010). Accordingly, the biosensor exhibited a 2-fold fluorescence turn-on *in vivo*, as analyzed by microscopy of *E. coli* co-expressing with the c-di-GMP synthase WspR. Kellenberger et al. also demonstrated that the biosensor's strong selectivity towards c-di-GMP could be altered by changing key binding residues in the riboswitch, thereby extending its utility towards studying properties of riboswitch-ligand binding.

Outlook

Over the past 30 years, advancements in c-di-GMP detection technologies have moved towards the coveted trifecta of high sensitivity, rapid detection, and *in vivo* applicability (for summary, see Figure 1.4). To that end, RNA-based fluorescent biosensors such as Vc2-Spinach (Kellenberger et al. 2013) hold tremendous potential, as they have demonstrated their ability to selectively detect sub-micromolar levels of c-di-GMP in live cells. Despite these triumphs, Vc2-Spinach still left room for improvements in brighness, speed, and sensitivity. Furthermore, the field of Spinach-based biosensors lacked a generalizable strategy towards developing next-generation sensors.

Herein I describe four new second-generation RNA-based biosensors for c-di-GMP that achieve the trifecta described above. Chapter 2 introduces these biosensors and describes the process by which they were developed, as improvements were made to all three domains –Spinach, the riboswitch aptamer, and the transducer domain. Chapter 3 then focuses on validating the sensor as a useful tool for *in vivo* detection. Importantly, I demonstrate that these RNA-based tools can function in anaerobic conditions, opening the door towards understanding c-di-GMP signaling in new environments such as the human gut or industrial bioreactors. Furthermore, I use our next-generation biosensors to track c-di-GMP dynamics as environmental stresses cause c-di-GMP levels to fluctuate.

The final chapter focuses on applications of the biosensor towards discovering new components in the signaling pathway of cyclic GMP-AMP (cGAMP), a related CDN. Known for roles in bacterial intestinal colonization and surface sensing, cGAMP presence was previously only verified in *V. cholerae* and *Geobacter* species (Kellenberger, Wilson, et al. 2015). We employed both c-di-GMP- and cGAMP-selective biosensors in high-throughput screens to identify novel cGAMP riboswitches and synthases. Furthermore, we used biochemical data from our high-throughput screens to bioinformatically extrapolate cGAMP signaling pathways in a number of other organisms, expanding the scope of this little-known molecule.

Meanwhile, the prominence of c-di-GMP only continues to grow. The number of c-di-GMP-related publications has steadily risen over the past decade, with the nascent 2016 figure continuing that trend (Figure 1.5). Recently, c-di-GMP effectors were discovered within the mammalian innate immune system (Burdette et al. 2011), expanding c-di-GMP signaling beyond bacteria, suggesting additional roles in maintaining host-microbe symbiosis and bacterial pathogenesis. These roles have only increased the need for c-di-GMP detection methods that are simple, sensitive, and accurate. The next-generation biosensors introduced here fulfill those needs, paving the way for a greater understanding of the microbial ecosystem.

FIGURES

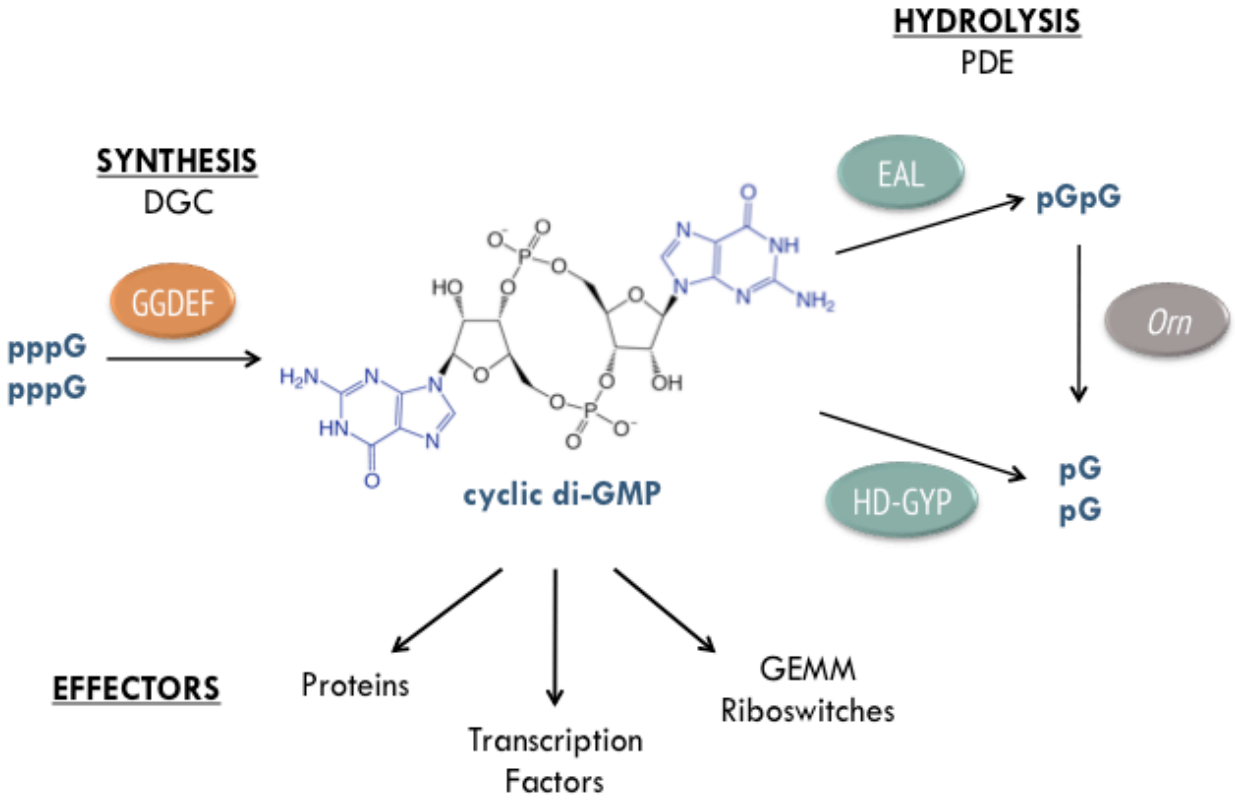


Figure 1.1 – c-di-GMP structure and pathways.

In cells, c-di-GMP is synthesized by GGDEF domain diguanylate cyclases, which take two GTP molecules (pppG) and cyclize them to create cyclic di-GMP. Two classes of enzymes then hydrolyze c-di-GMP into its constituent nucleotides: HD-GYP domain proteins directly break the molecule into two GMPs, while EAL domain proteins create pGpG, which are then further hydrolyzed into two GMPs by oligoribonucleases (*Orn*). Downstream, effectors such as proteins, transcription factors, and GEMM family riboswitches respond to changes in c-di-GMP levels with further cellular activity.

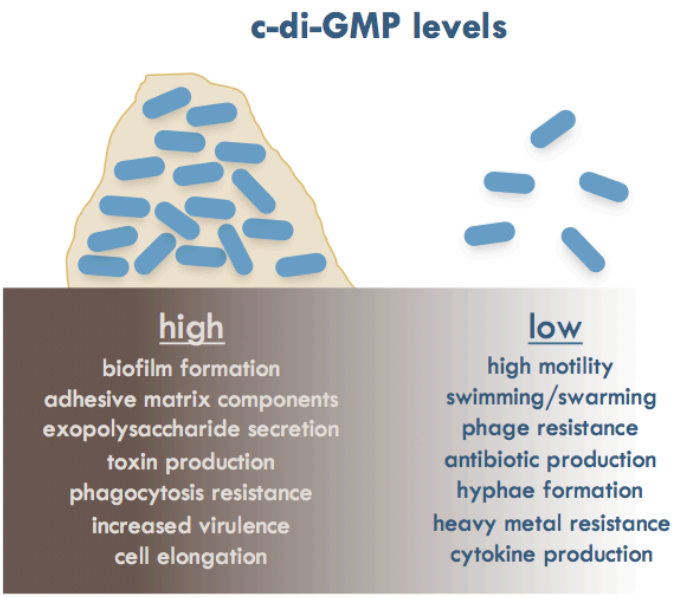


Figure 1.2 – c-di-GMP downstream signals.

In cells, high c-di-GMP concentration is associated with phenotypes promoting a sessile lifestyle, while low concentrations are associated with phenotypes promoting a motile, planktonic lifestyle. Some other downstream signals are listed above. Information adapted from Römling et al. 2013.

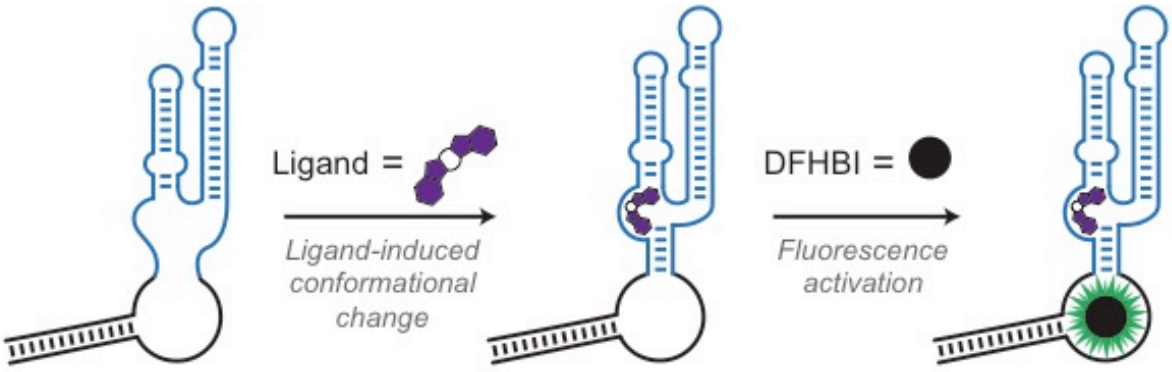


Figure 1.3 – RNA-based biosensor scheme.

RNA-based biosensors such as this Spinach-based biosensor depicted above consiste of three modules: a ligand-binding recognition domain (blue), a dye-binding domain (black), and a communication domain that transduces the ligand recognition signal to the dye-binding domain (stem between the blue and black regions that forms upon ligand binding).

		Direct ligand detection	Unlabeled substrate	Quantification of [c-di-GMP]	Dynamic	Suitable for HTS	DGC/PDE screening	Effector screening (affinity)	Sensitivity & dynamic range (M)
in vitro	**DRaCALA**	✔		✔*		✔	✔	Protein/ RNA (Y)	N.D.
	TLC	✔					✔		N.D.
	In-line probing	✔						RNA (Y)	N.D.
lysate	**G-quadruplex**	✔	✔				✔		$>10^{-6}$
	HPLC	✔	✔	✔			✔		$>10^{-9}$
	Pull-down	✔						Protein/ RNA (Y)	N.D.
in vivo	**Motility Assay**		✔				✔		N.D.
	Biofilm Assays		✔			✔	✔		N.D.
	Riboswitch Reporter		✔			✔	✔	RNA (N)	N.D.
	FRET sensor	✔	✔	✔*	✔	✔	✔		10^{-7}-10^{-6}
	RNA sensor	✔	✔	✔*	✔	✔	✔	RNA (Y)	10^{-8}-10^{-5}

* DRaCALA can approximate c-di-GMP concentrations via a ligand displacement assay. The FRET and RNA sensors can quantify c-di-GMP *in vitro*, but currently only offer approximate relative concentrations of c-di-GMP *in vivo*.

Figure 1.4 – Functional comparison of c-di-GMP detection methods.

The various c-di-GMP detection methods outlined in this chapter are summarized above. The designations at the left indicate whether the method requires purified components (*in vitro*), can be used to detect c-di-GMP in cell lysates, and can be fully genetically encodable (*in vivo*). Note that all lysate methods, as well as the FRET and RNA sensors, can also be used *in vitro*.

Direct ligand detection: do the assay results directly correalate with c-di-GMP levels?
Unlabeled substrate: can the assay proceed without covalent or radioactive ligand modifications?
Quantification of [c-di-GMP]: can the assay determine c-di-GMP concentrations?
Dynamic: can the assay detect changes in c-di-GMP levels over time?
Suitable for HTS: can the assay be adapted easily for high-throughput screening?
DGC/PDE screening: can the assay results identify DGCs and PDEs?
Effector screening (affinity): which effectors can the assay identify, and can it measure binding affinity?
Sensitivity & dynamic range: what concentrations can this assay detect?

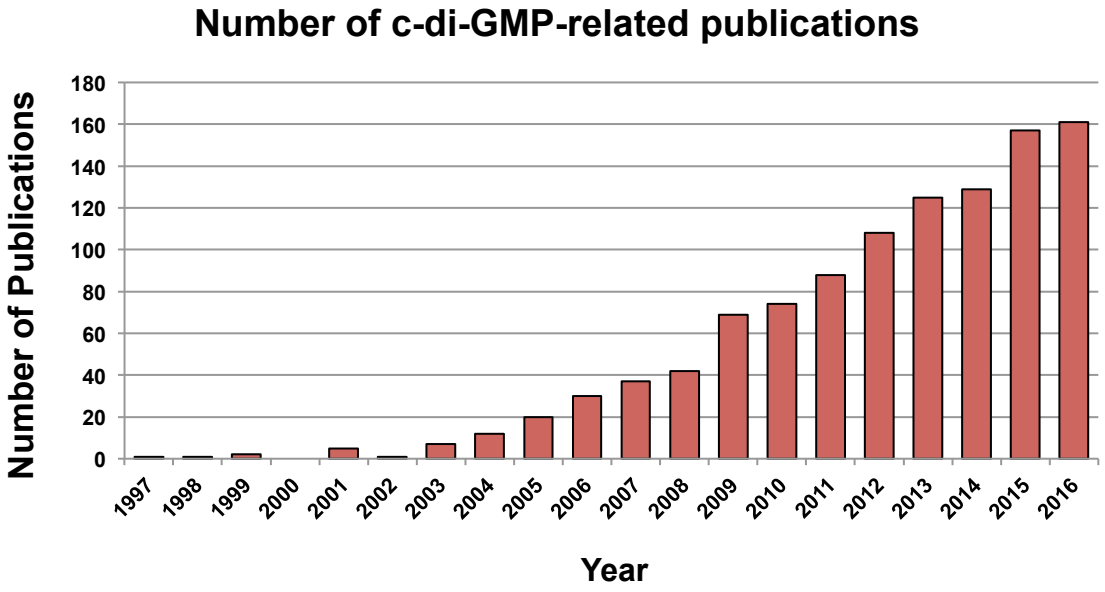


Figure 1.5 – Number of c-di-GMP-related publications, 1997-present.

Publications were counted based on searching for the following keywords on pubmed.gov: "cyclic di-gmp" OR "c-di-gmp" OR "cyclic di-guanosine" OR "cyclic diguanylate." The value for 2016 (161) was calculated based on a projection from 67 publications as of May 1, 2016.

Chapter Two

Engineering and optimizing an RNA-based biosensor for cyclic di-GMP

Portions of this chapter have been published in:

Wang, Xin C, Stephen C Wilson, Ming C Hammond. 2016. "Next-generation Fluorescent RNA Biosensors Enable Anaerobic Detection of Cyclic di-GMP." *Nucleic Acids Research,* in press.

INTRODUCTION

The bacterial second messenger cyclic di-GMP (c-di-GMP) was identified over 25 years ago as a mediator of intracellular signaling pathways. Predicted to be a signaling molecule in 75% of all sequenced bacteria (Seshasayee et al. 2010), this ubiquitous molecule regulates biofilm formation, host colonization, and bacterial virulence (Ryan 2013; Römling et al. 2013). However, the lack of a robust tool to measure c-di-GMP levels has prevented extensive and high-throughput analysis of c-di-GMP signaling pathways.

Previously, our lab developed Vc2-Spinach, a fluorescent biosensor based on the Spinach aptamer that responds to c-di-GMP (Kellenberger et al. 2013). To construct this biosensor, Kellenberger et al. replaced a stem loop of the Spinach aptamer with Vc2, a c-di-GMP-binding riboswitch aptamer from *Vibrio cholerae*. In this design, c-di-GMP binding stabilizes the Spinach binding pocket, allowing the DFHBI dye to bind and hence fluoresce. We and others have shown that Spinach-based biosensors follow a modular design consisting of three domains: the ligand-binding riboswitch aptamer, the Spinach dye-binding aptamer, and the transducer stem region at their interface (Kellenberger et al. 2013; Paige et al. 2012; Song et al. 2013). Varying any of these three domains creates biosensors with different binding and stability characteristics.

Vc2 belongs to the GEMM-I class naturally occurring c-di-GMP-binding riboswitches, a large class of over 2330 putative sequences from 641 organisms (Griffiths-Jones et al. 2003). These sequences are characterized based on their unique "tuning fork" secondary structure, with two distinct stem-loop "tines" that come together at a base (Weinberg et al. 2007; Sudarsan et al. 2008; Smith et al. 2009). GEMM-I riboswitches hence share structures but have little similarity in their nucleotide sequences, driving diversity in the pool of candidate biosensor sequences.

Here we report a series of second-generation biosensors selective for c-di-GMP that are up to 450% brighter and 13 times faster than Vc2. Together, they enable detection of c-di-GMP from picomolar to micromolar concentrations. In addition, they exhibit less sensitivity to temperature and magnesium concentration, characteristics critical for high fluorescence turn-on *in vivo*. Interestingly, our brightest biosensors have greater overall fluorescence than Spinach2, providing further insights towards biosensor design.

RESULTS

Phylogenetic screen identifies two second-generation biosensors

We began our search for improved biosensors by screening other natural riboswitches as the ligand-binding aptamer. Bioinformatics analysis of riboswitches from the c-di-GMP-binding GEMM-I class yielded a phylogenetic sequence library, from which 52 sequences were selected for screening. Corresponding riboswitch-Spinach fusions were designed, synthesized, and evaluated for fluorescence response to two concentrations of c-di-GMP (Figure 2.1). Ct256-Spinach, derived from Clostridium thermocellum, and Dp17-Spinach, derived from Deinococcus proteolyticus, were selected for high fluorescence turn-on (4.3- and 3.3-fold, respectively, with 50

μM c-di-GMP) and fast fluorescence activation. These fold activation values are impressive given the low concentrations of the biosensors (30 nM) and represent a significant improvement over Vc2-Spinach, which exhibits only 1.5-fold turn-on under identically stringent conditions (3 mM Mg^{2+} and 37 °C).

Spinach2 increases biosensor brightness proportionately

To further increase biosensor brightness, we replaced the dye-binding Spinach aptamer with the improved Spinach2 sequence, which has 1.5-fold higher fluorescence than the original Spinach aptamer (Strack et al. 2013). Maximal fluorescence for Vc2-Spinach2, Ct256-Spinach2 (referred to as Ct), and Dp17-Spinach2 (referred to as Dp) were enhanced to similar extents over their Spinach counterparts (between 1.5 and 1.9-fold) (Figure 2.2A) Furthermore, c-di-GMP binding affinity remained constant for Vc2 (Figure 2.2B). Vc2-Spinach2, Ct, and Dp showed 2.9, 6.7, and 6.0-fold fluorescence turn-on, respectively, at 30 nM biosensor and 50 μM c-di-GMP. In contrast, Spinach2 alone does not respond to c-di-GMP, nor do biosensors with mutations in the riboswitch P2' stem that disrupt its folding (Ct-M and Dp-M) (Figure 2.3).

Transducer stem screen yields two additional second-generation biosensors

The transducer pairing stem (P2) that lies at the interface between the Spinach and riboswitch aptamers affects ligand binding affinity and fluorescence turn-on (Paige et al. 2012). Thus, changing the transducer stem serves as another strategy to improve the biosensor. A survey of different P2 sequences of varying lengths and thermodynamic stabilities did not yield brighter biosensors than Ct and Dp (Figure 2.4). However, a similar survey for Pl156-Spinach, which was derived from *Paenibacillus lactis* and exhibited a 1.9-fold turn-on in the original phylogenetic screen, yielded two improved biosensors: Pl-A-Spinach2 (referred to as Pl-A), with a 7 base-pair transducer stem and 5.0-fold turn-on, and Pl-B-Spinach2 (referred to as Pl-B), with a 4 base-pair transducer stem and 4.2-fold turn-on (Figure 2.3, 2.4). Non-binding mutants of the biosensors (Pl-A-M and Pl-B-M) again showed minimal response to c-di-GMP (Figure 2.3).

Second-generation biosensors have improved in vitro chracteristics

Through screening the phylogenetic and P2 stem libraries (a total of 68 constructs), we identified four second-generation biosensors with improved brightness and fluorescence turn-on that span a range of binding affinities for c-di-GMP (Figure 2.3B, 2.5). This suite of biosensors is responsive from 10% to 90% signal from <400 pM to 2.5 μM c-di-GMP, which is about four orders of magnitude in c-di-GMP concentrations. Furthermore, these biosensors all activate in response to c-di-GMP at least 13 times more rapidly than Vc2-Spinach ($t_{1/2}$ = 1-1.5 minutes for new biosensors, $t_{1/2}$ = 20.25 min for Vc2-Spinach) (Figure 2.3C).

The biosensors are also selective towards c-di-GMP versus all other CDNs and related compounds (Figure 2.6). The two highest-affinity biosensors, Ct and Pl-B, respond weakly to high concentrations of pGpG, a linear cleavage product of c-di-GMP specific phosphodiesterases, but the biosensors maintain a >1,000-fold selectivity for c-di-GMP, as each biosensor's fluorescence response to 100 nM c-di-GMP is similar to that with 50 μM pGpG (Figure 2.6C).

These biosensors are also less affected by changes in temperature and Mg^{2+} concentration than Vc2-Spinach (Figure 2.7).

Second-generation biosensors are brighter than Spinach2

Impressively, our second-generation biosensors have maximal fluorescence levels brighter than that of Spinach2 itself (106 to 143%, Figure 2.3A). This is in stark contrast to other published Spinach-riboswitch fusion biosensors to date, all of which exhibit considerably lower fluorescence (Kellenberger et al. 2013; Kellenberger, Wilson, et al. 2015; Kellenberger, C. Chen, et al. 2015; Paige et al. 2012; Strack & Jaffrey 2013). We found that the new biosensors, when saturated with ligand, have tighter apparent binding affinities to DFHBI than Spinach2 alone (Figure 2.8), which has the effect of increasing dye occupancy. While increased brightness of Spinach2 was attributed to higher folding stability relative to Spinach (Ren et al. 2015), we did not find a consistent trend in analyzing the folding stabilities of the biosensors versus Spinach2 (Figure 2.9).

Taken together, these facts suggest that variations in maximal brightness derive from differences in extinction coefficients and quantum yields, as well as dye occupancy. We tested whether these effects were solely attributable to the transducer stem sequence by replacing the P2 stem of Spinach2 with stem sequences from the four biosensors and a GAAA loop in place of the riboswitch aptamer (Figure 2.10). The results were mixed, as the P2 variants for Ct and Dp are considerably less fluorescent than the corresponding biosensors, whereas P2 variants for Pl-A and Pl-B exhibit similar yet lower brightnesses than the biosensors (Figure 2.10C). It is likely that sequence determinants outside of the transducer stem and/or the global riboswitch fold dramatically impact biosensor performance.

DISCUSSION

This latest series of RNA-based fluorescent biosensors significantly improve upon the original Vc2 biosensor towards detecting c-di-GMP in live cells and further reinforces several design principles for creating riboswitch-based Spinach biosensors. Namely, our results demonstrate that sampling the phylogenetic diversity of natural riboswitches provides an efficient strategy for surveying functional sequence space, similar to natural enzyme variants forming the starting point for directed evolution in protein engineering (Lutz 2010). The aptamer portions of the four second-generation biosensors share only 37-66% nucleotide sequence identity and include several insertions and deletions, a level of diversity not easily achieved via random mutagenesis (Figure 2.11). This type of sequence diversity would be present in a random sequence library, but the phylogenetic approach is considerably more efficient, as the 68 sequences we assayed to efficiently arrive at four biosensors that all function *in vivo* is considerably less than the total number of unique sequences in a corresponding random library.

Our biosensors maintain high selectivity towards cyclic di-GMP against similar ligands. The highest-affinity biosensors, Ct and Pl-B, also respond to the phosphodiesterase product pGpG, a promiscuity that could interfere with c-di-GMP measurement accuracy in cells that have high levels of pGpG. The PA14 strain of *Pseudomonas aeruginosa*, for instance, has high μM

concentrations of pGpG and 20-fold lower concentration of c-di-GMP, while mutants deficient for oligoribonuclease (*orn*), which hydrolyze pGpG into two GMP molecules, have even higher concentrations of pGpG (Orr et al. 2015). However, other strains of *P. aeruginosa* have undetectable levels of pGpG (Cohen et al. 2015). Furthermore, our biosensors are also the first fluorescent biosensors for pGpG, which could prove useful as an *in vitro* tool for measuring the activity of *orns* and other pGpG-metabolizing enzymes.

We recently described first-generation RNA-based fluorescent biosensors for live cell imaging of other signaling molecules, including c-di-AMP (Kellenberger, C. Chen, et al. 2015) and c-AMP-GMP (Kellenberger, Wilson, et al. 2015). Our experience with designing, synthesizing, and now optimizing these types of biosensors have established the following general principles: 1) accurate knowledge of aptamer secondary structure is important for rational design of the transducer stem (Kellenberger, C. Chen, et al. 2015); 2) stringent screening conditions, which includes accurate measurement of RNA concentrations (Wilson et al. 2014), are critical to comparing and optimizing biosensor performance; and 3) sampling diverse sequences enables different solutions to the folding problem, yielding better biosensors (this study). The strong performance characteristics displayed by these second-generation c-di-GMP biosensors, along with the potential to engineer riboswitch- or aptamer-based biosensors for other small molecules or proteins, should inspire others to join in the use and development of this promising biosensor technology.

MATERIAL AND METHODS

General reagents and oligonucleotides

CDNs used in this study were purchased from Axxorra, LLC (Farmingdale NY). DFHBI and DFHBI-1T were synthesized as previously described (Kramer & Mills 1981; Paige et al. 2012; Song et al. 2014) and stored as a ~30 mM stock in DMSO. All GEMM-I-Spinach DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT, Coralville, IA), while other oligonucleotides were purchased from Elim Biopharmaceuticals (Hayward, CA). All oligos are listed in Table 2.1.

Bioinformatic analysis of GEMM-I variants

The GEMM-I riboswitch aptamer variants employed in the phylogenetic screen were selected as previously described (Kellenberger, Wilson, et al. 2015). Briefly, sequences were extracted from Rfam database (accession RF01051, http://rfam.xfam.org/), and were ranked, sorted, and selected with respect to various criteria including but not limited to: folding stabilization energies, presence of specific c-di-GMP binding pocket residues, host organism and source, evolutionary position, tractability of the P1 stem, and downstream genes. The phylogenetic sequences themselves are listed in Table 2.2, while the Spinach flanking sequences Table 2.3.

In vitro fluorescence assays

All biosensor and Spinach RNAs used for *in vitro* fluorescence activation assays were prepared as previously described (Kellenberger & Hammond 2015). Briefly, DNA templates were first amplified with the appropriate Spinach or Spinach2 primer pairs. Transcriptions were performed using T7 RNA polymerase and the RNA product purified by a denaturing (7.5 M urea) 6% PAGE gel. RNA was eluted from the gel, precipitated, dried, and resuspended in water. Accurate RNA quantitation was obtained by thermal hydrolysis.

Fluorescence activation assays were performed as previously described (Kellenberger et al. 2013; Kellenberger, Wilson, et al. 2015; Ren et al. 2015). Briefly, each reaction consisted of RNA, ligand, and DFHBI in a binding buffer consisting of 40 mM HEPES, 125 mM KCl, and 3 or 10 mM $MgCl_2$ at pH 7.5. RNA was refolded in binding buffer before being added to the binding reaction. The reaction plate was incubated at the appropriate temperature and fluorescent measurements were taken on a SpectraMax Paradigm plate reader (Molecular Devices) at 448 nm excitation / 506 nm emission. Reported fluorescence values are for reactions that have reached equilibrium, as defined by fluorescence level stabilization over time. For ligand selectivity experiments, fluorescence values were normalized to biosensor with c-di-GMP. For DFHBI titration experiments, fluorescence values were background-subtracted, with background defined as which is defined as fluorescence of buffer, ligand, and DFHBI without RNA. All other values are reported in raw form.

In vitro fluorescence turn-on kinetics

For kinetics experiments, a reaction containing 10 μM DFHBI and 50 μM c-di-GMP in binding buffer was pre-incubated in the dark to 30 ˚C. RNA was refolded in binding buffer and pre-incubated separately to the same temperature. Fluorescence measurements were taken every 15 seconds as described above, starting immediately after the addition of RNA to the reaction mixture. There was an approximate dead time of 15 seconds between RNA addition and the first fluorescence reading. Fluorescence values were then normalized against maximum fluorescence exhibited for each biosensor.

Melt curves

For thermostability measurements, each sample reaction consisted of 100 nM biosensor or Spinach2 RNA, 10 μM DFHBI, and 50 μM c-di-GMP in a buffer containing 40 mM HEPES, 125 mM KCl, and 3 or 10 mM $MgCl_2$ at pH 7.5. The temperature was then increased from 20˚C to 60˚C in 1˚C increments every 5 minutes, with fluorescence measured at each temperature with the SYBR Green channel of a CFX96 Thermal Cycler (Bio-Rad). Melting temperatures were calculated using the Bio-Rad CFX Manager software (Version 1.5.534.0511) and verified by calculating the inflection point of the first derivative in GraphPad Prism 6 software.

FIGURES

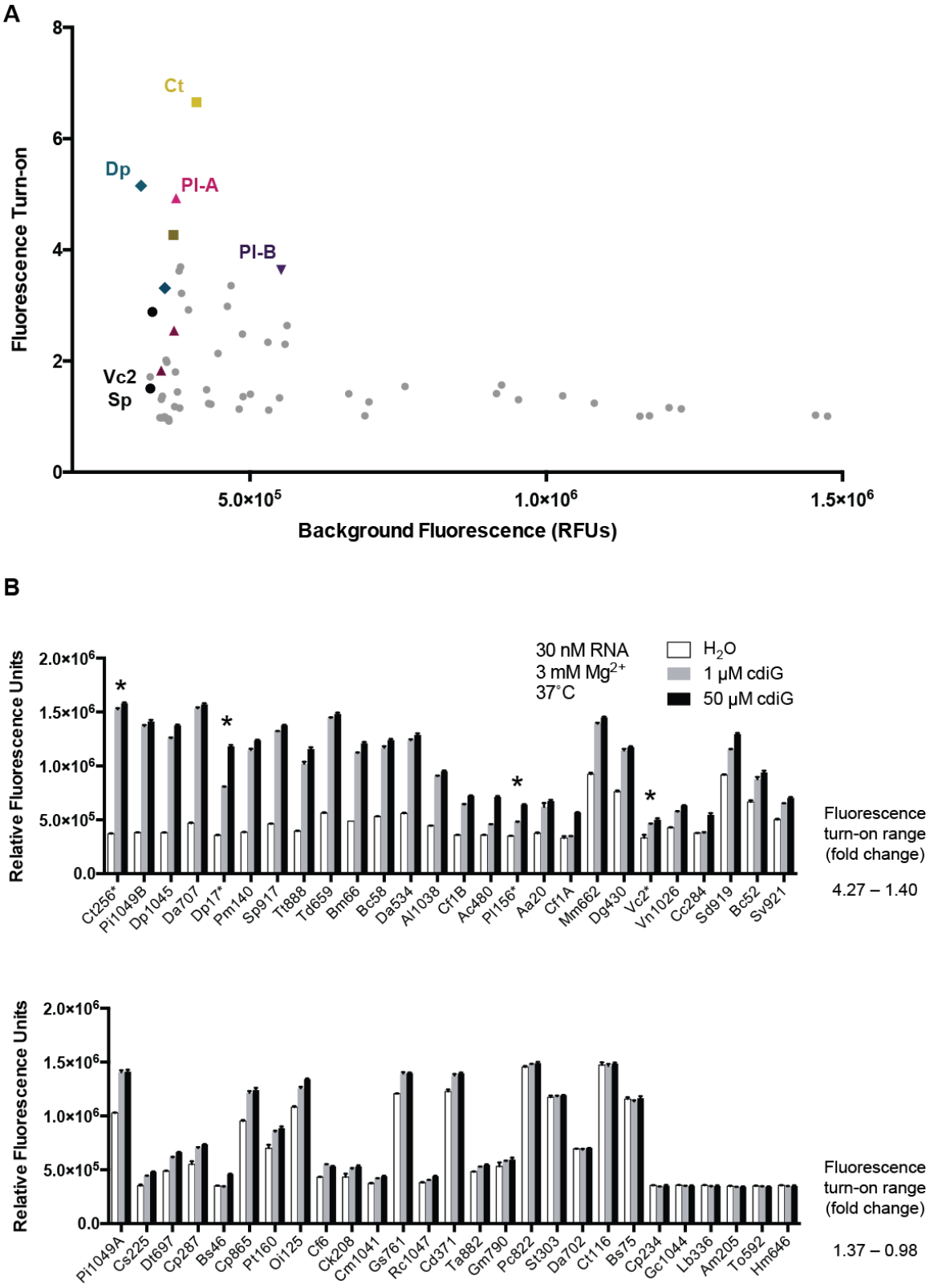

A
Fluorescence Turn-on
Background Fluorescence (RFUs)
Ct
Dp
PI-A
PI-B
Vc2
Sp
B
Relative Fluorescence Units
30 nM RNA
3 mM Mg2+
37°C
H2O
1 μM cdiG
50 μM cdiG
Fluorescence turn-on range (fold change)
4.27 – 1.40
1.37 – 0.98

Figure 2.1 – Phylogenetic screen of GEMM-I riboswitch aptamers fused to Spinach.

A. Select data from the original GEMMI-Spinach phylogenetic screen (Supplementary Figure S2) plotted with respect to background fluorescence (x-axis), defined as fluorescence with no c-di-GMP, versus fluorescence turn-on (y-axis), defined as ratio of fluorescence with 50 μM c-di-GMP over with no c-di-GMP. The original Vc2-Spinach and biosensors chosen for additional analysis and development (Ct256, Dp17, Pl156) are labeled as shown.
B. A total of 52 GEMM-I-Spinach biosensors were assayed *in vitro* for their relative fluorescence in the presence of 0 (H_2O), 1, and 50 μM c-di-GMP. The biosensors are displayed from left to right, top to bottom, in order of decreasing fluorescence turn-on, as defined by the relative fluorescence at 50 μM c-di-GMP divided by that at 0 μM c-di-GMP. The asterisks indicate biosensors upon which additional analysis was performed (Ct256, Dp17, Pl156), as well as the first-generation sensor Vc2. Error bars represent the standard deviation of three independent experiments with duplicate samples.

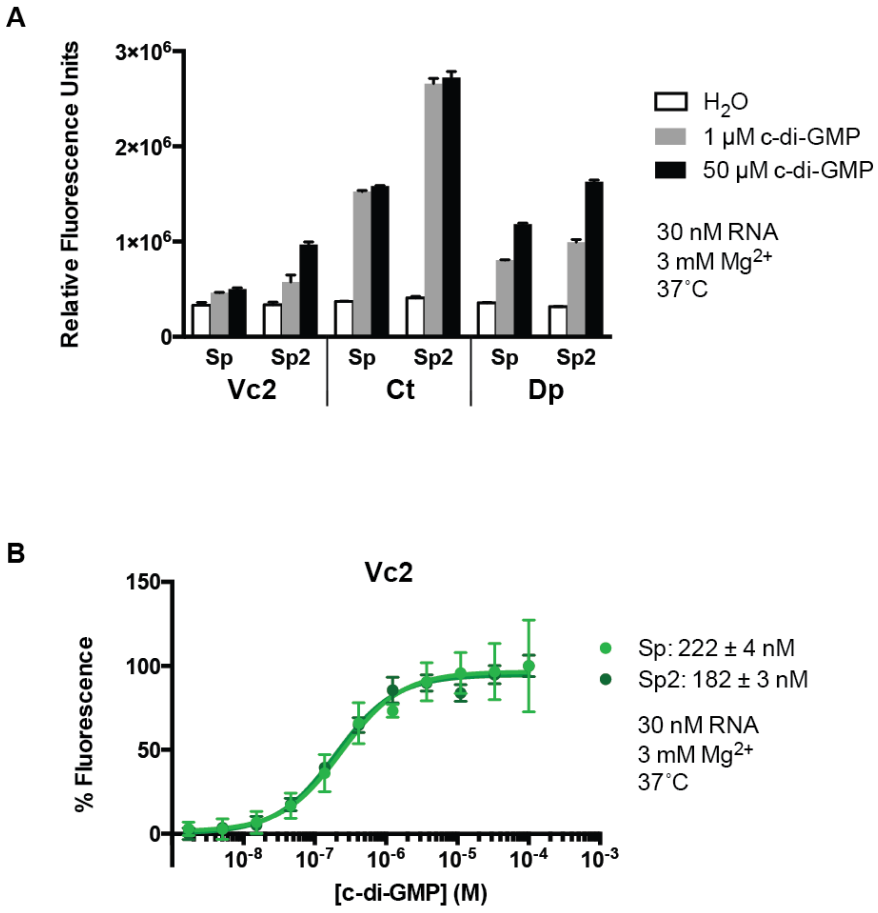


Figure 2.2 – Spinach vs. Spinach2 biosensors.

A. *In vitro* fluorescence activation of biosensors with different riboswitch aptamers (Vc2, Ct, and Dp) fused to either Spinach (Sp) or Spinach2 (Sp2), in the presence of 0 (H_2O), 1, and 50 μM c-di-GMP. Error bars represent the standard deviation of three independent experiments with duplicate samples.

B. Ligand (c-di-GMP) affinity of Vc2-Spinach (Sp, light green) vs. Vc2-Spinach2 (Sp2, dark green). Points and error bars represent averages and standard deviations, respectively, from three independent replicates with duplicate samples. Best-fit curves are also shown, and the K_Ds of each biosensor are listed in the legend.

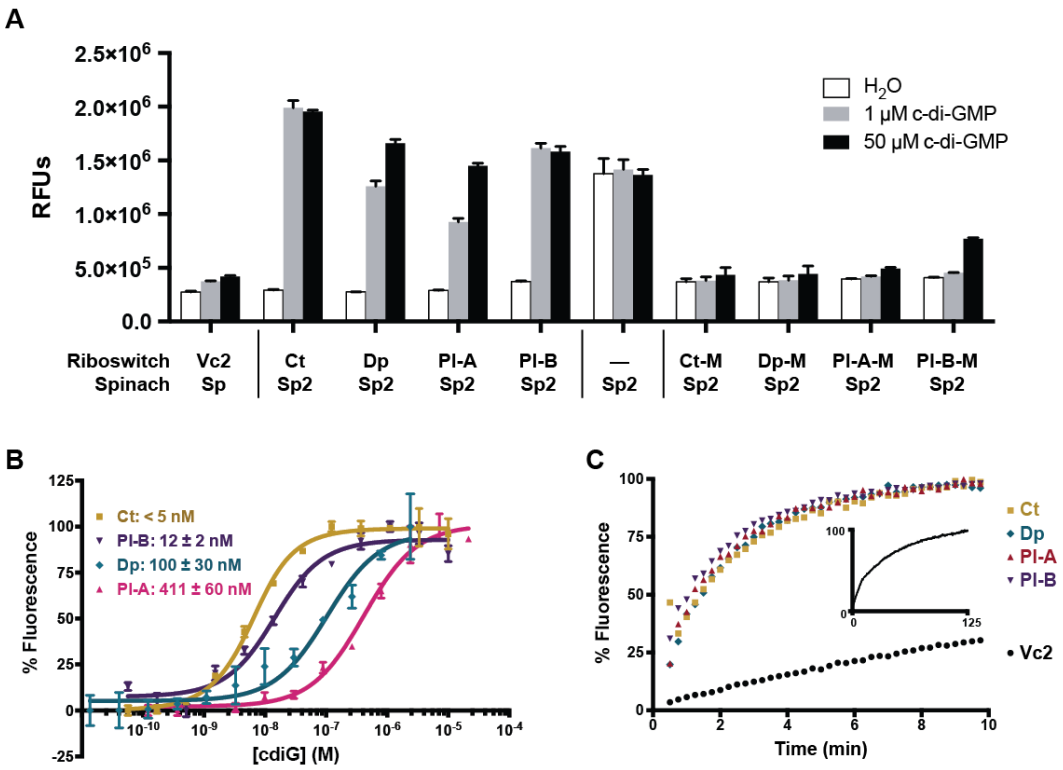


Figure 2.3 – *in vitro* fluorescence properties of second-generation biosensors.

A. Fluorescence activation of different biosensors constructs and Spinach2, with 0, 1 μM, and 50 μM c-di-GMP. Data are from 3 independent replicates represented as mean ± SD.

B. Biosensor binding affinity measurements are shown as % maximum fluorescence, which is normalized for each biosensor, with titration of c-di-GMP at 3 mM Mg^{2+}, 37 ˚C. Dissociation constant values are shown in the legend. Data are from 3 independent replicates represented as mean ± SD.

C. Biosensor fluorescence turn-on kinetics are shown as % maximum fluorescence, which is normalized for each biosensor. Data are from 3 independent replicates and the mean is shown. Error bars were omitted for clarity. Inset shows data for Vc2 over a longer time scale (125 minutes).

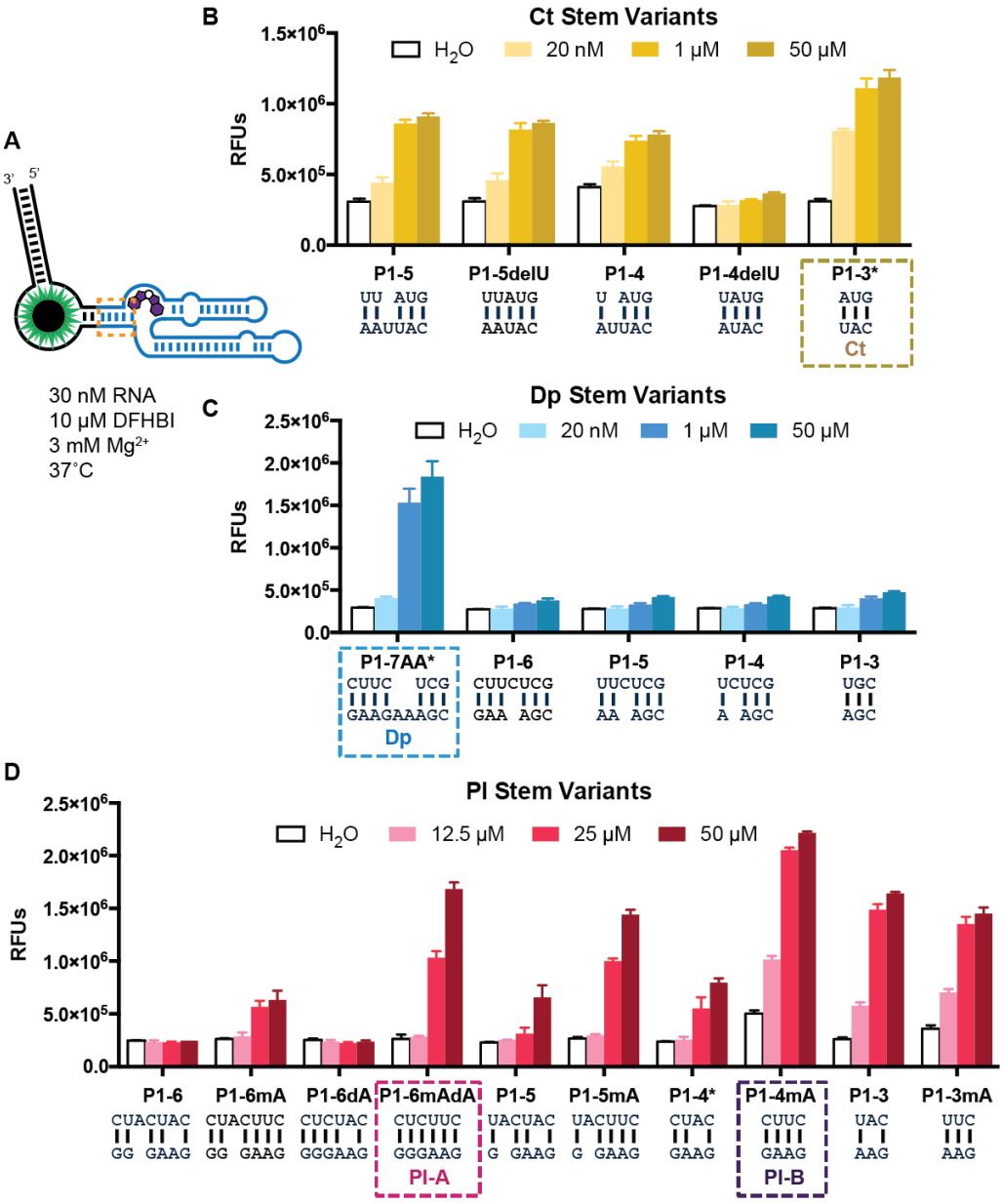


Figure 2.4 – Transducer stem variants.

A. Schematic diagram of biosensor, with Spinach segment in black, the riboswitch aptamer in blue, and the transducer stem region outlined with an orange dotted rectangle. The orientation of the RNA strand (5' and 3' ends) are shown. Experimental conditions for parts (**B**)-(**D**) are shown underneath.

B. Fluorescence turn-on of the Ct-Spinach2 biosensor stem variants with 0 (H_2O), 20 nM, 1 μM, and 50 μM c-di-GMP. The stem sequences are listed, with the top strand in the 5' to 3' direction, as if superimposed on the transducer stem region in (A). Base-paired interactions are denoted by a vertical line. The original WT stem is denoted with an asterisk, while the final Ct biosensor, P1-3, is boxed.

C. As (B), but with Dp-Spinach2. The final Dp biosensor, P1-7AA, is boxed.

D. As (B), but with Pl-Spinach2 and 0 (H_2O), 12.5 μM, 25 μM, and 50 μM c-di-GMP. Pl-A is denoted with a pink dotted box, and Pl-B is denoted with a purple dotted box.

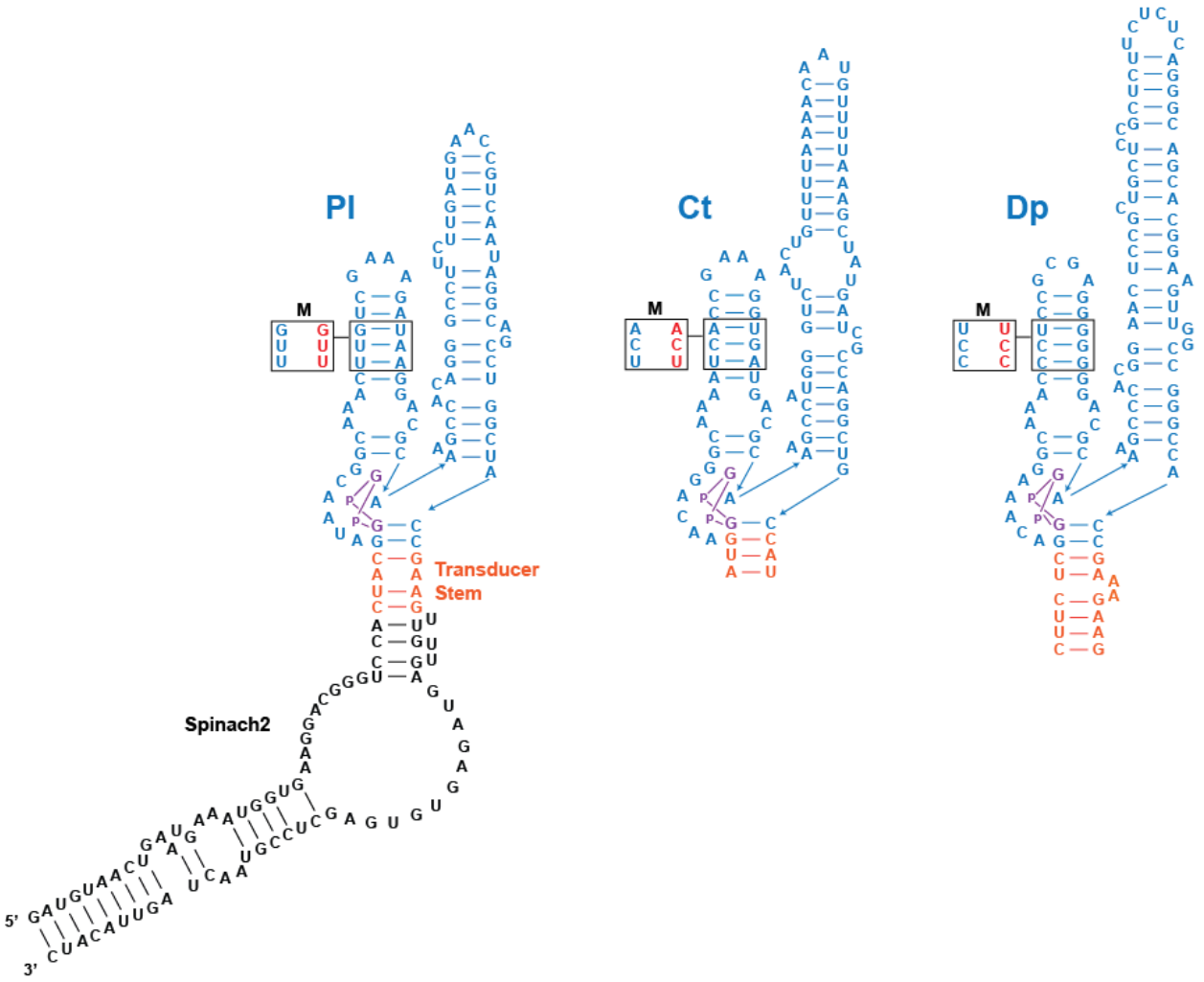


Figure 2.5 – Secondary structure diagrams.

Sequence and secondary structure of Spinach-riboswitch fusions. The Spinach2 sequence is in black, the transducer stem (depicted sequences are the WT sequences used in the original screen) in orange, and the riboswitch aptamer sequence in blue. The bound ligand is denoted in purple, and the identity of the non-binding M sequences are boxed. The Ct and Dp sequences are depicted without the Spinach2 aptamer sequence.

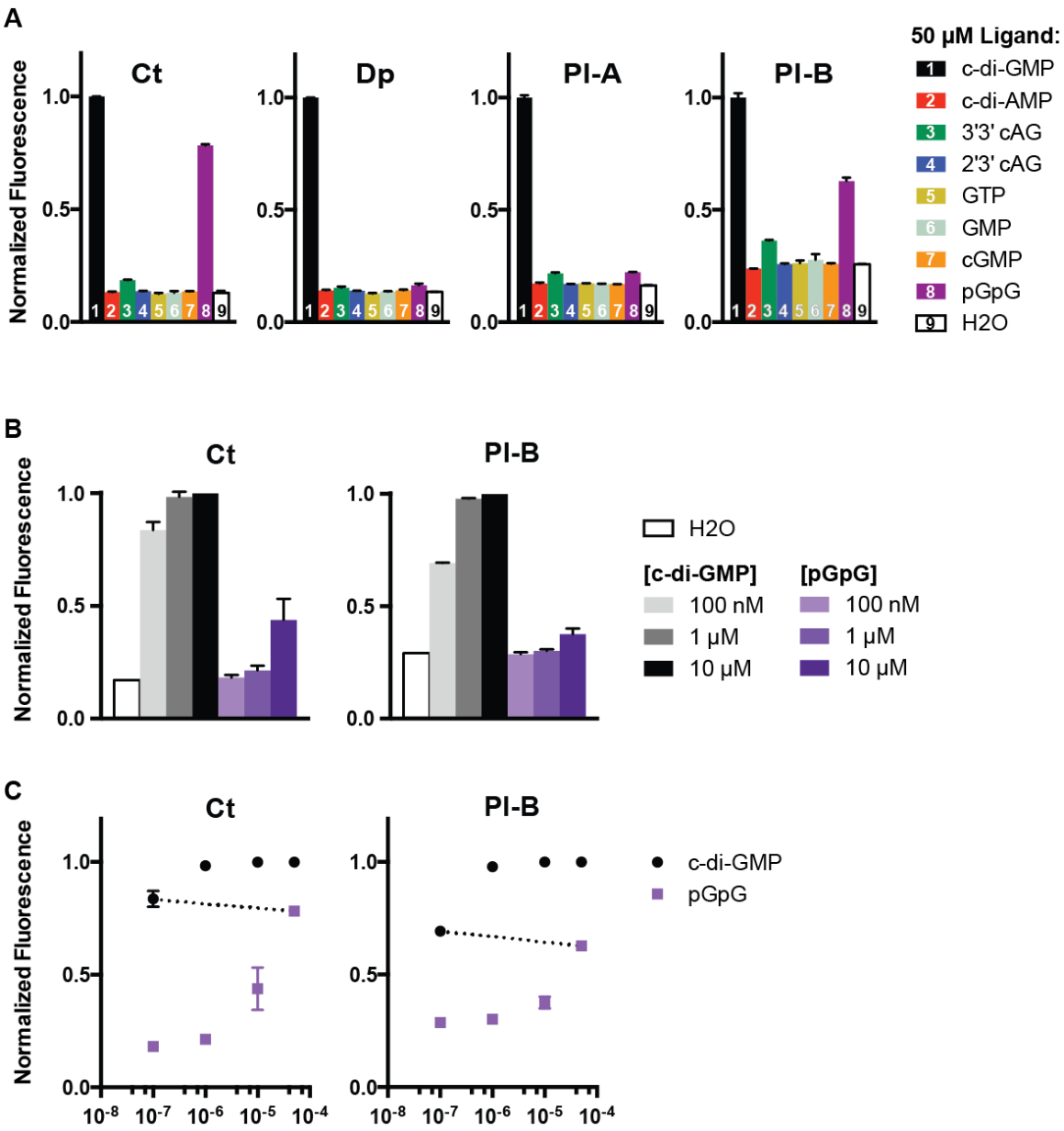
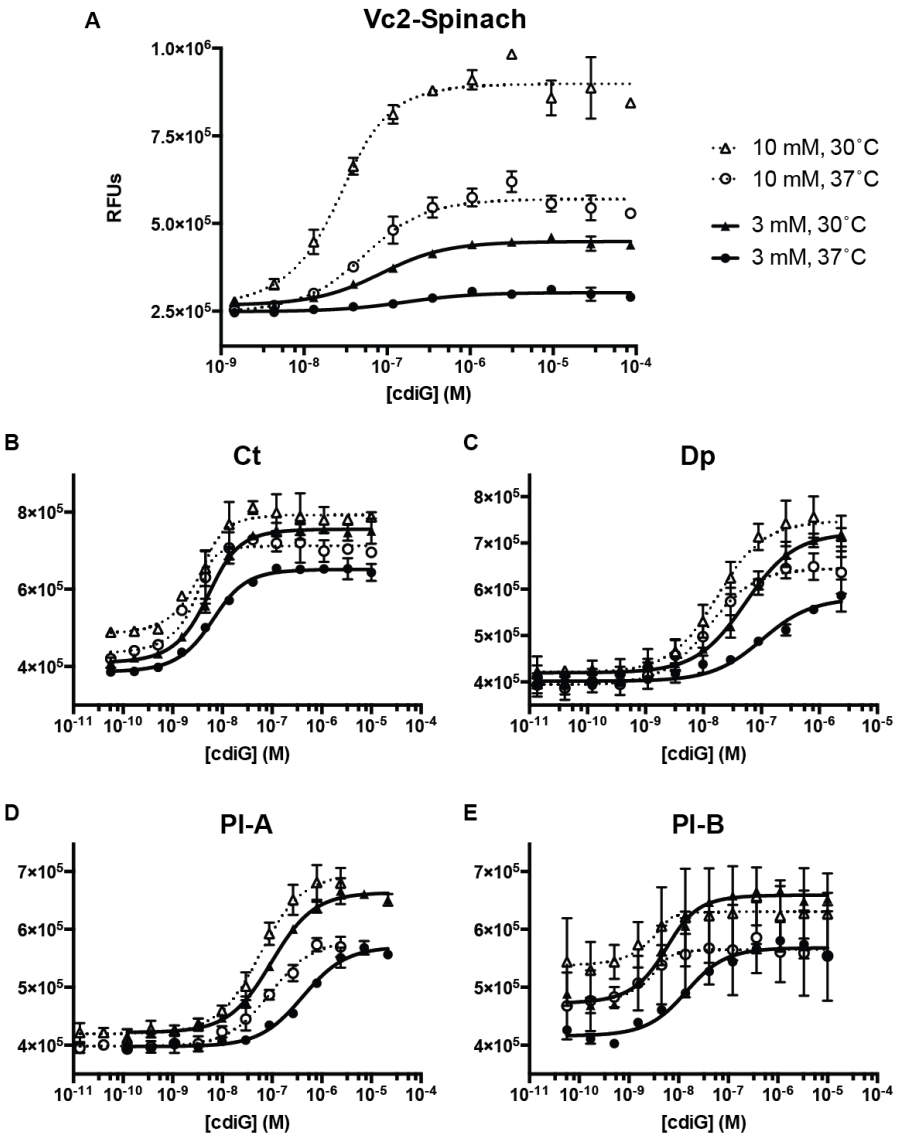


Figure 2.6 – Biosensor ligand selectivity.

A. Relative fluorescence of the four second-generation biosensors with 50 μM of various ligands: c-di-GMP (1, black), c-di-AMP (2, red), 3'3' cGAMP (3, green), 2'3' cGAMP (4, blue), GTP (5, goldenrod yellow), GMP (6, light green), cGMP (7, orange), pGpG (8, purple), and no ligand (9, white). For each biosensor, fluorescence levels were normalized to signal with c-di-GMP.

B. Relative fluorescence of Ct (left) and Pl-B (right) with 100 nM, 1 μM, and 50 μM c-di-GMP (light gray, dark gray, and black, respectively) and pGpG (light purple, purple, and dark purple, respectively). For each biosensor, fluorescence levels were normalized to signal with 50 μM c-di-GMP.

C. Same data as part b and 50 μM ligand data from part a, but with relative fluorescence of c-di-GMP (black circles) and pGpG (purple squares) on the same axis. The dotted lines connect the biosensor fluorescence response to 100 nM c-di-GMP and 50 μM pGpG.

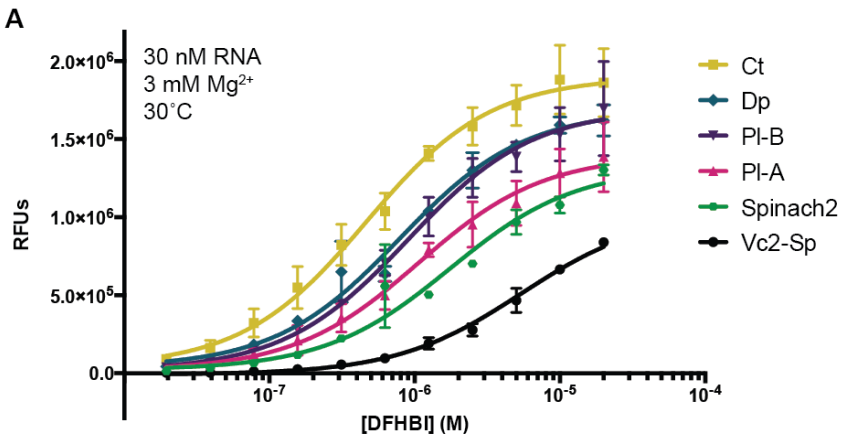


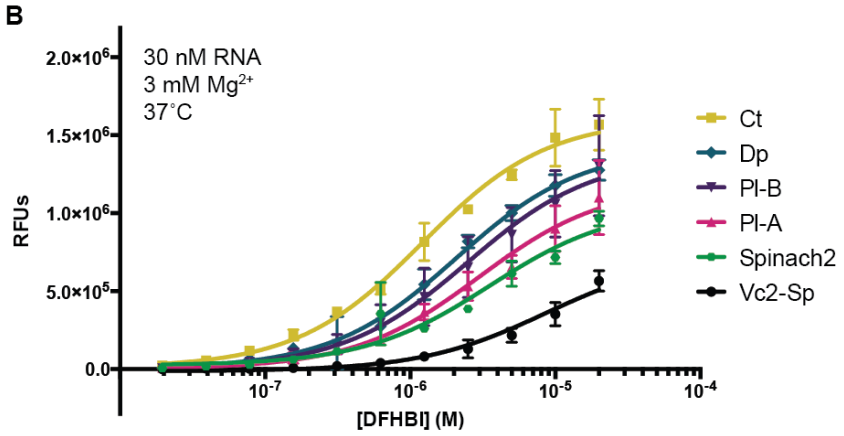
F

[Mg]	Temp	Vc2	Ct	Dp	Pl-A	Pl-B
10 mM	30˚C	12 ± 2	< 5	15 ± 6	55 ± 10	< 5
	37˚C	38 ± 7	< 5	10 ± 27	89 ± 58	< 5
3 mM	30˚C	69 ± 2	< 5	50 ± 4	94 ± 8	< 5
	37˚C	150 ± 4	< 5	100 ± 3	411 ± 15	12 ± 1

Figure 2.7 – Biosensor fluorescence with c-di-GMP titration.

Titration of c-di-GMP with the biosensor and DFHBI in varying temperature (30˚C – triangles; 37˚C – circles) and buffer (3 mM Mg^{2+} – filled shapes; 10 mM Mg^{2+} – empty shapes) conditions. Each panel displays a separate biosensor (**A.** Vc2; B**.** Ct; **C.** Dp; **D.** Pl-A; **E.** Pl-B). Points and error bars represent averages and standard deviations of three independent trials; best-fit curves are also shown (3 mM Mg^{2+} – solid lines; 10 mM Mg^{2+} – dotted lines). All assays included 10 μM DFHBI and 30 nM (Vc2) or 5 nM (Ct, Dp, Pl-A, Pl-B) RNA. A table of the K_D values, in nM, are displayed in **(F)**.





C

	Ct	Dp	Pl-B	Pl-A	Spinach2	Vc2-Sp
30˚C	448 ± 2	770 ± 2	850 ± 2	1051 ± 3	1736 ± 4	> 5422
37˚C	1190 ± 2	2055 ± 2	1370 ± 4	> 2972	> 3530	> 8570

Figure 2.8 – Biosensor fluorescence with DFHBI titration.

Relative fluorescence of biosensors (Ct: goldenrod squares; Dp: teal diamonds; Pl-B: purple downward-pointing triangles; Pl-A: pink upward-pointing triangles; Spinach2: green hexagons; Vc2: black circles) with varying concentrations of DFHBI, assessed at **(A)** 30˚C and **(B)** 37˚C. Values displayed are background-subtracted, as determined by a no-RNA control sample. Points and error bars represent averages and standard deviations of three independent trials; best-fit curves are also shown. A table of K_D values (nM) of each biosensor for DFHBI, at 30˚ and 37˚C, is displayed in **(C)**.

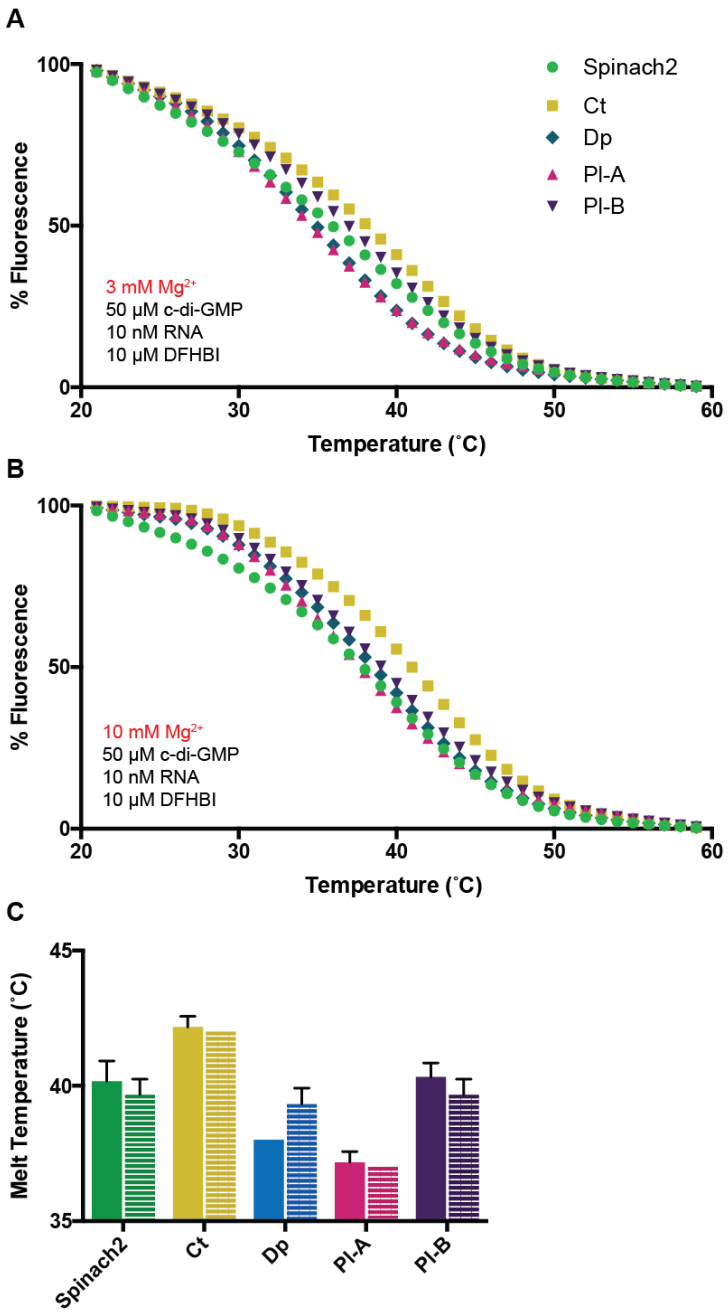


Figure 2.9 – Biosensor fluorescence with temperature ramp

Relative fluorescence of Spinach2 and biosensors (Spinach2: green circles; Ct: goldenrod squares; Dp: teal diamonds; Pl-A: pink upward-pointing triangles; Pl-B: purple downward-pointing triangles) under saturating conditions of c-di-GMP and DFHBI, assessed over a temperature range between 20-60˚C with **(A)** 3 mM Mg^{2+} and **(B)** 10 mM Mg^{2+}. Values displayed are normalized and are from one representative trial. The actual melt temperatures, as defined by the inflection point in the melt curve, are graphed in **(C)**, with data from 3 mM Mg^{2+} (solid bars, left) and 10 mM Mg^{2+} (with horizontal lines, right). Values represent averages and standard deviations of three independent trials.

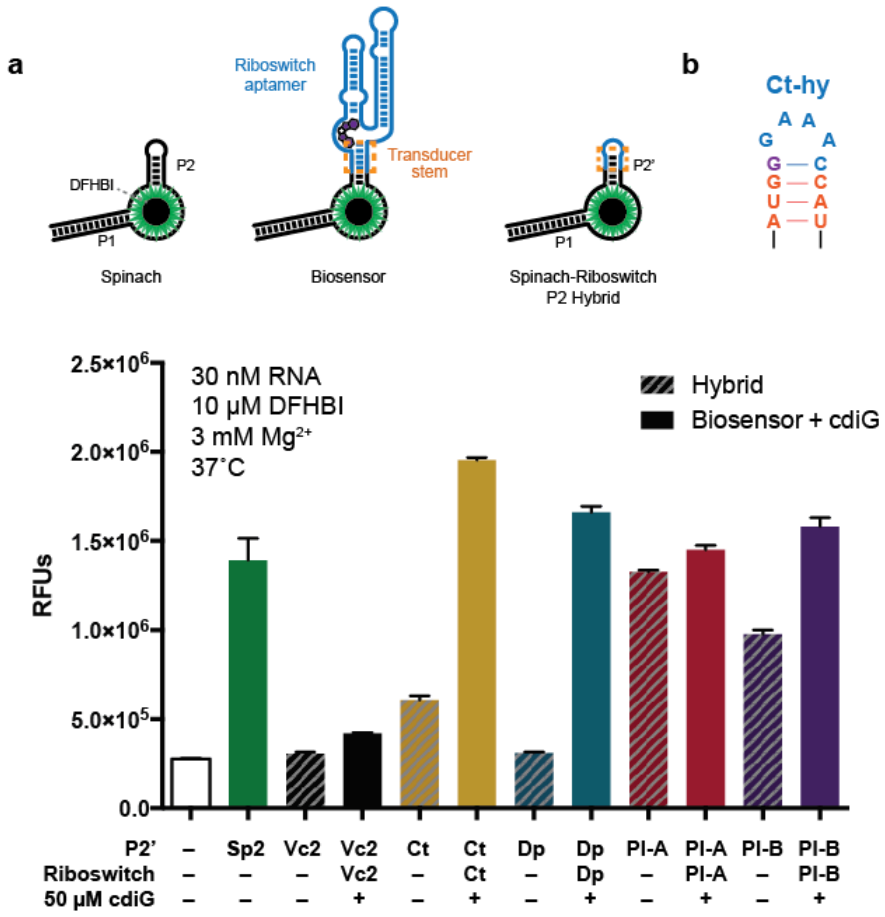


Figure 2.10 – Spinach-P2 hybrid variants

A. Schematic comparing Spinach (left), the biosensor (middle), and the Spinach-riboswitch P2' hybrid RNAs. The P2' hybrids are constructed by replacing the Spinach P2 stem with the riboswitch aptamer P1 stem.
B. Secondary structure and sequence of the Ct hybrid, Ct-hy. The transducer stem nucleotides are in orange, and the GAAA loop is in blue. The G-C base pair between the transducer stem and the GAAA loop derives from the base pair between the ligand G and aptamer C, and was included in each of the hybrid designs. The black lines at the bottom indicate the positions where Spinach was attached.
C. Relative fluorescence of P2' hybrid RNAs (striped) against a no RNA control (white), Spinach2 (green), and the biosensor at maximal fluorescence (solid bars). All data are averages of three independent replicaties.

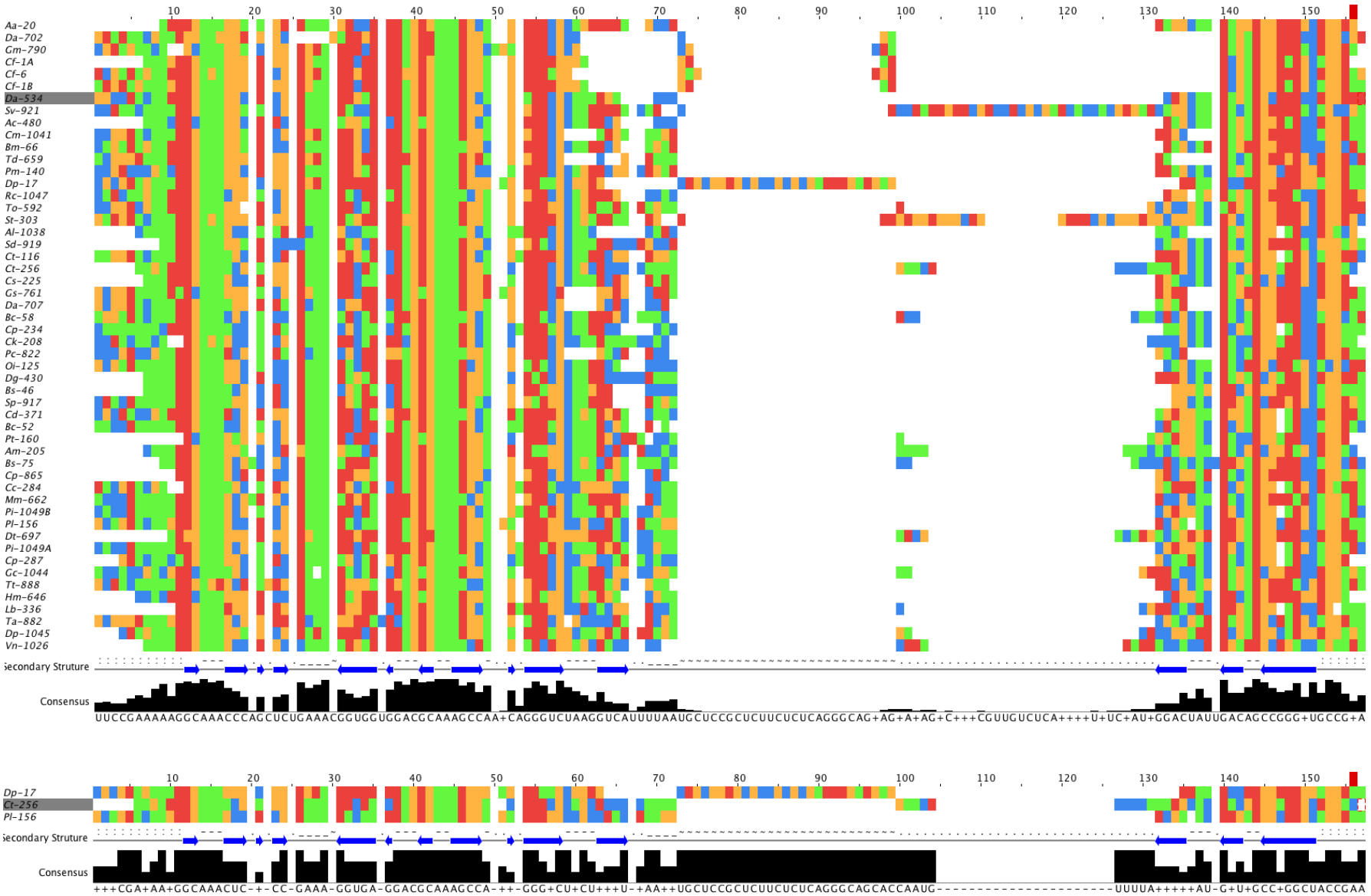

10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
Aa-20
Da-702
Gm-790
Cf-1A
Cf-6
Cf-1B
Da-534
Sv-921
Ac-480
Cm-1041
Bm-66
Td-659
Pm-140
Dp-17
Rc-1047
To-592
St-303
Al-1038
Sd-919
Ct-116
Ct-256
Cs-225
Gs-761
Da-707
Bc-58
Cp-234
Ck-208
Pc-822
Oi-125
Dg-430
Bs-46
Sp-917
Cd-371
Bc-52
Pt-160
Am-205
Bs-75
Cp-865
Cc-284
Mm-662
Pi-1049B
Pl-156
Dt-697
Pi-1049A
Cp-287
Gc-1044
Tt-888
Hm-646
Lb-336
Ta-882
Dp-1045
Vn-1026
Secondary Struture
Consensus
UUCCGAAAAAGGCAAACCCAGCUCUGAAACGGUGGUGGACGCAAAGCCAA+CAGGGUCUAAGGUCAUUUUAAUGCUCCGCUCUUCUCUCAGGGCAG+AG+A+AG+C+++CGUUGUCUCA++++U+UC+AU+GGACUAUUGACAGCCGGG+UGCCG+A
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
Dp-17
Ct-256
Pl-156
Secondary Struture
Consensus
+++CGA+AA+GGCAAACUC-+-CC-GAAA-GGUGA-GGACGCAAAGCCA-++-GGG+CU+CU+++U-+AA++UGCUCCGCUCUUCUCUCAGGGCAGCACCAAUG----------UUUUA+++++AU-G+U+GCC+GGCUACCGAA

Figure 2.11: Sequence analysis of GEMM-I riboswitches

The phylogenetic variants used are presented aligned by secondary structure and colored by nucleotide. Red = A, blue = T, green = C, yellow = G, white = no base. Arrows denoting the secondary structure are shown in blue at the bottom, with opposing arrows indicating paired regions. All 52 sequences are indicated in the upper diagram, while the 3 sequences that were used for the second-generation sensors are isolated in the lower diagram.

Sequences were aligned by secondary structure with Infernal (Paul et al. 2010; Nawrocki & Eddy 2013) using the GEMM-I covariance model provided in Rfam, accession RF01051 (Griffiths-Jones et al. 2003)). Visualization was performed with JalView (Waterhouse et al. 2009).

TABLES

Table 2.1 – Flanking Sequences.

All riboswitch-Spinach fusions were flanked by the following Spinach or Spinach2 sequences on their 5' and 3' sides.

Sp 5′	GACGCGACTGAATGAAATGGTGAAGGACGGGTCCA
Sp 3′	TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTCGCGTC
Sp2 5′	GATGTAACTGAATGAAATGGTGAAGGACGGGTCCA
Sp2 3′	TTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC

Table 2.2 – Phylogenetic Sequences.

Riboswitch sequences used in the phylogenetic analysis of GEMM-I riboswitch-Spinach fusions. Each sequence listed here was ordered as an ultramer flanked by the Sp sequences listed in Table 1.1.

Name	Sequence (5'→3')	Accession ID	Range
Aa-20	GCGTTTCAGGGCAAACCAACGGAAACGTTGGGACGCAAAGCTACGGGTCTACGGGGACTTGGACCTAAGACCGCCGGGCTGCCGC	ACCS01000010.1	47422-47506
Ac-480	TTTGAAATGGTAAACCTGGTGAAAACCAGTGACACAAAGCTACGGGTCTAAGGTCTTTGACTAAGACAGCCGAGTTGCCGAA	AEDB01000061.1	5163-5244
Al-1038	GATTAAAAGGCAAACTTAAGGTAACTTAAGGACGCAAAACTAAAGGGTCTAATTAGTAATAGACAGCCAGTTGCATC	CP000896.1	1029331-1029407
Am-205	ATGTAAAGGCAAACCATTGCAAACAATGGGACGCAAAGCCAGGAACCTAAAGTGTGTTATAAAAAAATATACCAAGATCGTCCGACTGCCAT	CP000724.1	3839671-3839762
Bc-52	AAAATTGAAAAAGGCAAATTCATCGAAAGGTGGAGACGCAAAGCTAAAGGGACTAAAGTCAGATGACCATGTCAGCCAGTTACCGATTTT	CP003056.1	2189769-2189858
Bc-58	CCTACCAGATAAAGGCAAATCTATTGAAAAGTAGAGGCGCAAAACTACGGATCTAAGGGCTAAATGTTTAATGTCTATGATAGCCGGGTTACCTAGTAGG	CP002394.1	854052-854151
Bm-66	TGTACAGACAAGGGCAAACCAGTTGAAAGGCTGGGACGCAAAACCTCGGGTCTAAGGTCACAGGACTAGGACGGCCGGGTTTCCTGATACA	CP003017.1	3889771-3889861
Bs-46	TTAGAAAAGGCAAAATCTGTGAAATCAGATGACGCAAAGCCACGGACCTAACGGTTTTCCCACGGTCGCCGGGCTACCAAA	CP001791.1	2058948-2059028
Bs-75	TTATAGAAGGCAAACTCATCTGAAAAGGGAGGACGCAAAGCCACGGGCCTACATGCAAAATATTATTTGTATATTGGCAGCCGGGTTACCTGTAG	ABCF01000033.1	2270-2364
Cc-284	GATCGATCAGCAAAACTAGCGAAAGCTAGTGACGCAAAGCTACAGGGATTTCCCCTTTTAACAGGGATGTCAGCCAGC	ADLJ01000004.1	176355-176441

	TGCAGGATT		
Cd-371	TGGTATCTGATTCAGGGCAAAGTCGCCGAAAGGTGACGGCGCAAAACTAGAGGGGCTACAGCGATAATACGCCAAGCCAGCCAGTTGCCGGATATCA	CP000860.1	1941826-1941922
Cf-1A	CGACAAACGGCAAACCCGCCGCAAGGTGGGGACGCAAAGCCACGGGGCCCACGAGGTCAGCCGAGCTACCG	CP001964.1	2407234-2407304
Cf-1B	GTCAGCGACAAACGGCAAACCCGCCGCAAGGTGGGGACGCAAAGCCACGGGGCCCACGAGGTCAGCCGAGCTACCGAACGAC	CP001964.1	2407228-2407309
Cf-6	GGTCAGACAAGGGCACACCCGTCGCGAGGCGGGGCCGCAAAGCCACGGGACCCACGCGGTCAGCCGGGCTGCCGACC	CP002666.1	2316134-2316210
Ck-208	TTGATAATAGCACACTTATCGAAAGGTAGGGTCGCAAAGCTATGGGTCTTAAGAAAATTATTTTTCTATGATTGCCAGGTTGCCAA	CP000673.1	2377707-2377792
Cm-1041	TTCCTTTCCGATAAAGGCAAACCAGTCGCGAGGCTGGGACCGCAAAGCCACCGGTCAGCAAACGGGCTGACAGCGGGGTTACCGAAGAAAGGAA	FP565575.1	196189-196281
Cp-234	CTTTAAAAAAATGGGCAAAATTAGAGAAATCTAATGACGCAAAGCTATAGGGACTAAGGTTTATAACTATGTCAGCCAGTTGCCAAAG	CP000246.1	1695612-1695699
Cp-287	CGATAATAGCAAACCTAGTGAAAACTAGCGACGCAAAACTATAGGGTCTTCCTTAGATATTCTAAGATGATAGCCAGTTACCG	CP000885.1	1913589-1913671
Cp-865	AGTCATTTGGCAAACTGGTTGAAAGGCCAGGACGCAAAGCCTCCGGTCTAAAGACATGTCGTCCAGGATAGCGGGGTTGCCACAT	AC167560.3	117626-117710
Cs-225	CGACAAAGGGCAAACTTGCCGAAAGGTAAGGACGCAAAGCCGAGGGTCTAAAGTGCGAGAGCATTATGACAGTCTGGCTGCCG	CP002109.1	3402849-3402931
Ct-116	TATAAACCGATAAAGGCAAAACTGTGGAAACGCAGTGACGCAAAGCTACAGGGGCTAAGGTCCGCCAGGGCTATGCCAGCCAGCTACCGGTTTATG	AFCE01000088.1	13-108
Ct-256	ATGAAACAGGGCAAAATCACCGAAAGGTGATGACGCAAAGCCATGGGTCTACTGTTTTAAAACAATGTTTTAAAGCTATGATCGCCAGGCTGCCAT	CP002416.1	1366451-1366546
Da-534	CCTTGATAAGGGCAAACCTGTCGAAAGGCAGGGACGCAAAGCTACAGGTCTAAAGCATTTTGCTAAGACAGCTGGGTTGCCGGGG	ACJM01000030.1	1763-1847
Da-702	AGCCGAGAATCAAGCCAACCCGCCTCAGGCGGGACGGAAAGCCACGGGTCTTTCAGACAGCCGGGTTGCCTCGGTT	CP000112.1	841125-841200
Da-707	TCTCTCCGATAAAGGCAAAACCGGAGTAATCCGGTGACGCAAAGCCACGGGTCCTGTTTGACAGGATCGCCGGGTTCCCGAAGAGG	AAEW02000001.1	1141863-1141931
Dg-430	TCTGAAAAGGCAAACCTGTTGAAAGACAGGGACGCAAAGCCATGAGTCTAAGGTTTTTGAAAGGGCTATGACAGTCAGGCTGCCGGA	AGJQ01000018.1	12829-12915
Dp-1045	TGTGACAAAGGCAAACCACTCGAGAGGGTGGGACGCAAAGCCAAGGGACCTAACGAGGGACACATGTTCCAGGGTCAGCCTAGCCGCCACA	ACJG01018181.1	935-1025
Dp-17	CTTCTCGACAAAGGCAAACCCTCCGCGAGGGGGGGACGCAAAGCCCACGGAACTCCGCTGCTCCGCTCTTCTCTCAGGGCAGCACGGAAGTTGGCCGGGCCACCGAAAGAAG	CP002539.1	69968-70079

Dt-697	TATCCAGCCAAACCCGCCGCAAGGCGGGGACGGAAAGCCACGGGCCCCCGGGATTTAAAAGTCATAGCATACGCGGGCAGCCGGGTTGCCGGATG	ACJN02000003.1	250970-251064
Gc-1044	AATTGAATTATTCATCGGTAAAACTATTGAAGATAGTGACACAAAGCCAAGGGTCTAAGGTCCTTCCAAACGGGATTATGACAGTCCGGTTGCCACATT	ACYC01000191.1	495-593
Gm-790	TCGACAATACTAAACCATCCGCGAGGGTGGGACGGAAAGCCTACAGGGTCTCTCTGAGACAGCCGGGATGCCGA	CP000148.1	1079467-1079540
Gs-761	CTCCGAAAAGAGTAAACCCATCGCAAGGTGGGGACACAAAGCCGACGGGTGCCGCTGGAGCGGGACGGCCGGGTTGCCGGAG	CP002031.1	2659688-2659769
Hm-646	TGCTTATGGGTAAACCCGTTGAAAGACGGGGACACAAAGCCACCGACCTACAGCATCAATGCCATGGTAGCGGGGCCGCCA	ABRM01031829.1	2236-2316
Lb-336	TTTCATTTGGCAAAGCCGGCGAAAGCCGGTGACGCAAAGCTAGAGGGCCTTGTATCCGTTATTCGGTATGTGGCAGCCAGTTGCAAA	ACTP01000092.1	157432-157518
Mm-662	GCATTTGAAAAAGGCAAACTCAGCTGAAAAGCGAGGGCGCAAAATCACCGGTCTAAGGGGCGTAAGTTCTAAGATAGCGGGAGTACCAGATGT	CP001672.1	134554-134646
Oi-125	ATCCTCAGAAAAAGGCAAACCTATTGAAAGATGGGGACGCAAAGTCACAGATCTAAGGTATTTTTACTAAGATGGCTGGACTATCTGGAT	BA000028.3	2580241-2580330
Pc-822	ATTTTCGTTCAAGGCAAAGTCAGAGTAATCTGGCCACGCAAAACCACGGGTCCATGGTTCATGGATAGCCGGGTTGCCGAAAAT	CP000142.2	1774617-1774700
Pi-1049A	TATTGAAAAAGGCAAACTCATCGAAAGGTGAGGGCGCAAAGCTACAGGAGCTAAAGCGATTCAATCGCCATGCTAGCCAGCTACCAGTA	AATU01015137.1	1224-1312
Pi-1049B	TCCGTATTGAAAAAGGCAAACTCATCGAAAGGTGAGGGCGCAAAGCTACAGGAGCTAAAGCGATTCAATCGCCATGCTAGCCAGCTACCAGTAAAGA	AATU01015137.1	1220-1316
Pl-156	CTACGATAACGGCAAACTTGTCGAAAGATAAGGACGCAAAGCCACAGGGCCTTCTTGATGAACCGTCAATGGCAGCCTGGCTACCGAAG	AGIP01000013.1	159170-159258
Pm-140	CATTCGTTTCATGGCAAACCTGCCGAAAGGCAGGGACGCAAAGCTTAGGGTCTACGGTCCTGCAGGGACTATGACAGCCTGGCCGCCGAATG	CP002869.1	30538-30629
Pt-160	CGTTAATCGTAAACACCTCGAAAGTGGTGGACACAAAGCCATGGGTCTAAAGCTGGATTAAACAGCCATGATTGCCAGGTTGCCG	CP003107.1	4365888-4365972
Rc-1047	CGAAGTAGTAACAAGGGCAAATCCATCGAAAGATGGAGACGCAAAATCACCGGTCTACGGGCTTATGCCACGACAGCGGGATTGCCGGCTTCG	AASG02030143.1	245-337
Sd-919	CTGTGGAAGGCAAACCAGTTTTAAAGACTGGGACGCAAAGCCTCCGATCTAAAGGTTTGCTTGTACCTATGATAGCGGGGATCCCACAG	CP000302.1	2534230-2534318
Sp-917	CACTTGATGAAAAAGGCAAAACCTGTGAAAGCAGGTGACGCAAAGCATCCAGCCTAAGGGGACACCTATGGCAGTGGCGCTACCGCAAGTG	CP000851.1	4449054-4449144

St-303	CCCCAGCGATAAAGGCACACCCGCCGAAAGACGGGGCCGCAAAGCCACGGGGCTACAGGAAGCGGGCGCCGCCCTGCCGGGTCTCCCGCCTCCCATGCTAGCCGGGCTGCCGCTCGGG	AP006840.1	3227381-3227498
Sv-921	TTTGAATAAAAGGCAAACCAATCGAAAGATTGGGACGCAAAGCCTCCGGTCTAAGGGGATAGTGTTGAGGCGGTCGTTGTCTCAGTATTATCGCAGTACCTAAGATAGCGGGGATACTTCAGG	AP011177.1	990518-990640
Ta-882	GCTCCCGAAACGGCAAACTCCGGGTAACCGGATGTACGCAAAGCCACAGGTCCTTTTGAGTCATCATCAGGACAGCTGAGCTACCGAAGGAGC	CP001616.1	1034200-1034292
Td-659	ATCCCGAAAGGGCAAACCCGCCGCGAGGCGGGGGCGCAAAGCGACCGGTCTCGAAAGAGATAGCGGCGCTGCCGTGGAT	CP000116.1	532900-532978
To-592	TTCCGTCAAGGGCAAACCGTCGAAAGGTAGGGACGCAAAGCCACGGGCCTACGAAGAGACAAGCTCTTCATAGCAGCCGGGCTGCCGGGG	CP002131.1	1785218-1785307
Tt-888	CTCTGATAAAAAGGCACGCTGACGCGAAAGCCCAGCACGCAAAATTACCGGTCTAAGGACTCAAGGTCTACGACAGCGGGATCGCCAGAG	CP001614.2	634814-634903
Vn-1026	GCTCAAAAGGCAAACTTGCTTAAAAGCAGGGACGCAAAATGACAGTGCCACTGATTTTTGGAGAGGATTATTGGCAGGCTGCATTACAAGAGC	AFWJ01000289.1	14511-14603

Table 2.3 – Oligos.

The following oligos were used in this study to generate DNA templates for *in vitro* transcription (Sp, Sp2). All sequences are listed in 5' → 3'.

Sp-F	CCAAGTAATACGACTCACTATAGACGCGACTGAATGAAATGGTGAAGG
Sp-R	GACGCGACTAGTTACGGAGCTCACAC
Sp2-F	CCAAGTAATACGACTCACTATAGGATGTAACTGAATGAAATGGTGAAG
Sp2-R	GATGTAACTAGTTACGGAGCTCAC

Chapter Three

Validation and further optimization of GEMM-I-based RNA biosensors *in vivo*

Portions of this chapter have been published in:

Wang, Xin C, Stephen C Wilson, Ming C Hammond. 2016. "Next-generation Fluorescent RNA Biosensors Enable Anaerobic Detection of Cyclic di-GMP." *Nucleic Acids Research,* in press.

INTRODUCTION

Biochemistry research has traditionally employed a reductionist approach by observing molecules of interest in isolation from the larger system. Technological advances such as fluorescent proteins and high-resolution microscopy, however, have allowed us to observe complex processes *in vivo* without harming the system at hand. To that end, genetically encoded fluorescent biosensors have become invaluable tools in cell biology research, allowing researchers to detect and monitor cellular conditions in the native environment (Crone et al. 2013). In addition to detecting the ligand of choice, well-behaved sensors must satisfy two basic requirements: that 1) they express adequately and stably inside a cell, and 2) they do not sequester or perturb the equilibrium of any cellular components.

In the previous chapter, I introduced four second-generation RNA-based fluorescent biosensors that respond to the bacterial second messenger cyclic di-GMP (c-di-GMP) (Kellenberger et al. 2013). Predicted to be a signaling molecule in 75% of all sequenced bacteria (Seshasayee et al. 2010), c-di-GMP regulates biofilm formation, host colonization, and bacterial virulence (Ryan 2013; Römling et al. 2013). Here, I validate the efficacy of these biosensors *in vivo* and survey different promoters and scaffolds to boost biosensor expression and stability. Furthermore, I demonstrate that these biosensors do not perturb downstream c-di-GMP signaling, and I use these sensors to monitor c-di-GMP levels in response to zinc exposure.

One major limitation of fluorescent protein-derived biosensors is their inability to be used in anaerobic environments, as fluorescent proteins require oxygen for chromophore maturation. Because Spinach is an RNA aptamer that binds an exogenous chromophore, it bypasses the oxygen requirement. Here I demonstrate the efficacy of both Spinach and the biosensor in oxygen-free conditions, a critical property as c-di-GMP exists as a signaling molecule in many anaerobic organisms.

RESULTS

Second-generation biosensors detect c-di-GMP in vivo

In the previous chapter, I introduced four second-generation fluorescent biosensors that respond to the bacterial second messenger cyclic di-GMP (c-di-GMP). These biosensors exhibited improved fluorescence, kinetics, and less sensitivity to temperature and magnesium concentration, suggesting significant improvement over the first-generation Vc2 *in vivo*.

To perform this comparison, we co-expressed the new biosensors in aerobically grown *E. coli* with enzymes that affect cellular c-di-GMP concentrations, and analyzed live cell fluorescence by flow cytometry (Figure 3.1). While Vc2-expressing cells showed no significant fluorescence turn-on upon co-expression of the constitutively active cyclic di-GMP synthase WspR D70E, all cells expressing second-generation biosensors exhibited fluorescence turn-on, indicating robust responses to high c-di-GMP concentrations. Furthermore, consistent with the in vitro measurements, maximal fluorescence levels for the new biosensors were higher than that for

Spinach2 (111 to 175%) analyzed under identical conditions and co-expressed with an inactive WspR to equalize the expression load. The two most sensitive biosensors, Ct and Pl-B, also displayed significant fluorescence activation in response to endogenous c-di-GMP levels, as indicated by higher fluorescence levels upon co-expression of an inactive enzyme, WspR G249A, versus YhjH, a c-di-GMP-specific phosphodiesterase ($p = 0.043$ and 0.053, respectively). This is particularly notable as YhjH contains an EAL domain and produces pGpG as the degradation product. Both Ct and Pl-B exhibit a weak fluorescence response to pGpG *in vitro*, but the lack of signal in YhjH-expressing cells affirms their strong ligand preference for c-di-GMP. In addition, the histograms display a uniform distribution of c-di-GMP levels in each population, with little heterogeneity.

Constitutive promoters reduce fluorescence levels

Expressing and maintaining high levels of well-folded RNA *in vivo* is challenging due to the high number of native ribonucleases and the inherent instability of RNA. Our current system employed the T7 promoter system, which yields extremely high RNA levels but also requires expression in specific strains that encode the T7 promoter under an inducible lac expression system. We attempted to universalize expression by switching to a high-performance constitutive promoter. To that end, we evaluated fluroescence in TOP10 cells expressing Spinach2 under a variety of constitutive promoters previously validated for protein expression in *E. coli*: cp25 (Jensen & Hammer 1998), pFAB708 (Goldbeck et al. 2013), and pFAB324 (Mutalik et al. 2013). In addition, we also tested a promoter region engineered from the RIL plasmid (Stratagene), which constitutively expresses extra copies of rare tRNAs, as well as inserting additional spacer regions between the promoter our desired promoter start site (Figure 3.2).

Unfortunately, none of these promoters gave any noticeable fluorescence in our initial trials, and a time course also did not reveal any noticeable trends in fluorescence (Figure 3.2b). We hence chose to discontinue further efforts in finding suitable constitutive promoters. Interestingly, the pFAB708 promoter exhibited low fluorescence in BL21 Star cells, suggesting that the lack of fluorescence comes not from low expression but rather high RNase degradation. Another hypothesis is that transcripts created under these constitutive promoters are less well-folded and hence not competent for fluorescence. This could be because the inducible promoter system uses T7 RNA polymerase, a polymerase known for its rapid extension rates (Golomb & Chamberlin 1974). Because RNA begins folding as nascent chains begin to form (Kramer & Mills 1981), having a well-folded biosensor might depend on T7's rapid transcription rates. Interestingly, the pool of RNA aptamers used in the initial selection for Spinach was also synthesized *in vitro* with T7 RNA polymerase, corroborating the notion that T7 might be involved in aptamer folding. Further work, however, would be needed to confirm these ideas.

Modified scaffolds improve fluorescence intensity

Another strategy to prevent RNase degradation is to encase the RNA aptamer in a tRNA scaffold (Ponchon & Dardel 2007). In this design, the biosensor sequence is inserted into the loop of a structured human lysine tRNA (*H. sapiens* $tRNA^{Lys}$), a practice that started and continued with the original Spinach (Paige et al. 2011). To further improve *in vivo* stability, we screened five

additional tRNA scaffolds from different organisms (*P. arsenaticum* $tRNA^{Val}$, *D. kamchatkensis* $tRNA^{Val}$, *C. korarchaeum* $tRNA^{Val}$, *O. sativa* $tRNA^{Ala}$, *E. coli* $tRNA^{Met}$), as well as a triple helix segment from the Malat1 gene, which has been shown to improve transcript longevity (Brown et al. 2014).

Our flow cytometry screens of Spinach2 expressing these scaffolds proved that the original *H. sapiens* $tRNA^{Lys}$ scaffold did have a significant improvement over non-scaffolded RNAs. Of the other scaffolds tested, only the *O. sativa* $tRNA^{Ala}$ scaffold exhibited an improved fluoresence of over 3-fold (Figure 3.3a). Interestingly, the performance of these scaffolds bore no relationship to their folding stabilities, as quantified by the ΔG of their predicted secondary structures (Zuker 2003) (Figure 3.3b, with Cindy Lam and Wanda Thi). Another interesting point is that the only native *E. coli* scaffold we tested also did not improve fluorescence – on the contrary, we saw no fluorescence signal.

Expression of RNA-based biosensor does not affect motility phenotype

Two general criticisms leveled at biosensors for live-cell applications are the saturation effect, in which excess target saturates biosensor signal, and the observer effect, in which biosensor binding sequesters the cellular pool of target ligands, thereby altering the phenotype (Haugh 2012). We intended to circumvent the saturation effect by making a suite of four sensors with different binding affinities, thus ensuring a wide dynamic range of almost four orders of magnitude. To assay for potential observer effect, we examined the impact of biosensor expression on cellular motility, a phenotype regulated by cellular c-di-GMP levels in *E. coli* (Figure 3.4A). High levels of c-di-GMP lead cells to transition from motile to sessile states. Accordingly, cells expressing the active synthase WspR D70E exhibit little to no motility in soft agar, while cells expressing the phosphodiesterase YhjH are more motile than control cells expressing the inactive WspR G249A. Furthermore, cells expressing only Ct or the non-binding mutant Ct-M both exhibit similar motilities as the WspR G249A control.

We then co-expressed Ct, our highest-affinity biosensor, with either WT WspR or YhjH off a single plasmid (Figure 3.4B). These cells maintained similar motilities as their enzyme only-expressing counterparts, with the YhjH-expressing cells traveling roughly 2.1 times farther than WspR-expressing cells (Figure 3.4C). This lack of observer effect could be due to faster turnover of RNA-based biosensors, which alleviates ligand sequestration. In contrast, cells expressing the c-di-GMP binding protein YcgR exhibit reduced motility, as this c-di-GMP effector acts to bind flagellar proteins and stop flagellar motion (Paul et al. 2010).

Anaerobic growth affects GFP but not RNA aptamer fluorescence

To confirm that Spinach-based systems could function anaerobically, we expressed the aptamer in anaerobically cultured *E. coli* BL21 Star cells. Bacteria expressing the second-generation Spinach2 (Strack et al. 2013) exhibited 4.0-fold brighter fluorescence than bacteria expressing the oxygen-dependent GFP-LVA (Figure 3.5). The LVA degron tag was added to ensure that measured fluorescence came from recently synthesized GFP (Andersen et al. 1998). The fluorescence levels of cells expressing Spinach2, as well as those expressing the flavin-binding iLOV, remained consistent after cells were exposed to ambient oxygen. On the other hand, the

fluorescence of GFP-expressing cells increased 13.2-fold after oxygen exposure, confirming its oxygen sensitivity and incompatibility with anaerobic applications.

RNA-based fluorescent biosensors function under anaerobic conditions

We next tested the ability of Ct, the brightest and highest affinity biosensor, to detect c-di-GMP under anaerobic conditions (Figure 3.6). The biosensor maintained responsiveness to c-di-GMP, as fluorescence turn-on was observed with WspR D70E co-expression. Intriguingly, fluorescence levels were the same for cells co-expressing the inactive WspR G249A or YhjH phosphodiesterase, which is in contrast to the observations under aerobic growth conditions. This result likely reflects lower endogenous c-di-GMP levels under anaerobic conditions, as oxygen activates the *E. coli* c-di-GMP synthase DosC (Tuckerman et al. 2009) and anaerobic promoters up-regulate expression of the *E. coli* c-di-GMP-specific phosphodiesterase YfgF (Lacey et al. 2010). Also consistent with these expected effects, a 1.54-fold increase in fluorescence was observed for cells co-expressing WspR D70E after the anaerobic culture was exposed to ambient oxygen for 2 h at 4 ˚C.

The fluorescence signal for the control, Spinach2, increases by 1.35-fold after oxygen recovery as well. Therefore, the fluorescence increase upon oxygen recovery may also be partially attributed to differences in biosensor expression or steady-state levels under these conditions. Nevertheless, it is clear that compared to the fluorescent protein GFP-LVA (13.2-fold), RNA-based fluorescent biosensors are less dependent on oxygen for fluorescence, and thus have great potential utility for measuring endogenous signals or metabolites *in vivo* under anaerobic conditions.

Time-resolved fluorescence of zinc-responsive c-di-GMP dynamics

We next exploited the fast kinetics of our biosensors to monitor c-di-GMP production *in vivo* in response to zinc. Zinc exposure has been shown to dispel biofilm formation, suggesting that zinc leads to decreased c-di-GMP levels. To confirm this phenomenon, we expressed our two highest-affinity biosensors, Pl-B (K_D = 12 ± 2 nM) and Ct (K_D < 5 nM), in *E. coli* grown in media with and without zinc (Figure 3.7, with Jongchan Yeo). Cells grown in zinc had lower fluorescent levels than those grown without zinc, suggesting that zinc exposure leads to lower c-di-GMP levels. We then monitored c-di-GMP dynamics by assaying zinc-exposed cells that were then moved into zinc-free media. The fluorescence levels of these cells increased over time, suggesting a direct relationship between zinc exposure and c-di-GMP levels. Control cells expressing Spinach2 alone and cells expressing non-binding mutant biosensors had similar fluorescence levels whether grown with or without zinc, and fluorescence levels remained stable after zinc removal. Interestingly, the Spinach2 zinc-removed cells had overall lower fluorescence levels, suggesting that changing media conditions might also affect RNA transcription levels or stability.

In *E. coli*, one enzyme responsible for this is DgcZ (formerly YdeH), a GGDEF with a zinc-sensing CZB (chemoreceptor zinc binding) domain at its N-terminus (Zähringer et al. 2013). A crystal structure of DgcZ showed that zinc binding induces a conformational strain that occludes the GTP binding sites, thus preventing c-di-GMP synthesis (Zähringer et al. 2013). We generated a DgcZ knockout strain from a BL21 Star background and expressed the biosensor under the same with zinc, without zinc, and zinc-removed conditions (Figure 3.7**).** This time, zinc removal

did not elicit any change in fluorescence, confirming that the fluorescence increase in the WT cells came from Zn^{2+}-dependent DgcZ activation.

DISCUSSION

The ubiquity of c-di-GMP as a bacterial signaling molecule has elicited the development of myriad assays for its detection. Very few of these assays, however, function both *in vitro* and *in vivo*, especially as a live imaging tool that can detect c-di-GMP with both spatial and temporal resolution. Our next-generation RNA-based c-di-GMP biosensors can detect c-di-GMP in live *E. coli* cells via flow cytometry, a high-throughput method that can assay thousands of cells per second. Furthermore, these biosensors can monitor changes in c-di-GMP levels over time in response to Zn^{2+} and other environmental signals. While more work is needed to correlate fluorescence levels with exact c-di-GMP concentrations, these new biosensors open the door to studying c-di-GMP signaling in live bacteria under a variety of dynamic environmental conditions.

Importantly, we demonstrate the ability of these biosensors to function *in vivo* under anaerobic conditions. This is particularly relevant to functional studies of gut microbiota, which occupy partial to fully anaerobic niches. In addition, we demonstrate that Spinach2 itself can function under anaerobic conditions, extending the known utility of this aptamer, which has been validated for live imaging of RNA and protein expression *in vivo* (Strack et al. 2013; Song et al. 2013). We propose that Spinach2 or related aptamer-dye pairs can be used to observe changes in gene expression under differential oxygen conditions, a known phenomenon yet to be monitored in real time (Bueno et al. 2012). Taken together, our results highlight the potential of aptamer-dye pair based methods for tracking RNAs, metabolites, or soluble signals under variable oxygen conditions, which can provide new insights into how cells adapt to changes in environmental oxygen availability.

One major advantage afforded by the high fluorescence signal of our improved biosensors is the ability to detect signal with lower biosensor levels. This, in turn, allows for more robust signals *in vivo*, as expressing and maintaining high levels of well-folded RNA is challenging given the high rate of RNA degradation *in vivo*. Vc2, for instance, exhibited a 5-fold fluorescence turn-on *in vitro* with 100 nM RNA (Kellenberger et al. 2013), but only a 1.5-fold turn-on with 30 nM RNA and a correspondingly undetectable fluorescence signal with flow cytometry *in vivo*. We hence screened for next-generation biosensors specifically under low RNA conditions to yield biosensors that could perform well *in vivo*. Accordingly, our next-generation biosensors have upwards of 7-fold turn-on *in vitro* and 11-fold turn-on *in vivo*. Furthermore, lower RNA levels also allow for lower dye concentrations, mitigating potential challenges in dye toxicity or delivery that could arise in adapting the biosensor towards other organisms.

Thus far, we have focused our *in vivo* efforts on *E. coli*, an organism known for its hardiness and versatility in laboratory conditions. However, preliminary results from collaborators studying other organisms, including pathogenic bacterial strains, show promise that our biosensors can be adapted for *in vivo* assays in other organisms. Optimizing the biosensor assay for these new

organisms requires finding highly efficient expression systems that promote transcript folding and longevity. While these expression systems will vary from organism to organism, we have outlined several variables that factor into fluorescence, namely promoter sequences, polymerase speeds, and stabilizing scaffolds. We hope that this generalizable strategy will guide the adaption of RNA-based biosensors for *in vivo* study in a plethora of organisms.

MATERIALS AND METHODS

Molecular cloning

For *in vivo* expression, biosensors were flanked by a tRNA scaffold and cloned into the BglII and XhoI sites of pET31b(+) as previously described (Kellenberger et al. 2013; Kellenberger, Wilson, et al. 2015; Ren et al. 2015) using the tSp2- and p31b- primers listed in Table 3.1. YhjH was amplified from *E. coli* BL21 Star genomic DNA, and its sequence and WspR alleles and YhjH were cloned into the NdeI and XhoI sites of pCOLADUET-1. The two plasmids encoding each biosensor-enzyme pair were co-transformed into *E. coli* BL21 (DE3) Star cells (Life Technologies).

Fluorescent proteins were cloned into pET31b(+). GFP-LVA has a LVA degradation tag at the C-terminal end and was a gift from the Guillemin Lab at the University of Oregon (Eugene, OR); the protein was cloned into pET31b(+) between the NdeI and XhoI restriction sites. The iLov sequence (Drepper et al. 2007) was ordered as a gBlock from IDT (Coralville, IA) and cloned into pET31b(+) between the NdeI and SalI restriction sites.

For motility assays, the Ct biosensor and enzymes were both cloned into pETDuet-1 via Gibson assembly. Ct was amplified with the dSens- primers, while enzymes were amplified with the dWspR- and dYhjH- primers. The vector was a fragment of the pETDuet-1 vector cut at the NdeI and XhoI positions, and the three pieces were assembled with Gibson Assembly Master Mix (NEB).

For the constitutive promoter and tRNA scaffold screens, the tRNA-Spinach2 sequence was amplified from the pET31b-T7-Spinach2 plasmid and subcloned into the BglII and XhoI sites of pET31b. The constitutive promoter sequences used the p31b-R reverse primer. In many instances, cloning required multiple rounds of nested PCR and are marked by numerical designations (-1, -2, etc.) corresponding to the order in which they were amplified. The final sequences are listed in Tables 3.2 and 3.3.

Flow cytometry analysis of in vivo fluorescence

Flow cytometry experiments were carried out as previously described (Kellenberger, Wilson, et al. 2015). Briefly, fresh LB/Carb/Kan cultures were started from overnight cultures, and cells were grown to an OD_{600} ~0.3 and biosensor and enzyme expression was induced with 1 mM IPTG at 37˚C for 3 hrs. Cells were then diluted 1:30 in 1x PBS containing 50 μM DFHBI and fluorescence was measured for at least 10,000 events using a BD Fortessa X20 flow cytometer equipped with a

488 nM laser and a 530/30 filter in the Flow Cytometry Core Facility at the University of California at Berkeley. Data was analyzed with FlowJo (version 10.0.7).

For the tRNA scaffold screen, constructs were transformed into BL21 Star cells and grown in ZYP-5052 autoinduction media (Studier 2005) supplemented with Carb. Cells were shaken at 37 ˚C for 16 hours and diluted 1:70 into 70 µL of 1x PBS pH 7.5 with 100 µM DFHBI-1T and fluorescence of 50,000 events analyzed on an Attune NxT flow cytometer (Life Technologies) equipped with a 488 nm laser for excitation and 515/15 filter for emission.

Flow cytometry analysis (anaerobic growth and oxygen recovery)

Anaerobic cultures were grown in Balch tubes in ZYP-5052 autoinduction media (Studier 2005) supplemented with Carb/Kan. Cultures were inoculated aerobically and the tubes were capped with a chlorobutyl rubber stopper and crimped with an aluminum seal. The headspace was then sparged with argon for >15 minutes to ensure anaerobic conditions. Cells were grown while shaking for ~15 hr at 37 ˚C until cells reached an OD_{600} > 3. Each tube was then individually opened and the cultures immediately diluted 1:70 into 70 µL of 1x PBS pH 7.5 with 100 µM DFHBI-1T and fluorescence of 50,000 events analyzed on an Attune NxT flow cytometer (Life Technologies) equipped with a 488 nm laser for excitation and 515/15 filter for emission.

For oxygen recovery, cultures were briefly agitated after breaking the anaerobic seal, then placed at 4 ˚C, loosely capped, for 2 hrs before taking an identical fluorescence measurement in the flow cytometer. Data was analyzed with FlowJo (version 10.0.7).

Flow cytometry analysis (Zinc-responsive time course)

1.5-mL *E. coli* BL21 Star cultures expressing biosensor from their respective pET31b-T7-tRNA constructs were grown in ZYP-5052 autoinduction media while shaking at 37 ˚C for 20 hours. The "with Zn^{2+}" samples were grown in media supplemented with 1 mM $ZnCl_2$.

A small aliquot of each culture was then saved, with the rest centrifuged at 13,000 rpm for 1 min and the supernatant extracted as spent media. DFHBI-1T was then added to both aliquots of spent media to a final concentration of 25 µM. Each sample was prepared as thus:

Sample	Culture (10 µL)	Spent Media (500 µL)
Without Zn	Without	Without
With Zn	With	With
Zn Removed	With	Without

Samples were incubated at room temperature for 5 minutes for first flow cytometry reading; following that, samples were shaken at 37 ˚C between each reading. Fluorescence measurements were taken every 5 minutes from 50,000 events on an Attune NxT flow cytometer (Life Technologies) equipped with a 488 nm laser for excitation and 515/15 filter for emission. Data was analyzed with FlowJo (version 10.0.7).

DgcZ knockout strain generation

The DgcZ knockout strain was generated in a BL21 Star (DE3) background using TargeTron® Gene Knockout System following the manufacturer's protocol (Sigma Aldrich). The DgcZ-specific IBS, EBS1d and EBS2 primers were generated by the TargeTron algorithm (http://sigma-aldrich.com/targetronaccess) (Table 3.1) with the DgcZ gene template (Accession NC_000913.3, Gene ID 946075. Following knockout generation, the knockouts were confirmed by colony PCR with the EBS2 and DgcZ-co primers (Table 3.1).

Motility Assays

E. coli MG1655 cells were transformed with constructs by electroporation and grown overnight in liquid culture. Culture densities were adjusted to OD600 = 4, and 3 μL of culture was spotted onto agar plates containing 0.5% tryptone, 0.5% NaCl, 0.3% agar, and 20 μM IPTG (Wolfe & Berg 1989). Plates were allowed to air dry at room temperature for 30 minutes before being inverted and placed at 28 ˚C for 15 hours.

FIGURES

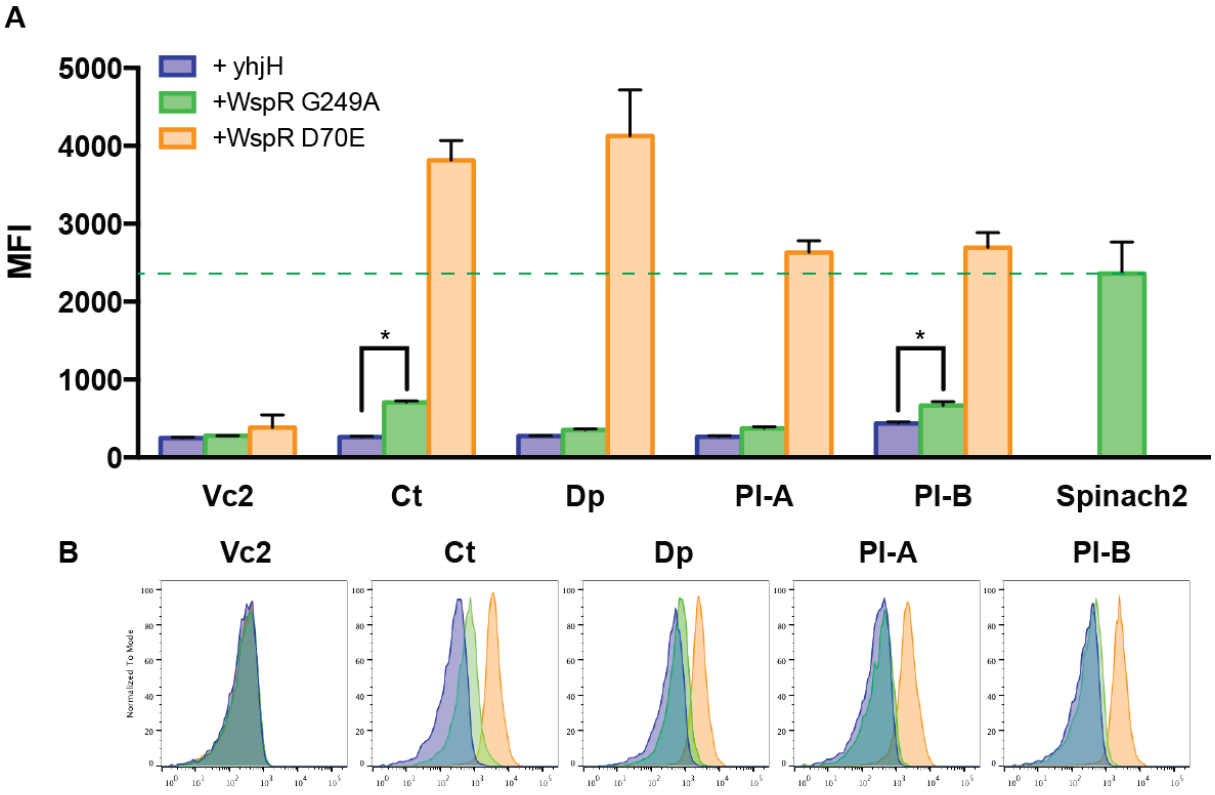


Figure 3.1 – Second-generation biosensors are brighter than Spinach2 *in vivo*.

A. Average mean fluorescence intensity (MFI) measured by flow cytometry of *E. coli* BL21 (DE3) Star cells co-expressing indicated biosensor constructs or Spinach2, along with various enzymes. Blue denotes phosphodiesterase YhjH; green denotes inactive mutant of diguanylate cyclase WspR (G249A); orange denotes constitutively active mutant of WspR (D70E). Data are from 3 independent replicates (10,000 cells/run) represented as mean ± SD.

B. Representative flow cytometry graphs for samples shown in A.

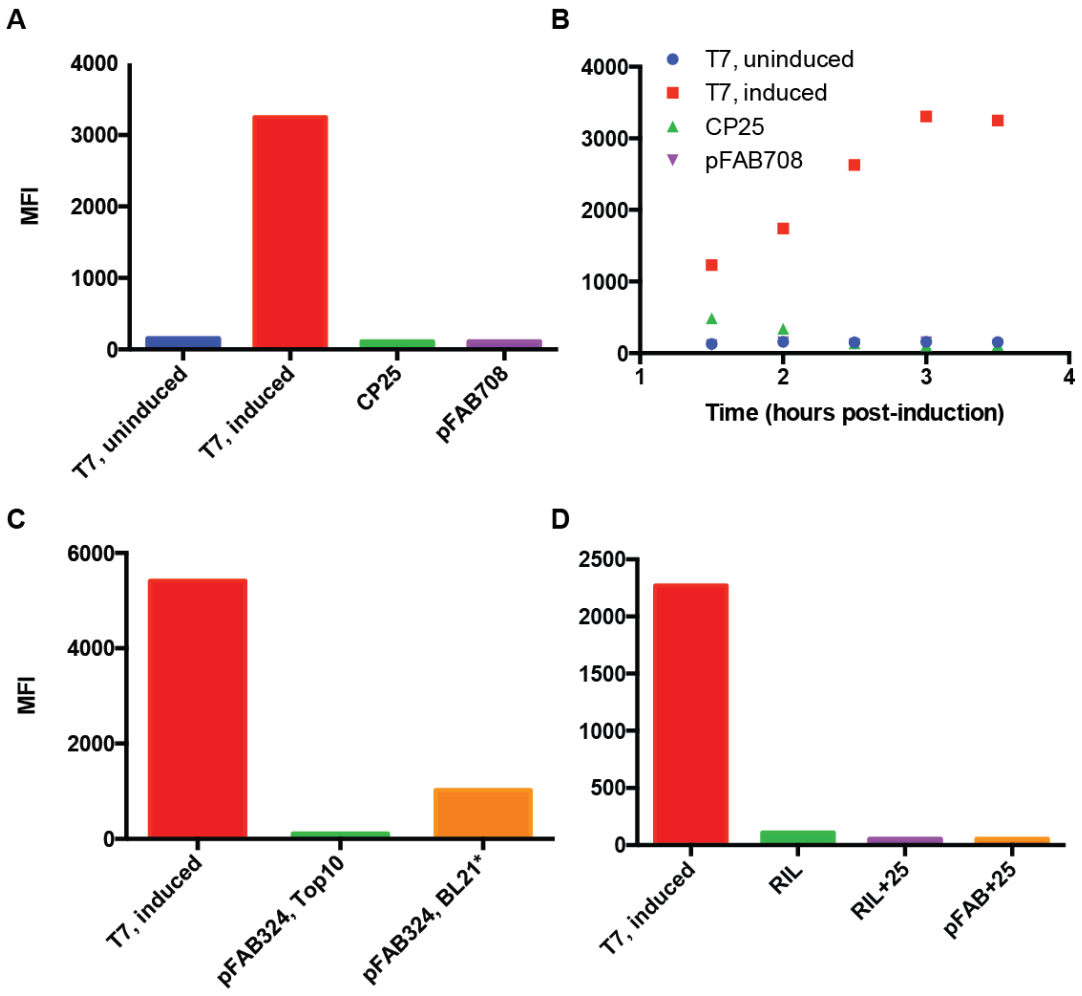


Figure 3.2 – Spinach2 has little to fluorescence when expressed under constitutive promoters.

A. Mean fluorescence intensity (MFI) of *E. coli* cells expressing Spinach2 under the control of the T7, CP25, and pFAB708 promoters. T7 constructs were expressed the BL21 Star strain, while others were expressed in the Top 10 strain. Data was taken 3.5 hours post T7 promoter induction. Data are from 1 independent replicate (10,000 cells/run).
B. A time course of data from part A.
C. Comparison of pFAB324-Spinach2 expressed in Top10 vs. BL21 Star cells. Data are from 1 independent replicate (10,000 cells/run).
D. Comparison of T7, RIL, RIL+25, and pFAB324+25 promoters. The T7 construct was expressed in BL21 Star cells, while the constitutive promoter constructs were expressed in Top10 cells. Data are from 1 independent replicate (10,000 cells/run)

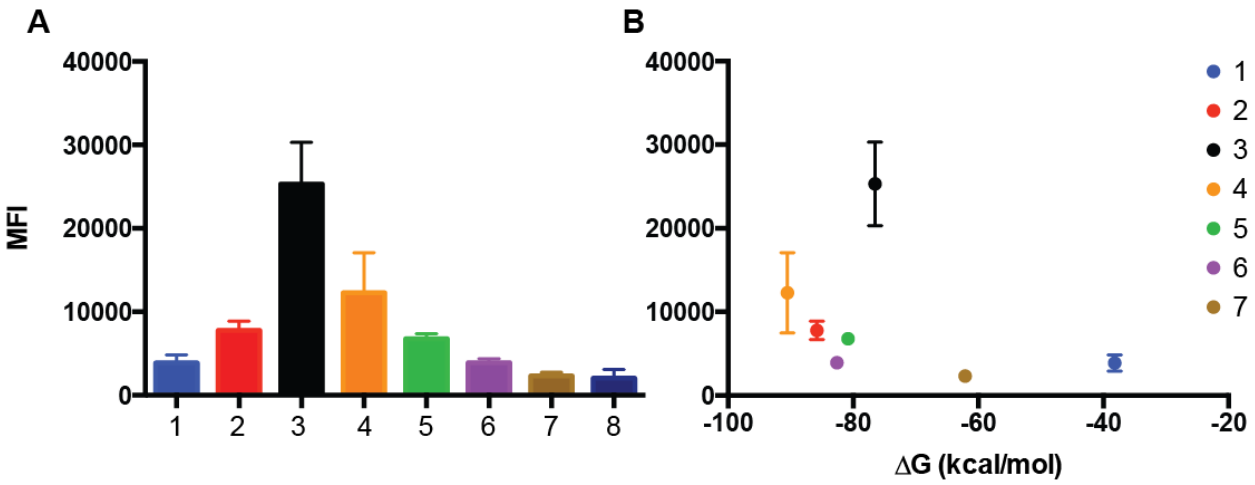


Figure 3.3 – Scaffolds improve Spinach2 fluorescence *in vivo*.

A. Mean fluorescence intensity (MFI) of *E. coli* cells expressing Spinach2 expressed with different scaffolds: 1. No scaffold; 2. *H. sapiens* $tRNA^{Lys}$; 3. *O. sativa* $tRNA^{Ala}$; 4. *C. korarchaeum* $tRNA^{Val}$, 5. *P. arsenaticum* $tRNA^{Val}$, 6. *D. kamchatkensis* $tRNA^{Val}$; 7. *E. coli* $tRNA^{Met}$, 8. Malat1. Data are from 3 independent replicates (10,000 cells/run) represented as mean ± SD.

B. MFIs (vertical axis) are plotted against ΔG of formation (horizontal axis) of the scaffold-Spinach2 construct, as determined by mFold.

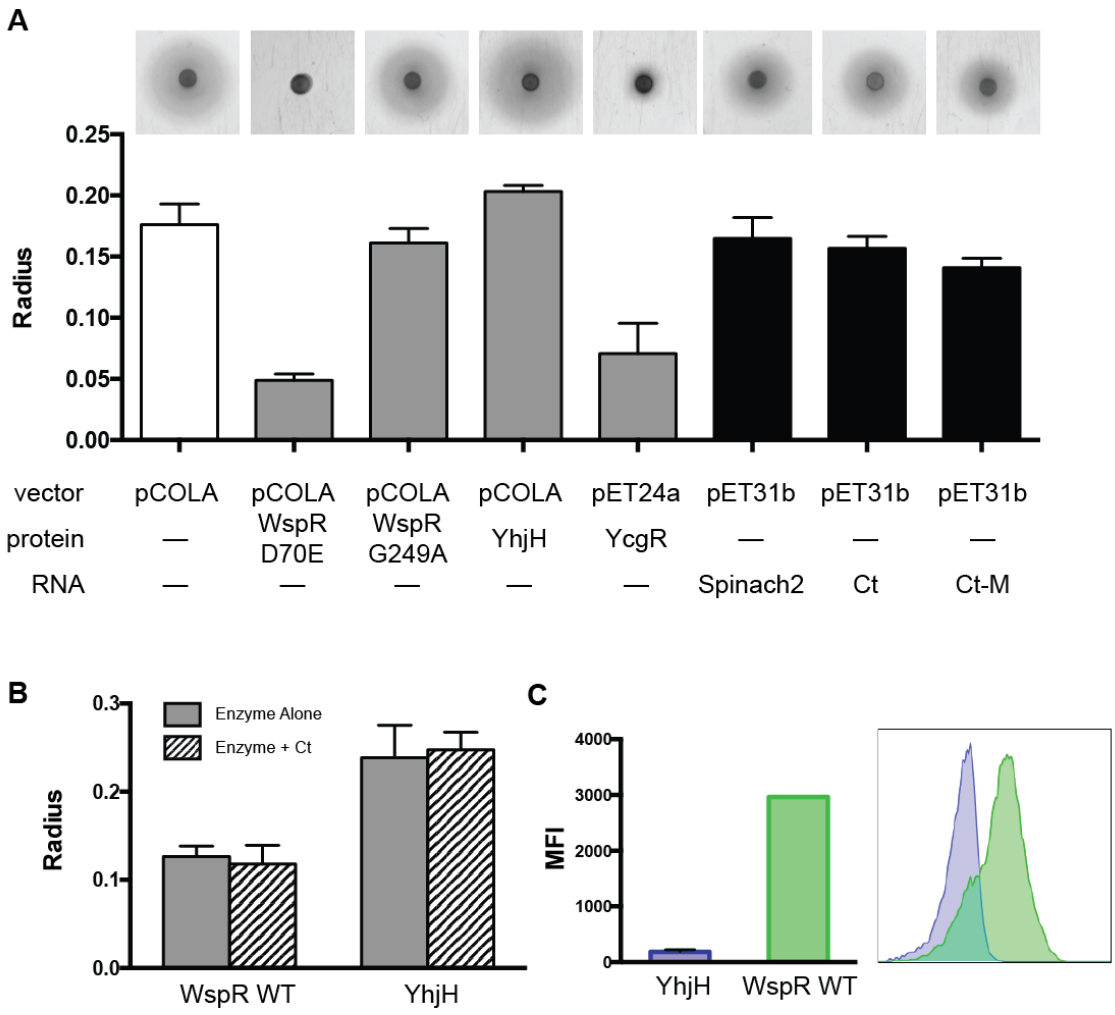


Figure 3.4 – Biosensor expression does not disrupt motility phenotype

A. Representative photos of motility zones for *E. coli* MG1655 cells in soft agar are shown above the analysis graph of the calculated radii. Data are from 3 independent replicates represented as mean ± SD.

B. Calculated radii of cells co-expressing enzyme and Ct biosensor (solid) or enzyme alone (striped). Data are from 2 independent replicates represented as a mean ± SD.

C. Mean fluorescence intensity (MFI) of cells expressing Ct biosensor with YhjH (blue) and WspR WT (green) in BL21 Star cells.

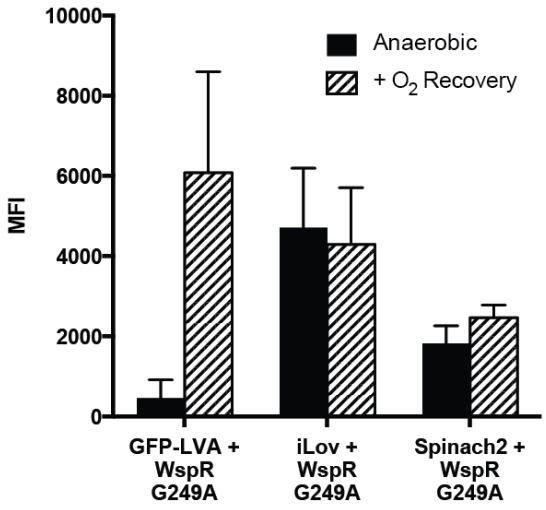


Figure 3.5 – Fluorescence under anaerobic growth and oxygen recovery.

Average mean fluorescence intensity (MFI) measured by flow cytometry of *E. coli* BL21 (DE3) Star cells co-expressing GFP-LVA, iLOV, or Spinach2 constructs. These cells also co-express an inactive enzyme WspR G249A to control for expression load with later experiments (see Figure 3.6), where cells express both biosensor and enzyme. Blue denotes after anaerobic growth; red denotes after anaerobic growth followed by oxygen recovery. Data are from 3 independent replicates (50,000 cells/run) represented as mean ± SD.

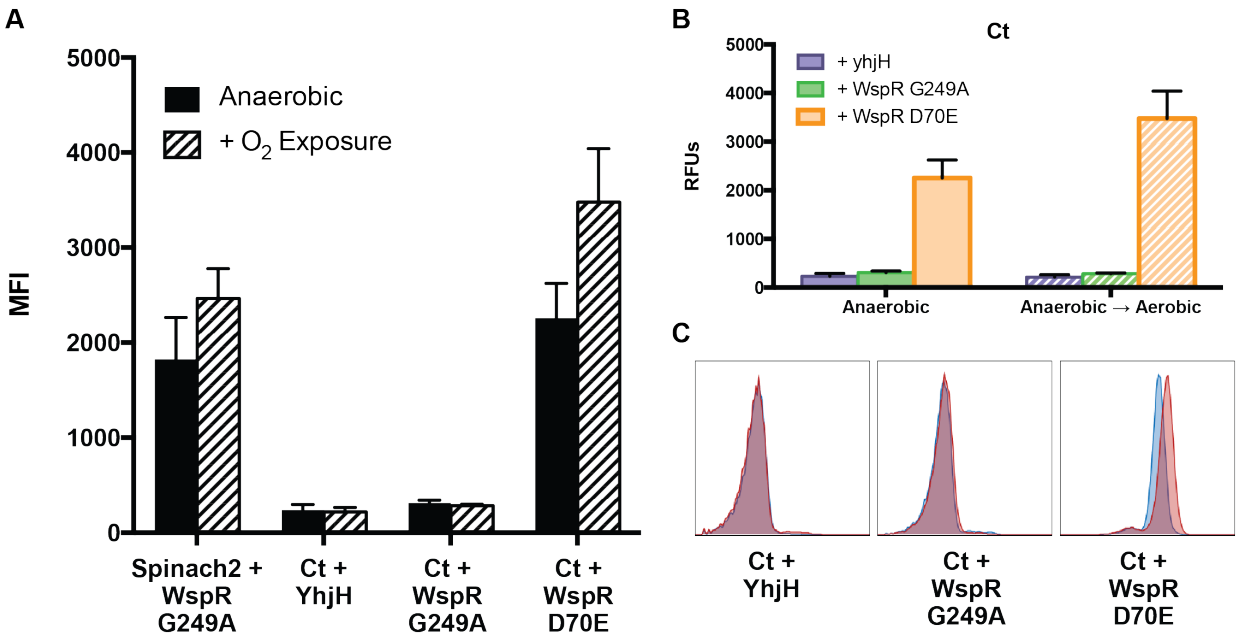


Figure 3.6 – Cyclic di-GMP biosensors function under anaerobic conditions.

A. Average mean fluorescence intensity (MFI) measured by flow cytometry of *E. coli* BL21 (DE3) Star cells co-expressing indicated biosensor constructs or Spinach2, along with various enzymes. Blue denotes phosphodiesterase YhjH; green denotes inactive mutant of diguanylate cyclase WspR (G249A); orange denotes constitutively active mutant of WspR (D70E). Data are from 3 independent replicates (10,000 cells/run) represented as mean ± SD.

B. Same data from **A** rearranged and shown using the coloring scheme from Figure 3.1.

C. Representative histograms of the data in **A**. Ct + WspR D70E cells have a small population of nonfluorescent cells that persist after oxygen recovery.

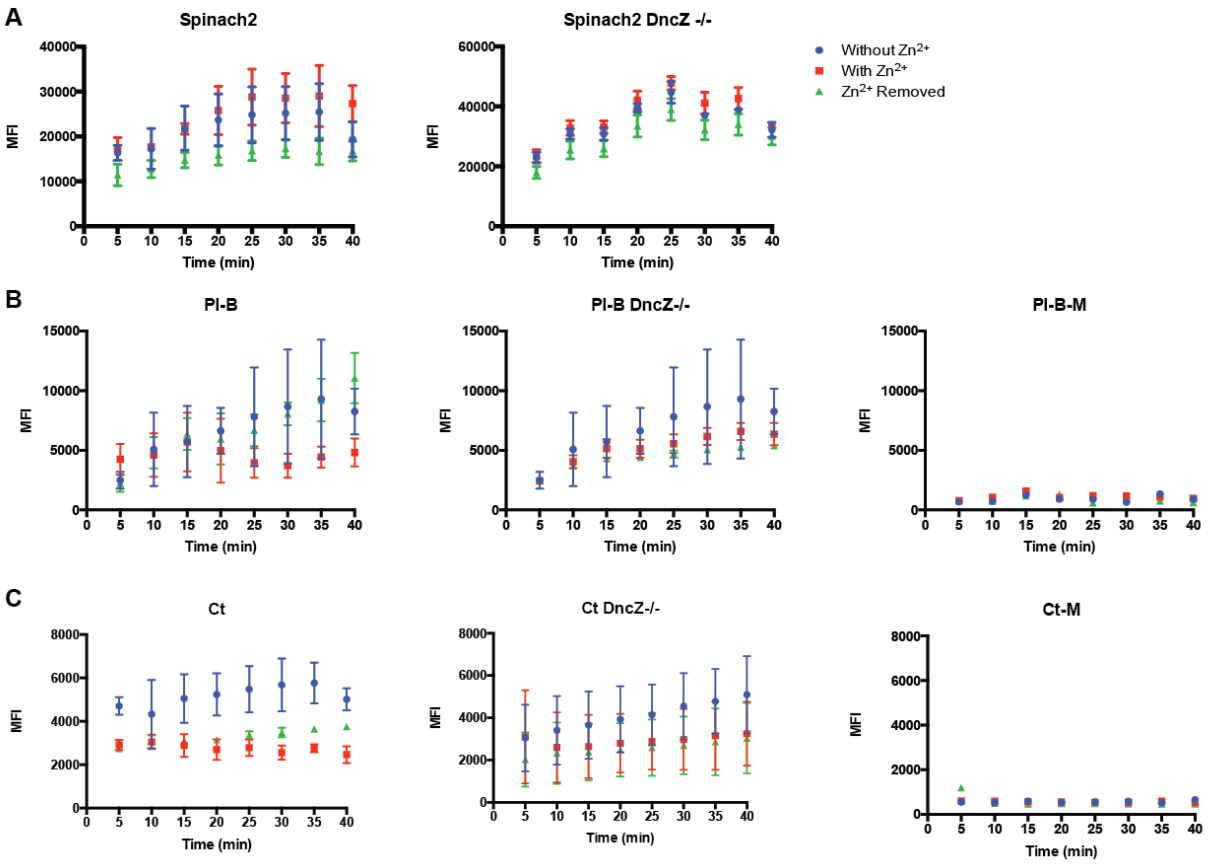


Figure 3.7 – Zinc-responsive c-di-GMP dynamics in *E. coli*.

A. Mean fluorescence intensities (MFI) of Spinach2-expressing cells grown without Zn^{2+} (blue circles), with Zn^{2+} (red squares), and with Zn^{2+} removed (green triangles) over time, with Zn^{2+} removal at time 0. Data are from 3 independent replicates (10,000 cells/run) represented as mean ± SD.

B. As A, but with BL21 Star cells expressing the biosensor Pl-B (left) or the nonbinding mutant Pl-B-M (right), or the biosensor Pl-B in a DncZ knockout strain (center).

C. As B, but with the biosensor Ct.

TABLES

Table 3.1 – Oligos.

The following oligos were used in this study to generate DNA templates to clone biosensors or fluorescent proteins into pET31b (tSp2, p31b, iLov, GFP), and to clone YhjH into pETDUET-1. The T7 promoter (in -F) and terminator (in -R) sequences are in red, the tRNA scaffold sequence is underlined, and restriction sites are italicized. All sequences are listed in 5' → 3'.

Sp-F	CCAAGTAATACGACTCACTATAGACGCGACTGAATGAAATGGTGAAGG
Sp-R	GACGCGACTAGTTACGGAGCTCACAC
Sp2-F	CCAAGTAATACGACTCACTATAGGATGTAACTGAATGAAATGGTGAAG
Sp2-R	GATGTAACTAGTTACGGAGCTCAC
tSp2-F	GAAATTAATACGACTCACTATAGGGGCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCGGATGTAACTGAATGAAATGGTG
tSp2-R	CCCCAAGGGGTTATGCTATGGCGCCCGAACAGGGACTTGAACCCTGGACCCGCGGCCGGATGTAACTAGTTACGGAGCTC
p31b-F	CAGTCA***AGATCT***CGATCCCGCGAAATTAATACGACTCACTATAGGG
p31b-R	CATCAG***CTCGAG***CAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTA
yhjH-F	CATG***CATATG***ATAAGGCAGGTTATCCAGCG
yhjH-R	CATG***CTCGAG***TTATAGCGCCAGAACCGCC
iLov-F	GATC***CATATG***GCGTCGTTCCAGTCGTTC
iLov-R	GATC***GTCGAC***CTCGAGCAGCTTTTCATATTCCTTCTGC
GFP-F	GATC***ATTAAT***ATGCGTAAAGGAGAAGAACTTTTCACT
GFP-R	CATG***CTCGAG***TCAGTGGTGGTGGTGGTGGTGAGCTACTAAAGCGTAGTTTTCGTC
dSens-F	tgcttaagtcgaacagaaagtaatcgtattgtacacggccgcataatcgaaat
dSens-R	ggtttctttaccagactcgaCAAAAAACCCCTCAAGACCC
dWspR-F	ctttaagaaggagatatacctatgcacaaccctcatgagagcaagaccgacctggg
dWspR-R	ctttctgttcgacttaagcatcagcccgccggggccggcg
dYhjH-F	ctttaagaaggagatataccTATGATAAGGCAGGTTATCCAGCG
dYhjH-R	ctttctgttcgacttaagcaTTATAGCGCCAGAACCGCCG
CP25-F	ctttggcagtttattcttgacatgtagtgaggggctggtataatcacatagtactgttGCCCGGATAGCTCAGTC
pFAB708-F	ACGT***AGATCT***TTGACAATTAATCATCCGGCTCGTAATGTTTGTGGAGCCCGGATAGCTCA
pFAB+25-F1	TTTGTGGAGATTACACCGCATTGCGGTATCAACGCCCGGATAGCTCA
pFAB+25-F2	GTAC***AGATCT***TTGACAATTAATCATCCGGCTCGTAATGTTTGTGGAGATTAC
RIL-F	GTAC***AGATCT***CA**TTGACT**CAGCAAGGGTTGACCGT**ATAATTCA**CGCGCCCGGATAGCTCA
RIL+25-F1	TTGACCGT**ATAATTCA**CGCGATTACACCGCATTGCGGTATCAACGCCCGGATAGCTCA
RIL+25-F2	GTAC***AGATCT***AGATCAAAAGCCA**TTGACT**CAGCAAGGGTTGACCGTATAATTCACGC
pFAB324-F1	TCTAACCTATAGGTATAATGTGTGGAT GCCCGGATAGCTCAGT
pFAB324-F2	GATCAGAG AAAAAATTTATTTGCTTTAAAG TCTAACCTATAGG
IBS	AAAAAAGCTTATAATTATCCTTAAATTGCTGCCATGTGCGCCCAGATAGGGTG
EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCTGCCATCTTAACTTACCTTTCTTTGT
EBS2	TGAACGCAAGTTTCTAATTTCGGTTCAATTCCGATAGAGGAAAGTGTCT
DgcZ-co	CACTGGTAGTGGGCATCGATAGCCTTATTGAGATTTAACAAG

Table 3.2 – Promoter sequences

Promoter sequences used in the constitutive promoter screen. All sequences are 5'→3'. The tRNA scaffold was cloned immediately 3' of the last nucleotide listed.

T7	CGATCCCGCGAAATTAATACGACTCACTATAGGG
CP25	CTTTGGCAGTTTATTCTTGACATGTAGTGAGGGGGCTGGTATAATCACATAGTACTGTT
pFAB708	TTGACAATTAATCATCCGGCTCGTAATGTTTGTGGA
pFAB+25	TTGACAATTAATCATCCGGCTCGTAATGTTTGTGGAGATTACACCGCATTGCGGTATCAAC
RIL	AGATCAAAAGCCATTGACTCAGCAAGGGTTGACCGTATAATTCACGC
RIL+25	AGATCAAAAGCCATTGACTCAGCAAGGGTTGACCGTATAATTCACGCGATTACACCGCATTGCGGTATCAAC
pFAB324	AAAAAATTTATTTGCTTTAAAGTCTAACCTATAGGTATAATGTGTGGAT

Table 3.3 – Alternate tRNA scaffold sequences

The Spinach2 sequence insertion site is designated by a /-/. All sequences are 5'→3'.

H. sapiens	$tRNA^{Lys}$	GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCGG /-/ CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGCCA
O. sativa	$tRNA^{Ala}$	GGGGGATGTAGCTCAAATGGTAGAGCGCTCGC /-/ GCGAGAGGCACGGGGTTCGATCCCCCGCATCTCCA
C. korarchaeum	$tRNA^{Val}$	GGGGCCGGTGGTCTAGAGGCTATGACGCCGCC /-/ GGCGGAGGTCGTGGGTTCGAGTCCCACCCGGCCCA
P. arsenaticum	$tRNA^{Val}$	GGGCCCGTAGTCTAGCGGTATGATGCCCGC /-/ GCGGGTGGTCCCGGGTTCAAATCCCGGCGGGCCCA
D. kamchatkensis	$tRNA^{Val}$	GGGGCCCGTCGTCTAGCTTGGTTAGGACGCCGCC /-/ GGCGGAGGCCCGGGGTTCAAGTCCCCGCGGGCCCA
E. coli	$tRNA^{Met}$	GGCTACGTAGCTCAGTTGGTTGAGCAGCGGCCGATATCCGCG /-/ GGTCACAGGTTCAAGTCCCGTCGTAGCCACCA
–	Malat1	/-/ TAGGGTCATGAAGGTTTTTCTTTTCCTGAGAAAACAACACGTATT GTTTTCTCAGGTTTTGCTTTTTGGCCTTTTTCTAGCTTAAAAAAA AAAAAAGCAAAA

Chapter Four

Applying GEMM-I-based RNA biosensors towards c-AMP-GMP signaling

Portions of this chapter have been previously published in:

Ren, Aiming, Xin C Wang, Colleen A Kellenberger, Kanagalaghatta R Rajashankar, Roger A Jones, Ming C Hammond, and Dinshaw J Patel. 2015. "Structural Basis for Molecular Discrimination by a 3',3'-cGAMP Sensing Riboswitch.." *Cell Reports* **11** (1): 1–12.

Hallberg, Zachary F, Xin C Wang, Todd A Wright, Beiyan Nan, Omer Ad, Jongchan Yeo, and Ming C Hammond. 2016. "Hybrid Promiscuous (Hypr) GGDEF Enzymes Produce Cyclic AMP-GMP (3', 3'-cGAMP)" *Proceedings of the National Academy of Sciences* **113** (7): 1790–95.

INTRODUCTION

Recent research has expanded the repertoire of CDN signaling molecules beyond c-di-GMP to include the mixed hybrid 3′,3′-cyclic AMP-GMP (cGAMP) (Figure 4.1). First discovered as the product of the enzyme DncV in the El Tor strain of *Vibrio cholerae*, cGAMP was found to regulate of *V. cholerae* motility and intestinal colonization in mammalian hosts. Little else was known about cGAMP signaling until our lab engineered RNA-based biosensors for cGAMP based on the GEMM-I riboswitch scaffold (Kellenberger, Wilson, et al. 2015). This cGAMP biosensor led to the discovery of a new riboswitch subclass selective for cGAMP, GEMM-Ib, in *Geobacter*, a genus of Deltaproteobacteria. Furthermore, we found that *Geobacter* use cGAMP-sensing riboswitches to regulate genes associated with extracellular electron transfer (Kellenberger, Wilson, et al. 2015; Nelson et al. 2015), an extraordinary activity that involves bacterial colonization on metal oxide surfaces (Reguera et al. 2005).

In this chapter, I apply our GEMM-I-based biosensors, including the c-di-GMP sensors introduced in the previous chapters, towards expanding our knowledge of cGAMP signaling on both the effector and synthase fronts. On the effector front, we used bioinformatic and mutational analyses, along with structural data, to identify factors that determine ligand discrimination in GEMM-I riboswitches and used it to predict the presence of such riboswitches in other organisms. On the synthase front, we performed a high-throughput *in vivo* screen of *Geobacter* GGDEF enzymes to discover a cGAMP synthase, demonstrating unprecedented activity from an enzyme class previously known to only make c-di-GMP. Further analysis of this GGDEF subclass revealed structural and sequence characteristics that we then found in a number of other organisms. These vignettes not only highlight advancements in our understanding of bacterial signaling, but perhaps more importantly, they also demonstrate the utility and potential of our biosensors towards making new biological discoveries.

Structural basis for molecular discrimination by a cGAMP-sensing riboswitch

Background & Results

The GEMM-I class of c-di-GMP-binding riboswitches spans 2000+ members from 600+ species (Griffiths-Jones et al. 2003). Crystal structures of the GEMM-I riboswitch Vc2 revealed a tuning fork architecture, in which a short helical base (P1) extends into two hairpins (P2 and P3) that form the tines (Smith et al. 2009) (Figure 4.2A). In this structure, the ligand binds at the base of the three-way junction between the three stems, with three highly conserved residues forming direct interactions with the ligand bases: a G20 residue at the bottom of the P2 stem that forms a Hoogstein base pair with the ligand Gα, a C92 at the bottom of the P3 stem that forms a canonical Watson-Crick base pair with the ligand Gβ, and an A41 between the two stems that intercalates between the two ligand bases (Figure 4.2B). The strong hydrogen bonding interactions present between Vc2 and its ligand rendered it extremely selective for c-di-GMP, with a 1000-fold higher binding affinity than any related ligands (K_D ~ 1 nM) (Sudarsan et al. 2008).

Accordingly, modifying the ligand-contacting residues on the riboswitch also changes the ligand selectivity. The single mutant C92U binds cGAMP (K_D = 19 ± 1.7 nM) (Shanahan et al. 2011) and the double mutant G20A/C92U binds c-di-AMP (K_D = 1,200 ± 130 nM) (Smith et al. 2009), albeit both variants still retain some c-di-GMP binding and have much lower affinity for their respective ligands than the WT for c-di-GMP. In addition, the single mutant G20A also binds c-di-GMP (K_D = 0.21 ± 0.07 nM) (Smith et al. 2010), and our Vc2-Spinach biosensors verified that the G20A variant binds c-di-GMP and cGAMP at similar affinities (Kellenberger et al. 2013). Intriguingly, prior phylogenetic analyses revealed that 23% of GEMM-I class riboswitches naturally have an A20 that presumably allows for cGAMP recognition (Smith et al. 2010). A fluorescent biosensor screen of several A20 phylogenetic variants found that although the majority of A20 GEMM-I were either promiscuous or still selective for c-di-GMP, most riboswitches from *Geobacter* were highly selective for cGAMP (Kellenberger, Wilson, et al. 2015). These riboswitches, termed GEMM-Ib, formed the basis for two cGAMP-selective Spinach fusion biosensors: Gm0970, from *G. metallireducens*, and Gs1761, from *G. sulfurreducens*.

The multiplicity of selectivity traits harbored by A20 riboswitches indicated that the G20A mutation was not the sole determinant of ligand selectivity. To understand this selectivity from a structural perspective, we collaborated with Dinshaw Patel's group to solve the crystal structure of the cGAMP-selective *Gs1761* bound to both cGAMP (Figure 4.3) and c-di-GMP (Ren et al. 2015) (Table 4.1). These structures revealed that a subtle shift in the contact between the ligand and the riboswitch A42, which rests at the hinge of the 3-way junction connecting the three stems, propagates to a detectable change in orientation of P2 and P3 stems (Figure 4.4). Meanwhile, the promiscuous Vc2 G20A lacks this shift, suggesting that perhaps the energetic strain induced by this shift disfavors c-di-GMP binding in the cGAMP-selective riboswitch (Figure 4.5).

We then returned to the results from the fluorescent biosensor screen to look for sequence patterns that might inform this selectivity. We grouped the sequences based on functional assessment of selectivity or promiscuity for binding cGAMP versus c-di-GMP. This revealed a trend: selective sequences exhibit conservation of nucleotide identity in the two base pairs of the P2 stem that stack directly on top of Aα (Figure 4.6, 4.7). In contrast, promiscuous sequences exhibit different types of base pairs at those positions (Figure 4.9A).

A series of mutants of the cGAMP riboswitch focused on this conserved P2a region were evaluated for their ability to bind either cGAMP or c-di-GMP. The mutations were chosen to swap the nucleotide or base pair with the corresponding sequence from the Vc2 riboswitch, which is selective for c-di-GMP. The mutations show the same effects for both Gm0970 (Figure 4.6C) and Gs1761 (Figure 4.7B) riboswitches, but the former exhibits higher fluorescence signal and binding affinity when fused to the Spinach aptamer.

It was found that changing only two nucleotides in the Gm0970/Gs1761 riboswitch aptamers was sufficient to switch ligand recognition from cGAMP to c-di-GMP (Figure 4.6C and 4.7B). Specifically, U16C makes these aptamers more promiscuous for binding c-di-GMP and is further additive to the effects of A14G (nucleotides are numbered in correspondence to the Gs1761 structure). Interestingly, U16C has a more pronounced effect than switching the orientation of

the C-G base pair that directly stacks with either Aα or Gα (Figure 4.6C). Based on the structure of Gs1761, this may be due to its proximity to the two stapling interactions at the base of the P2 and P3 stems. Additional mutations focused on the U16•G39 wobble pair show that replacement with Watson-Crick pairs generally increases recognition of c-di-GMP (Figure 4.8A). We found that either moving or reversing the wobble pair, or replacing it with unpaired bases, also makes the riboswitch more promiscuous and leads to further loss of fluorescence signal, suggesting that these mutations may destabilize the RNA fold.

In contrast, performing the reverse set of mutations to the c-di-GMP selective Vc2 riboswitch did not result in a full switch of ligand specificity (Figure 4.8B). At most, mutations led to promiscuous binding of cGAMP and c-di-GMP. Complete loss of fluorescence signal was observed upon full conversion of the conserved nucleotides to match the Gm0970/Gs1761 sequence, likely due to misfolding of the RNA. These results show that the extent of ligand discrimination is dependent on the overall sequence context as well, and that the conserved nucleotides in the P2a region are not sufficient to define whether a GEMM-I riboswitch discriminates between cGAMP and c-di-GMP. In a related example, we found that both wild-type Gs1761 and the U72C/C73U construct exhibit selectivity for cGAMP, but mutations of the P2a region showed different effects depending on the identity of those nucleotides (Figure S3C).

Nevertheless, we used the five conserved nucleotides to search for additional GEMM-I sequences that are selective for cGAMP (Figure 4.9). Most representatives were found in the *Geobacter* genus, which have been validated as harboring many cGAMP riboswitches (Kellenberger, Wilson, et al. 2015). However, four matching sequences were found in the genome of *Pelobacter propionicus*. Through the Spinach fusion assay, three of these riboswitch candidates were confirmed to be selective for cGAMP, while the fourth did not show signal in the assay (Figure 4.9B).

Discussion

The cGAMP riboswitch for which we solved the structure was identified in *Geobacter sulfurreducens*, an organism that was shown to produce all three bacterial CDNs: c-di-GMP, cGAMP, and c-di-AMP (Kellenberger, Wilson, et al. 2015). Thus, discrimination between these CDNs is critical to the regulatory function of this riboswitch, and we have shown that the wild-type Gm0970 and Gs1761 riboswitches from *Geobacter* are over ~1,600-fold selective for cGAMP versus other CDNs (Kellenberger, Wilson, et al. 2015). C-di-GMP has been shown to bind with comparable affinity as cGAMP to the G20A Vc2 riboswitch (Kellenberger et al. 2013) and to the U16C Gm0970 riboswitch in this study. The discrepancy in ligand selectivity between these riboswitches and the wild-type cGAMP riboswitches is remarkable because all of these RNAs harbor identical nucleotides at the positions that base pair with the ligand. This observation raises the important question of what other sequence and structural aspects dictate whether binding is selective or promiscuous for these two CDNs.

Our structural and biochemical experiments revealed that recognition of Aα or Gα is particularly sensitive to changes in the P2a stem (Figure 4.8). This region is involved directly in stacking and is connected to the critical intercalating nucleotide A41 (Figure 4.3D). It also is connected to

A14, the nucleotide that base pairs with Aα or Gα (Figure 4.4B). Finally, it forms an A-minor triple with A43, so the P2a stem makes interactions on both sides of the A42 "hinge." However, we also showed that the sequences of the base pairing nucleotides and the P2a stem are not always sufficient to dictate ligand selectivity, as we were unable to convert c-di-GMP Vc2 riboswitch into a cGAMP-selective riboswitch. This latter result indicates that there are other important differences between the RNA scaffolds for cGAMP and class-I c-di-GMP riboswitches.

On the outset, it was surprising that c-di-GMP and cGAMP could ever bind with similar affinities, given that Aα in the cGAMP complex should form two hydrogen bonds to A14, whereas Gα in the c-di-GMP complex should form one hydrogen bond to A14. There are several possible explanations for how these riboswitches can be promiscuous. One possibility is that the second hydrogen bond between Aα and A14 is weak and does not make a significant contribution to the stability of the complex. Another possibility is that Gα actually forms two hydrogen bonds to A14, which is effected either by a perturbation of the pKa, such that A14 is protonated at N1 (Smith et al. 2010), or by perturbation of the keto:enol equilibrium, such that Gα binds as the enol tautomer (Figure 4.4B, 4.10) (Singh et al. 2015; Westhof 2014). Both our structure of the cGAMP riboswitch bound to c-di-GMP and the published structure of the G20A Vc2 riboswitch bound to c-di-GMP (Smith et al. 2010) show that the orientation of A14 (or A20, its counterpart in Vc2) is consistent with two hydrogen bonds. A final possibility is that differences in the stacking interactions can compensate for the difference in hydrogen bonds. In fact, each of these factors may be influenced by changes to the proximal P2a stem that instead lead to selective binding of cGAMP over c-di-GMP (Figure 4.10).

The three crystal structures solved in this study, along with the published structure of c-di-GMP complexed with G20A Vc2 riboswitch, provide a useful comparison of how a selective versus promiscuous riboswitch binds cGAMP and c-di-GMP. It is striking that the selective riboswitch (Gs1761) displays different orientations of the P2 and P3 stems depending on the identity of the bound ligand (Figure 4.4A), whereas the promiscuous riboswitch (G20A Vc2) adopts nearly identical orientations when bound to either ligand (Figure 4.5A). This correlation does not distinguish between cause and effect, however, as cGAMP either binds a lower energy conformation or induces a lower energy conformation of the Gs1761 riboswitch.

Our structural and biochemical data have also shown how non-contacting nucleotides can impact ligand nucleobase selectivity for the GEMM-I riboswitch scaffold. These insights are necessary to accurately predict whether a given sequence is a c-di-GMP Vc2 riboswitch or a cGAMP riboswitch, and would assist in ongoing studies of the prevalence of 3',3'-cGAMP signaling in bacteria. Finally, the observation that subtle mutations outside of the binding pocket can change ligand recognition, as previously observed for the 2'-deoxyguanosine riboswitch (Kim et al. 2013; Edwards & Batey 2009; Pikovskaya et al. 2011), demonstrates the potential, as well as the challenge, of re-engineering natural riboswitch scaffolds for new sensing functions.

Author Contributions

X-ray structure determination ligand-riboswitch complexes was performed by A.R. under the supervision of D.J.P. The solution to the phase problem related to structure determination of the

ligand-riboswitch complex was provided by K.R.R. c-GAMP linkage isomers were provided by R.A.J. Bioinformatics, design of riboswitch-Spinach constructs, and fluorescence binding assays were performed by X.C.W. and C.A.K. under the supervision of M.C.H.

Hybrid promiscuous (Hypr) GGDEF enzymes produce cGAMP

Background & Results

The discovery that *Geobacter* produce cGAMP and harbor cGAMP-sensing riboswitches confirmed the presence of a cGAMP signaling pathway in *Geobacter* (Kellenberger, Wilson, et al. 2015). However, the synthase responsible for creating cGAMP in the cell remained a mystery. Geobacter genomes have no homologs to the cGAMP synthases DncV or cGAS, which harbor oligoadenylate synthase (OAS)-like domains and produce structurally distinct isomers of cGAMP (3', 3'-cGAMP and 2',3'-cGAMP, respectively) (Diner et al. 2013; Ablasser et al. 2013; P. Gao et al. 2013; Kranzusch et al. 2013). cGAMP-sensing riboswitches in *Geobacter* gained function via adapting the ligand binding pocket of GEMM-I riboswitches, which typically bind c-di-GMP. Thus, we considered that cGAMP signaling may have evolved in *Geobacter* by co-opting components from the c-di-GMP signaling pathway.

The *G. sulfurreducens* genome encodes 29 GGDEF domain-containing enzymes that are assigned as DGCs (Figure 4.11). These enzymes comprise nearly 1% of the coding sequences. Whereas the existence of GGDEF domains in a genome is considered sufficient proof for c-di-GMP signaling and has been used to establish the presence of c-di-GMP signaling in the greater majority of sequenced bacterial species (Seshasayee et al. 2010), the redundancy of GGDEF enzymes could have permitted at least one of these enzymes to evolve new functions. This led us to hypothesize that one or more GGDEF domains had gained cGAMP synthase activity.

To test this hypothesis, an *in vivo* flow cytometry screen of all 29 GGDEF enzymes was performed using fluorescent riboswitch-based biosensors that respond selectively to c-di-GMP or cGAMP (Figure 4.11A) (Kellenberger et al. 2013; Kellenberger, Wilson, et al. 2015). We posited that overexpression may drive dimerization of the GGDEF protein, thus enabling enzymatic activity to be assayed even in the absence of activating signal. Sixteen enzymes exhibited significant fluorescence turn-on with the c-di-GMP biosensor and little to no turn-on with the cGAMP biosensor, in line with the results for WspR, a well-characterized diguanylate cyclase (Hickman et al. 2005; De et al. 2008). Twelve others exhibited little to no-turn on with both biosensors, indicating that these enzymes were not functional or were poorly expressed under the assay conditions in a heterologous host. However, the GGDEF enzyme encoded by the gene GSU1658 displayed a markedly different signal profile, namely significant fluorescence turn-on with the cGAMP biosensor and little turn-on with the c-di-GMP biosensor. Similar results were observed for DncV, the cGAMP synthase from *V. cholerae*, although DncV and GSU1658 share no sequence homology.

To validate the results of the biosensor screen, we performed cell extract analysis of *E. coli* expressing the candidate cGAMP synthase GSU1658, a candidate diguanylate cyclase GSU1656, or empty plasmid. LC-MS and MS/MS data showed that *E. coli* do not inherently produce

cGAMP, but overexpression of GSU1658 leads to high cGAMP levels (Figure 4.12B, 4.13). Furthermore, this activity requires an active GGDEF domain, as no cGAMP was observed with the GLDEF mutant of GSU1658. In contrast, overexpression of GSU1656 leads to high c-di-GMP levels, but no cGAMP. Together with the *in vivo* screening results, these data reveal GSU1658 as a GGDEF enzyme with unprecedented cGAMP synthase activity.

Sequence alignment and analysis of the X-ray crystal structure of PleD (Chan et al. 2004) revealed that GSU1658 may harbor a Ser residue (S347) in place of the Asp residue (D344) that interacts with the nucleobase of the GTP substrate (Figure 4.14B, 4.15). The ability of the side chain hydroxyl to serve as either hydrogen bond donor or acceptor could allow both GTP and ATP to serve as substrates (Figure 4.14C). Biochemical analyses of several mutants confirmed this residue in conferring ligand specificity, as an S347D mutant only makes c-di-GMP, while an S347T mutant makes all three bacterial CDNs. However, the protein background also appears important, as the inverse D344S mutant of PleD was inactive (Figure 4.16), rather than making all three CDNs as predicted.

Previously, GGDEF enzymes were uniformly assigned as DGCs if they have the active motif [G/A/S]G[D/E]E[F/Y] (Römling et al. 2013). Our results reveal that GGDEF enzymes are a family of dinucleotide cyclases, in which DGCs are the major sub-family. GSU1658 is the founding member of a distinct sub-family of GGDEF enzymes that make hybrid CDNs and are promiscuous for ATP and GTP substrates. To survey this newfound sub-family, which we termed Hypr (hybrid, promiscuous) GGDEF enzymes, we performed a bioinformatics analysis of 32,587 predicted active GGDEF enzymes to identify sequences that harbor the D-to-S or D-to-T variation at the specificity position (Table 4.2). These two variants, which we predict give rise to Hypr activity, are rare and comprise only 0.17% of all GGDEF domains. All sequenced *Geobacter* and *Pelobacter* species have at least one Hypr enzyme (Figure 4.15) and have riboswitch effectors that regulate genes in response to cGAMP (Kellenberger, Wilson, et al. 2015; Nelson et al. 2015).

Intriguingly, bacteria that do not harbor cGAMP-selective riboswitches also appear to encode candidate Hypr enzymes in their genomes (Table 4.2). We tested three such candidates, two from *Myxococcus xanthus* and one from *Bdellovibrio bacteriovorus*, by expressing the proteins in *E. coli* and analyzing the by cell extract by LC-MS. In all tested cases that expressed well (Figure 4.17), we observed the production of cGAMP, although to varying levels (Figure 4.18). Furthermore, we isolated cGAMP from surface-grown *M. xanthus* and showed that cGAMP levels are modulated by solution versus solid growth conditions, which correlates with a proposed role in surface sensing. Taken together, our results reveal that Hypr activity is more widespread in bacteria than the distribution of cGAMP riboswitches, expanding the potential scope of cGAMP signaling (Figure 4.19).

Discussion

These vignettes highlight new frontiers in our understanding of bacterial cGAMP signaling. By examining the specific binding interactions between cGAMP and c-di-GMP-selective riboswitches, we discovered sequence motifs and structural elements that determined for ligand selectivity. By screening GGDEF domain proteins from *G. sulfurreducens*, we discovered Hypr-

GGDEFs, a subclass of enzymes that can produce cGAMP. In both cases, our biochemical approaches allowed us to identify unique sequence motifs that conferred selectivity, and our subsequent bioinformatics analyses found cGAMP signaling components in several organisms outside of *Geobacter*. Moreover, we identified Hypr-GGDEFs and verified cGAMP production in organisms that lack any cGAMP-selective GEMM-Ib riboswitches, opening the door to new questions surrounding the identity of cGAMP effectors in those organisms.

These studies were both enabled by our riboswitch-Spinach fusion biosensors, which were essential in designing high-throughput screening assays that were fast and accurate. The vastly different natures of the screening assays performed in each study – an *in vitro* screen of different riboswitch mutants to identify cGAMP effectors, vs. an *in vivo* screen of different DGCs to identify cGAMP synthases – attest to the biosensor's versatility. Alternative screening strategies such as fractionation (Sun et al. 2013; Wu et al. 2013), *in vitro* enzyme screening (Corrigan et al. 2013), or phenotype-based assays in the native organism (Lori et al. 2015) are relatively more time- and resource-intensive.

Following our original hypothesis that components of the c-di-GMP signaling pathway were co-opted for cGAMP signaling, we predict that variant PilZ domains and potentially novel effectors exist in these cGAMP-producing organisms. Our results reveal that the large abundance and redundancy of GGDEF genes in bacterial genomes have allowed this enzyme family to diverge and evolve towards new synthase activity; the distribution of HD-GYP and EAL phosphodiesterase domains also are expanded in Deltaproteobacteria (Seshasayee et al. 2010), and at least one variant HD-GYP domain has been shown to degrade cGAMP (J. Gao et al. 2015). Besides synthesizing cGAMP, we also showed that GGDEF domains can make c-di-AMP, an activity that had been speculated (Nelson et al. 2015), but only now proven. We conceive that larger distortions of the substrate-binding pocket, including in the signature GGDEF motif, could accommodate synthesis of pyrimidine-containing CDNs, which would expand the palette of CDN signaling molecules in nature.

Author contributions

Z.F.H. and M.C.H. designed research; X.C.W. developed new reagents/analytic tools, performed the GGDEF screen, and contributed bioinformatics analysis. Z.F.H., T.A.W., B.N., O.A., and J.Y. performed all other experiments.

MATERIALS AND METHODS

General reagents and oligonucleotides

All oligonucleotides were purchased from Elim Biopharmaceuticals (Hayward, CA) or IDT (Coralville, IA). *Geobacter sulfurreducens* PCA was obtained from the laboratory of John Coates at UC Berkeley. Genomic DNA from *G. sulfurreducens* was isolated using the Purelink Genomic DNA mini kit (Invitrogen). Genomic DNA from *Myxococcus xanthus* was obtained from the laboratory of David Zusman at UC Berkeley. Additional GGDEF domain-containing synthase genes were purchased as gBlocks from IDT (Table 4.5). CDNs were purchased from Axxorra

(Farmingdale, NY) or enzymatically synthesized. DFHBI was chemically synthesized following literature protocols (Paige et al. 2011).

RNA preparation, purification and complex formation for crystallography

The aptamer domain of the 3',3'-cGAMP riboswitch followed by the hammerhead ribozyme was transcribed *in vitro* using T7 RNA polymerase (Pikovskaya et al. 2011). To facilitate the *in vitro* transcription, a G1-G2 step was introduced instead of the G1-A2 step at the 5'-end of the native riboswitch. In addition, our sequence contained U72-C73 and it was only shown later that the natural riboswitch sequence should be corrected to C72-U73. The transcribed RNA was purified by denaturing polyacrylamide gel electrophoresis (PAGE), followed by anion-exchange chromatography and ethanol precipitation. The complex was generated by annealing the purified 3',3'-cGAMP riboswitch at 70 ˚C with 3',3'-cGAMP or c-di-GMP in a 1:2 molar ratio for 5 min in a buffer containing 100 mM K-acetate, pH 6.8, and 5 mM $MgCl_2$, followed by incubation at 37 ˚C for 5 min and then cooling on ice for 1 h before setting up crystallization trials.

The G20A mutant c-di-GMP riboswitch was transcribed *in vitro* using T7 RNA polymerase similar to the wild-type c-di-GMP riboswitch (Smith et al. 2009). A U1A protein binding RNA loop was also introduced into stem P3 of the G20A mutant to facilitate the crystallization. To improve the crystal resolution, the last G at the 3'-end was removed. The RNA-ligand complex was generated by annealing the purified c-di-GMP riboswitch mutant G20A at 70 ˚C with 3',3'-cGAMP in a 1:2 molar ratio for 5 min in a buffer containing 100 mM K-acetate, pH 6.8, and 5 mM $MgCl_2$ followed by incubation at 37 ˚C for 5 min and then cooling on ice for 1 h. U1A protein in the same buffer was added in a 1:1 molar ratio to the complex and incubated for half an hour before setting up crystallization trials.

Crystallization

The crystals of the aptamer domain of the 3',3'-cGAMP riboswitch with bound 3',3'-cGAMP or c-di-GMP were grown at 20 ˚C over a period of 1 week using the sitting-drop vapor diffusion approach after mixing the complex at an equimolar ratio with the reservoir solution containing 0.1 M Na/K-phosphate, pH 6.2-6.6, 0.2 M NaCl and 40-45% PEG400. For data collection, the crystals were quickly flash-frozen in liquid nitrogen.

The crystals of the complex of the G20A mutant c-di-GMP riboswitch with 3',3'-cGAMP were grown from the condition of 0.1 M Na-citrate pH 5.5, 5% PEG1000, 35% iso-propanol over 2 weeks. To collect the x-ray diffraction data, the crystals were transferred to 0.1 M Na-citrate pH 5.5, 5% PEG1000, 35% MPD and quickly flash-frozen in liquid nitrogen.

X-ray data collection and refinement

All the data were collected at 100 K at the NE-CAT beamline ID24C at the Advanced Photon Source, Argonne National Laboratory and processed with XDS programs. The structure of the 3',3'-cGAMP riboswitch bound with 3',3'-cGAMP (space group: $P2_1$) was solved by molecular replacement method based on the structure of c-di-GMP bound to the c-di-GMP Vc2 riboswitch (PDB code: 3IRW) as starting model. The initial RNA model was traced and built in COOT

(Emsley & Cowtan 2004) and refined in PHENIX (Adams et al. 2002). Metal ions and their coordinated waters were identified based on *2Fo-Fc* and *Fo-Fc* maps guided by the coordination geometries. 3',3'-cGAMP molecules were added to the model at the last stage based on the experimental and refined maps, coupled with electrostatic analysis. The structure of the 3',3'-cGAMP riboswitch with bound c-di-GMP (space group: C2) was solved and refined using 3',3'-cGAMP-bound structure as a starting model. The x-ray statistics of the crystals of 3',3'-cGAMP riboswitch with bound 3',3'-cGAMP and with bound c-di-GMP are listed in Table S1.

The structure of the G20A mutant c-di-GMP Vc2 riboswitch with bound 3',3'-cGAMP (space group: $P2_1$) was solved and refined with the structure of the c-di-GMP riboswitch with bound c-di-GMP (PDB code: 3IRW) as starting model. The x-ray statistics of the crystal of G20A mutant c-di-GMP Vc2 riboswitch bound with bound 3',3'-cGAMP are listed in Table S3.

Spinach-based fluorescence assays for ligand binding

DNA templates (Table 4.3) corresponding to riboswitch-Spinach constructs were either purchased as single-stranded oligonucleotides (IDT) or generated via QuikChange site-directed mutagenesis (Agilent Technologies) (Table 4.4) from a template plasmid following manufacturer instructions. After constructs were confirmed by sequencing, templates were amplified via PCR with primers F and R and the RNA was transcribed *in vitro* using T7 RNA polymerase and purified by denaturing (7.5 M urea) 6% PAGE gel. After elution from the gel, the RNA was precipitated with ethanol, dried, resuspended in water, and the stock concentration was determined by thermal hydrolysis (Wilson et al. 2014).

Fluorescence activation assays to determine ligand selectivity of different riboswitch-Spinach constructs were performed as previously described (Kellenberger et al. 2013). Briefly, all assays were performed in 96-well plates (Corning Costar 3915) and analyzed in a SpectraMax Paradigm plate reader (Molecular Devices). RNAs were refolded in binding buffer (40 mM HEPES, 125 mM KCl, and 10 mM $MgCl_2$ at pH 7.5) and added to wells containing binding buffer, 10 μM DFHBI, and ligand at given concentrations or no ligand (control). The reaction plate was incubated at 30 ˚C in the plate reader and fluorescence measurements were taken over time using 460 nm excitation / 500 nm emission or 448 nm excitation / 506 nm emission (for Figure 4.8A inset). Reported fluorescence values are for reactions that have reached equilibrium and with background fluorescence subtracted, which is defined as fluorescence of buffer without DFHBI.

Molecular cloning

For untagged constructs used for flow cytometry screening with fluorescent biosensors, gene sequences were amplified from genomic DNA and inserted into the MCS2 region of pCOLADuet-1. For C-terminal 6x-His-tagged constructs, gene sequences were inserted between NdeI and XhoI restriction sites of pET24a or pET31b. For N-terminal 6xHis-MBP-tagged constructs, gene sequences were inserted between BamHI and XhoI restriction sites of a custom pET16-derived vector (Kranzusch et al. 2013). pET28a containing *E. coli* BL21 (DE3)-derived yhdH between the NdeI and EcoRI cut sites was provided by the M. Chang lab at the University of California at Berkeley. Primers are listed in Table 4.6.

Fluorescent biosensor screening assay

Chemically competent *E. coli* BL21 (DE3) Star cells (Life Technologies) were co-transformed with different combinations of biosensor plasmid (pET31b with GM0970 P1-4delA-Spinach or Dp-Spinach2) and enzyme construct plasmid (pCOLA-Duet1, various enzymes). Single colonies from LB/Carb/Kan plates were picked for overnight starter cultures, which were used to inoculate fresh liquid cultures. Cells were grown aerobically to an OD_{600} ~ 0.3, then biosensor and enzyme expression was induced with 1 mM IPTG at 37 ˚C for 6 h. 2 μL of each culture was diluted into 60 μL of 1xPBS containing 50 μM DFHBI. Cellular fluorescence was measured for at least 10,000 cells using a BD Fortessa X20 flow cytometer with BD FACSDiva Software (Version 1.0.0.650) located in the Flow Cytometry Core Facility at the University of California at Berkeley. Flow cytometry data was then analyzed by FlowJo (Version 10.0.7).

Fluorescence turn-on was analyzed by the Student's t-test using 1 tail and 2 sample equal variance parameters, $p<0.01$ was the cut-off for significant turn-on. For the c-di-GMP biosensor, the significance test was between candidate GGDEF signal and pCOLA signal. The cGAMP biosensor is ~100-fold selective for cGAMP over c-di-GMP, but some fluorescence above background is still observed for c-di-GMP synthases. Thus, for the cGAMP biosensor, the significance test was between candidate GGDEF signal and WspR signal.

Liquid culture growth of E. coli BL21 (DE3) Star for nucleotide extraction

Overnight starter cultures of BL21 (DE3) Star cells containing the pRARE2 plasmid (Invitrogen) and dinucleotide cyclase enzymes in pET24a (or pET31b for GSU1656; pET-MBP for ACP_2467, Calni_1629, and DEFDS_0689) were inoculated into LB media and grown aerobically to an OD_{600} ~ 0.3. Cultures were then induced with 1 mM IPTG at 28 ˚C for 4 h. Cells were harvested by centrifugation at 4,700 rpm for 15 min at 4 ˚C, and pellets were stored at -80 ˚C.

Cell extraction from E. coli

CDNs were extracted as described previously (Spangler et al. 2010), with the following modifications. A frozen cell pellet from 100 mL of liquid culture was thawed and resuspended in 1.4 mL extraction buffer (40% methanol, 40% acetonitrile, 20% ddH_2O). The cell solution was incubated at ambient temperature with agitation for 20 min. After centrifugation for 5 min at 13,200 rpm, the supernatant was carefully removed and stored on ice. The remaining pellet was extracted twice more as described, with 700 μL extraction solvent each time. The combined supernatants were evaporated to dryness by rotary evaporation, and the dried material was resuspended in 300 μL ddH_2O. The extract was filtered through a 3 kDa MWCO Amicon Ultra-4 Protein Concentrator (Millipore) and used immediately or stored at -20 ˚C.

LC-MS analysis of *E. coli* cell extracts was performed using an Agilent 6120 Quadrupole LC-MS with an Agilent 1260 Infinity liquid chromatograph equipped with a diode array detector. Sample volumes of 20 μL were separated on a Poroshell 120 EC C18 column (50 mm length x 4.6 mm internal diameter, 2.7 μm particle size, Agilent) at a flow rate of 0.4 mL/min. For analysis of cell extracts and purified protein, a linear elution program of 0 to 10% B over 20 min. Solvent A was H_2O + 0.05 % TFA and solvent B was MeCN + 0.05 % TFA. Under the former conditions,

the CDNs in extracts were found to always elute in the order of c-di-GMP (7.3±0.3 min), cGAMP (7.6±0.3), and c-di-AMP (7.9±0.4 min). Due to slight variability in retention times, the assignment of CDN identity was made through analysis of the mass spectra. Shown in figures are the MS spectra from integrating the retention time region containing all three CDNs (6 to 8 min).

Extract samples were analyzed by MS in the positive ion mode using the range of m/z = 600 to 800. When a broader range of 100 to 1000 m/z was used, the expected mass for the corresponding CDN was present, but was not the most abundant ion peak observed, even with the standards. This observation suggests that the relative ionization of CDNs is low under these conditions, and furthermore the CDNs may not be fully resolved from other small molecules present in the extract. Thus the UV absorbance peaks detected at 254 nm may not be solely attributable to CDNs.

For high-resolution and tandem MS/MS, lysate was first fractionated on a Agilent 1260 Infinity liquid chromatograph equipped with a diode array detector and analytical-scale fraction collector as previously described (Kellenberger, Wilson, et al. 2015). High-resolution mass spectrometry (HRMS) and tandem mass spectrometry (MS/MS) measurements of collected fractions were performed as previously described (Kellenberger, Wilson, et al. 2015) 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 (ESI) source (Thermo Fisher Scientific). This instrumentation is located in the QB3/Chemistry Mass Spectrometry Facility at UC Berkeley.

In vitro activity assay of dinucleotide cyclases using radiolabeled NTPs

In vitro activity assays were performed as previously described by Kranzusch et al., with slight modifications (25). 10 μM enzyme was incubated in a solution of 50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 10 mM $MgCl_2$, and 5 mM dithiothreitol with the indicated amounts of ATP and GTP and ~0.1 μCi radiolabeled [α-^{32}P]-ATP or [α-^{32}P]-GTP (Perkin Elmer) as indicated. Reactions were incubated at 28 °C for 1 h. The total concentration of radiolabeled nucleotide did not exceed 66 nM, and so we expect that this does not significantly affect the results of any ratio-based experiments performed. Following incubation, the reaction was treated with 20 units of Calf Intestinal Alkaline Phosphatase (NEB) at 28 °C for 30 min to digest the unincorporated NTPs. Reactions were terminated by heating to 95 °C for 30 s. The reaction mixture (1 μL) was then spotted onto a PEI-cellulose F Thin-Layer Chromatography plate (Millipore), and allowed to dry for 15 min at room temperature. TLC plates were developed using 1 M KH_2PO_4, pH 3.6. Plates were dried overnight post-development, and radiolabeled products were detected using a Phosphor-image screen (GE Healthcare) and a Typhoon scanner (GE Healthcare).

In vitro activity assay of dinucleotide cyclases using LC-MS

In vitro activity assays were performed as described above, with omission of both radiolabeled nucleotides and digestion with Calf Intestinal Alkaline Phosphatase. After termination of the reactions by heating to 95 °C for 30 s, reactions were filtered using a 0.45 μm filter, and analyzed by LC-MS. For LC-MS analysis, an elution program of 0% B for 5 minutes, followed by a linear

gradient from 0 to 5% B over 10 min, was used. Solvent A was H_2O + 0.05 % TFA and solvent B was MeCN + 0.05 % TFA. Under these conditions, the CDNs in extracts were found to always elute in the order of c-di-GMP (8.7 ± 0.3 min), cGAMP (10.6 ± 0.3 min), and c-di-AMP (11.0 ± 0.5 min). Due to slight variability in retention times, the assignment of CDN identity was made through analysis of the mass spectra. Shown in figures are the MS spectra from integrating the retention time region containing all three CDNs (8 to 12 min).

Bioinformatic Analysis of GGDEF Variants

A Python-based program was developed to extract alignment data for a library of 42,747 putative GGDEF domain-containing proteins from the Pfam database (accession PF00990, http://pfam.xfam.org/, accessed 06/05/2014). In particular, positions critical for catalytic activity (i.e. the GG[D/E]EF sequence) and selectivity (i.e. positions 344 and 326 in PleD) were identified and analyzed for each sequence. Given previous results with some DGCs possessing altered signature motifs, we assigned any diguanylate cyclase with a [G/A/S]G[D/E][F/Y] motif to be active.

FIGURES

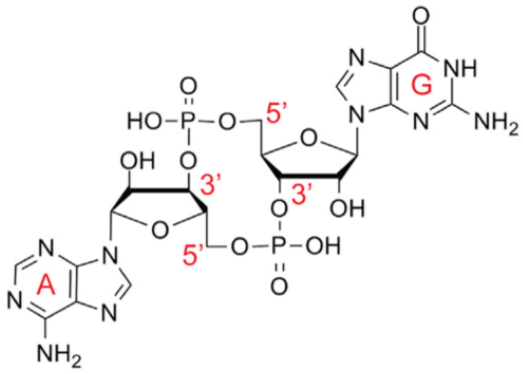


Figure 4.1 – structure of 3',3' cyclic AMP-GMP (cGAMP). The linkage numbering and base identities are written in red.

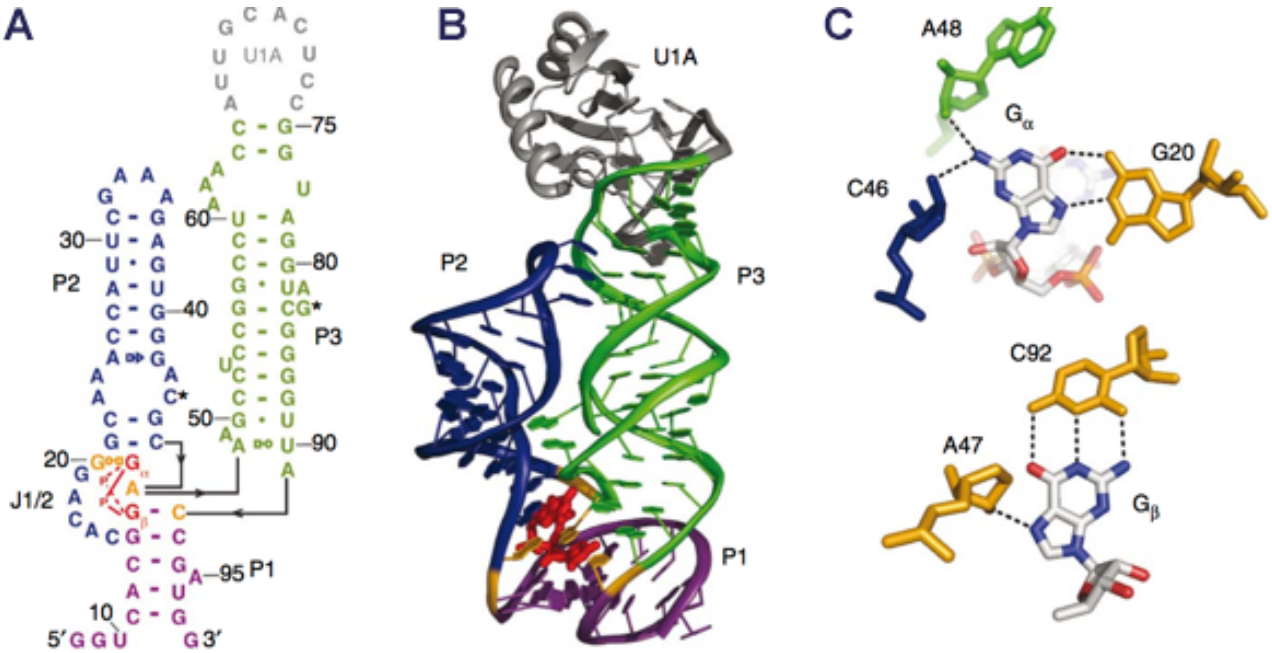


Figure 4.2 – Crystal structure of Vc2 with c-di-GMP bound (PDB: 3IRW).

A. Secondary structure representation of the crystallized c-di-GMP aptamer. Helices P1, P2 and P3 are colored purple, blue and green, respectively. c-di-GMP is shown in red. Nucleotides that directly contact c-di-GMP are shown in orange. The U1A cocrystallization sequence is shown in gray. Numbering is the same as previously published for this sequence8. Watson-Crick base pairs are denoted by dashes. * indicates a Watson-Crick base pair in the tertiary structure. The two guanines of c-di-GMP are denoted Gα and Gβ for clarity.
B. Representation of the crystal structure of the Vc2 c-di-GMP riboswitch aptamer from *V. cholerae*. Coloring is the same as used in **a**.
C. Top, the Hoogsteen face of Gα is recognized by G20 and the N2 of Gα is contacted by C46 and A48. Bottom, Gβ forms a canonical Watson-Crick base pair with C92. The N7 of Gβ is contacted by A47.

Figure and legend adapted from (Smith et al. 2009).

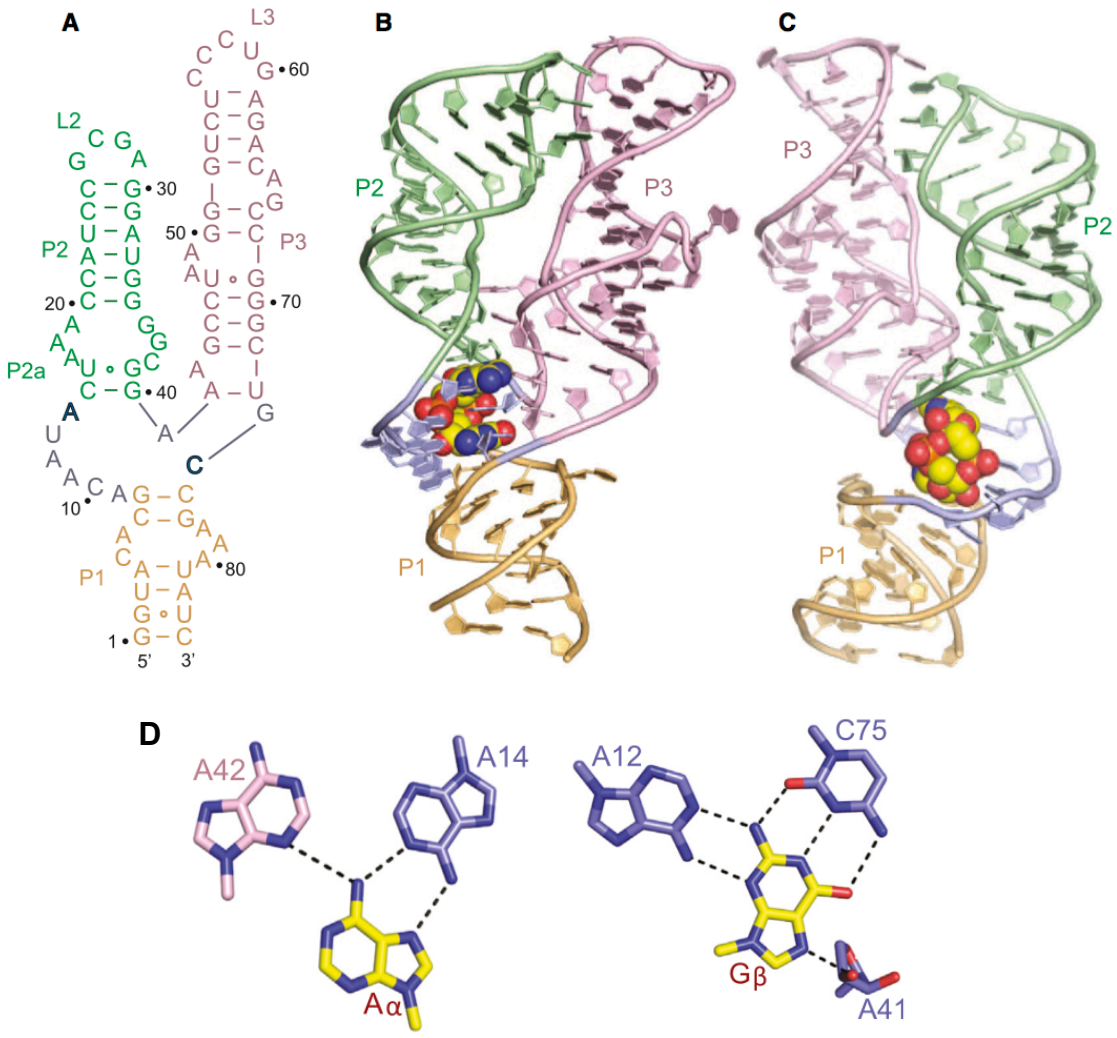


Figure 4.3 – Crystal structure of Gs1761 with cGAMP bound.

A. Schematic of the secondary structure of the cGAMP-selective riboswitch Gs1761. This sequence differs from the natural riboswitch by containing A2G, U72C, and C73U substitutions. Residues A14 and C75 that interact through base pairing with the bound cGAMP are shown in bold.
B. Views of the 2.05 Å structure of Gs1761 bound to cGAMP. The riboswitch RNA is shown in a ribbon representation and color-coded by segments, while the bound cGAMP is shown in a space-filling representation.
C. As B, but an alternate view.
D. Intermolecular recognition the cGAMP bases (in yellow) by base-base hydrogen bonding with riboswitch RNA bases (in pink and blue, with same color coding as A-C) in the complex.

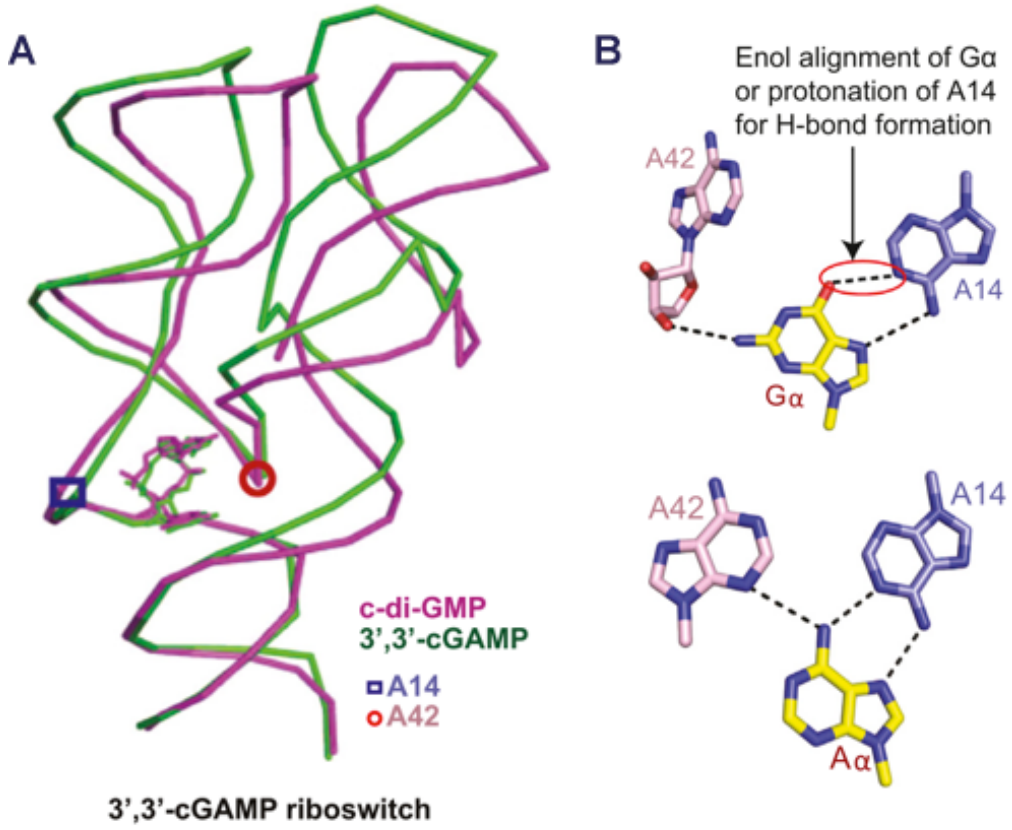


Figure 4.4 – Gs1761 bound to cGAMP vs. c-di-GMP

A. Superposition of the structures of the complexes of the cGAMP riboswitch Gs1761 with bound cGAMP (in green) and c-di-GMP (in magenta). The positions of G42 and A14 are indicated by a closed circles and squares in the two complexes.
B. Molecular interactions of the Gs1761 riboswitch bound to c-di-GMP Gα (top) or cGAMP Aα (bottom) with riboswitch bases A14 and A42. Note the different orientations of A42 in either structure. A potential hydrogen bonding interaction between the enol form of Gα and a protonated A14 is shown.

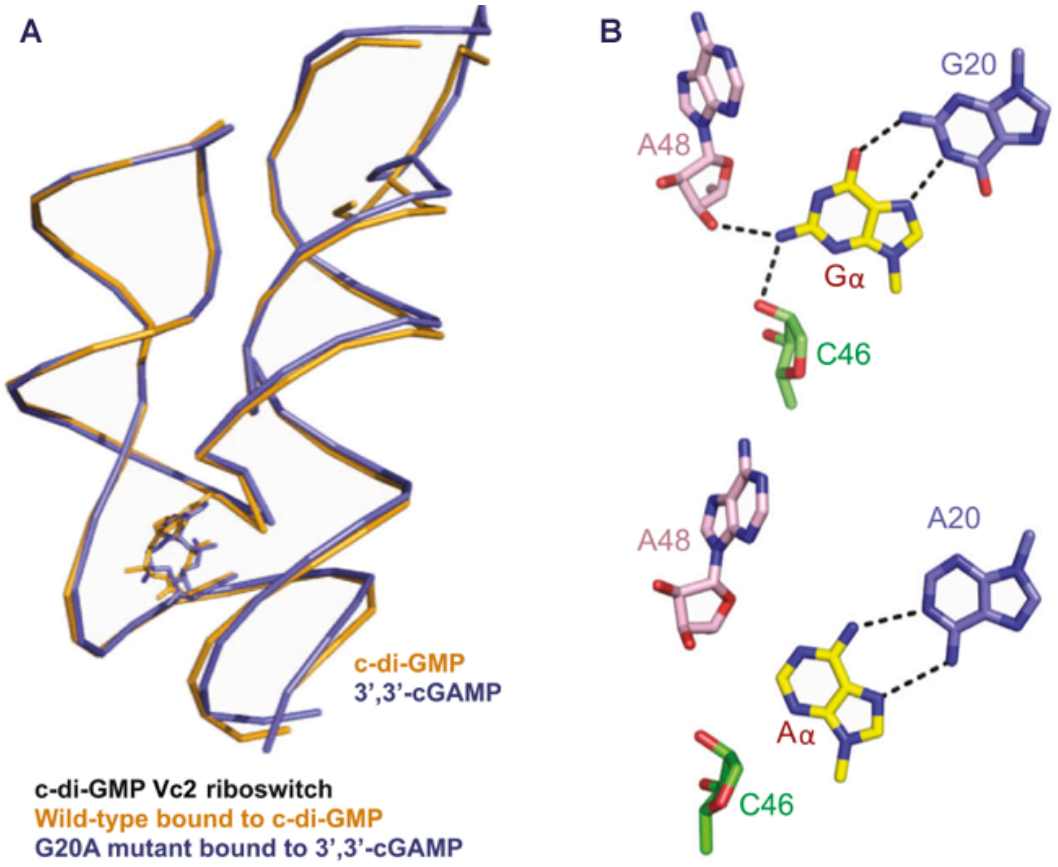


Figure 4.5 – Vc2 bound to cGAMP vs. c-di-GMP

A. Superposition of the structures of the complexes of the wild type Vc2 bound to c-di-GMP (goldenrod) and the promiscuous mutant Vc2 G20A bound to cGAMP (blue). The ligands are shown as sticks.
B. Molecular interactions of wild-type Vc2 bound to c-di-GMP (top), and Vc2 G20A bound to cGAMP (bottom).

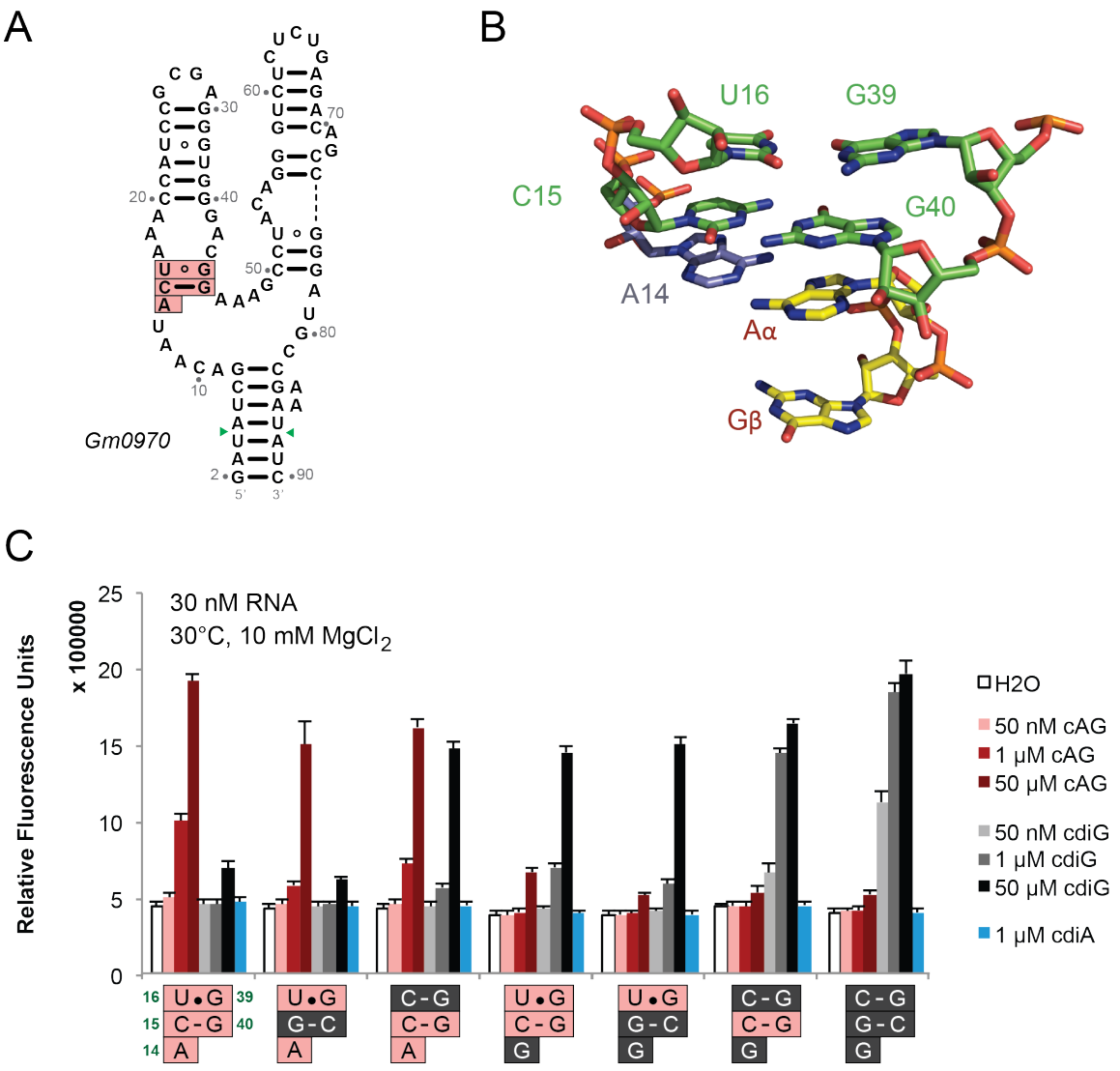


Figure 4.6 – The P2a region Affects ligand selectivity of the Gm0970 cGAMP riboswitch

A. Secondary structure of Gm0970, another cGAMP-selective riboswitch. The nucleotide numbering is set to match that of Gs1761 in the P2a region (pink boxes). Green arrows indicate the positions in the P1 stem to which the Spinach aptamer was fused.
B. Structure of the P2a region for Gs1761 with bound cGAMP.
C. Spinach-based selectivity screen of Gm0970 riboswitch constructs with mutations to the P2a region shown. Fluorescence activation was measured in the presence of no ligand or different CDNs at the indicated concentrations. The nucleotides from Gm0970 (pink) were changed to those from the c-di-GMP-selective riboswitch, Vc2 (gray).

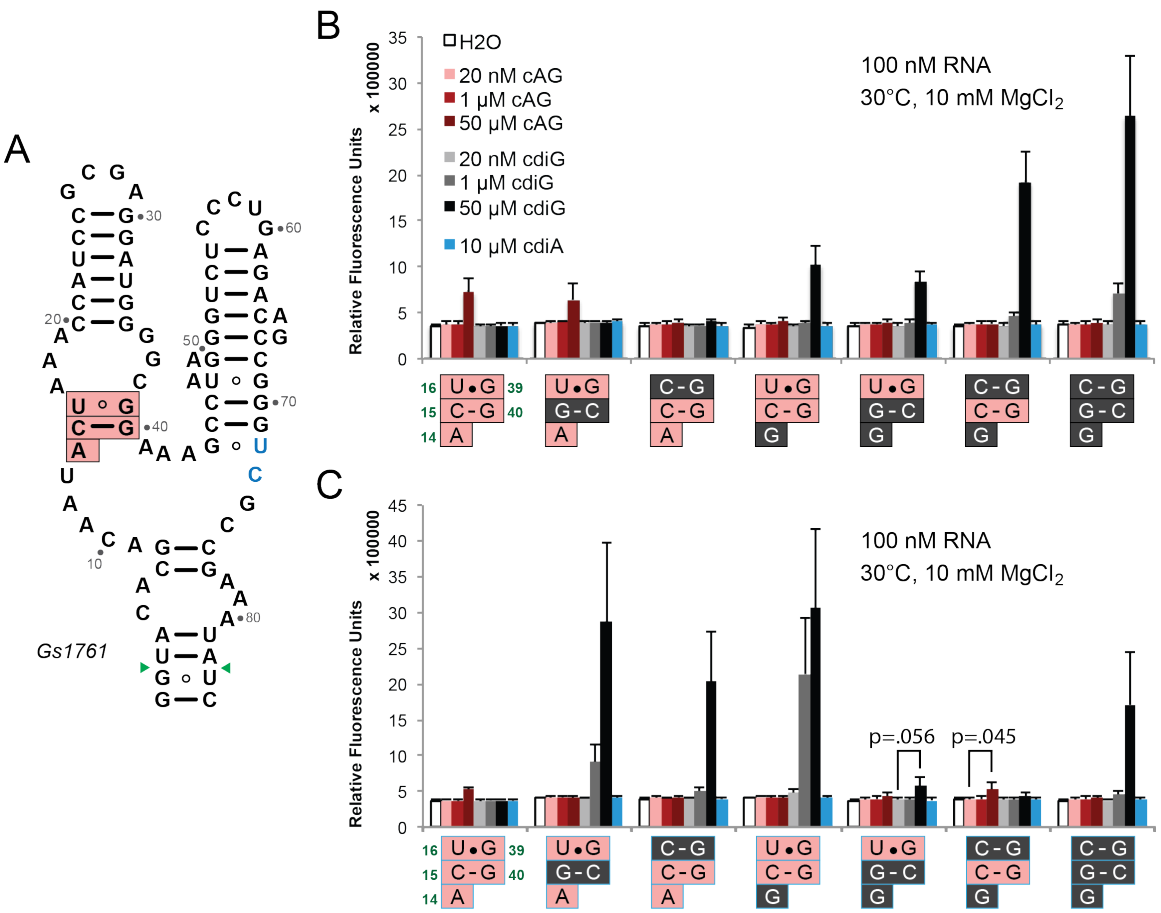


Figure 4.7 – The P2a region affects ligand selectivity of the Gs1761 3',3'-cGAMP riboswitch

A. Secondary structure of Gs1761, the 3',3'-cGAMP-selective riboswitch whose structure was elucidated by x-ray crystallography in this study. The P2a region is indicated by pink boxes. Green arrows indicate the positions in the P1 stem to which the Spinach aptamer was fused. The crystallography construct has the two blue nucleotides reversed (U72C, C73U).

B. Spinach-based selectivity screen of wild-type Gs1761 riboswitch constructs with mutations to the P2a region shown. Fluorescence activation was measured in the presence of no ligand or different CDNs at the indicated concentrations. The nucleotides from Gs1761 (pink) were changed to those from the c-di-GMP-selective riboswitch, Vc2 (gray).

C. Same as part (B) for the Gs1761 crystallography construct, with the mutations U72C and C73U.

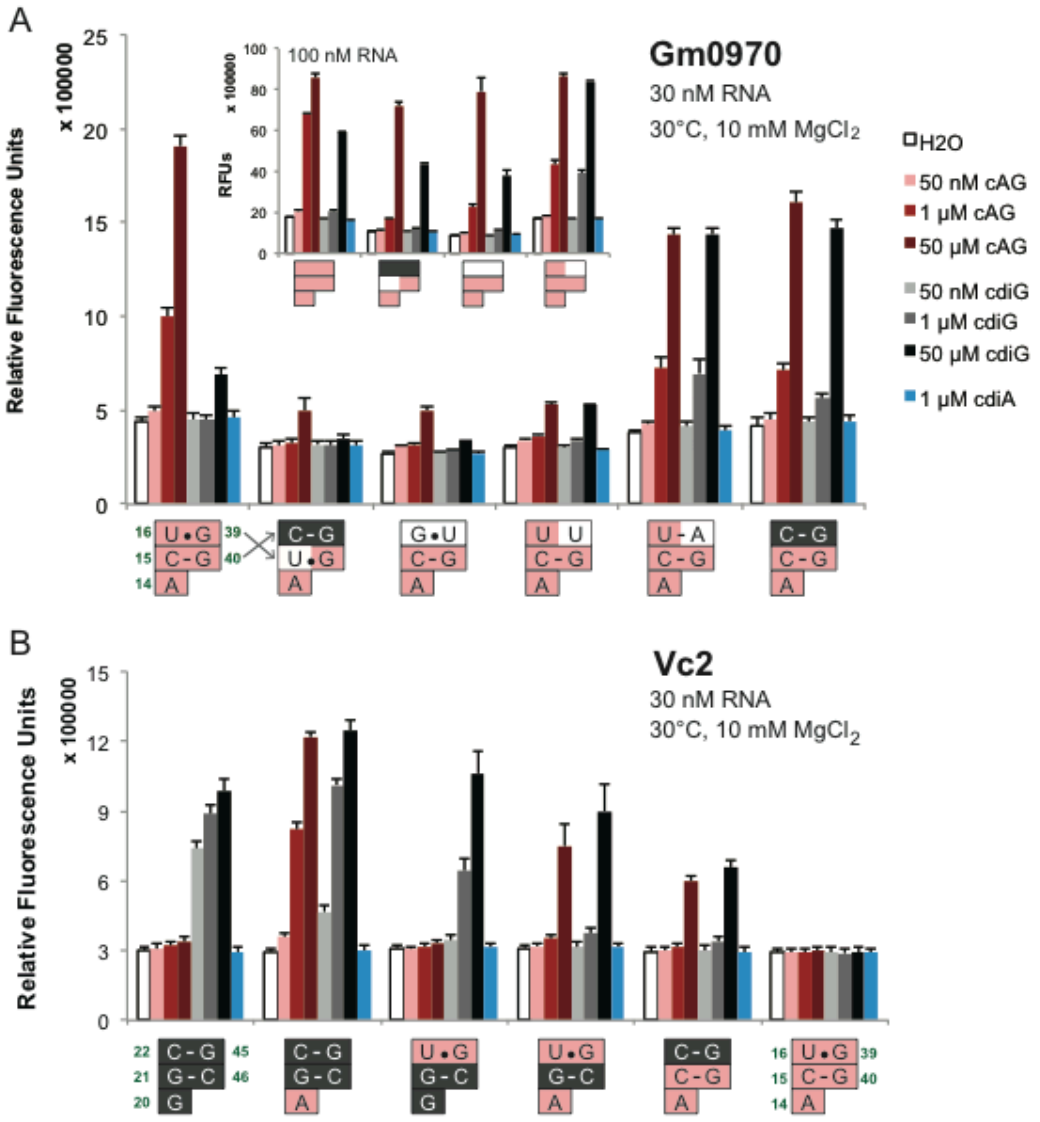


Figure 4.8 – Changes to the wobble base pair affects ligand selectivity, while corresponding mutations to Vc2 does not result in a cGAMP-selective riboswitch.

A. Spinach-based selectivity screen of Gm0970 riboswitch constructs with mutations to the P2a region shown. Nucleotide colors indicate a match to the sequence from Gm0970 (pink), Vc2 (gray), or neither (white). The inset shows data for constructs related to the first four shown in the main graph (see Methods, Table S4), except analyzed at higher RNA concentrations.

B. Spinach-based selectivity screen of Vc2 riboswitch constructs with mutations to the P2a region shown. Fluorescence activation was measured in the presence of no ligand or different CDNs at the indicated concentrations. The nucleotides from Vc2 (gray) were changed to those from cGAMP-selective riboswitches (pink). The different nucleotide numbering schemes for Vc2 and Gs1761 are shown for comparison.

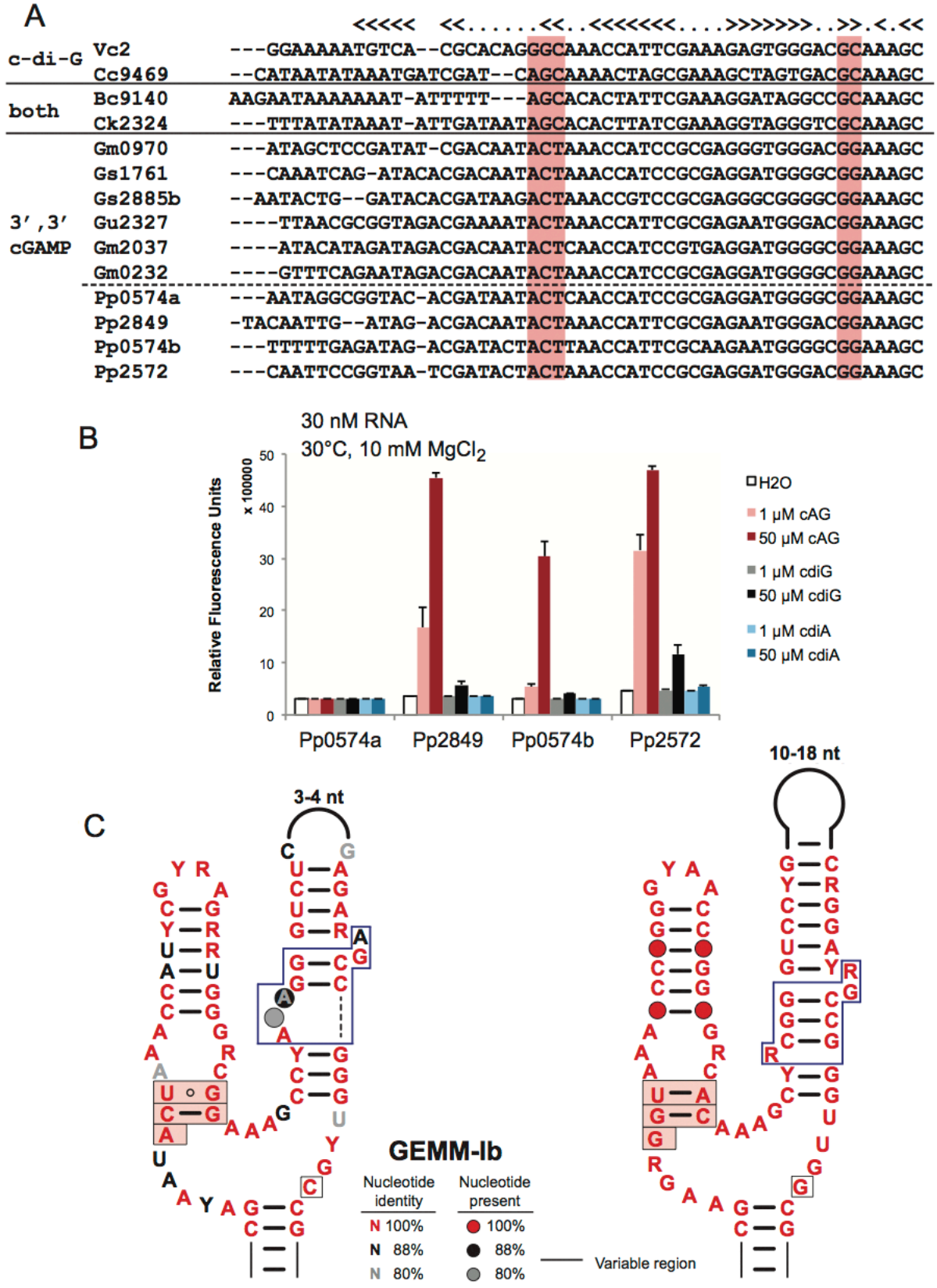

A
<<<<< <<.....<<..<<<<<<<....>>>>>>>..>>.<.<<
c-di-G
Vc2 ---GGAAAAATGTCA--CGCACAGGGCAAACCATTCGAAAGAGTGGGACGCAAAGC
Cc9469 --CATAATATAAATGATCGAT--CAGCAAAACTAGCGAAAGCTAGTGACGCAAAGC
both
Bc9140 AAGAATAAAAAAAT-ATTTTT---AGCACACTATTCGAAAGGATAGGCCGCAAAGC
Ck2324 ---TTTATATAAAT-ATTGATAATAGCACACTTATCGAAAGGTAGGGTCGCAAAGC
Gm0970 ---ATAGCTCCGATAT-CGACAATACTAAACCATCCGCGAGGGTGGGACGGAAAGC
Gs1761 ---CAAATCAG-ATACACGACAATACTAAACCATCCGCGAGGATGGGGCGGAAAGC
Gs2885b --AATACTG--GATACACGATAAGACTAAACCGTCCGCGAGGGCGGGGCGGAAAGC
3′,3′ Gu2327 ----TTAACGCGGTAGACGAAAATACTAAACCATTCGCGAGAATGGGACGGAAAGC
cGAMP Gm2037 ----ATACATAGATAGACGACAATACTCAACCATCCGTGAGGATGGGGCGGAAAGC
Gm0232 ----GTTTCAGAATAGACGACAATACTAAACCATCCGCGAGGATGGGGCGGAAAGC
Pp0574a ---AATAGGCGGTAC-ACGATAATACTCAACCATCCGCGAGGATGGGGCGGAAAGC
Pp2849 -TACAATTG--ATAG-ACGACAATACTAAACCATTCGCGAGAATGGGACGGAAAGC
Pp0574b ---TTTTTGAGATAG-ACGATACTACTTAACCATTCGCAAGAATGGGGCGGAAAGC
Pp2572 ---CAATTCCGGTAA-TCGATACTACTAAACCATCCGCGAGGATGGGACGGAAAGC
B
30 nM RNA
30°C, 10 mM MgCl2
Relative Fluorescence Units
x 100000
50
40
30
20
10
0
Pp0574a
Pp2849
Pp0574b
Pp2572
H2O
1 μM cAG
50 μM cAG
1 μM cdiG
50 μM cdiG
1 μM cdiA
50 μM cdiA
C
3-4 nt
10-18 nt
GEMM-Ib
Nucleotide identity
Nucleotide present
N 100%
N 88%
N 80%
100%
88%
80%
Variable region

Figure 4.9 – The P2a region is conserved in 3',3'-cGAMP riboswitches.

A. Partial sequence alignment of GEMM-I riboswitch sequences that were found to selectively bind c-di-GMP, cGAMP, or be promiscuous for both (Kellenberger, Wilson, et al. 2015). Except for Vc2, all sequences harbor an A at the nucleotide position predicted to interact with Aa or Ga of the ligand. The predicted base-pairs of the pairing stem P2 are denoted by < and >. The P2a region is highlighted in pink. Four sequences with conservation of this region from *Pelobacter propionicus* are shown. All full sequences are shown in Table S4.
B. Spinach-based selectivity screen of the four *Pelobacter propionicus* GEMM-I riboswitches.
C. Consensus sequence and secondary structure models for cGAMP (also called GEMM-Ib) riboswitches that harbor an A (left panel) or G (right panel) at the nucleotide position predicted to interact with Aα or Gα of the ligand. These models are based on functionally characterized sequences (28 and 4 representatives for left and right panels, respectively) from this study and from Kellenberger et al. 2015. Consistent with other figures, the P2a region is indicated by pink boxes and the bulges in P3 are indicated by a blue box. The nucleotide position predicted to interact with Gb of the ligand also is boxed.

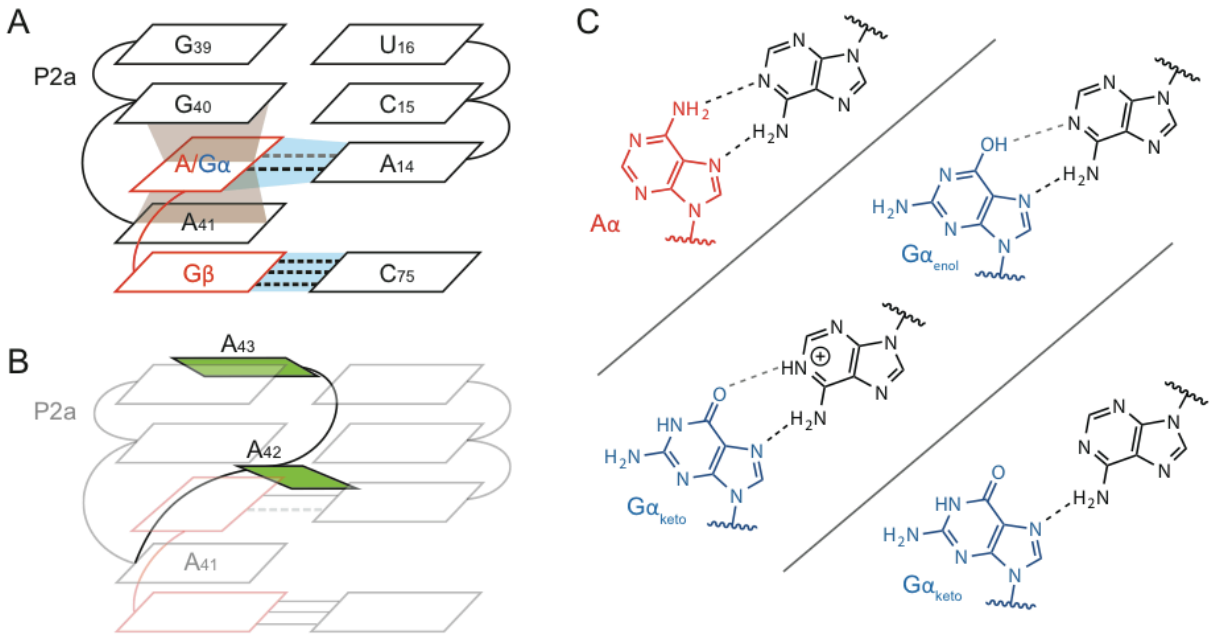


Figure 4.10 – The P2a stem and the ligand binding pocket for the 3',3'-cGAMP riboswitch.

A. A schematic depicting the proximity of the P2a stem to riboswitch nucleotides involved in stacking interactions (brown) and hydrogen-bonding (blue) to the CDN ligand.
B. Same view as in part A with inclusion of the stapling interaction between the P2a stem and A43 and showing A42.
C. Chemical structures of different possible hydrogen-bonding modes for Gα versus Aα.

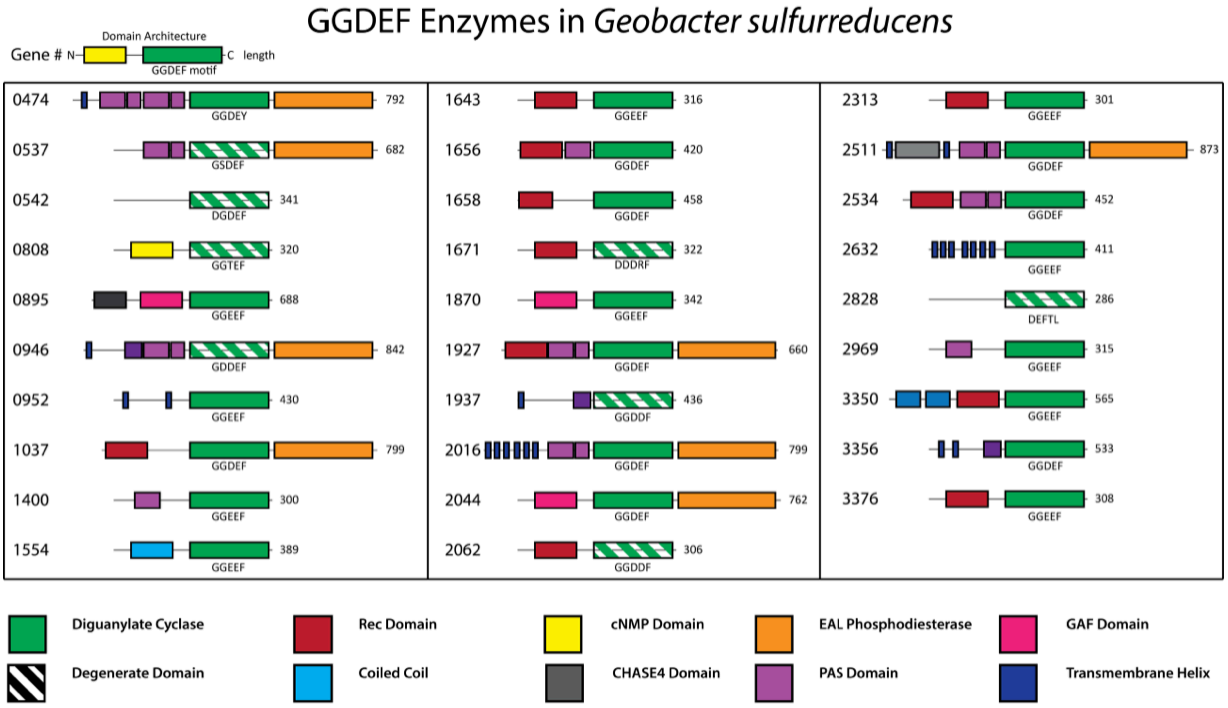


Figure 4.11 – Domain architecture of GGDEF-domain containing proteins in *Geobacter sulfurreducens PCA*

Proteins tested for c-di-GMP- and cGAMP-synthase activity are shown. REC, response receiver regulator domain found in two-component regulatory systems; cNMP, cyclic nucleotide monophosphate binding domain; EAL, c-di-GMP-specific phosphodiesterase domain; GAF, domain present in cGMP phosphodiesterases, adenylyl cyclases, and FhlA, sometimes associated with phytochromes; CHASE4, cyclase/histidine kinase associated extracellular sensor domain; PAS, PER/ARNT/SIM domain involved in oxygen, light, and redox state sensing. The residues corresponding to the "signature" "GGDEF" motif are shown below the GGDEF domain for each.

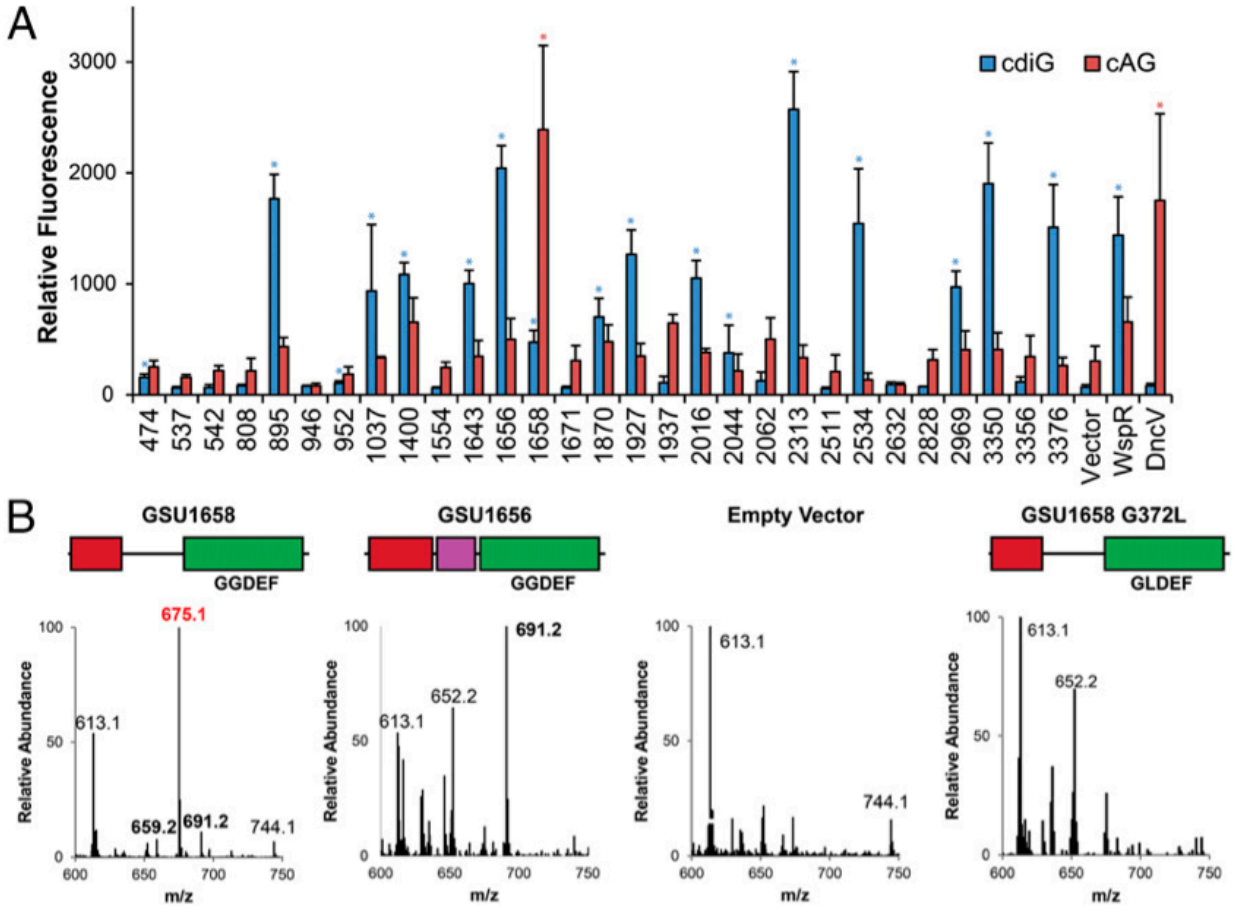


Figure 4.12 – *In vivo* fluorescent biosensor screen of 29 *Geobacter* GGDEF genes reveals a cGAMP synthase

A. Average fluorescence measured by flow cytometry (n = 3; 10,000 cells per run) of *E. coli* BL21 (DE3) Star cells coexpressing the c-di-GMP-selective biosensor Dp-Spinach2 (blue) or cGAMP-selective biosensor Gm0970-p1-4delA-Spinach (red) along with GGDEF domain proteins from *G. sulfurreducens* strain PCA, diguanylate cyclase WspR, cGAMP synthase DncV, or empty vector. Blue and red stars denote significant ($P < 0.01$) fluorescence turn-on by Student's t test above control signal (i.e., significant signal above pCOLA background for the c-di-GMP sensor; above WspR for the cGAMP sensor).

B. LC/MS analysis of *E. coli* cell extracts overexpressing constructs shown or empty vector; see Figure 4.10 for protein domain color scheme. Shown are the MS spectra from integrating the retention time region containing all three CDNs. Expected masses are for c-di-GMP (m/z = 691), cGAMP (m/z = 675), and c-di-AMP (m/z = 659).

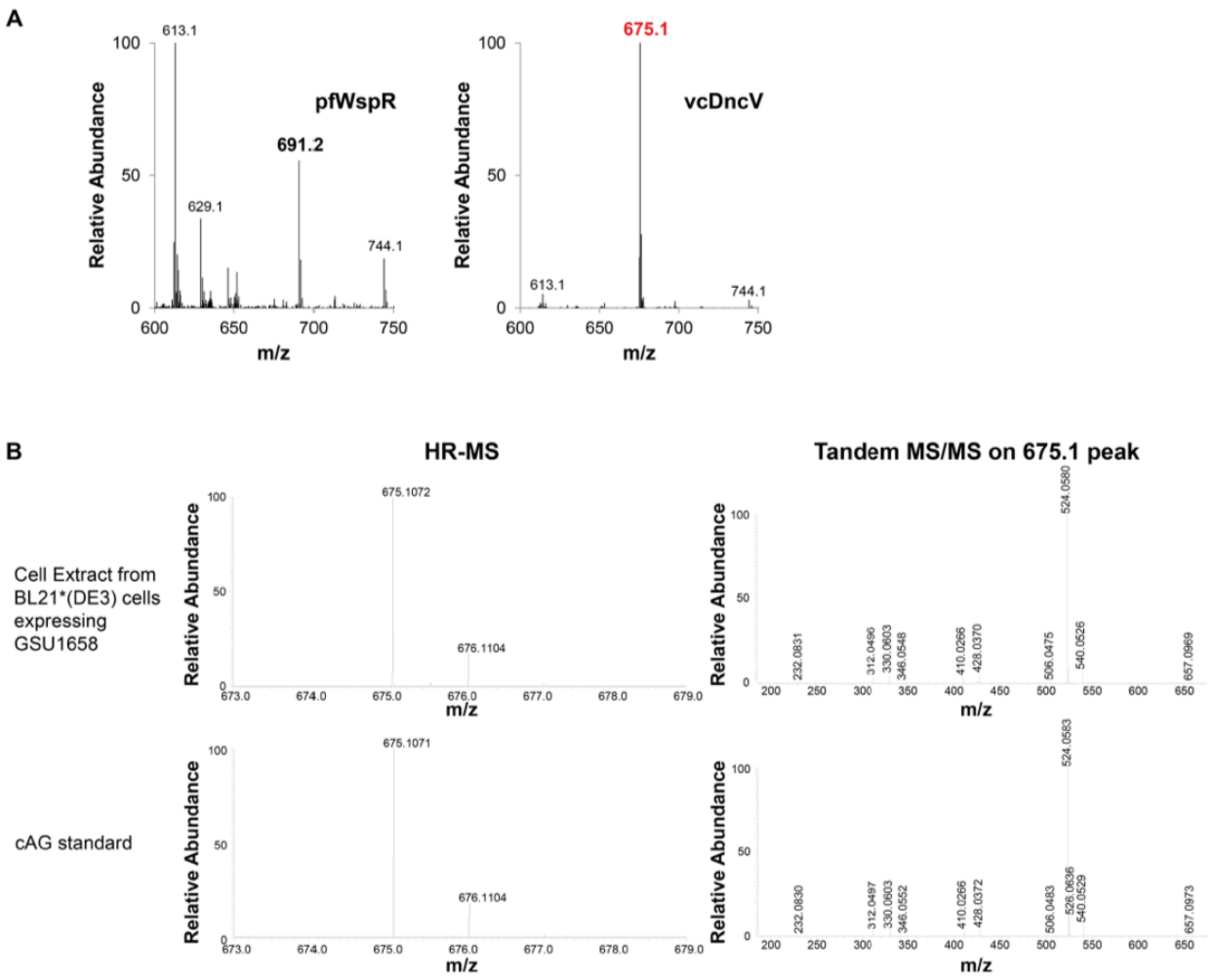


Figure 4.13 – Cell extraction of enzyme standards and GSU1658

A. HPLC-MS of lysates from *E. coli* BL21 Star (DE3) cells expressing control enzymes. Left: WspR, a c-di-GMP (m/z = 691.1) synthase; right: DncV, a cGAMP (m/z = 675.2) synthase.
B. Mass spectrometry analysis of lysate from BL21 star (DE3) cells expressing GSU1658 or synthetic 3',3'- cGAMP standard. Left: High-resolution mass-spectrometry; Right: Tandem MS/MS of the 675.1 peak observed.

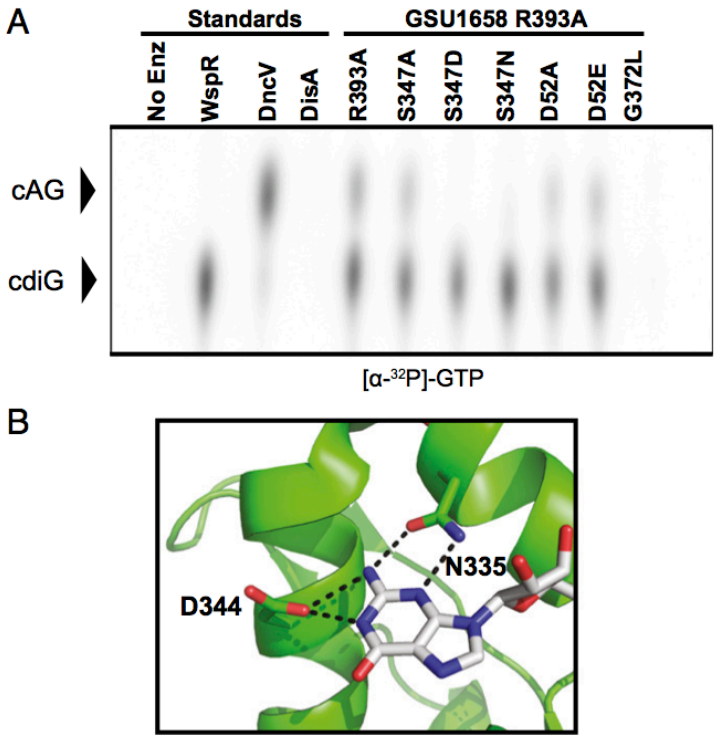


Figure 4.14 – Identification of specificity position in Hypr GGDEF active site

A. Cellulose TLC of radiolabeled products from enzymatic reactions with 1:1 ATP- to-GTP substrates in excess and doped with trace amounts of α-^{32}P-labeled GTP. Residue R393 is located in the putative I-site, S347 is located in the nucleotide binding site, and D52 is the putative phosphorylation site in the Rec domain.
B. Nucleotide binding region of PleD in complex with non-hydrolyzable GTP analog (PDB: 2V0N). Hydrogen bonding contacts between the guanine base and key protein residues are shown as dotted lines.

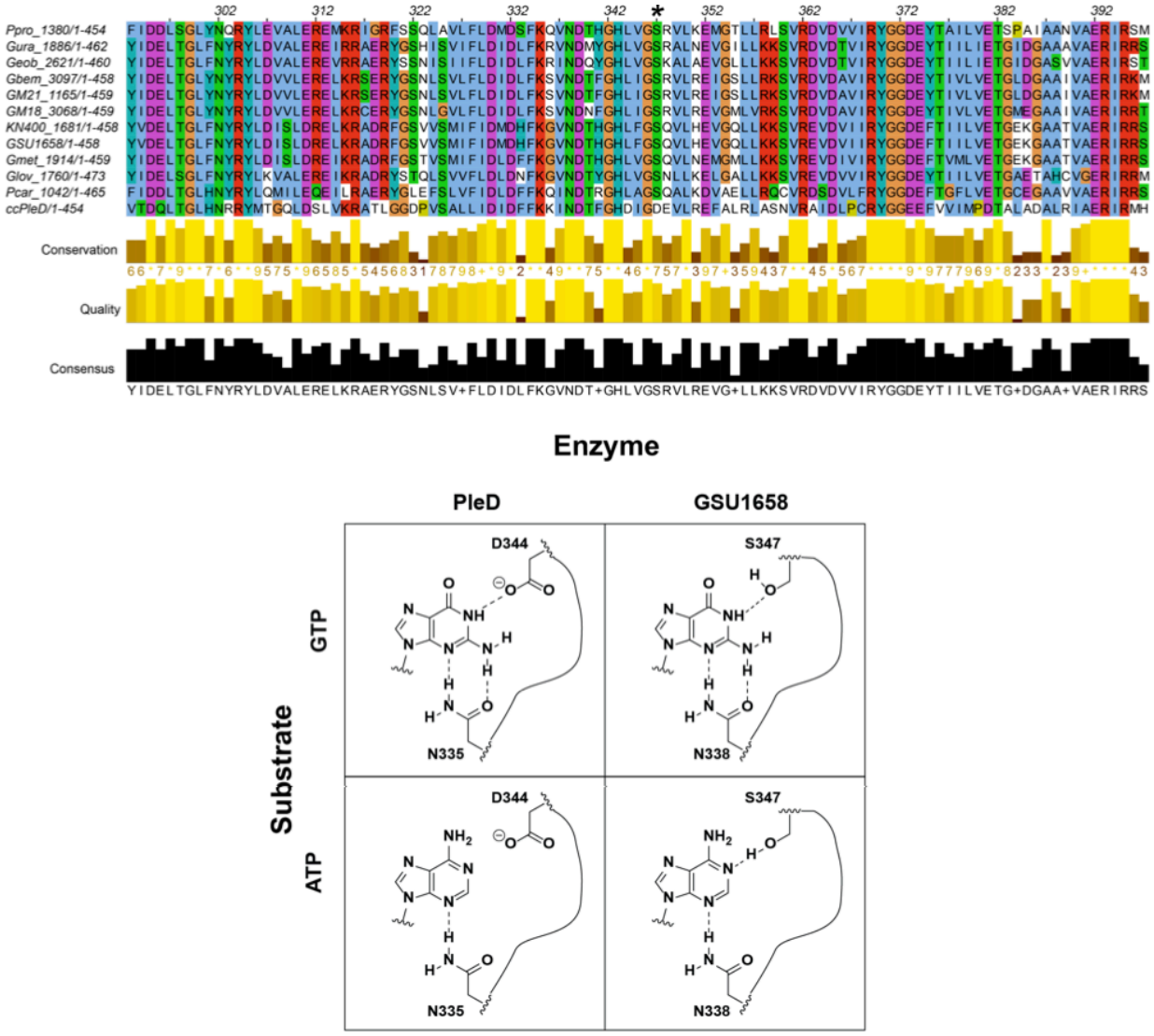


Figure 4.15 – Alignment of representative *Geobacter* GGDEF domains and proposed model for nucleotide recognition

A. Sequence alignment of the GGDEF domain of PleD, a canonical diguanylate cyclase, with all HyprA GGDEF domains in sequenced *Geobacter* and *Pelobacter* species. The encoding gene is conserved and found in the same genomic location, 5' to the histidyl tRNA synthetase gene hisS. The position of the substrate-binding aspartate, D344 in PleD, and its S347 counterpart in GSU1658 is marked with an asterisk. Ppro, *Pelobacter propionicus*; Gura, *Geobacter uraniireducens*; Geob, *Geobacter daltonii* FRC-32; GM21, *Geobacter* sp. (Strain M21); GM18, *Geobacter* sp. M18; KN400, *Geobacter sulfurreducens* KN400; GSU, *Geobacter sulfurreducens* PCA; Gmet, *Geobacter metalireducens*; Glov, *Geobacter lovleyi*; Pcar, *Pelobacter carbinolicus*. Alignments were performed using the MUSCLE alignment program with the standard settings in JalView (Waterhouse et al. 2009).

B. Proposed model for purine nucleotide recognition by PleD versus GSU1658.

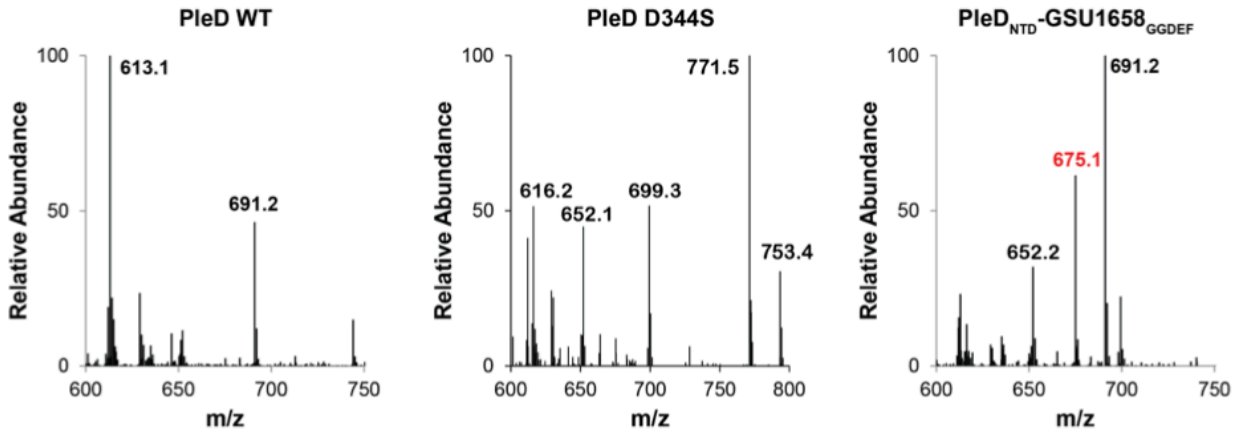


Figure 4.16 – HPLC-MS analysis of PleD and associated mutants. LC/MS analysis of *E. coli* cell extracts overexpressing PleD variants as shown. PleD$_{NTD}$-GSU1658$_{GGDEF}$ is a fusion between residues 1-293 of PleD and residues 297-458 of GSU1658. Shown are the MS spectra from integrating the retention time region containing all three CDNs (6 to 8 min). Expected masses are for c-di-GMP (m/z = 691) and cGAMP (m/z = 675).

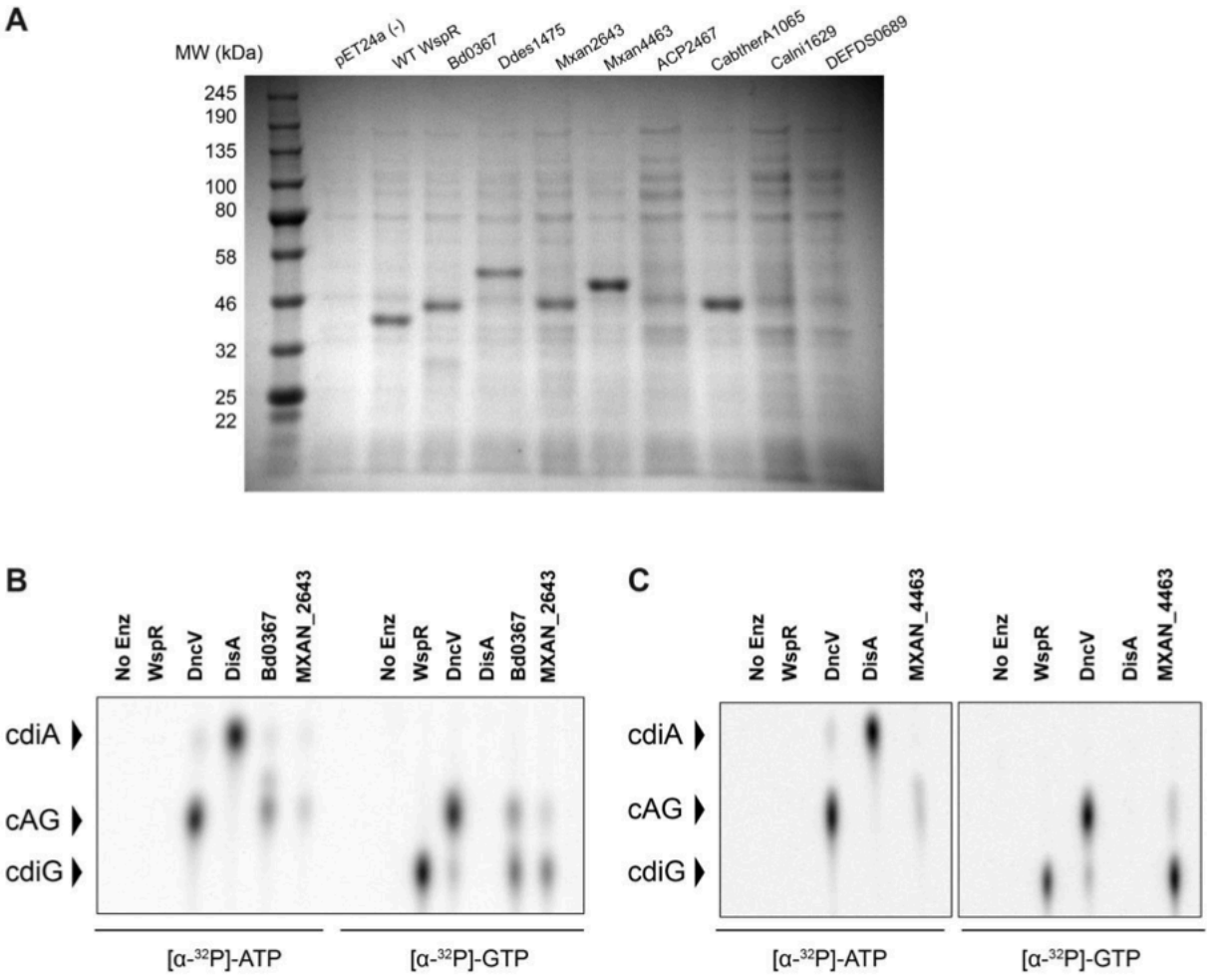


Figure 4.17 – Activity assay for additional Hypr GGDEF domains.

A. SDS-PAGE gel analysis of lysates from cells expressing C-terminal 6x-His constructs in pET24a, or (for ACP2467, Calni1629, and DEFDS0689) N-terminal 6x-His-MBP constructs. Gel was stained with GelCode Blue (Thermo Scientific).
B. Cellulose thin layer chromatography showing CDN region of radiolabeled products from enzymatic reactions of MBP-tagged I-site mutations of Bd0367 (R260A) or MXAN_2643 (R292A) with 1:1 ATP to GTP substrates in excess and doped with trace amounts of α-^{32}P-labeled GTP or ATP.
C. As (B), with wild-type C-terminal 6x-His tagged MXAN_4463.

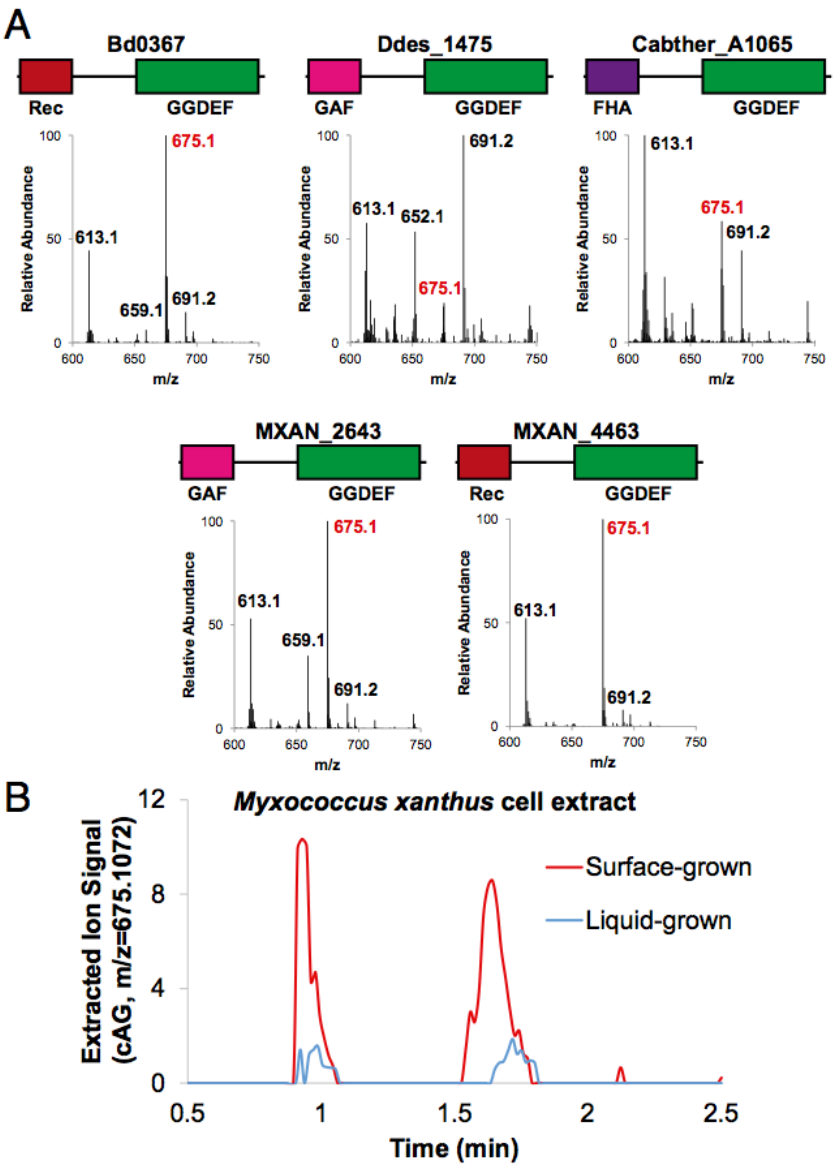


Figure 4.18 – Validation of Hypr activity in select *Deltaproteobacteria* and *Acidobacteria*.

A. LC/MS analysis of *E. coli* cell extracts overexpressing candidate Hypr enzymes; see Figure 4.11 for protein domain color scheme and Figure 4.16 for corresponding protein gel. Bd, *Bdellovibrio bacteriovorus*; Ddes, *Desulfovibrio desulfuricans*; Mxan, *Myxococcus xanthus*; Cabther, *Candidatus Chloracidobacterium thermophilum*. Shown are the MS spectra from integrating the retention time region containing all three CDNs. Expected masses are for c-di-GMP (m/z = 691), cGAMP (m/z = 675), and c-di-AMP (m/z = 659).

B. LC/MS analysis of *M. xanthus* cell extracts from surface- or liquid-grown samples. Shown is the extracted ion trace for cGAMP (m/z = 675.1072; ppm < 10 cutoff) normalized to the weight of extracted cells.

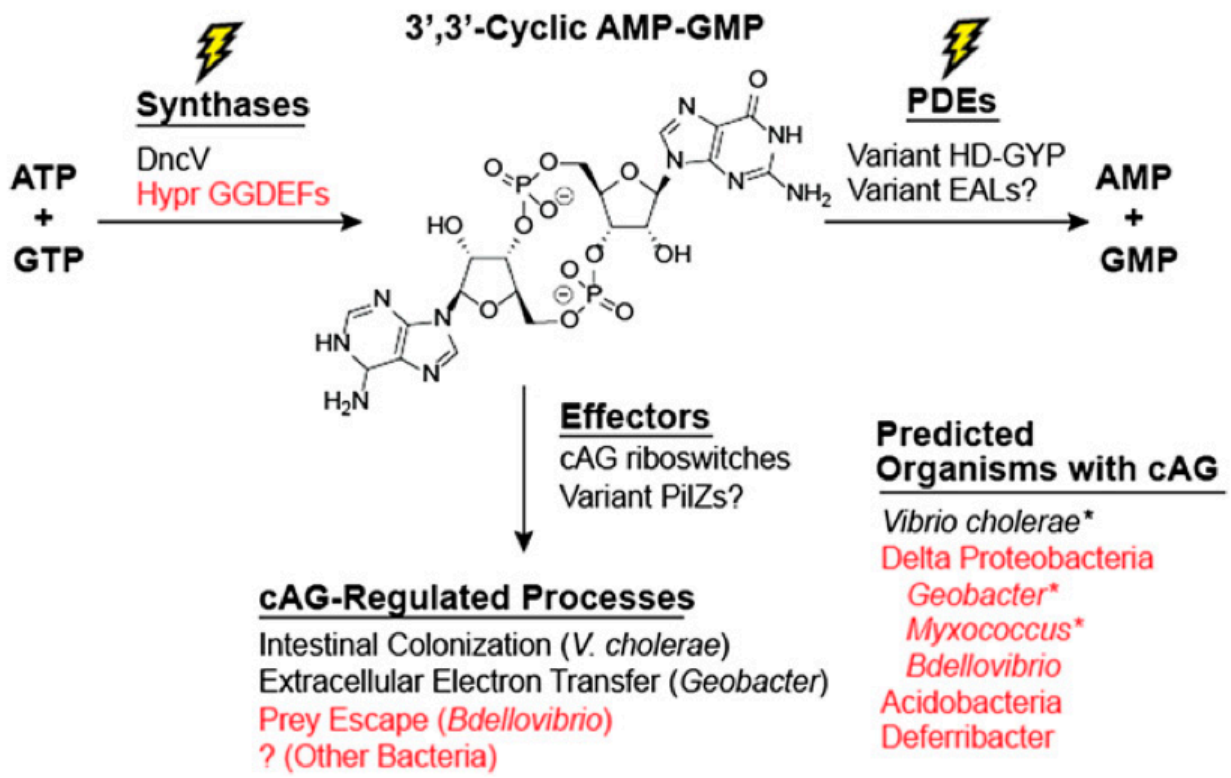


Figure 4.19 - Expanded understanding of the function and evolution of cGAMP signaling.

Levels of the second messenger cGAMP are regulated by the activity of cGAMP synthases and phosphodiesterases (PDEs) in response to primary environmental signals. Effectors that bind cGAMP then propagate downstream effects on bacterial physiology. In *V. cholerae*, DncV serves as the synthase and variant HD-GYP domains as the PDEs, but no effectors are known. We have shown that diverse *Deltaproteobacteria* and *Acidobacteria* contain Hypr GGDEF enzymes that can act as cGAMP synthases and control various processes through cGAMP riboswitches (*Geobacter*) and other unidentified effectors. An asterisk indicates organisms confirmed to have endogenous cGAMP. Text in red indicates new information put forth in this section.

TABLES

Table 4.1 – Crystallographic statistics of structures obtained.

Crystal	Gs1761 with cGAMP	Gs1761 with c-di-GMP	Vc2 G20A with cGAMP
Data collection	24-ID-C	24-ID-C	24-ID-C
Space group	$P2_1$	C2	$P2_1$
Cell dimensions			
a, *b*, *c* (Å)	66.5, 50.4, 78.4	174.8, 44.9, 68.2	51.7, 46.8, 80.8
a, *b*, *g* (º)	90, 91.7, 90	90, 103.3, 90	90, 90.4, 90
Wavelength	0.9792	0.9792	0.9792
Resolution (Å)	78.3-2.05 (2.16-2.05)	85.1-2.12 (2.24-2.12)	46.8-2.08 (2.14-2.08)
R_{merge}	0.051 (0.789)	0.044 (0.909)	0.059 (0.679)
I/sI	13.5 (1.6)	16.3 (1.3)	11.3 (0.6)
Completeness (%)	98.3 (99.0)	98.1 (97.9)	75.2 (42.2)
Redundancy	3.3 (3.2)	3.3 (3.4)	2.0 (1.2)
Unique reflections	32313 (4741)	28937 (4171)	17467 (763)
Refinement			
Resolution (Å)	42.5-2.0	85.1-2.12	43.7-2.08
No. reflections	29171	28916	17405
R_{work}/R_{free}	0.21/0.25	0.22/0.25	0.22/0.30
No. atoms			
RNA	3606	3605	1954
ligand	90	92	45
Cations	12	10	1
Water	196	48	11
B-factors			
RNA	34.0	69.5	53.8
ligand	26.6	52.2	64.0
Cations	52.7	77.6	74.2
Water	33.6	64.9	50.8
R.m.s deviations			
Bond lengths (Å)	0.013	0.027	0.012
Bond angles (º)	1.292	1.236	1.223

Values for the highest-resolution shell are in parentheses.

Table 4.2 – GGDEFs containing variant residues at the PleD 344 alignment position.

Genes listed are only those which possess a functional GGDEF motif, which we consider [G/A/S]G[D/E]EF.

UniProt KB Accession	Gene names	Organism	Selectivity Residue
A3IYP6	CY0110_05007	*Cyanothece sp. CCY0110*	Y
A1ALA3	Ppro_0492	*Pelobacter propionicus (strain DSM 2379)*	T
A5G6K7	Gura_3266	*Geobacter uraniireducens (strain Rf4) (Geobacter uraniumreducens)*	T
B5EC68	Gbem_3531	*Geobacter bemidjiensis (strain Bem / ATCC BAA-1014 / DSM 16622)*	T
C6E666	GM21_3597	*Geobacter sp. (strain M21)*	T
D3PC46	DEFDS_0689	*Deferribacter desulfuricans (strain DSM 14783 / JCM 11476 / NBRC 101012 / SSM1)*	T
E1WZI7	BMS_1301	*Halobacteriovorax marinus (strain ATCC BAA-682 / DSM 15412 / SJ) (Bacteriovorax marinus)*	T
E8WQD1	GM18_0558	*Geobacter sp. (strain M18)*	T
Q46S25	Reut_B4711	*Cupriavidus pinatubonensis (strain JMP 134 / LMG 1197) (Ralstonia eutropha (strain JMP 134))*	T
A1ANS6	Ppro_1380	*Pelobacter propionicus (strain DSM 2379)*	S
A5GF71	Gura_1886	*Geobacter uraniireducens (strain Rf4) (Geobacter uraniumreducens)*	S
A7HAD5	Anae109_1474	*Anaeromyxobacter sp. (strain Fw109-5)*	S
B3E1R0	Glov_1844	*Geobacter lovleyi (strain ATCC BAA-1151 / DSM 17278 / SZ)*	S
B3EB82	Glov_1760	*Geobacter lovleyi (strain ATCC BAA-1151 / DSM 17278 / SZ)*	S
B4UJZ3	AnaeK_1471	*Anaeromyxobacter sp. (strain K)*	S
B5E8T5	Gbem_3097	*Geobacter bemidjiensis (strain Bem / ATCC BAA-1014 / DSM 16622)*	S
B8EMQ6	Msil_3853	*Methylocella silvestris (strain BL2 / DSM 15510 / NCIMB 13906)*	S
B8J0V0	Ddes_1475	*Desulfovibrio desulfuricans (strain ATCC 27774 / DSM 6949)*	S
B8J555	A2cp1_1566	*Anaeromyxobacter dehalogenans (strain 2CP-1 / ATCC BAA-258)*	S
B9M0W8	Geob_2621	*Geobacter daltonii (strain DSM 22248 / JCM 15807 / FRC-32)*	S
C0QB51	HRM2_17460	*Desulfobacterium autotrophicum (strain ATCC 43914 / DSM 3382 / HRM2)*	S
C1F1G0	ACP_2467	*Acidobacterium capsulatum (strain ATCC 51196 / DSM 11244 / JCM 7670 / NBRC 15755 / NCIMB 13165 / 161)*	S
C6E353	GM21_1165	*Geobacter sp. (strain M21)*	S
C7QHC7	Caci_0111	*Catenulispora acidiphila (strain DSM 44928 / NRRL B-24433 / NBRC 102108 / JCM 14897)*	S
E3FYP4	STAUR_3377	*Stigmatella aurantiaca (strain DW4/3-1)*	S
E3T615		*uncultured bacterium 293*	S
E4TFG3	Calni_1629	*Calditerrivibrio nitroreducens (strain DSM 19672 / NBRC 101217 / Yu37-1)*	S
E6PWY9	CARN3_0369	*mine drainage metagenome*	S
E8RE90	Despr_2994	*Desulfobulbus propionicus (strain ATCC 33891 / DSM 2032 / 1pr3)*	S
E8V865	AciPR4_3292	*Terriglobus saanensis (strain ATCC BAA-1853 / DSM 23119 / SP1PR4)*	S
E8WHW1	GM18_3068	*Geobacter sp. (strain M18)*	S
E8WYM1	AciX9_0547	*Granulicella tundricola (strain ATCC BAA-1859 / DSM 23138 / MP5ACTX9)*	S

F2NID6	Desac_2520	*Desulfobacca acetoxidans (strain ATCC 700848 / DSM 11109 / ASRB2)*	S
F8CEK8	LILAB_20895	*Myxococcus fulvus (strain ATCC BAA-855 / HW-1)*	S
F8CQQ7	LILAB_30450	*Myxococcus fulvus (strain ATCC BAA-855 / HW-1)*	S
G2LH77	Cabther_A1065	*Chloracidobacterium thermophilum (strain B)*	S
H8MHV0	pleD2 COCOR_03316	*Corallococcus coralloides (strain ATCC 25202 / DSM 2259 / NBRC 100086 / M2) (Myxococcus coralloides)*	S
H8MXI3	cph2C COCOR_05401	*Corallococcus coralloides (strain ATCC 25202 / DSM 2259 / NBRC 100086 / M2) (Myxococcus coralloides)*	S
Q08TQ2	STIAU_4749	*Stigmatella aurantiaca (strain DW4/3-1)*	S
Q08YB4	STAUR_4818 STIAU_0908	*Stigmatella aurantiaca (strain DW4/3-1)*	S
Q1D3Y9	MXAN_4463	*Myxococcus xanthus (strain DK 1622)*	S
Q1D911	MXAN_2643	*Myxococcus xanthus (strain DK 1622)*	S
Q1IKE0	Acid345_3659	*Koribacter versatilis (strain Ellin345)*	S
Q1JVE0	Dace_0065	*Desulfuromonas acetoxidans DSM 684*	S
Q2IKI3	Adeh_2393	*Anaeromyxobacter dehalogenans (strain 2CP-C)*	S
Q39UD1	Gmet_1914	*Geobacter metallireducens (strain GS-15 / ATCC 53774 / DSM 7210)*	S
Q3A5R5	Pcar_1042	*Pelobacter carbinolicus (strain DSM 2380 / Gra Bd 1)*	S
Q5ZPC6		*Angiococcus disciformis*	S
Q6MQU2	pleD Bd0367	*Bdellovibrio bacteriovorus (strain ATCC 15356 / DSM 50701 / NCIB 9529 / HD100)*	S
Q74CL4	GSU1658	*Geobacter sulfurreducens (strain ATCC 51573 / DSM 12127 / PCA)*	S
A0AK16	lwe1930	*Listeria welshimeri serovar 6b (strain ATCC 35897 / DSM 20650 / SLCC5334)*	N
A0AK17	lwe1931	*Listeria welshimeri serovar 6b (strain ATCC 35897 / DSM 20650 / SLCC5334)*	N
A0Q1B8	NT01CX_2347	*Clostridium novyi (strain NT)*	N
A0YR26	L8106_11667	*Lyngbya sp. (strain PCC 8106) (Lyngbya aestuarii (strain CCY9616))*	N
A1S6Z8	Sama_1949	*Shewanella amazonensis (strain ATCC BAA-1098 / SB2B)*	N
A1SR97	Ping_0142	*Psychromonas ingrahamii (strain 37)*	N
A1SZB6	Ping_3143	*Psychromonas ingrahamii (strain 37)*	N
A1UF51	Mkms_2261	*Mycobacterium sp. (strain KMS)*	N
A3DC33	Cthe_0273	*Clostridium thermocellum (strain ATCC 27405 / DSM 1237 / NBRC 103400 / NCIMB 10682 / NRRL B-4536 / VPI 7372) (Ruminiclostridium thermocellum)*	N
A3IWY5	CY0110_23131	*Cyanothece sp. CCY0110*	N
A3YE50	MED121_21460	*Marinomonas sp. MED121*	N
A4B9U6	MED297_20957	*Reinekea blandensis MED297*	N
A4BH36	MED297_14975	*Reinekea blandensis MED297*	N
A4E847	COLAER_00587	*Collinsella aerofaciens ATCC 25986*	N
A4U2M2	MGR_1840	*Magnetospirillum gryphiswaldense*	N
A5VMQ1	Lreu_1888	*Lactobacillus reuteri (strain DSM 20016)*	N
A5ZSV1	RUMOBE_02079	*Blautia obeum ATCC 29174*	N
A6CN78	BSG1_01135	*Bacillus sp. SG-1*	N
A6TIC3	KPN_pKPN3p05967	*Klebsiella pneumoniae subsp. pneumoniae (strain ATCC 700721 / MGH 78578)*	N

A6VRU9	Mmwyl1_0236	*Marinomonas sp. (strain MWYL1)*	N
A8DJI0	YS_M60-F11.073	*Chloracidobacterium thermophilum*	N
A8DW58	v1g49651	*Nematostella vectensis (Starlet sea anemone)*	N
A8G1K2	Ssed_4373	*Shewanella sediminis (strain HAW-EB3)*	N
A8GYS2	Spea_0130	*Shewanella pealeana (strain ATCC 700345 / ANG-SQ1)*	N
A8GZL9	Spea_0428	*Shewanella pealeana (strain ATCC 700345 / ANG-SQ1)*	N
A8RMS1	CLOBOL_02006	*Clostridium bolteae (strain ATCC BAA-613 / WAL 16351)*	N
A8S3F1	CLOBOL_06594	*Clostridium bolteae (strain ATCC BAA-613 / WAL 16351)*	N
A8UBZ9	CAT7_10515	*Carnobacterium sp. AT7*	N
A8UU60	HG1285_16780	*Hydrogenivirga sp. 128-5-R1-1*	N
A9AXC8	Haur_4216	*Herpetosiphon aurantiacus (strain ATCC 23779 / DSM 785)*	N
A9D5G8	KT99_02136	*Shewanella benthica KT99*	N
B0C8G1	AM1_5154	*Acaryochloris marina (strain MBIC 11017)*	N
B0JT77	MAE_12990	*Microcystis aeruginosa (strain NIES-843)*	N
B0TJH3	Shal_2411	*Shewanella halifaxensis (strain HAW-EB4)*	N
B0TMZ0	Shal_4188	*Shewanella halifaxensis (strain HAW-EB4)*	N
B1BA96	CBC_A0861	*Clostridium botulinum C str. Eklund*	N
B1KNY3	Swoo_4765	*Shewanella woodyi (strain ATCC 51908 / MS32)*	N
B1LZ02	Mrad2831_2428	*Methylobacterium radiotolerans (strain ATCC 27329 / DSM 1819 / JCM 2831)*	N
B1MWP5	LCK_00134	*Leuconostoc citreum (strain KM20)*	N
B1WSQ7	cce_4288	*Cyanothece sp. (strain ATCC 51142)*	N
B1XL77	SYNPCC7002_A2587	*Synechococcus sp. (strain ATCC 27264 / PCC 7002 / PR-6) (Agmenellum quadruplicatum)*	N
B2A0U2	Nther_2407	*Natranaerobius thermophilus (strain ATCC BAA-1301 / DSM 18059 / JW/NM-WN-LF)*	N
B2A818	Nther_2224	*Natranaerobius thermophilus (strain ATCC BAA-1301 / DSM 18059 / JW/NM-WN-LF)*	N
B2J5G5	Npun_R3941	*Nostoc punctiforme (strain ATCC 29133 / PCC 73102)*	N
B3PET3	CJA_1657	*Cellvibrio japonicus (strain Ueda107) (Pseudomonas fluorescens subsp. cellulosa)*	N
B5JXV5	GP5015_1671	*gamma proteobacterium HTCC5015*	N
B5U200		*uncultured bacterium*	N
B6ARH3	CGL2_11390004	*Leptospirillum sp. Group II '5-way CG'*	N
B6BGA1	SMGD1_1002	*Sulfurimonas gotlandica (strain DSM 19862 / JCM 16533 / GD1)*	N
B6BIR1	SMGD1_1897	*Sulfurimonas gotlandica (strain DSM 19862 / JCM 16533 / GD1)*	N
B6WPU3	DESPIG_00058	*Desulfovibrio piger ATCC 29098*	N
B7KYX3	Mchl_0308	*Methylobacterium extorquens (strain CM4 / NCIMB 13688) (Methylobacterium chloromethanicum)*	N
B8CH18	swp_0168	*Shewanella piezotolerans (strain WP3 / JCM 13877)*	N
B8D025	Hore_21320	*Halothermothrix orenii (strain H 168 / OCM 544 / DSM 9562)*	N
C0BQ78	BBPC_0428 BIFPSEUDO_02518	*Bifidobacterium pseudocatenulatum DSM 20438 = JCM 1200 = LMG 10505*	N
C0BWF0	CLOHYLEM_04111	*[Clostridium] hylemonae DSM 15053*	N
C0D7P9	CLOSTASPAR_05296	*[Clostridium asparagiforme] DSM 15981*	N

C0QXE8	BHWA1_00306	*Brachyspira hyodysenteriae (strain ATCC 49526 / WA1)*	N
C0WEP4	ACDG_01935	*Acidaminococcus sp. D21*	N
C0YZG5	HMPREF0535_1180	*Lactobacillus reuteri MM2-3*	N
C2EFU2	HMPREF0545_0514	*Lactobacillus salivarius DSM 20555 = ATCC 11741*	N
C2EUE6	HMPREF0549_1082	*Lactobacillus vaginalis DSM 5837 = ATCC 49540*	N
C2JZI8	HMPREF0539_2323	*Lactobacillus rhamnosus LMS2-1*	N
C5EJH1	CBFG_00456	*Clostridiales bacterium 1_7_47FAA*	N
C6WGK8	Amir_0355	*Actinosynnema mirum (strain ATCC 29888 / DSM 43827 / NBRC 14064 / IMRU 3971)*	N
C6WVQ3	Mmol_1094	*Methylotenera mobilis (strain JLW8 / ATCC BAA-1282 / DSM 17540)*	N
C7XH20	HMPREF5045_00103	*Lactobacillus crispatus 125-2-CHN*	N
C9A618	ECBG_00198	*Enterococcus casseliflavus EC20*	N
C9AAA1	ECBG_01681	*Enterococcus casseliflavus EC20*	N
C9CHJ5	ECAG_00228	*Enterococcus casseliflavus EC10*	N
C9CIF0	ECAG_00128	*Enterococcus casseliflavus EC10*	N
C9CKC7	ECAG_01191	*Enterococcus casseliflavus EC10*	N
C9YL14	CDR20291_1266	*Peptoclostridium difficile (strain R20291) (Clostridium difficile)*	N
D0DEJ0	HMPREF0508_00079	*Lactobacillus crispatus MV-3A-US*	N
D1C7L1	Sthe_0405	*Sphaerobacter thermophilus (strain DSM 20745 / S 6022)*	N
D1CCW0	Tter_1719	*Thermobaculum terrenum (strain ATCC BAA-798 / YNP1)*	N
D2RKZ4	Acfer_1386	*Acidaminococcus fermentans (strain ATCC 25085 / DSM 20731 / VR4)*	N
D2RKZ5	Acfer_1387	*Acidaminococcus fermentans (strain ATCC 25085 / DSM 20731 / VR4)*	N
D4CHU0	CLOM621_09029	*Clostridium sp. M62/1*	N
D4IYS5	CIY_00120	*Butyrivibrio fibrisolvens 16/4*	N
D4LR02	CK5_18170	*Blautia obeum A2-162*	N
D4LW73	CK5_02480	*Blautia obeum A2-162*	N
D4MNW8	CL3_00230	*butyrate-producing bacterium SM4/1*	N
D5Q675	HMPREF0220_2407	*Peptoclostridium difficile NAP08*	N
D5S2D4	HMPREF0219_2715	*Peptoclostridium difficile NAP07*	N
D5U625	Bmur_0413	*Brachyspira murdochii (strain ATCC 51284 / DSM 12563 / 56-150) (Serpulina murdochii)*	N
D5XCW0	TherJR_2804	*Thermincola potens (strain JR)*	N
D6DM36	CLS_36110	*[Clostridium] cf. saccharolyticum K10*	N
D6XT12	Bsel_1436	*Bacillus selenitireducens (strain ATCC 700615 / DSM 15326 / MLS10)*	N
D6XZ65	Bsel_2859	*Bacillus selenitireducens (strain ATCC 700615 / DSM 15326 / MLS10)*	N
D7CN39	Slip_1361	*Syntrophothermus lipocalidus (strain DSM 12680 / TGB-C1)*	N
D8IA76	BP951000_2233	*Brachyspira pilosicoli (strain ATCC BAA-1826 / 95/1000)*	N
D9S411	FSU_2241	*Fibrobacter succinogenes (strain ATCC 19169 / S85)*	N
D9T0X8	Micau_1801	*Micromonospora aurantiaca (strain ATCC 27029 / DSM 43813 / JCM 10878 / NBRC 16125 / INA 9442)*	N
D9T2P9	Micau_6102	*Micromonospora aurantiaca (strain ATCC 27029 / DSM 43813 / JCM 10878 / NBRC 16125 / INA 9442)*	N
E0RAP7	PPE_01378	*Paenibacillus polymyxa (strain E681)*	N

E0S298	bpr_I1183	*Butyrivibrio proteoclasticus (strain ATCC 51982 / DSM 14932 / B316) (Clostridium proteoclasticum)*	N
E0SCN2	yddV Dda3937_02950	*Dickeya dadantii (strain 3937) (Erwinia chrysanthemi (strain 3937))*	N
E1IH26	OSCT_2627	*Oscillochloris trichoides DG-6*	N
E1QZ42	Olsu_0541	*Olsenella uli (strain ATCC 49627 / DSM 7084 / CIP 109912 / JCM 12494 / VPI D76D-27C) (Lactobacillus uli)*	N
E2SNJ1	HMPREF0983_02712	*Erysipelotrichaceae bacterium 3_1_53*	N
E3EFH1	PPSC2_07110	*Paenibacillus polymyxa (strain SC2) (Bacillus polymyxa)*	N
E3R434	LBKG_01059	*Lactobacillus crispatus CTV-05*	N
E3YRK3	NT05LM_2248	*Listeria marthii FSL S4-120*	N
E3YRK4	NT05LM_2249	*Listeria marthii FSL S4-120*	N
E3ZRW6	NT03LS_2247	*Listeria seeligeri FSL N1-067*	N
E3ZRW7	NT03LS_2248	*Listeria seeligeri FSL N1-067*	N
E4NFT5	KSE_45840	*Kitasatospora setae (strain ATCC 33774 / DSM 43861 / JCM 3304 / KCC A-0304 / NBRC 14216 / KM-6054) (Streptomyces setae)*	N
E5WJ43	HMPREF1013_02467	*Bacillus sp. 2_A_57_CT2*	N
E5Y1N9	HMPREF0179_00110	*Bilophila wadsworthia 3_1_6*	N
E6QNG4	CARN6_2294	*mine drainage metagenome*	N
E6TZP4	Bcell_3107	*Bacillus cellulosilyticus (strain ATCC 21833 / DSM 2522 / FERM P-1141 / JCM 9156 / N-4)*	N
E6U113	Bcell_2063	*Bacillus cellulosilyticus (strain ATCC 21833 / DSM 2522 / FERM P-1141 / JCM 9156 / N-4)*	N
E6VA14	Varpa_3212	*Variovorax paradoxus (strain EPS)*	N
E7GSI6	HMPREF9474_03881	*[Clostridium] symbiosum WAL-14163*	N
E8WZ00	AciX9_2892	*Granulicella tundricola (strain ATCC BAA-1859 / DSM 23138 / MP5ACTX9)*	N
E9USG8	NBCG_01687	*Nocardioidaceae bacterium Broad-1*	N
F0EFX9	HMPREF9087_0076	*Enterococcus casseliflavus ATCC 12755*	N
F0EJB5	HMPREF9087_1553	*Enterococcus casseliflavus ATCC 12755*	N
F0EPR5	HMPREF9087_3407	*Enterococcus casseliflavus ATCC 12755*	N
F0RXB1	SpiBuddy_0117	*Sphaerochaeta globosa (strain ATCC BAA-1886 / DSM 22777 / Buddy) (Spirochaeta sp. (strain Buddy))*	N
F0SZ73	Sgly_2922	*Syntrophobotulus glycolicus (strain DSM 8271 / FlGlyR)*	N
F2F2E2	SSIL_0971	*Solibacillus silvestris (strain StLB046) (Bacillus silvestris)*	N
F2M179	LAB52_07155	*Lactobacillus amylovorus (strain GRL 1118)*	N
F3LIZ8	IMCC1989_1671	*gamma proteobacterium IMCC1989*	N
F3MTT0	AAULR_17209	*Lactobacillus rhamnosus MTCC 5462*	N
F3S3A2	SXCC_00522	*Gluconacetobacter sp. SXCC-1*	N
F4AAH9	CbC4_1639	*Clostridium botulinum BKT015925*	N
F4BN82	yhcK CAR_c18940	*Carnobacterium sp. (strain 17-4)*	N
F4FH21	VAB18032_07575	*Verrucosispora maris (strain AB-18-032)*	N
F5JFM0	AGRO_3972	*Agrobacterium sp. ATCC 31749*	N
F5LAL3	CathTA2_2948	*Caldalkalibacillus thermarum TA2.A1*	N
F5LCV7	HMPREF9413_3556	*Paenibacillus sp. HGF7*	N
F6B3T5	Desca_2416	*Desulfotomaculum carboxydivorans (strain DSM 14880 / VKM B-2319 / CO-1-SRB)*	N

F6CN01	Desku_3271	*Desulfotomaculum kuznetsovii (strain DSM 6115 / VKM B-1805 / 17)*	N
F7QUH5	LSGJ_00968	*Lactobacillus salivarius GJ-24*	N
F7S1E1	A28LD_2356	*Idiomarina sp. A28L*	N
F7UXQ1	EGYY_28350	*Eggerthella sp. (strain YY7918)*	N
F8CDF6	LILAB_26855	*Myxococcus fulvus (strain ATCC BAA-855 / HW-1)*	N
F8FBD9	KNP414_03968	*Paenibacillus mucilaginosus (strain KNP414)*	N
F8I0J4	WKK_01730	*Weissella koreensis (strain KACC 15510)*	N
F8KDS1	LRATCC53608_0991	*Lactobacillus reuteri ATCC 53608*	N
F9DPS8	HMPREF9372_0808	*Sporosarcina newyorkensis 2681*	N
F9S3C5	VII00023_19474	*Vibrio ichthyoenteri ATCC 700023*	N
F9UBV0	ThimaDRAFT_2402	*Thiocapsa marina 5811*	N
G0EP50	Bint_1301	*Brachyspira intermedia (strain ATCC 51140 / PWS/A) (Serpulina intermedia)*	N
G0VQR3	MELS_1543	*Megasphaera elsdenii DSM 20460*	N
G1V667	HMPREF0178_03014	*Bilophila sp. 4_1_30*	N
G2DZG9	ThidrDRAFT_1432	*Thiorhodococcus drewsii AZ1*	N
G2LDT7	Cabther_A2208	*Chloracidobacterium thermophilum (strain B)*	N
G2LE99	Cabther_A0561	*Chloracidobacterium thermophilum (strain B)*	N
G2MXY0	Thewi_2388	*Thermoanaerobacter wiegelii Rt8.B1*	N
G2ZC65	LIV_1891	*Listeria ivanovii (strain ATCC BAA-678 / PAM 55)*	N
G2ZC66	LIV_1892	*Listeria ivanovii (strain ATCC BAA-678 / PAM 55)*	N
G3J2F6	Mettu_3032	*Methylobacter tundripaludum SV96*	N
G4L0E2	OBV_17720	*Oscillibacter valericigenes (strain DSM 18026 / NBRC 101213 / Sjm18-20)*	N
G4Q567	Acin_0817	*Acidaminococcus intestini (strain RyC-MR95)*	N
G5FI24	HMPREF1020_04120	*Clostridium sp. 7_3_54FAA*	N
G5HCX5	HMPREF9469_00437	*[Clostridium] citroniae WAL-17108*	N
G6B9S8	HMPREF1122_02608	*Peptoclostridium difficile 002-P50-2011*	N
G6BFS5	HMPREF1123_00856	*Peptoclostridium difficile 050-P50-2011*	N
G6FMV2	FJSC11DRAFT_0199	*Fischerella sp. JSC-11*	N
G6XN69	ATCR1_00425	*Agrobacterium tumefaciens CCNWGS0286*	N
G7M2F7	CDLVIII_2446	*Clostridium sp. DL-VIII*	N
G7RV34	PUUH_pUUH2392p0067	*Klebsiella pneumoniae*	N
G7VXU3	HPL003_15520	*Paenibacillus terrae (strain HPL-003)*	N
G8PE99	PECL_22	*Pediococcus claussenii (strain ATCC BAA-344 / DSM 14800 / JCM 18046 / KCTC 3811 / P06)*	N
G8QI44	Dsui_3113	*Azospira oryzae (strain ATCC BAA-33 / DSM 13638 / PS) (Dechlorosoma suillum)*	N
G8QIN9	Dsui_3167	*Azospira oryzae (strain ATCC BAA-33 / DSM 13638 / PS) (Dechlorosoma suillum)*	N
G8QR99	SpiGrapes_0926	*Sphaerochaeta pleomorpha (strain ATCC BAA-1885 / DSM 22778 / Grapes)*	N
G9X349	HMPREF9629_00806	*Peptostreptococcaceae bacterium ACC19a*	N
G9XC84	HMPREF9628_00292	*Peptostreptococcaceae bacterium CM5*	N

H1G8Q6	HMPREF0557_00376	*Listeria innocua ATCC 33091*	N
H1G8Q7	HMPREF0557_00377	*Listeria innocua ATCC 33091*	N
H1LGS0	HMPREF9104_01798	*Lactobacillus kisonensis F0435*	N
H1WNM7	LEUCOC10_01345	*Leuconostoc citreum LBAE C10*	N
H2J665	Marpi_0688	*Marinitoga piezophila (strain DSM 14283 / JCM 11233 / KA3)*	N
H5UVE1	MOPEL_132_00660	*Mobilicoccus pelagius NBRC 104925*	N
H6CGY3	WG8_1543	*Paenibacillus sp. Aloe-11*	N
H6NB78	PM3016_5343	*Paenibacillus mucilaginosus 3016*	N
H7F339	KKC_02784	*Listeria fleischmannii subsp. coloradonensis*	N
H7F340	KKC_02789	*Listeria fleischmannii subsp. coloradonensis*	N
H8FVX9	PHAMO_40078	*Phaeospirillum molischianum DSM 120*	N
H8N0H1	pleD3 COCOR_04267	*Corallococcus coralloides (strain ATCC 25202 / DSM 2259 / NBRC 100086 / M2) (Myxococcus coralloides)*	N
H9UHF0	Spiaf_0851	*Spirochaeta africana (strain ATCC 700263 / DSM 8902 / Z-7692)*	N
I0BPW9	B2K_27610	*Paenibacillus mucilaginosus K02*	N
I0IKW4	LFE_0186	*Leptospirillum ferrooxidans (strain C2-3)*	N
I0IQK6	LFE_1877	*Leptospirillum ferrooxidans (strain C2-3)*	N
I0JPE6	HBHAL_3671	*Halobacillus halophilus (strain ATCC 35676 / DSM 2266 / JCM 20832 / NBRC 102448/ NCIMB 2269) (Sporosarcina halophila)*	N
I0X6Y9	MSI_21000	*Treponema sp. JC4*	N
I0X855	MSI_16880	*Treponema sp. JC4*	N
I0XA21	MSI_9180	*Treponema sp. JC4*	N
I1B2E0	C357_01298	*Citreicella sp. 357*	N
Q03VX4	LEUM_1556	*Leuconostoc mesenteroides subsp. mesenteroides (strain ATCC 8293 / NCDO 523)*	N
Q04DU8	OEOE_1517	*Oenococcus oeni (strain ATCC BAA-331 / PSU-1)*	N
Q08T21	STAUR_4235 STIAU_6437	*Stigmatella aurantiaca (strain DW4/3-1)*	N
Q18BU0	CD630_14190	*Peptoclostridium difficile (strain 630) (Clostridium difficile)*	N
Q1D603	MXAN_3735	*Myxococcus xanthus (strain DK 1622)*	N
Q1JW04	Dace_0162	*Desulfuromonas acetoxidans DSM 684*	N
Q1WTG6	LSL_1024	*Lactobacillus salivarius (strain UCC118)*	N
Q221T2	Rfer_0469	*Rhodoferax ferrireducens (strain ATCC BAA-621 / DSM 15236 / T118) (Albidiferax ferrireducens)*	N
Q2B6K1	B14911_06261	*Bacillus sp. NRRL B-14911*	N
Q2RLY1	Moth_0223	*Moorella thermoacetica (strain ATCC 39073)*	N
Q2VZS3	amb4098	*Magnetospirillum magneticum (strain AMB-1 / ATCC 700264)*	N
Q3MFD3	Ava_0679	*Anabaena variabilis (strain ATCC 29413 / PCC 7937)*	N
Q5FJ86	LBA1413	*Lactobacillus acidophilus (strain ATCC 700396 / NCK56 / N2 / NCFM)*	N
Q6ALR6	DP1980	*Desulfotalea psychrophila (strain LSv54 / DSM 12343)*	N
Q7CZY1	Atu1119	*Agrobacterium fabrum (strain C58 / ATCC 33970) (Agrobacterium tumefaciens (strain C58))*	N
Q8Y5Z1	lmo1912	*Listeria monocytogenes serovar 1/2a (strain ATCC BAA-679 / EGD-e)*	N
Q8Y5Z2	lmo1911	*Listeria monocytogenes serovar 1/2a (strain ATCC BAA-679 / EGD-e)*	N

Q8YMN7	all4896	*Nostoc sp. (strain PCC 7120 / UTEX 2576)*	N
Q8YPG9	all4225	*Nostoc sp. (strain PCC 7120 / UTEX 2576)*	N
Q92A96	lin2026	*Listeria innocua serovar 6a (strain CLIP 11262)*	N
Q92A97	lin2025	*Listeria innocua serovar 6a (strain CLIP 11262)*	N
Q9K8H0	BH3036	*Bacillus halodurans (strain ATCC BAA-125 / DSM 18197 / FERM 7344 / JCM 9153 / C-125)*	N
A0YP32	L8106_11277	*Lyngbya sp. (strain PCC 8106) (Lyngbya aestuarii (strain CCY9616))*	E
A5CYJ5	PTH_2757	*Pelotomaculum thermopropionicum (strain DSM 13744 / JCM 10971 / SI)*	E
D8IYM5	Hsero_2714	*Herbaspirillum seropedicae (strain SmR1)*	E
Q2BMC3	MED92_12931	*Neptuniibacter caesariensis*	E

Table 4.3 – Sequences of riboswitch & riboswitch-Spinach constructs, and primers

Name	Sequence	Accession / Notes
Gm0970 (crystallography construct)	GGTATCGACAATACTAAACCATCCGCGAGGGTGGGACGGAAAGCCTAC AGGGTCTCTCTGAGACAGCCGGGATGCCAGAATATC	CP000148.1: 1079466-1079541
Gs1761 (crystallography construct)	GGTACACGACAATACTAAACCATCCGCGAGGATGGGGCGGAAAGCCTA AGGGTCTCCCTGAGACAGCCGGGCTGCCGAAATATC	AE017180.2 1922650-1922729; Red nt show U72C/C73U
Gm0970-Spinach	gacgcgactgaatgaaatggtgaaggacgggtccaATCGACAATACTAAACCATCCGCGAGGGTGGGACGGAAAGCCTACAGGGTCTCTCTGAGACAGCCGGGATGCCGAATttgttgagtagagtgtgagctccgtaactagtcgcgtc	CP000148.1: 1079466-1079541; Construct w/ red nt deleted used in Fig. S3 only
Gs1761-Spinach	gacgcgactgaatgaaatggtgaaggacgggtccaTACACGACAATACTAAACCATCCGCGAGGATGGGGCGGAAAGCCTAAGGGTCTCCCTGAGACAGCCGGGTCGCCGAAATAttgttgagtagagtgtgagctccgtaactagtcgcgtc	AE017180.2 1922650-1922729
Vc2-Spinach	gacgcgactgaatgaaatggtgaaggacgggtccaCACGCACAGGGCAAACCATTCGAAAGAGTGGGACGCAAAGCCTCCGGCCTAAACCAGAAGACATGGTAGGTAGCGGGGTTACCGATGttgttgagtagagtgtgagctccgtaactagtcgcgtc	CP007634.1 1329924-1330010
Cc9469	gacgcgactgaatgaaatggtgaaggacgggtccaTCGATCAGCAAAACTAGCGAAAGCTAGTGACGCAAAGCTACAGGGATTTCCCCTTTTAACAGGGATGTCAGCCAGCTGCAGGttgttgagtagagtgtgagctccgtaactagtcgcgtc	ADLJ01000004.1 176358-176439
Bc9140	gacgcgactgaatgaaatggtgaaggacgggtccaAATATTTTTAGCACACTATTCGAAAGGATAGGCCGCAAAGCTTAGAGTCTACGGTAATATATTGGTTACTAAGATCGTCTGGTTGCACATTttgttgagtagagtgtgagctccgtaactagtcgcgtc	ACMP01000037.1 9235-9325
Ck2324	gacgcgactgaatgaaatggtgaaggacgggtccaTTGATAATAGCACACTTATCGAAAGGTAGGGTCGCAAAGCTATGGGTCTTAAGAAAATTATTTTTCTATGATTGCCAGGTTGCCAAttgttgagtagagtgtgagctccgtaactagtcgcgtc	CP000673.1 2377707-2377792
Gs2885b	gacgcgactgaatgaaatggtgaaggacgggtccaCGATAAGACTAAACCGTCCGCGAGGGCGGGGCGGAAAGCCTAGGGTCTCCTAGAGACAGCCGGGATGCCGttgttgagtagagtgtgagctccgtaactagtcgcgtc	AE017180.1 3168890-3168959
Gu2327	gacgcgactgaatgaaatggtgaaggacgggtccaCGAAAATACTAAACCATTCGCGAGAATGGGACGGAAAGCCTAAAGGGTCTCACCGAGACAGCCGGGTCGCCGttgttgagtagagtgtgagctccgtaactagtcgcgtc	CP000698.1 2691964-2692035
Gm2037	gacgcgactgaatgaaatggtgaaggacgggtccaCGACAATACTCAACCATCCGTGAGGATGGGGCGGAAAGCCTATTGGGTCTCACCGAGACAGCCGGGTTGCCGttgttgagtagagtgtgagctccgtaactagtcgcgtc	CP000148.1 2280695-2280766
Gm0232	gacgcgactgaatgaaatggtgaaggacgggtccaCGACAATACTAAACCATCCGCGAGGATGGGGCGGAAAGCCCATAGGGTCTCACCGAGACAGCCGttgttgagtagagtgtgagctccgtaactagtcgcgtcGGTTGCCG	CP000148.1 264999-265062
Pp0574a-Spinach	gacgcgactgaatgaaatggtgaaggacgggtccaCACGATAATACTCAACCATCCGCGAGGATGGGGCGGAAAGCCTACAGGGTCTCACCGAGACAGCCGGGTTGCCGAAATGttgttgagtagagtgtgagctccgtaactagtcgcgtc	CP000482.1: 609390-609468
Pp2849-	gacgcgactgaatgaaatggtgaaggacgggtccaTAGACGACAATACTAAACCATTCGCGAGAATGGGACGGAAAGCCTACAGGGTCTCAC	CP000482.1:

Spinach	CGAGACAGCCGGGTCGCCGAAATAttgttgagtagagtgtgagctc ccgtaactagtcgcgtc	3117926-3118006
Pp0574b-Spinach	gacgcgactgaatgaaatggtgaaggacgggtccaTAGACGATACT ACTTAACCATTCGCAAGAATGGGGCGGAAAGCCTAAGGGTCTTACT GAGACAGCCGGGTTGCCGAAATAttgttgagtagagtgtgagctcc gtaactagtcgcgtc	CP000482.1: 609610-609689
Pp2572-Spinach	gacgcgactgaatgaaatggtgaaggacgggtccaATCGATACTAC TAAACCATCCGCGAGGATGGGACGGAAAGCCCACAGGGTCT CCAGAAGACAGCCGGGTCGCCGAAATttgttgagtagagtgtgagc tccgtaactagtcgcgtc	CP000482.1: 2787032-2787109
Primer F	ccaagtaatacgactcactataGACGCGACTGAATGAAATGGTGAA GG	Extended T7 promoter (lowercase)
Primer R	GACGCGACTAGTTACGGAGCTCACAC	

Table 4.4 – Sequences of mutant riboswitch-Spinach constructs, and Quikchange primers

#	Sequence	Mutations List	P2a	Source Plasmid	QuikChange Primers (if applicable)
1	Gm0970	(none)	U.G C-G A	IDT	
2	Gm0970	C15G G40C	U.G G-C A	IDT	
3	Gm0970	U16C	C-G C-G A	IDT	
4	Gm0970	A14G	U.G C-G G	IDT	
5	Gm0970	A14G C15G G40C	U.G G-C G	IDT	
6	Gm0970	A14G U16C	C-G C-G G	IDT	
7	Gm0970	A14G C15G U16C G40C	C-G G-C G	IDT	
8	Gm0970	G39A	U-A C-G A	IDT	
9	Gm0970	U16G G40U	G.U C-G A	IDT	
10	Gm0970	G39U	U-U C-G A	1	cctgtaggctttcagtcccaccctcgc gcgagggtgggactgaaagcctacagg
11	Gm0970	C15U U16C	C-G U.G A	1	ccctcgcggatggtttcgtattgtcgattggacc ggtccaatcgacaatacgaaaccatccgcgaggg
12	Gs1761	(none)	U.G C-G A	IDT	
13	Gs1761	C15G G40C	U.G G-C A	12	cgggtccatacacgacaatagtaaaccatccgcg cgcggatggtttactattgtcgtgtatggacccg acccttaggctttgcgccccatcctcg cgaggatggggcgcaaagcctaagggt
14	Gs1761	U16C	C-G C-G A	12	tcctcgcggatggtttggtattgtcgtgtatgga tccatacacgacaataccaaaccatccgcgagga
15	Gs1761	A14G	U.G C-G G	12	cctcgcggatggtttagcattgtcgtgtatggacc ggtccatacacgacaatgctaaaccatccgcgagg
16	Gs1761	A14G C15G G40C	U.G G-C G	12	gggtccatacacgacaatggtaaaccatccgcgaggat atcctcgcggatggtttaccattgtcgtgtatggaccc acccttaggctttgcgccccatcctcg cgaggatggggcgcaaagcctaagggt

17	Gs1761	A14G U16C	C-G C-G G	12	tcgcggatggtttggcattgtcgtgtatggacccgtcc ggacgggtccatacacgacaatgccaaaccatccgcga
18	Gs1761	A14G C15G U16C G40C	C-G G-C G	12	ctcgcggatggtttgccattgtcgtgtatggacccgtcc tt aaggacgggtccatacacgacaatggcaaaccatccgcg ag acccttaggctttgcgccccatcctcg cgaggatggggcgcaaagcctaagggt
19	Gs1761	U72C C73U	U.G C-G A	12	ctcaacaatatttcggcagcccggctgtctcaggga tccctgagacagccgggctgccgaaatattgttgag
20	Gs1761	C15G G40C U72C C73U	U.G G-C A	13	ctcaacaatatttcggcagcccggctgtctcaggga tccctgagacagccgggctgccgaaatattgttgag
21	Gs1761	U16C U72C C73U	C-G G-C A	14	ctcaacaatatttcggcagcccggctgtctcaggga tccctgagacagccgggctgccgaaatattgttgag
22	Gs1761	A14G U72C C73U	U.G C-G G	15	ctcaacaatatttcggcagcccggctgtctcaggga tccctgagacagccgggctgccgaaatattgttgag
23	Gs1761	A14G C15G G40C U72C C73U	U.G G-C G	16	ctcaacaatatttcggcagcccggctgtctcaggga tccctgagacagccgggctgccgaaatattgttgag
24	Gs1761	A14G U16C U72C C73U	C-G C-G G	17	ctcaacaatatttcggcagcccggctgtctcaggga tccctgagacagccgggctgccgaaatattgttgag
25	Gs1761	A14G C15G U16C G40C U72C C73U	C-G G-C G	18	ctcaacaatatttcggcagcccggctgtctcaggga tccctgagacagccgggctgccgaaatattgttgag
26	Vc2	(none)	C-G G-C G	IDT	
27	Vc2	G20A	C-G G-C A	IDT	
28	Vc2	C22U	U.G G-C G	IDT	
29	Vc2	G20A C22U	U.G G-C A	IDT	
30	Vc2	G20A G21C C46G	C-G C-G A	IDT	
31	Vc2	G20A G21C C22U C46G	U.G C-G A	IDT	

Table 4.5 – List of genes tested for Hypr activity. Genes codon-optimized for *Escherichia coli* K12 strains have an asterisk next to the gene. All codon-optimized genes were ordered from IDT.

Gene	UniProt ID	Nucleotide Sequence (5'→3')
GSU1658	Q74CL4	ATGGAACGGATTCTCGTTGTCGAAGATGACCGTTTTTTTCGTCAGATGTATGTTGATCTCCTGAAA GAGGAGGGATACGAGGTCGATACCGTGGCATCGGGCACCGAGGGGTTGAAGCGGCTTGAGAAGCAA GAATACCACCTCGTCATTACCGACCTGGTCATGCCCGGAATGAGCGGTATCGAGGTGTTGTCCCGC GTCAAGCAGAAAGCTCCGAACGTCGATGTCATCCTCGTCACCGGTCACGCCAACCTCGAATCGGCC GTCTATGCCCTCAAGAATGGTGCCCGCGATTATATTCTCAAACCGTTCAACCATGATGAATTCAAG CACACCGTGGCACTTTGCTTTGAGCAGCGGAGGCTTATCAACGAAAACTACGAGCTCAAGGAGCTG CTGAATCTTTTTCAAGTTGGGCAGAACATAGCCAACTGTATCGACTTGGAACGGCTCTCTGCGGTT GTGGTCGATGCTTTCTGCAAGGAGGTCGGAGTTTCACGCGCTATCGGCCTCTTTCCCGAAAAGAGC GAACCCCACGCCCTCAAGGAGCTGAGGGGGCTTGAGCCTGAAGTTGCAGCCGCTCTTGCCGAAAAA GCTCTTACCCTTTGCAGTGACGCCGCGGAGACGGCAGGGGGCTTTCGACGGCTCGACGGTTCCCAT TTTTCCGATGGTCTCCTGCGAACTGCGGGGATTAATGGCGCCCTTGTGGTTAGCATCCGCCAGCGT ACGCTCCTGCAGGGAGTGCTTCTGCTGGTCAATGACCAGGGCAAGCCGTTCCCTGCCGTGTTCAAA CATAAAAGCATCCAGTTTTTGCTGGAGCAGGCATCGCTTGCCTTCGACAACGCCCTGCGTTACTCC AGCGCCCGCGACATGCTCTATGTTGACGAACTCACGGGACTCTTCAACTACCGTTACCTTGACATC TCGCTGGACCGGGAGTTGAAGCGGGCTGACCGATTCGGCTCGGTAGTTTCCATGATCTTCATCGAC ATGGACCACTTCAAGGGAGTCAACGACACCCACGGCCATCTTTTTGGGAGCCAGGTCCTCCATGAA GTAGGTCAATTGCTCAAGAAGTCGGTCCGTGAGGTCGATGTAATCATTCGCTACGGTGGCGACGAG TTCACCATAATTCTGGTGGAAACCGGTGAAAAGGGCGCTGCAACCGTGGCTGAAAGGATTCGTCGC TCCATCGAGGACCACCACTTTCTGGCCTCTGAAGGGCTCGATGTCCGGCTCACCGCAAGTCTCGGC TACGCCTGTTATCCCCTTGACACCCAGTCCAAAATGGAACTTCTCGAACTGGCGGACAAAGCCATG TATAGGGGCAAGGAAGAGGGCAAAAACCGTGTATTCCGGGCAACGGCAATCCGTTGA
Mxan2643	Q1D911	ATGAATCCCGCGGACCTCCTGTCGGCCATGAAGCGGACAGTGGAGCAGTTGGCCGCCTTCAATGAG ATGGCGAAGGCCCTGACGTCCACGCTCGAGCTCCGCGAGGTGCTGGCGCTGGTGATGCAGAAGGTC AGCAGCCTGCTGCTGCCTCGCAACTGGTCGCTCATCCTCCAGGACGAGCGCACCGGAAAGCTCTAC TTCGAAATCGCGGTGGGTGACGGCGCGGACGTGCTCAAGGGCCTCCAGCTCAACCCGGGCGAGGGC ATTGCCGGCGCCGTCTTCACGTCCGGCGCGGCGCGGCTCGTCCATGACGTGGGTGGGGACCCCAGC TTCTCGCCACGCTTCGATGAAGCCTCCGCCTTCCACACCCGCTCCATCCTCGCGGTGCCGCTGCTG GCCCGGGGCCGGGTCCTGGGCATCATCGAACTGGTGAACGGGCCCATGGACCCCCCCTTCACCAAC GAGGACCTCACCATTCTCACCGCCATCGCGGACTACGCGGCCATCGCGATTGAGAACGCGCGCAAC TTCCGGCGGGTGCAGGAGTTGACGATTACGGACGAGCACACCGGCTGCTACAACGCCCGGCACCTG CGCGCCTTGCTGGACCAGGAGGTGAAGCGCTCGGAGCGCTTCAGCCACCCGCTGTCGCTCGTCTTC CTGGACCTGGACCACTTCAAGAGCATCAACGACACCCATGGGCACCTGGTGGGTAGCGCCACCTTG AAGGAAGTGGGGGACCTGCTGATGACCCTGGGCCGGCAGAACCTGGACGCCGTCTTCCGCTACGGC GGCGACGAGTTCGCCATGTTGCTGGTGGAGACGGACCCGGAGGGCGCGGCCGTCATCGGCCAGCGC GTCTGCGAGGCCTTTCGGGGGCGGGGCTTCCTCCTGGAGCAGGGCCTGGACGTGCGCCTCACCGCC AGCGTGGGCGTGGCCACCTACCCGGACCATGCCTCGTCCGCGCTGGACCTCATCCGCGCGGCGGAC TTCGCCATGTACGCGGCCAAGGCCCGGGGCCGGGACGCGCTCTGCATCGCCGAGCCCATTGCTCCG AACGGCGGCACAGGCTCCCACGAGTTCCCGGAGCGGTAG
Mxan4463	Q1D3Y9	ATGGCGCGAATCCTCCTCGTCGACGACGAAAAGATCGCCCGCACCCTGTACGGCGACTACCTCACC GCCGTGGGACACGCCGTCACGGCGGTGGGCACGCTACAAGAGGCAAAGGAAGCACTCGCAGGCGAC CGTTTCGACGCGGTGGTGACGGACCTCATCCTCCCCGGTGGTGACGGCATGGAGGTCCTGCGGCAC GTGCGGGAACATCACCCGGGCGTGGAGGTGGTGGTCATCACTGGCCTGGAGAAGGTGGACCCCGCC GTGCGCGCCATCAAGAGCGGCGCCGCGGAGTACCTCGTCAAGCCGGTGGCCCCGGAGGCCCTGCAG CACGCCGTGCGCCGAGCGCTCACCACGCGCGACCTGATGCAGGAGAACGCGTCGCTGCGCCGCCAT GTGGCCATGTTGGAGGCGGGGCAACGCATCGCCACCACCCTGGACCGCGAGAAGCTGGCCTCGGCC ACCGCCAGCGCGCTGCAGAGCATGGCCTCCGCCAGCGCCGTGGTCCTGCTGGAGCGCGACTCTGCC TTCGCGCTGCGGCGCCACGGCACCAGCGGCCTGTCCACCGCGCTGGAAGAGCCGCTCATCGCCGAG CTCATCGAACGCCTGACGAACGAACGCGGTCCGCGCGAGCTGGACGGCATGGACGCGCCCTTTCCT CGCGCAATCTCCTTCCCCGCGCTGGAGGGTGACGCCGTGCTGGGACACGCGGTGCTCTTCTTCGGC GGCACGGGCGCGGAGTGGGCGGGCGAGACGGCCAGCTTCCTGGTTCGCAACTGGGCGCTCGCGCTG CGCAACCTCGGCCGCTTCGCCGCGGTGGAGGACCTGGCGTACGTCGACGACCTCACGCGCCTGTTC AACACCCGCTACCTGCACCTGGTGGTGGACCGCGAGGTCCAGGACGCGCTCCAGTCACAGCGCACC TTCAGCCTGCTGTTCCTGGACCTGGACCACTTCAAGTCCATCAACGATACCCATGGCCACCTCGTG GGCTCCAAGGTGCTGGTGGAGGCGGCGCGCGTGGTGAAGGGCTGCGTGAGAGACCACGACGTCGTC GCGCGCTACGGCGGAGACGAATACGTGGTGGTGCTGCGCAACACCGACTCCGGCGGCGCGCTCAAG GTGGCCGAGCGCATCCGACGCACCATGGAGACGCACAACTTCCTGGCGCGCGAAGGCCTGTCGCTC AAGCTCACCACGTGTATCGGCGTGGCCAGCTTCCCCGAGCACGCCCAGGACAAGGCCACGCTGTTG GACCTGTCGGACCGGGCCATGTACCGCGGCAAGCGGGGCTCGCGGAACGTCGTCTACATGGCGGCG

		AAGGACCTGGAGGCCCCACCGGCCGAGCGCCGGCAGGCCCACTCCGCGTCCTGA
Ddes1475	B9J0V0	ATGCTGAACAAGTCAAGCATCATACCAGAACATATACAGCTCGAATCACACGACCCTGTCTGGGAG TGGCACACGACATCCGACAGGCTTTTCATGAGCGTAGGCGCCCTTGCCCAGCTACGCATGGACGGC AAACCGCCGCGCAGCATGAAAGATTATCTGGAGCACTGCCCCCTCGAAAGCCTGGCTCCCCTTCTT GAATCTATGGAAAAAGCGCTCAACGGCTCCCACGGGCCGCACCTTGAAGTGTTTTATCCTTTTGAC AGTTTTCTGGTACGGTCTCAGATACTGGTCTTGCGGCGCGACGTTTTCGGTCGCGGAACCCTGGTA ACAGGCTGCAACGTGGCTATGGACAGACAAAGGCTTGCACCCACTGCTGCCGCCGCCCCCGTGCCG GCACCCCAGCCCCCGCCCCGAAGCCTGGCCGAAGCCGCCGTTCCTTCCACGGCCCGCAGCGACGCC AGCCGCCTCATGCTGGCCCTCAACGCCGCCAGCGATGGCCTGTGGGACTGGGACCCCAGCACCAAT GCCATTTATTTCAGTCCCCGCTACCTCGACATGCTTGGCTACACCAGCGAAGAATTCCCCCCCCTG TCCACATCATGGACCAGCAAGGTACACCCCGACGATTACGACAACATCGTTCCCATGCAGATTGAA TTCATCAACAACCCCAAAATGGGCGACAGCTTTGAATGCACCTACCGGATGCAGCGCGGCGACGGC ACCTGGGCATGGATTCTCAGCCGGGGCTATGTGACTCACCGCGACGCAAGCGGCAAGGCCATCCGC GTTGTGGGCCTGCACACAGACGTCAGCGCGAGCCAGGGCGACAGGGCACGGCTTGAAGAGCTGGTG CGTAACGACGCCCTTACCGGGCTGCGCAGCCGCACCTATTATGGAATGACTGTCGACAAGCTGGAA CAGCAGCAGATGCGGCCCGTCAGCATCATCATCGCCGACATGGACGGACTCAAAATGGTCAACGAC CATGTAGGCCATACCGAAGGCAGCGAAATGCTCTGCCAGGCAGCCATCATACTGCGCGGCAGTCTC AATGCCACCGACTGTATCGCCCGCATGGGCGGCGATGAATTTGCCGCCATTGTGCCGGGTTGCGCC AAGGAAGACCTTGAGGCGCTCATCCAGCGGGTCAGAGACGCTTTTGATGCCTATAATGCCGACCCG GACCATGTGCCGACACACATGTCTGTGGGCGGAGCATGCGCTGACGACATGAACACCACCCTGGCC CAGGCCCTGTCGGAGGCGGATCGCAACATGCTGGCCGTCAAGCACGAAAGCAGCCCAAAGTGGCGC CTGCGCATAAAAAACTGGATAGAAAACCGGACCGGCAAAACAATTCAGCTTGAAGACAGCCGCTAC AGGATGTCCCCCACGCACGACGACTCTTGA
Bd0367 WT *	Q6MQU2	ATGTCGCGCGCCGAAGTGACGCTCGTATGTAAAATGAGCTTTGAAGTATCGCCGAAGCAACCAAAG AGCCGCCGTATCCTGGTTATCGACGACGATAAGGACTCATTAGAAATTTTATTGGAACCCCTGCGC TGGGAAGGTTATGACGCGCGTGGCGTGACTACCGAAGCGGAGGCGCATAAATTAATCGAGTCATGG ATTCCGCATATCGTGATCCTGGATTGGATGGCCCCGTCAATGGCCGGCCTGCGCGTTCTGAAATCC GTACGCGAACGCCTGAGTCATGTCTCGTGTGTCTTTGTATCGGAAAATTCTTCCACAGAGGCTATT ATTGAGGCTTTGGATTCGGGCGCCGACGATTATATTGTAAAGCCATTCGTGCCATTAGAGTTGTTA GCACGCATCCGCTCTCAACTGCGCATCCGCGATCTGCACGAGCAGCTGCTGTTTGCCAACGAAAAA TTAAAGGAACTGGTTGATACCGACGATTTAACCGGTTTATATAATATGCGTAGCTTATACCAGCGT CTGGATTTTGAAATGGAACGTGGCCGCCGCTTCCACCGCGACGTGTGCGTGGTCATGATGGACATG GACTATTTCAAAACCGTGAATGATGGACACGACCACTTATTCGGGAGTTATGTGCTGAGCGAAGTT GGTAAAATCATTCGCGCCAACACTCGTAACATCGATATCCCGGCACGTTATGGGGGGGATGAGTTT CTGATGGTCCTGACCGAAACTAATCATGCGGGCGCTATGTATTTTTGCGAGCGCCTGCGCGAAAAT ATTGAAAAAACAACCTTTCGTAACGGCGAGGACAGCATGAAATTGACAGCCTCACTGGGCTTTGCG ATCACCATCCCCGGCGAAAACATCAGCGCGCGTGAACTGGTTCGCCGCGCCGACCACGCTCTGTAT CAGGCAAAACGCGCCGGGCGCAACCAGGTGGCGCATTACAAACCGGAGAGCGCGCCCGTAGTTGAG ATCAAGTCGGCAGTGCACAAACGCCGTAAAGCCGCCGGTTAA
CabtherA_1065 WT *	G2LH77	ATGAACCTTAAACTGGGCGCCATCTTACGTCCGGTTAATAGCCTCAACCAAACACAGAAACTGCAA GCCAACCCACTTGCCCGCCCGCGTCAAGCGCGTCCGGCACTGGTGCACATGCGTGGAGATTATCTG GGCTCAAGCTTTCGTATTGAACATGCCATTACGCGCATTGGACGCGGATCAGACGCAGAGTTACGT TTAGAAAATGATGACGAAGCAAGTCGCTTACACGCCCGTATTGAGCGCCTGGAAACACCTACAGGG CATTTCCAATATTGGTTGACCGATCTGCGCTCTACCAACGGGACCCAACTGAATGGTATTCCGCTG GTGCCGGGCGAGGCAGTGTTGCTGCATGATGGCGATAAATTTAGTATCGGCCGTCATATCCTCAAG TTCACTTTTTTAGACGATATTGATGAGGAGTTTCATCGTCGTATCACCGAACTCATCACTCATGAT GACTTAACCGGTCTGCTGAACCGCAAATCGTTCATCCTGGAAATGCAGCGTGAGATGGCCCGTAGT AACCGCTACGGTCACCCATTTGGCCTGCTGATGATGGATATTGATCATTTTAAGCGTGTCAATGAT ACCTATGGTCACCTGGTTGGTTCTCAGGTATTACGCGAGGTGGCTACCGTTATCCGCGAAACACTG CGTGACTCTGACATTGCAGGTCGTTATGGTGGAGAAGAATATATTGCCCTCTTACCAGAAACCGAT CGCCTGCGCGCACACGAAGCGGCCGAGCGCATTCGTCAAGCAATCGAACGCACCCCGTTCACAGCA AGCCTCAACGTGCCGCACCACAAGTTACGTCTGACCATTAGTATTGGGATCGCGAGTTATCCGGGG GACGCAGCCCAAATTAATGATCTGATCCAGCGCGCGGATGAAGCGATGTATGAAGCAAAACGCCGC GGTCGTAATCTGGTGCAGACGACGGGCCAATCGGCGGCCAATCGCGCGACCCCCCCTTCACTTCCA TTGCCGCCGCCGTCTGGAGATGACAGTCCCACGGAGCATCTGACCGTGGAGCAGCCGCAGCCTGTC AAACCTTAG
DEFDS_0689 R248A *	D3PC46	ATGTATGAAAGCCTGAAACGCAACATCTTCGTCATTCTGACAAGCATTCTCCTTATTTATGAAACC TATAACAAAACAAATGAGAATTTGCTGCTGCTGACTTCTTTACTGCTCACGTGCTATATTGCTGCA ACGTTGATCAAAAAGGTTGAACTGGATGAAGTACTGTTTGCTCTGTTCGTGATTTTAATCGGTTAT CTGAGCATCGCAAACCGTGAGTTCATTTATTTTCAAATTCTCGCGATTACATTTTTGGTATTTGAT TCGAAGTTCTACGTGATTAAAGTGATCCTGGCGATTCTGTTGATTCTCTTCGATTTATTCTACTTG AACATTTCGATCCTTTCCACGTTTAGCCTGATGATTCTCTATTCCTTATTCTTTTCTATCTTCATC AAATTGCTGATTGATCGTTTGGAAGAAGAAATCGACGAACTGTCCATTACGGACGACCTCACGGGT CTGCTGAACCAAAAAGGATTCCTGAAAAAGTTTGAGGAAGAATATTATCGTAGCGTTCGCTACAAG

		AAAAATTTTACCGTTATCATGTTGGATAGCGATGATCTGAAGAAAGTTAATGACACTTATGGGCAC AAATACGGGACCAAAGTTATTCTGTTCATCGCGGATGAAATTAAGAAGAACATTCGCCGTACCGAC TTTGCTTGCCGCTACGGCGGTGACGAGTTTATGATCTGTCTGGTTGAAACACCTATCAACAACGGC AAAATTTTCGCGGAAGCGCTGAAAAACAACATTGCAATGAAACCGGTATTTACCGATAAAGGCCGT GGTTTCAATGTGACAGTGTCGGTGGGGGTTGTTGGTTATCCGCACACAAGCGAAAAGTCGTTCGAG CTGCTTGATCTGGTTGACAAAGCGCTGTACGAAGCAAAAAACAAAGGCAAAAATCGCGTTGAGATC CTGACCAAAAATTCTTCCTTA
Calni_1629 R268A *	E4TFG3	ATGATTGATAACAAGATTAAACACTTTCAGTATAAGATTGCAGAAGTCTTATATCTCTTTGCGTTT ACTATCATCATCGCCTCCATTAAATTTCTTGATTCTACCAGTAACAAAGCAAACTACGCGATCTTA GTGTTTTTCCTGATCTTCATCATCCTGAAATTTAGCATTGACGATAATCTGTTCAGCTCAAAAATC CTGTCTTATTATCTGCTGTTCCAGTCACAAAACATTTTTGCCGCGTTCATTAACGGGTCAACTCCA AATCTGTTAATCTTTATGTCGATGCTGGGCTTGCTGATTTTTAGCATTGTGCTGTATGATAAAAAA TATCTCGTCGTCCACTTTATTGTGACCGGTATCCTTGCGTTCGTTTTCTTCACAACATTTGATTCG AAAGAGAGTTTTGTGTTCTTCATCTCACTTCCTTTCATCTTTATTATTTCGTTAAACTTTAATAAG ATCTACATTACTACCCGTAACCTGATCACCGAGTTATCTATTACTGATGAGATGACCGGCCTCCTC AACCAGAGCGGGTTTATGAAGAAGATCGAAGAAGAATTTTATCGCAGCCAGCGTTACCAGAAAACG TTTTCAGTTCTGATGATCGATTCAGACAATTTAAAACTGATTAATGATACTTATGGCCATAAATAT GGGAGCATCGTCATCAAATCCATTGCGGAAGTCATTAAGACTAACATTCGCCGTACAGACTTCGCG GCGCGCTATGGCGGAGATGAATTCATTCTGTGCCTGGTGGAAACTGATTTAGACGGGGCTCTGGAA GTGGCAGAGGCGATCCGTAAGCAGTTCGAGCTGAAAAGCTTCTTTACCAAAGATGAGAAGAAGTTC ACAATCACGATTAGTATCGGAGTAAGCAACTATCCTAAAAGCGGCGATTCTCTGATGGATGTGATT GAACTTGCGGACAAGGCCATGTACCATAGCAAGAACAGCGGTAAGAACAAGACGAGTTTCCTGCTG AAGAAC
ACP_2467 WT *	C1F1G0	ATGGACGCCCACACTATCGTCAGTCTGCCGCCCACTTGGAACCAAGGGATGTCTGCCGAAGCGCGC AATCAGAACTGGAAGGATTTGGTGGTCTTCCATAACTTAGCACGCGCTCTGACCTCCTCCCTGGAG CTTGATTCGGTGCTGCATGCAATCATGGAACAGATGCGTCAATTCTTCGAACCGGAGACCTGGTCG TTGCTTATCCTGGATGAAACAACCCAGGAATTGTATTACGCGGTTGCAGTCGGACAGTCCGAAGCG GCTCTGCGTAATGTGCGTGTGCCGCTGGGAGAAGGCATGGCGGGTTGGGTGGCCCAACATGGCGAG TCCCTCATCGTGCCGGATCTGGAACAAGATCCGCGCTTCGCCGCGACCTCGGATGCCCGCACCCCA ATGCGTAGCGCGATCTGCATGCCACTGCTCTCACGCCAACGCACCCTGGGCGTGATTCAACTGTTT AACTGCCGCCTGGAAAGCATGACCGAATACACCATTAGCTTCCTGCATATCCTGTGCGACTATGCG GCGATTGCAATCGAAAATGCACGTGCAGTGGAGAAAATCCAGGCCCTGACGATTACGGATGACTGT ACCGGCTTATACAACCAACGTCATCTCCAGCAGAAGATCGAAGAAGAGGTCACCCGTGCTCGTCGT CACCACCATCCATTCTCAGTCATCTTTCTGGATCTTGACCATTTCAAACAAATCAATGACCAACAC GGGCACTTAATCGGGAGCCGCCTTCTGGCGGGTATTGGCCAGTGCCTCCGCCTGCACATTCGCCCG GGAGACCATGCCTTTCGCTATGGTGGCGATGAATTTATCTTACTGCTTCCAGAAACCACAAAAGCG GAAGCCGAGCAGATTGCGCGCAACCTGCGTCAAAAACTGCGTAGCCATGTCTTCGAGATGGGCAGC GATCTCCGTTTGCAGGTTTCGGCCTCATTCGGTGTCGCCAGTTTTCCGGAGGATGGCCGTACGGGC CATCAGATCATTCGTATGGCCGATGCAATGATGTATTTGGTGAAAGGCTCTACGCGCGACGACGTG GCAGTTGCGGACCGTAATACCGAACTGCTCCGCAACTCA

Table 4.6 – List of primer sequences used in Hypr GGDEF study. Restriction sites are denoted by an underline.

#	Nucleotide Sequence (5'→3')	Purpose
1	GAG AGA CAT ATG GAT TTC ACA AAA ATC TCC G	For Primer for GSU0474 into pCOLA
2	GAG AGA CTC GAG TTA CGC TGT AAC GCG GCA G	Rev Primer for GSU0474 into pCOLA
3	GAG AGA CAT ATG CCC TTG CGC AAG AA	For Primer for GSU0537 into pCOLA
4	GAG AGA CTC GAG TTA CGG TTG AAG TGA CCT GAG C	Rev Primer for GSU0537 into pCOLA
5	GAG AGA CAT ATG TCC GGC GAC ATT CTG	For Primer for GSU0542 into pCOLA
6	GAG AGA CTC GAG CTA TTT CAC GAC AAC CTT GTT CTT G	Rev Primer for GSU0542 into pCOLA
7	GAG AGA CAT ATG TCC AGG AAC CAC CTG C	For Primer for GSU0808 into pCOLA
8	GAG AGA AGA TCT CTA ACG GGA AAC GGT GTT GC	Rev Primer for GSU0808 into pCOLA
9	GAG AGA CAT ATG CCC CAT GTG AAC CTG	For Primer for GSU0895 into pCOLA
10	GAG AGA CTC GAG TCA TGG CAG GTT GAG CG	Rev Primer for GSU0895 into pCOLA
11	GAG AGA CAT ATG AAG ATT CGG AGC ACC CT	For Primer for GSU0946 into pCOLA
12	GAG AGA CTC GAG CTA CCC CTC TTC GGC CCT	Rev Primer for GSU0946 into pCOLA
13	GAG AGA CAT ATG TCG GCA GAA AAA GAA CAG AC	For Primer for GSU0952 into pCOLA
14	GAG AGA CTC GAG CTA ACC TTT GAC GGC CTC CAG	Rev Primer for GSU0952 into pCOLA
15	GAG AGA CAT ATG GGC AGG GAG GGC	For Primer for GSU1037 into pCOLA
16	GAG AGA CTC GAG TCA CCT TCC CCG CGC	Rev Primer for GSU1037 into pCOLA
17	GAG AGA CAT ATG AAG CCT GAC ACC ACC TTC	For Primer for GSU1400 into pCOLA
18	GAG AGA CTC GAG CTA TGC GCA GGT GAC GC	Rev Primer for GSU1400 into pCOLA
19	GAG AGA CAT ATG CCG CGA AAG AAG AAA AC	For Primer for GSU1554 into pCOLA
20	GAG AGA CTC GAG TCA GAC GTC GGC GCG	Rev Primer for GSU1554 into pCOLA
21	GAG AGA CAT ATG ACG GAT GAA CAG AGA CAA TG	For Primer for GSU1643 into pCOLA
22	GAG AGA AGA TCT TCA GAG TTG TTC GCT GCA CAC	Rev Primer for GSU1643 into pCOLA
23	GAG AGA CAT ATG CCC CCT CCG CTT C	For Primer for GSU1656 into pCOLA, pET16b
24	GAG AGA CTC GAG TTA TGC AGG TAA TAC GCA GCA TTT TTT A	Rev Primer for GSU1656 into pCOLA, pET16b, pET-MBP

25	GAG AGA CAT ATG GAA CGG ATT CTC GTT GTC	For Primer for GSU1658 into pCOLA, pET24a
26	GAG AGA CTC GAG TCA ACG GAT TGC CGT TGC	Rev Primer for GSU1658 into pCOLA
27	GAG AGA CAT ATG ACA GAT GCC ATT ACG GAT G	For Primer for GSU1671 into pCOLA
28	GAG AGA CTC GAG TCA ATG AAG CTG GAC TCC CTT G	Rev Primer for GSU1671 into pCOLA
29	GAG AGA CAT ATG GAA CTC AGC CCC GAG	For Primer for GSU1870 into pCOLA
30	GAG AGA CTC GAG TCA TGG CTC ATC CTC TCT TCT G	Rev Primer for GSU1870 into pCOLA
31	GAG AGA CAT ATG CGA AAA GAG GGC AAG G	For Primer for GSU1927 into pCOLA
32	GAG AGA AGA TCT CTA GCG CGA CCG AGC G	Rev Primer for GSU1927 into pCOLA
33	GAG AGA CAT ATGACCCTCGCCGAAG	For Primer for GSU1937 into pCOLA
34	GAG AGA CTC GAG TCAGGGGTGCATTGACAG	Rev Primer for GSU1937 into pCOLA
35	GAG AGA CAT ATG GCC CAG ACT TCA TTG AC	For Primer for GSU2016 into pCOLA
36	GAG AGA AGA TCT TTA CGG GGC TGA GTT CAG ACT G	Rev Primer for GSU2016 into pCOLA
37	GAG AGA CAT ATG GCG AAT CTC AAG CGA TAT AAT	For Primer for GSU2044 into pCOLA
38	GAG AGA CTC GAG TCA GCA CCA GGT TCC GAA AC	Rev Primer for GSU2044 into pCOLA
39	GAG AGA CAT ATGAGATCTGACCTGAGAATAGCC	For Primer for GSU2062 into pCOLA
40	GAG AGA CTC GAG TCAGTACTTACGTCGGTCGAC	Rev Primer for GSU2062 into pCOLA
41	GAG AGA CAT ATG CGA ATT CTC ATC GCC	For Primer for GSU2313 into pCOLA
42	GAG AGA CTC GAG TCA TGG TGA TCC CGC CTG	Rev Primer for GSU2313 into pCOLA
43	GAG AGA CAT ATG GTT GCG TTC TTC ACA CAG TA	For Primer for GSU2511 into pCOLA
44	GAG AGA CTC GAG TCA TTC CCT CGG CGC	Rev Primer for GSU2511 into pCOLA
45	GAG AGA CAT ATG GCC GAA TCA CGT CC	For Primer for GSU2534 into pCOLA
46	GAG AGA CTC GAG CTA GCA CGG GGA TCC GG	Rev Primer for GSU2534 into pCOLA
47	GAG AGA CAT ATG AAC ACC CTG ACG GCA	For Primer for GSU2632 into pCOLA
48	GAG AGA CTC GAG TCA GGT GCT CAC CTG GTT GC	Rev Primer for GSU2632 into pCOLA
49	GAG AGA CAT ATG ACT GAA TTG ACG GAG TTC GTA G	For Primer for GSU2828 into pCOLA
50	GAG AGA CTC GAG TCA TCC GTT CAC TGC GCC	Rev Primer for GSU2828 into pCOLA
51	GAG AGA CAT ATG CCC AAC AAC GAC AGC	For Primer for GSU2969 into pCOLA

52	GAG AGA CTC GAG TCA GGG TGA CGC GGA C	Rev Primer for GSU2969 into pCOLA
53	GAG AGA CAT ATG ACG CGC CGG C	For Primer for GSU3350 into pCOLA
54	GAG AGA CTC GAG TCA ATC GGT TCC GTC CG	Rev Primer for GSU3350 into pCOLA
55	GAG AGA CAT ATG AGA CGA GCA AGC CTG AAA	For Primer for GSU3356 into pCOLA
56	GAG AGA AGA TCT TCA GGA GGC CGA AAC GG	Rev Primer for GSU3356 into pCOLA
57	GAG AGA CAT ATG GCG ATG ACA GCC CTC	For Primer for GSU3376 into pCOLA
58	GAG AGA CTC GAG TTA TGT CGA GCC TGA CAT GAG CTC	Rev Primer for GSU3376 into pCOLA
59	GAG AGA CTC GAG ACG GAT TGC CGT TGC	Rev Primer for GSU1658 into pET24a
60	CAACCGTGGCTGAAGCCATTCGTCGCTCCATC	For Primer to Quickchange GSU1658 R393A
61	GATGGAGCGACGAATGGCTTCAGCCACGGTTG	Rev Primer to Quickchange GSU1658 R393A
62	CCATCTTTTTGGGGCCCAGGTCCTCCATG	For Primer to Quickchange GSU1658 S347A
63	CATGGAGGACCTGGGCCCCAAAAAGATGG	Rev Primer to Quickchange GSU1658 S347A
64	CGGCCATCTTTTTGGGGATCAGGTCCTCCATGAAG	For Primer to Quickchange GSU1658 S347D
65	CTTCATGGAGGACCTGATCCCCAAAAAGATGGCCG	Rev Primer to Quickchange GSU1658 S347D
66	GGCCATCTTTTTGGGAATCAGGTCCTCCATGAAG	For Primer to Quickchange GSU1658 S347N
67	CTTCATGGAGGACCTGATTCCCAAAAAGATGGCC	Rev Primer to Quickchange GSU1658 S347N
68	GTAATCATTCGCTACCTTGGCGACGAGTTCACC	For Primer to Quickchange GSU1658 G371L
69	GGTGAACTCGTCGCCAAGGTAGCGAATGATTAC	Rev Primer to Quickchange GSU1658 G371L
70	CCACCTCGTCATTACCGCACTGGTCATGCCCGGAAT G	For Primer to Quickchange GSU1658 D52A
71	CATTCCGGGCATGACCAGTGCGGTAATGACGAGGTG G	Rev Primer to Quickchange GSU1658 D52A
72	CGTCATTACCGAGCTGGTCATGCCC	For Primer to Quickchange GSU1658 D52E
73	GGGCATGACCAGCTCGGTAATGACG	Rev Primer to Quickchange GSU1658 D52E
74	GAGAGAGGATCCATGGACGCCCACACTATC	For Primer for ACP_2467 into pET-MBP
75	GAGAGACTCGAGTCATGAGTTGCGGAGCAGTTC	Rev Primer for ACP_2467 into pET-MBP
76	AGAGACAT ATGTCGCGCGCCG	For Primer for Bd0367 into pET24a
77	AGAGACTCGAGACCGGCGGCTTTACG	Rev Primer for Bd0367 into pET24a
78	AGAGACATATGAACCTTAAACTGGGCGC	For Primer for Cabther_A1065 into pET24a
79	AGAGACTCGAGAGGTTTGACAGGCTGCG	Rev Primer for Cabther_A1065 into pET24a
80	GAGAGAGGATCCATGATTGATAACAAGATTAAACAC	For Primer for Calni_1629 into pET-MBP
81	AGAGAGCTCGAGTCAGTTCTTCAGCAGGAAACTC	Rev Primer for Calni_1629 into pET-MBP
82	GAGAGAGGATCCATGTATGAAAGCCTGAAACG	For Primer for DEFDS_0689 into pET-MBP
83	AGAGAGCTCGAGTCATAAGGAAGAATTTTTGGTCAG G	Rev Primer for DEFDS_0689 into pET-MBP
84	GAGAGACATATGCTGAACAAGTCAAGCATC	For Primer for Ddes_1475 into pET24a

85	GAGAGACTCGAGAGAGTCGTCGTGCGTGG	Rev Primer for Ddes_1475 into pET24a
86	AGAGACATATGAATCCCGCGGACCTC	For Primer for Mxan_2643 into pET24a
87	AGAGAGCGGCCGCTCCGGGAACTCGTGG	Rev Primer for Mxan_2643 into pET24a
88	AGAGACATATGGCGCGAATCCTCC	For Primer for Mxan_4463 into pET24a
89	AGAGACTCGAGGGACGCGGAGTGGGC	Rev Primer for Mxan_4463 into pET24a
90	GAGAGACATATGAGCGCCCGGATCCTCG	For Primer for ccPleD into pET24a
91	GAGAGACTCGAGTCAGGCGGCCTTGCCG	Rev Primer for ccPleD into pET24a
92	CTGGTCGGTGACGGC	LIC primers for GSU1658-PleD fusion (rev; primes to PleD 293)
93	GCCGTCACCGACCAGCTCACGGGACTCTTCAACTAC	Lic primers for GSU1658-PleD fusion (overhang with PleD NTD; primes to 1658 f 297)
94	CTTCGGTCACGATATCGGCAGTGAGGTGCTGCGCGAGTTC	For primer to Quickchange PleD D344S
95	GAACTCGCGCAGCACCTCACTGCCGATATCGTGACCGAAG	Rev primer to Quickchange PleD D344S
96	GGGCACCAGATGGGAAGCGACCTCCTCAAAATG	Forward primer for Round-the-horn of GSU1400 D195S
97	GAAACTGTCGTTGATCTCCTTGAAG	Reverse primer for Round-the-horn of GSU1400 D195S
98	CATCAGACCGGAAGCGAGGTGCTGTGC	Forward primer for Round-the-horn of GSU2313 D194S
99	ACCGTAACGGTCGTTTACCC	Reverse primer for Round-the-horn of GSU2313 D194S
100	CCACGACGCCGGCAGTGTGCTCCTGATGG	Forward primer for Round-the-horn of GSU2534 D325S
101	CCGAAGACATCGTTCACTCC	Reverse primer for Round-the-horn of GSU2534 D325S
102	CACCTCCGGGGCAGCGAGGTCCTCAG	Forward primer for Round-the-horn of GSU3350 D471S
103	GCCGTGGCAGTCGTTG	Reverse primer for Round-the-horn of GSU3350 D471S
104	GAGAGGATCCATGAATCCCGCGGACCTC	For primer for Mxan_2643 into pET-MBP
105	GAGAGAGTCGACTCATCCGGGAACTCGTGG	Rev primer for Mxan_2643 into pET-MBP
106	GAGAGGATCCATGTCGCGCGCCG	For primer for Bd0367 into pET-MBP
107	GAGAGACTCGAGTCAACCGGCGGCTTTACG	Rev primer for Bd0367 into pET-MBP
108	GCAGGGCGGGGCTTCCTC	For primer to Quickchange Mxan_2643 R292A
109	AAAGGCCTCGCAGACG	Rev primer to Quickchange Mxan_2643 R292A
110	GCGCTGCGCGAAAATATTGAAAAAAC	For primer to Quickchange Bd0367 R260A
111	CTCGCAAAAATACATAGCG	Rev Primer to Quickchange Bd0367 R260A

REFERENCES

Ablasser, A. et al., 2013. cGAS produces a 2′-5′-linked cyclic dinucleotide second messenger that activates STING. *Nature*, 498(7454), 380–384.

Adams, P.D. et al., 2002. PHENIX: building new software for automated crystallographic structure determination. *Acta crystallographica. Section D, Biological crystallography*, 58(Pt 11), 1948–1954.

Amikam, D. & Galperin, M.Y., 2006. PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics*, 22(1), 3–6.

An, S., Wu, J. & Zhang, L.-H., 2010. Modulation of Pseudomonas aeruginosa biofilm dispersal by a cyclic-Di-GMP phosphodiesterase with a putative hypoxia-sensing domain. *Applied and environmental microbiology*, 76(24), 8160–8173.

Andersen, J.B. et al., 1998. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Applied and environmental microbiology*, 64(6), 2240–2246.

Antoniani, D. et al., 2010. Monitoring of diguanylate cyclase activity and of cyclic-di-GMP biosynthesis by whole-cell assays suitable for high-throughput screening of biofilm inhibitors. *Applied microbiology and biotechnology*, 85(4), 1095–1104.

Ausmees, N. et al., 2001. Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. *FEMS Microbiology Letters*, 204(1), 163–167.

Babendure, J.R., Adams, S.R. & Tsien, R.Y., 2003. Aptamers switch on fluorescence of triphenylmethane dyes. *Journal of the American Chemical Society*, 125(48), 14716–14717.

Brown, J.A. et al., 2014. Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nature Structural & Molecular Biology*, 21(7), 633–640.

Buckley, A.M. et al., 2015. LOV-based reporters for fluorescence imaging. *Current opinion in chemical biology*, 27, 39–45.

Bueno, E. et al., 2012. Bacterial Adaptation of Respiration from Oxic to Microoxic and Anoxic Conditions: Redox Control. *Antioxidants & Redox Signaling*, 16(8), 819–852.

Burdette, D.L. et al., 2011. STING is a direct innate immune sensor of cyclic di-GMP. *Nature*, 478(7370), 515–518.

Chan, C. et al., 2004. Structural basis of activity and allosteric control of diguanylate cyclase. *Proceedings of the National Academy of Sciences*, 101(49), 17084–17089.

Chapman, S. et al., 2008. The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proceedings of the National Academy of Sciences*, 105(50), 20038–20043.

Chen, Z.-H. & Schaap, P., 2012. The prokaryote messenger c-di-GMP triggers stalk cell differentiation in Dictyostelium. *Nature*, 488(7413), 680–683.

Christen, M. et al., 2010. Asymmetrical Distribution of the Second Messenger c-di-GMP upon Bacterial Cell Division. *Science*, 328(5983), 1295–1297.

Cohen, D. et al., 2015. Oligoribonuclease is a central feature of cyclic diguanylate signaling in Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences*, 112(36), 11359–11364.

Corrigan, R.M. et al., 2013. Systematic identification of conserved bacterial c-di-AMP receptor proteins. *Proceedings of the National Academy of Sciences*, 110(22), 9084–9089.

Craggs, T.D., 2009. Green fluorescent protein: structure, folding and chromophore maturation. *Chemical Society Reviews*, 38(10), 2865–2875.

Crone, D.E. et al., 2013. GFP-based biosensors. In T. Rinken, ed. *State of the Art in Biosensors - General Aspects*. InTech.

De, N. et al., 2008. Phosphorylation-independent regulation of the diguanylate cyclase WspR. *PLoS Biology*, 6(3), p.e67.

Diner, E.J. et al., 2013. The Innate Immune DNA Sensor cGAS Produces a Noncanonical Cyclic Dinucleotide that Activates Human STING. *Cell Reports*, 3(5), 1355–1361.

Dolgosheina, E.V. et al., 2014. RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking. *ACS Chemical Biology*, 9(10), 2412–2420.

Donaldson, G.P. et al., 2012. A rapid assay for affinity and kinetics of molecular interactions with nucleic acids. *Nucleic Acids Research*, 40(7), e48–e48.

Drepper, T. et al., 2007. Reporter proteins for in vivo fluorescence without oxygen. *Nature Biotechnology*, 25(4), 443–445.

Duerig, A. et al., 2009. Second messenger-mediated spatiotemporal control of protein degradation regulates bacterial cell cycle progression. *Genes & Development*, 23(1), 93–104.

Düvel, J. et al., 2012. A chemical proteomics approach to identify c-di-GMP binding proteins in Pseudomonas aeruginosa. *Journal of Microbiological Methods*, 88(2), 229–236.

Edwards, A.L. & Batey, R.T., 2009. A structural basis for the recognition of 2'-deoxyguanosine by the purine riboswitch. *Journal of Molecular Biology*, 385(3), 938–948.

Emsley, P. & Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological Crystallography*, 60(Pt 12 Pt 1), 2126–2132.

Enterina, J.R., Wu, L. & Campbell, R.E., 2015. Emerging fluorescent protein technologies. *Current Opinion in Chemical Biology*, 27, 10–17.

Erapaneedi, R. et al., 2015. A novel family of fluorescent hypoxia sensors reveal strong heterogeneity in tumor hypoxia at the cellular level. *The EMBO Journal*, (35), 102–113.

Fowler, C.C., Brown, E.D. & Li, Y., 2010. Using a riboswitch sensor to examine coenzyme B(12) metabolism and transport in *E. coli*. *Chemistry & Biology*, 17(7), 756–765.

Furukawa, K. et al., 2012. Identification of ligand analogues that control c-di-GMP riboswitches. *ACS Chemical Biology*, 7(8), 1436–1443.

Galperin, M.Y. et al., 1999. A specialized version of the HD hydrolase domain implicated in signal transduction. *Journal of Molecular Microbiology and Biotechnology*, 1(2), 303–305.

Gao, J. et al., 2015. Identification and characterization of phosphodiesterases that specifically degrade 3′3′-cyclic GMP-AMP. *Cell Research*, 25(5), 539–550.

Gao, P. et al., 2013. Cyclic [G(2′,5′)pA(3′,5′)p] Is the Metazoan Second Messenger Produced by DNA-Activated Cyclic GMP-AMP Synthase. *Cell*, 153(5), 1094–1107.

Gao, S. et al., 2014. Genomic analysis of cyclic-di-GMP-related genes in rhizobial type strains and functional analysis in Rhizobium etli. *Applied Microbiology and Biotechnology*, 98(10), 4589–4602.

Goldbeck, C.P. et al., 2013. Tuning promoter strengths for improved synthesis and function of electron conduits in Escherichia coli. *ACS Synthetic Biology*, 2(3), 150–159.

Golomb, M. & Chamberlin, M., 1974. Characterization of T7-specific ribonucleic acid polymerase. IV. Resolution of the major in vitro transcripts by gel electrophoresis. *The Journal of Biological Chemistry*, 249(9), 2858–2863.

Griffiths-Jones, S. et al., 2003. Rfam: an RNA family database. *Nucleic Acids Research*, 31(1), 439–441.

Hallberg, Z.F. et al., 2016. Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3“, 3-”cGAMP). *Proceedings of the National Academy of Sciences*, 113(7), 1790–1795.

Haugh, J.M., 2012. Live-cell fluorescence microscopy with molecular biosensors: what are we really measuring? *Biophysical Journal*, 102(9), 2003–2011.

Hickman, J.W., Tifrea, D.F. & Harwood, C.S., 2005. A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. *Proceedings of the National Academy of Sciences*, 102(40), 14422–14427.

Ho, C.L. et al., 2013. Visualizing the Perturbation of Cellular Cyclic di-GMP Levels in Bacterial Cells. *Journal of the American Chemical Society*, 135(2), 566–569.

Jensen, P.R. & Hammer, K., 1998. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Applied and Environmental Microbiology*, 64(1), 82–87.

Kellenberger, C.A. & Hammond, M.C., 2015. In vitro analysis of riboswitch-Spinach aptamer fusions as metabolite-sensing fluorescent biosensors. *Methods in Enzymology*, 550, 147–172.

Kellenberger, C.A. et al., 2013. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messengers Cyclic di-GMP and Cyclic AMP-GMP. *Journal of the American Chemical Society*, 135, 4906–4909.

Kellenberger, C.A., Chen, C., et al., 2015. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messenger Cyclic di-AMP. *Journal of the American Chemical Society*, 137(20), 6432–6435.

Kellenberger, C.A., Sales-Lee, J., et al., 2015. A minimalist biosensor: Quantitation of cyclic di-GMP using the conformational change of a riboswitch aptamer. *RNA Biology*, 12(11), 1189–1197.

Kellenberger, C.A., Wilson, S.C., et al., 2015. GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. *Proceedings of the National Academy of Sciences*, 112(17), 5383–5388.

Kim, S. et al., 2013. Anticancer flavonoids are mouse-selective STING agonists. *ACS Chemical Biology*, 8(7), 1396–1401.

Koh, J.T., 2002. Engineering Selectivity and Discrimination into Ligand-Receptor Interfaces. *Chemistry & Biology*, 9(1), 17–23.

Kramer, F.R. & Mills, D.R., 1981. Secondary structure formation during RNA synthesis. *Nucleic Acids Research*, 9(19), 5109–5124.

Kranzusch, P.J. et al., 2013. Structure of Human cGAS Reveals a Conserved Family of Second-Messenger Enzymes in Innate Immunity. *Cell Reports*, 3(5), 1362–1368.

Kumagai, A. et al., 2013. A bilirubin-inducible fluorescent protein from eel muscle. *Cell*, 153(7), 1602–1611.

Lacey, M.M., Partridge, J.D. & Green, J., 2010. Escherichia coli K-12 YfgF is an anaerobic cyclic di-GMP phosphodiesterase with roles in cell surface remodelling and the oxidative stress response. *Microbiology*, 156(9), 2873–2886.

Liaw, Y.C. et al., 1990. Cyclic diguanylic acid behaves as a host molecule for planar intercalators. *FEBS letters*, 264(2), 223–227.

Lieberman, O.J. et al., 2014. High-throughput screening using the differential radial capillary action of ligand assay identifies ebselen as an inhibitor of diguanylate cyclases. *ACS Chemical Biology*, 9(1), 183–192.

Liem, P.H. et al., 2015. A simple and highly sensitive method of measuring heme oxygenase activity. *Biological Chemistry*, 396(11), 1265–1268.

Lori, C. et al., 2015. Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication. *Nature*, 523(7559), 236–239.

Lozupone, C.A. et al., 2012. Diversity, stability and resilience of the human gut microbiota. *Nature*, 489(7415), 220–230.

Luo, Y. et al., 2012. Differential binding of 2′-biotinylated analogs of c-di-GMP with c-di-GMP riboswitches and binding proteins. *Molecular BioSystems*, 8(3), 772–778.

Lutz, S., 2010. Beyond directed evolution—semi-rational protein engineering and design. *Current Opinion in Biotechnology*, 21(6), 734–743.

Mitra, R.D., Silva, C.M. & Youvan, D.C., 1996. Fluorescence resonance energy transfer between blue-emitting and red-shifted excitation derivatives of the green fluorescent protein. *Gene*, 173(1), 13–17.

Mutalik, V.K. et al., 2013. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods*, 10(4), 354–360.

Nakayama, S., Kelsey, I., Wang, J. & Sintim, H.O., 2011. c-di-GMP can form remarkably stable G-quadruplexes at physiological conditions in the presence of some planar intercalators. *Chemical Communications*, 47(16), 4766–4768.

Nakayama, S., Kelsey, I., Wang, J., Roelofs, K., et al., 2011. Thiazole orange-induced c-di-GMP quadruplex formation facilitates a simple fluorescent detection of this ubiquitous biofilm regulating molecule. *Journal of the American Chemical Society*, 133(13), 4856–4864.

Nakayama, S., Luo, Y., et al., 2012. Nanomolar fluorescent detection of c-di-GMP using a modular aptamer strategy. *Chemical Communications*, 48(72), 9059–9061.

Nakayama, S., Roelofs, K., et al., 2012. A C-di-GMP-proflavine-hemin supramolecular complex has peroxidase activity--implication for a simple colorimetric detection. *Molecular BioSystems*, 8(3), 726–729.

Nawrocki, E.P. & Eddy, S.R., 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29(22), 2933–2935.

Nelson, J.W. et al., 2015. Control of bacterial exoelectrogenesis by c-AMP-GMP. *Proceedings of the National Academy of Sciences*, 112(17), 5389–5394.

Nicholson, J.K. et al., 2012. Host-gut microbiota metabolic interactions. *Science*, 336(6086), 1262–1267.

O'Toole, G.A. et al., 1999. Genetic approaches to study of biofilms. *Methods in Enzymology*, 310, 91–109.

Orr, M.W. et al., 2015. Oligoribonuclease is the primary degradative enzyme for pGpG in Pseudomonas aeruginosa that is required for cyclic-di-GMP turnover. *Proceedings of the National Academy of Sciences*, 112(36), E5048–57.

Paige, J.S. et al., 2012. Fluorescence imaging of cellular metabolites with RNA. *Science*, 335(6073), 1194–1194.

Paige, J.S., Wu, K.Y. & Jaffrey, S.R., 2011. RNA mimics of green fluorescent protein. *Science*, 333(6042), 642–646.

Pan, Y. et al., 2014. High-Throughput Electrophoretic Mobility Shift Assays for Quantitative Analysis of Molecular Binding Reactions. *Analytical Chemistry*, 86(20), 10357–10364.

Paul, K. et al., 2010. The c-di-GMP binding protein YcgR controls flagellar motor direction and speed to affect chemotaxis by a "backstop brake" mechanism. *Molecular Cell*, 38(1), 128–139.

Pikovskaya, O. et al., 2011. Structural principles of nucleoside selectivity in a 2'-deoxyguanosine riboswitch. *Nature Chemical Biology*, 7(10), 748–755.

Ponchon, L. & Dardel, F., 2007. Recombinant RNA technology: the tRNA scaffold. *Nature Methods*, 4(7), 571–576.

Ransom, E.M., Ellermeier, C.D. & Weiss, D.S., 2015. Use of mCherry Red fluorescent protein for studies of protein localization and gene expression in Clostridium difficile. *Applied and Environmental Microbiology*, 81(5), 1652–1660.

Reguera, G. et al., 2005. Extracellular electron transfer via microbial nanowires. *Nature*, 435(7045), 1098–1101.

Reid, B.G. & Flynn, G.C., 1997. Chromophore formation in green fluorescent protein. *Biochemistry*, 36(22), 6786–6791.

Ren, A. et al., 2015. Structural basis for molecular discrimination by a 3",3-"cGAMP sensing riboswitch. *Cell Reports*, 11(1), 1–12.

Roelofs, K.G. et al., 2011. Differential radial capillary action of ligand assay for high-throughput detection of protein-metabolite interactions. *Proceedings of the National Academy of Sciences*, 108(37), 15528–15533.

Roembke, B.T. et al., 2014. A cyclic dinucleotide containing 2-aminopurine is a general fluorescent sensor for c-di-GMP and 3",3-"cGAMP. *Molecular BioSystems*, 10(6), 1568–1575.

Ross, P. et al., 1987. Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. *Nature*, 325(6101), 279–281.

Römling, U., Galperin, M.Y. & Gomelsky, M., 2013. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiology and Molecular Biology Reviews : MMBR*, 77(1), 1–52.

Ryan, R.P., 2013. Cyclic di-GMP signalling and the regulation of bacterial virulence. *Microbiology (Reading, England)*, 159(Pt 7), 1286–1297.

Seshasayee, A.S.N., Fraser, G.M. & Luscombe, N.M., 2010. Comparative genomics of cyclic-di-GMP signalling in bacteria: post-translational regulation and catalytic activity. *Nucleic Acids Research*, 38(18), 5970–5981.

Shanahan, C.A. et al., 2011. Differential analogue binding by two classes of c-di-GMP riboswitches. *Journal of the American Chemical Society*, 133(39), 15578–15592.

Simm, R. et al., 2004. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. *Molecular Microbiology*, 53(4), 1123–1134.

Simm, R. et al., 2005. Phenotypic Convergence Mediated by GGDEF-Domain-Containing Proteins. *Journal of Bacteriology*, 187(19), 6816–6823.

Simm, R. et al., 2009. Quantitative determination of cyclic diguanosine monophosphate concentrations in nucleotide extracts of bacteria by matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry. *Analytical Biochemistry*, 386(1), 53–58.

Singh, V., Fedeles, B.I. & Essigmann, J.M., 2015. Role of tautomerism in RNA biochemistry. *RNA*, 21(1), 1–13.

Skerker, J.M. & Laub, M.T., 2004. Cell-cycle progression and the generation of asymmetry in Caulobacter crescentus. *Nature Reviews Microbiology*, 2(4), 325–337.

Smith, K.D. et al., 2010. Structural and biochemical determinants of ligand binding by the c-di-GMP riboswitch . *Biochemistry*, 49(34), 7351–7359.

Smith, K.D. et al., 2009. Structural basis of ligand binding by a c-di-GMP riboswitch. *Nature Structural & Molecular Biology*, 16(12), 1218–1223.

Song, W. et al., 2014. Plug-and-play fluorophores extend the spectral properties of Spinach. *Journal of the American Chemical Society*, 136, 1198–1201.

Song, W., Strack, R.L. & Jaffrey, S.R., 2013. Imaging bacterial protein expression using genetically encoded RNA sensors. *Nature Methods*, 10(9), 873–875.

Spangler, C. et al., 2010. A liquid chromatography-coupled tandem mass spectrometry method

for quantitation of cyclic di-guanosine monophosphate. *Journal of Microbiological Methods*, 81(3), 226–231.

Srivastava, S., Sinha, R. & Roy, D., 2004. Toxicological effects of malachite green. *Aquatic toxicology*, 66(3), 319–329.

Stelitano, V. et al., 2013. Probing the activity of diguanylate cyclases and c-di-GMP phosphodiesterases in real-time by CD spectroscopy. *Nucleic Acids Research*, 41(7), e79–e79.

Stojanovic, M.N. & Kolpashchikov, D.M., 2004. Modular aptameric sensors. *Journal of the American Chemical Society*, 126(30), 9266–9270.

Strack, R.L. & Jaffrey, S.R., 2013. New approaches for sensing metabolites and proteins in live cells using RNA. *Current Opinion in Chemical Biology*, 17(4), 651–655.

Strack, R.L., Disney, M.D. & Jaffrey, S.R., 2013. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat–containing RNA. *Nature Methods*, 10(12), 1219–1224.

Studier, F.W., 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification*, 41(1), 207–234.

Sudarsan, N. et al., 2008. Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science*, 321(5887), 411–413.

Sun, L. et al., 2013. Cyclic GMP-AMP Synthase Is a Cytosolic DNA Sensor That Activates the Type I Interferon Pathway. *Science*, 339(6121), 786–791.

Topp, S. et al., 2010. Synthetic riboswitches that induce gene expression in diverse bacterial species. *Applied and Environmental Microbiology*, 76(23), 7881–7884.

Travascio, P. et al., 1999. A ribozyme and a catalytic DNA with peroxidase activity: active sites versus cofactor-binding sites. *Chemistry & Biology*, 6(11), 779–787.

Tuckerman, J.R. et al., 2009. An oxygen-sensing diguanylate cyclase and phosphodiesterase couple for c-di-GMP control. *Biochemistry*, 48(41), 9764–9774.

Tuerk, C. & Gold, L., 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 249(4968), 505–510.

Waterhouse, A.M. et al., 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189–1191.

Waters, C.M., 2010. Methods for Cyclic Di-GMP Detection, *from* The Second Messenger Cyclic Di-GMP, A. J. Wolfe and K. L. Visick, eds. ASM Press, Washington, DC, 68–75.

Weinberg, Z. et al., 2007. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Research*, 35(14), 4809–4819.

Weinhouse, H. et al., 1997. c-di-GMP-binding protein, a new factor regulating cellulose synthesis in Acetobacter xylinum. *FEBS Letters*, 416(2), 207–211.

Westhof, E., 2014. Isostericity and tautomerism of base pairs in nucleic acids. *FEBS Letters*, 588(15), 2464–2469.

Wilson, S.C. et al., 2014. A neutral pH thermal hydrolysis method for quantification of structured RNAs. *RNA*, 20(7), 1153–1160.

Wolfe, A.J. & Berg, H.C., 1989. Migration of bacteria in semisolid agar. *Proceedings of the National Academy of Sciences*, 86(18), 6973–6977.

Wu, J. et al., 2013. Cyclic GMP-AMP Is an Endogenous Second Messenger in Innate Immune Signaling by Cytosolic DNA. *Science*, 339(6121), 826–830.

Zähringer, F. et al., 2013. Structure and signaling mechanism of a zinc-sensory diguanylate cyclase. *Structure*, 21(7), 1149–1157.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13), 3406–3415.