Title

Broken Symmetry: Discovery, Characterization, and Structure of Hybrid Promiscuous (HyPr) GGDEF Enzymes

Author

Hallberg, Zachary

Publication Date

2017

Peer reviewed|Thesis/dissertation

Broken Symmetry:
Discovery, Characterization, and Structure of Hybrid Promiscuous (HyPr) GGDEF Enzymes

By

Zachary Francis Hallberg

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ming Chen Hammond, Chair
Professor Matthew Francis
Professor Russell Vance

Fall 2017

ABSTRACT

Broken Symmetry:
Discovery, Mechanistic Analysis, and Structure of
Hybrid Promiscuous (HyPr) GGDEF Enzymes
by
Zachary Francis Hallberg

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Ming C. Hammond, Chair

This past decade has witnessed an expanded diversity in the cast of nucleotide-based second messengers. The small lineup of cyclic AMP, cyclic di-GMP and ppGpp have been joined by new co-stars, including cyclic di-AMP, cyclic oligoA, and two flavors of cyclic AMP-GMP - with one possessing a unique 2'-5' phosphodiester linkage specific to eukaryotes. As each new signaling pathway has been uncovered, new enzymes classes have been attached to each specific second messenger, suggesting rigid evolutionary pathways in the development of new signal transduction pathways.

We have begun to unravel this paradigm with the discovery of a subclass of GGDEF enzyme - previously associated only with cyclic di-GMP synthesis - responsible for synthesis of a different cyclic dinucleotide, cyclic AMP-GMP. Intriguingly, this enzyme behaves nothing like the other two. Our discovery of this new activity from an ancient enzyme class pinpoints a set of deltaproteobacteria that have this "hybrid promiscuous" GGDEF activity, which preliminary work suggests controls transient surface attachment in diverse bacterial lifestyles. Furthermore, we have begun to unravel how Nature has allowed this newfound activity to act separately from other GGDEF-related pathways to allow precise control over electron acceptor choice in *Geobacter* species. Our work suggests that there is further underappreciated potential for individual components of nucleotide second messenger signaling enzymes, and establishes a footing for engineering new bioorthogonal signaling functions into this venerable enzyme class.

For Mary Dolan

TABLE OF CONTENTS

“It’s not about you.” - (now-Dr.) Cindy Wang

Per the usual, Cindy was right: I couldn't do this alone. I have so many fabulous colleagues who helped keep me sane through this process. They were especially instrumental the last few weeks, when I was on the verge of mental collapse due to lack of sleep and needed coherent people to make sure that what I was writing (on no sleep) (a) made sense and (b) were true statements. And I have a boss who is willing to put in near-superhuman hours to make sure that the things we write are of high quality. More than any other professor in chemistry that I've seen, my adviser is not ashamed to be right next to her students in the arena. That's a topic for later, though.

But that's the point: I can't do this alone. Any of this. None of us can.

I’m going to get religious at this point, so if that’s not your cup of tea, then please move on to page two. Audrey Assad (one of my favorite Christian artists) has an album titled "Fortunate Fall." The content is all based around the idea of the "Felix Culpa" (Happy Fault) as espoused by St. Augustine - his solution to the problem of evil is that "God judged it better to bring good out of evil than not to permit any evil to exist," and thus, the original sin of Adam was what led to the Incarnation of God in human form. The album itself is gorgeous (seriously, listen), in both lyrical and musical content.

If you want a stronger recommendation - I've spent the past week alternating between this and her album "Heart."

Anyways - different artists have different messages that they want to communicate with the world - being a semi-Platonist, I'd argue that the "sparks" that lead us to different forms of expression and creativity stem from us seeing different fundamental aspects of God. Think of it this way - God is multidimensional in His aspects, and by multidimensional I'd argue we're talking thousands if not millions of dimensions (most likely infinite. Word's still out on whether it's countably infinite or uncountable). While we are, ourselves, multidimensional in our personalities, we can only experience a finite amount of the dimensions of God at a time (the same way we can only perceive 3 dimensions spatially at a time). And it's those dimensions that we experience that we are most comfortable communicating as a message to others.

I suppose the message that comes through my experience of God I sum up as "Not By, Through." It's the fundamental idea that everything that I've ever accomplished has happened only because I've had the chance to stand on the shoulders of giants - friends, mentors, colleagues, historical figures, grad students from 50 years ago, or, well, God. There is nothing that I have done that can't be traced back to someone else.

I think that's an important exercise for everyone to do. Find something that you like about where you are in life, trace it back through, and figure out how many people helped make that into a reality.

I know, that's a super simple - almost stupidly simple - exercise that we all probably did in kindergarten. But do it. Give yourself the gift of remembering just how loved you are. What follows is my meditation upon this, and my thanks to each and every one of those people. I may miss some people - understand that it’s not personal, that I’m running on not-so-much sleep, and that there is a finite deadline of less than 8 hours before I have to submit this thing as a whole. Also as a quick note - if the name Dr. is in parentheses, it is because they became a PhD between when I met them and now. If the name Dr. is not in parentheses, it is because they were already a PhD when I had met them.

Although my PhD was earned studying the biochemistry of an anaerobe, my science has never existed in a vacuum. From the earliest days that I thought about becoming a scientist, to the present, I have been blessed beyond measure with communities of diverse and passionate people who have supported my endeavors. It is a daily reminder to me that often environment has a stronger effect on personal outcomes than genetics. So, without further adieu - these are the people through whom I found the strength, courage, energy, and passion to keep doing science (almost) every day for five and a half years.

To my immediate family - I know that I may put you to sleep talking about science. I would say that I'm sorry, but part of telling a young person that they should pursue their dreams and cultivate their love of science suggests to me that I shouldn't be that sorry. To my parents, who were the first to nurture that love of science - thank you for all that you have done. I see now how lucky I am to have had parents who made it a priority that I graduate from college debt free. I see now how lucky I am to have had parents who made building a stable home life a priority over having the most auspicious career or the most money. I see now how lucky I am to have been given a strong moral compass grounded in love, grace, and integrity. I see now how lucky I am to have been given gifts outside of science - primarily music - to cultivate when my spirit needs respite. Let's face it - the past two years were scary for our family. If there is one thing that I regret not saying often enough - it is that I love you both so much.

To my siblings, I am sorry that you had to go through school living in the midst of my shadow. I am not surprised, however, that you each became successful in finding your own trajectories. Know that, while I love you both, it would be really inappropriate for me to do anything other than tease you in an acknowledgements section - Ellen, I learn as I write this acknowledgments section that it's quite easy to accidentally miss someone. For someone as… brevity-challenged… as I am, I find the idea of having to condense acknowledgements to one page to be, well… impossible. Know that while I may give you a hard time for that valedictorian speech, it comes from love and the knowledge that, maybe as much as I have been your older brother, you as my younger sister have taught me more about resilience than I thought I'd be needing to learn in the last year of graduate school. Truth be told - the only reason I got through those two months on my broken foot were because I knew you were getting through your own battle with post-concussion syndrome.

Ben - what can I say, little brother… you're the only one willing to go to BWW's with me when I'm at home. That in and of itself is a miracle. I would apologize for beating you at all of those video games when we were younger, but that wouldn't have been fun now, would it? Thank you for being the one to bite the private Catholic undergrad institution bullet that the parents wanted one of us three to go through (even though IU is much superior). Thank you also for waiting a few months to get your first real-world job, so that I could win the race to employment as the eldest sibling.

This PhD is dedicated to an amazing woman, my maternal grandmother, Mary Dolan. To tell her story in brief, I tell the story of a woman who was ready to go to college, but denied at the last minute so her sister could go instead. She wanted her children to all have the chance to receive a quality post-high school education… and she succeeded. I am not the first PhD in my extended family, but I will never forget how fiercely she believed in my ability to succeed at whatever academic track I wanted to pursue. Her personality could be described by only one phrase - full of gratitude. I made my graduate school decision the same day that she passed away… and I find it only fitting that her name be attached to this.

To the many other aunts, uncles, and cousins - this list is by no means comprehensive, but does represent the majority that leave likes on my Facebook posts reasonably often: Mary and Kurt Angstman, Fran and Dana Miller (and their daughter Kirsten Campbell - I still owe you an apology for missing your wedding), Maggie and Greg Harp (and their children Elizabeth, Sara, and Eric), Pat and Tim Lawler (and children Sarah, Bridget, and Mike), and Dan and Betsy Hallberg (and their children Brandon, Brittney, and Kara). Your support means the world - I am especially grateful to my Aunt Margaret for (almost) always commenting when I'm posting about science - it's nice to have the reminder that people are reading and rooting for you even when they're as far afield as music is from biochemistry.

A super shoutout to James Angstman, my cousin on the East coast. I consider it a privilege to have a family member who is also a colleague, to not have to put much effort in remembering the name associated with "The Angstman Rules of CRISPR," and to remind me regularly that one is, in fact, "a number." To my cousin Beth Just, and her husband Thomas. Thank you for taking me into your home and letting me call it home as well all these times during my graduate career. And don't worry - Austin wins hands-down for food :) Also dogs are the best… pretty sure that that last week in Austin sold me on getting a dog once I'm graduated and everything.

To the teachers who taught me to find wonder in the world, but more importantly who taught me that it's nice to be important, but more important to be nice - with a special emphasis on two of them: Mrs. Merkel and Mrs. Dano. Also to my many teachers in music - the discipline that I learned from practicing (or learned that I had to develop from not having practiced!) is foundational to the discipline I had to cultivate throughout my PhD. Lissa McKinnis, I am so grateful to you after all of these years for trusting me to be one of your first flute students - know that if the walls of the Hammond Lab dark room could talk, they would either thank you for having taught me to play beautifully, or complain about how many hours I have spent doing long tones in there.

A special shoutout goes to Lynette Schneider - I didn't know for sure what I wanted to do until I got to your class. By the end, I knew I wanted to be a scientist. I am so grateful you let me engage in friendly competition with Ariel over who could finish our AP bio tests the fastest, and through four years of Quizbowl, I learned some form of professionalism in the competitive sphere - which turns out to be an important skill to have when one is competing with other professionals to be the first to get a publication, the first to get a grant, etc. Thank you for keeping us healthy with all of those cookies, sweets, and snack foods - that is how you keep a Quizbowl team fit, right? I am so grateful that I get to call you my science mom at home, and our yearly summer trips to watch professional tennis have been respites from some really difficult days. Thank you especially for helping me get Roger Federer's autograph, and helping me get a selfie with Grigor Dimitrov - what can I say, sometimes I need help from people to be brave.

To my church communities - who nurtured in me a faith that may not always stand on solid ground (what can I say, I struggle with doubts… which I think is a central tenet to a healthy faith life), but from which springs hope eternal. From my home parish of St. Thomas Aquinas in West Lafayette: Jane and Duane Sellers, Heather Cooper, Beth Berger, and the Kogers. From St. Paul's in Bloomington: Jill Switzer-Wolf, Annie Schroeder, her late mother Mary Jane Hanser, Claire Tafoya, Cindy Maceda-Berin, Kyle Barker, and my dear oboe friend and woodwind partner-in-crime of three years, Angela Quatrini (You have now been peacified!). Also to my St. Jarlath colleagues in music - thank you Dr. Robin Estrada for finding me again in Ben & Jerry's and reminding me that home was not that far away. Thank you Dr. Hiyas Hila for being an amazing partner to improvise with - pianists who one can improvise with on the spot live are rare to come by, and refreshing to play with. Thank you Paula for your weekly rides and being

willing to roll with my out-of-control schedule. Fathers Jim, Rich, Stan, and Bob - in a world where the pendulum of the church has seemed to move in a very conservative direction, and where many priests have cultivated a 'holier-than-thou' attitude, you have reminded me often that a central part about being Catholic is about fighting for what is right, not about looking holy while doing it.

To my friends I've picked up along the way here at Berkeley, from undergrad, and from high school who have kept me sane throughout this process, and who continue to accept me no matter how much I've changed from the person they met years ago: Carrie, Maria, and Stephanie - my first thesis draft sent to my committee was 165 pages… and yes, you were the first people I thought of. Rebekah Sims - promise me you'll never make that Buzzfeed quiz asking "Am I a hermit," because the answer to that question will always be "no." Rachel Cheeseman - thank you for having the best quotes possible on everything that is going on in this nation right now. Daniela and Adriana Giuliani - you are two incredibly strong and brave people who light up my life whenever I see you - know that wherever I go, it is home for you both as well. To Megann Rogers - Llamas. And also thanks for letting a boy one day out of his walking boot trek the entire length of Manhattan with you in a moments notice :) To parts of my IFLE family including Katie Mika,and (Dr.) Gabby Sell - I found my first community of research scientist-peers in you, and it has been humbling to see where all of us have ended up and the careers we have led up to this point. I also include here Alex Kerl - thank you for playing host to me whenever I end up in Seattle. I will, however, make sure that I watch Elf in your honor in the next week, with a nice up of hot cocoa by my side.

To the family that I have here at Berkeley - the Balewski family. It is an honor to call myself one of your four daughters, Ewa and Jan - thank you for always welcoming me into your home even though I'm terrible at keeping in touch and often very absent-minded. Ewa - you are one of the kindest people I know (I'm sure that your other daughters may disagree with me, but that's their loss) - thank you for your patience and for always being ready to put an extra plate of amazing Polish food out when I come by. Jan - I suspect that our scientific collaborations are only beginning. In the midst of grad school, I have found strength in seeing that, once all is said and done, one can still be as excited about science as you are. Zuzanna - thank you for choosing the best coast. I'm looking forward to being next door to you when I start my postdoc! And Weronika - you are one of my favorite people to make music with. I'm super excited for the Boston Early Music Festival :)

A sea-change in my life that occurred in graduate school was a pivot to focus on physical activity as a way to relieve stress, a mechanism to daily undertake (and succeed at) small, but doable, challenges, and make friends. The instructors who have enriched my body, my mind, and my spirit at some of these classes deserve special mention: Mitch and Noel, you both have taught me how to dance, and in the process, how to love myself doing it. Carli - 'Gorgeous' may not be the best of Taylor's new songs, but thank you for letting me sing along in Barré class as a way to not think about how my legs were literally on fire. Finally - Shola. The greatest personal challenges I faced in graduate school are the ones that came from your class. It turns out that none of those challenges were the weekly sets of burpees you give that I have grown to love. Thank you for giving me a biweekly home to remember that no matter what the world throws at me, I am #ToughNeverTired when I'm surrounded by people who lift me up. Thank you also to Rebecca Johnson of the Sarpong group for first introducing me to this amazing class almost five years ago.

Morgan and Erika (Adre) - I am so grateful that I met you two through Stronger… but more importantly, thank you for getting me to dance. If you are reading this acknowledgements

section, you probably know how self conscious I am in so many areas of life. One of those used to be dancing. You two taught me to love myself through dance. And more importantly, your caring personalities give me reason to dance. To Erika - thank you for teaching me to be brave. To Morgan - thank you for giving me the courage to go to Stronger on your last day - my first day back even though I was still in my walking boot. Thank you for bringing me back to one of the many places I can call home. Thank you also to Alexandra Courtois, the only one of my dancing buddies who is no longer gone. You are the critic whose opinion matters the most when I choreograph new things, because you care the most about how movement matches the music. Thank you to all of you for making me feel beautiful every Monday and Wednesday.

The “Lih”-nguists. How could I comprehensively thank everyone in graduate school if I didn’t thank you six? In my last year here, through you all I was reminded about how beautiful the world is outside of my own field - something I knew but hadn’t experienced in years. More importantly, as my field of mentors and peers have begun to dwindle (by that wonderful thing known as graduate), you have given me a new peer group where I find belonging. Andrew Cheng - you challenged me first to see how far I could push myself in the gym. Then you challenged me to see how far I could push myself to become stronger as a person. I will always look fondly upon our friendly rivalry at Stronger. Thank you. Tyler Lau - your smile is infectious, and if it looks like I’m angry when you say the way I pronounce something is “cute,” I’m actually just laughing inside. Thank you for trusting a biologist with no linguistic experience to test out the language phylogeny practice you made for your students. Eric Wilbanks - you work too hard. No, seriously, you work too hard. Try not to center those “C”s too often, and remember to take breaks once in a while to avoid burning out. Edwin Ko - I hope someday I’ll be able to convince you that your biggest critic is yourself, because you have both natural talent and a strong work ethic (I’m exceedingly jealous of the latter, because unlike you I cannot just “do work” for fun) - you are going to do amazing things. Lih Seng Goh - you are always optimistic, always supportive, and have made me experience more new things in the past year than most people on this list put together (Example for those who don’t understand just how large of a rock I lived under the past 5 years: I had never been to Dolores park until Lih entered my life). And Abel Ferrel - even though you’re from that other school down South, I would be stuck listening to linguistics 24/7 if you weren’t around. Thank you for being that other biologist in the room when you’re around so I don’t feel too weird on my own!

And also - none of you have played Set with me at game night. This has to change.

To the many scientists from my undergraduate career at IU who gave me the firm background required to succeed as a graduate student. Professors Silas Cook, Zachary Aron, and Kevin Brown - I am able to talk comfortably with synthetic chemists at Berkeley because of you three, and I think the foundational knowledge you gave me was what led to the opportunity to teach graduate physical organic chemistry in my third year. To Professor Kent Orr of the Math department - thank you for steering me towards that second degree in math; it ended up coming in handy throughout most of chapter 3. To Professor James Drummond - thank you for your advice over the years. It was the pinnacle of my graduate career to be labeled a “chemistry demigod” in front of a class of 30 fellow chemistry majors (much to Martha’s chagrin… I’m sure I had to be taken down a few notches after that!). To Professor David Giedroc, when I entered your lab as a freshman who had little to no clue how "real science" was performed, I was fully intent on going to medical school. Your research group (including two dear mentors, (Dr.) Sarah Keane and (Dr.) Alfredo Guerra) of four years ago changed that; having an open and friendly environment to talk about sciencewhet my scientific appetite and showed me that there were many options I hadn't yet considered. Namely research. I can confidently say that had I not

started research early, and (more to the point) had I not been exposed to your research group when I was, I would probably be a medical doctor right now.

To Professor Martha Oakley - I cannot thank you enough for the opportunity to work in your lab on bacterial chromosome condensation. You were the first person who trusted me with a project that was, to some degree, my own. You are truly a testament to how amazing scientists can also be stellar human beings. Thank you for giving me the chance to play in the big leagues. Thank you also to (Dr.) Nichole Stewart - I learned how to clone, make competent cells, perform experiments… and how to make sure your transformations work (using voodoo!) from you.

Most people begin graduate school having to rebuild their friend groups. I was lucky enough that not one, but two of my best friends started graduate school with me. Erika Anderson, my eternal running buddy and, moreover, confused comrade in biology. Thank you for asking me if I was that Zach who played the piccolo all those years ago. Through you I didn't just get a friend to complain about epsilonics with, you also introduced me to amazing people like Kate Sanders. I owe the both of you lifetime supplies of gratitude for the many nights agonizing over my writing (Ming, if you're reading this, the worst of my writing you have read pales in comparison to what they have edited for me!). Kimberly Long (aka Korra, my past life) - every time I eat cheese, I think of you and the theme song you wrote for me. There is no one I know who can push me harder on a track, and no one whose roast chicken I would rather eat. You were the first person who showed me that I can, in fact, be tough never tired, even if that wasn't the mantra you developed. Thank you to both of you, Erika and Kim, for keeping me alive through my first year - there were some days that first year that I would not have eaten had you not made me. You are two amazing human beings, and more importantly still some of the best scientists I have seen - I cannot wait to see you both do crazy-awesome things after your doctorates. Thank you for walking this journey of life with me for the past 9 years.

I started graduate school in the Sarpong lab - and I have Professor Richmond Sarpong to thank for helping me make my decision to come to Berkeley (you still owe me a tennis match… and as I will be here for a few more years I intend fully to collect on this). I also have a body of mentors to thank from this lab:

(Dr.) Jessica (the Awesome) Kisunzu - before you became an amazing professor at Colorado College, and before you went abroad to Switzerland to study abroad with Helma, you were here, at Berkeley, in our dear Latimer 844. Thank you for being my mentor in the ways of synthesis - from teaching me how to dry a round-bottom flask to training me on pyrophorics - while I may never succeed at total synthesis, your mentorship - and the mentorship I received from your undergrad Kyle Claag - gives me confidence that if I ever need to make something myself, I have the training necessary to do so. I am so thankful that you took my "Best Days Ever" seriously enough to start counting them, and have always been that gentle reminder that, "Even when we can't imagine the ending or way out, it's not forgotten. There's a way."

(Dr.) Sidney Hill - I remember wanting to be just like you when I joined the Sarpong lab. You were always so methodical, so knowledgable, and so in love with science (and - bonus - had fashion sense!). Of course, at some point I needed to admit to myself that I was never going to be type-A enough to be you, but that doesn't mean I have not continued to look up to you any less. I'm so excited to have had the opportunity to witness your life story in real time and… someday… let's talk Vitamin B12 ;) Also, I should mention - of all of my friends whenever I ask how things are going, you always want to switch to talking via the phone. That means a lot.

(Dr.) Paul Leger, my cohort-mate in the Sarpong group - Our (sometimes daily) boba dates have been a huge source of strength for me - I always got at least one chance when I needed to reset

my emotions… and you know me well enough to know that I often needed that reset. Thank you for letting me be a part of your life - your patience and ability to listen without judgement is something I need to keep developing. Looking forward to seeing what you'll make me do next… like in May in Santa Clara (…are you ready for it? ;)) I will add in here my thanks to Nikki Doering - you have more than earned your place as my newest boba buddy. Your insights into the world come from a completely different perspective than the one I'm used to, and for the perspective you have given me on things from politics to Taylor Swift, I am grateful. Thank you also to your fiancée Grace for putting up with our boba antics, and good luck to her in her startup endeavors - I can't wait to see how Garden grows!

Next, the people in other labs who are much more knowledgeable than I am about a great many things. (Dr.) Omer Ad, Jase Gehrig, and Vivian Yu from Michelle Chang's lab all donated significant amounts of time (VY taught me how to Western Blot) and/or effort (OA did hella mass spec) and/or reagents (JG gave me a negative control plasmid for an experiment). In addition, you have given me a place next door to the lab to find friends. Thank you especially to Omer - for teaching me the art of the TRX, for our weekly bagel mornings, and for literally giving me a shoulder to cry on when we finally found cAG in *Myxococcus*. You're going to do amazing things with Professor Schepartz.

Our lab is not a protein lab, but the Doudna lab is: (Dr.) James Nuñez taught me how to purify proteins, but more importantly was sitting in the right place at the right time - right next to Dr. Philip Kranzusch, who taught me a lot about protein crystallography and who along with (Dr.) Kevin Doxzen were instrumental in helping us get the first crystal structure of a HyPr GGDEF. (Dr.) Megan Hochstrasser taught me how to perform ITC, and then did one better by trying to fit my curves before I came upstairs to check the results. All of you are amazing scientists, and I am so grateful I got to have you as mentors, if only for short whiles.

One Doudna-ite gets special mention - (Dr.) Spencer Knight. I can only imagine that you thought I was crazy when you got an e-mail six years ago that read something along the lines of "Hi. You seem sane. Are you going to Berkeley? Want to be roommates?" Despite House-pocalypse (honestly, we probably should have stayed…), going through graduate school with you has a friend was all worth it. The status quo in chemistry is that if you switch groups, you're a failure. Thank you for helping me prove all the haters (that are gonna hate, hate, hate, hate, hate) wrong, and for reminding me that everyone gets out of the woods. I'm looking forward to seeing where all those data science skills you're acquiring are going to take you… but know that I'm still secretly pulling for you to end up back in academia (so that I can pit my future 'Golden Boy' against yours).

Thanks to my kids. And by kids, I mean the now-fourth years that I GSI'd for. You all (I don't have room to list all of you) inspire me to keep doing science and to keep teaching pretty much every day. I know that the science world, and more important the academic world, are doing everything they can to dash your spirits - whether it's failed reactions, departmental politics, or work-life balance. Please don't give up on science - I see you all and I know just how bright you can all shine. The opportunity to walk with all of you during the early years of your graduate education is worth more to me than any paper, and has made me a much better scientist because of it. Special shoutouts to Julia Lazzari-Dean, Kristin Wucherer, Rachel Bisiewicz, Helen Hobbs, Tracey Hinder, Effie Zhou, Emily Hartman, Alec Christian, Also Robert Nichols, Sean Higgins, Emily Powers, Tom Laughlin, and Lynda Truong: my life may not be that different if you guys hadn't entered it, but darn it would be a lot less fun.

There comes a point where there are left only a few people who have been around as long as or longer than you have been and who are infinitely wiser than you are. To Dr. Ariel Furst, Allegra Aron, Katie Klymko, and (Dr.) Michelle Bloom, my thanks. Ariel, our morning coffee dates have always given me hope to start off the day. You are going to be an amazing professor wherever you are at. Thank you for being my mentor, even if my constant dragging you down the hill to the RSF is the reason you can't feel your legs (I currently can't feel my quads, so we are even). Allegra and Katie - you both are always able to find positive things in the midst of a department that likes to focus on the negatives - thank you for always finding a way to cheer me up even on the worst of days. I am so thankful to have someone with your indefatigable spirit as a friend in graduate school, and - to Allegra, I look forward to seeing you around as a fellow member of the microbiome community. Michelle, we proved that if we can do 4 minutes of burpees, we can, in fact, do anything. Thank you for letting me be your cheerleader while you finished your thesis, and thank you for being my cheerleader while I finished mine. Those spot checks of chapters and daily reminders that I am not, in fact, insane have meant the world.

(Dr.) Cheri Ackerman - I knew that you would be an important part of my graduate school life - I just didn't know how until I became a 6th year. You are, without a doubt, one of the most gifted individuals I know. Thank you for teaching me the art of empathy. Thank you for making space for my emotions. Thank you for holding me to the scientific standards that only a Bergman GSI could hold someone to. Thank you so much for your friendship and for being, both emotionally and scientifically, my other half this past year. I look forward to having you as a colleague someday, and to holding our future students to the fire… nicely :) Thank you also for giving me wonderful friends like Lindsey Osimiri - I will always remember the importance of crispy M&Ms in maintaining emotional health because of you.

To the many wonderful professors and lecturers I have had the opportunity to work with or learn from over the past five and a half years: Professor Robert Bergman - the moment I took your physical organic chemistry class, the one thing I wanted was to be able to teach that class. And I am so grateful that you placed your trust for teaching first year graduate students in me and my colleague (Dr.) Rebecca Triano. In learning how to teach phys org, you taught me so much about mentoring and about science. Professors Carolyn Bertozzi and Thomas Maimone - thank you for not completely failing me during my qual despite the fact that I probably deserved to be held even more to the fire. Professor Matthew Francis - part of me still never wants to forgive you for your response when I demanded halfway through my qual to know why I wasn't being asked about physical organic chemistry ("Because we know you know it") - and the another part of me is a Berkeley grad student who loves to be challenged. Professor Russell Vance - thank you for glossing over my not having known that mechanism paper… it taught me the ever-important lesson of "just because you don't like the professor who wrote it doesn't mean you shouldn't read the paper."

Finally, I would like to thank the most important people in the making of this thesis - my lab.

(Dr.) Scott Hickey and (Dr.) Colleen Kellenberger were the first two brave souls to start graduate work in the Hammond lab. While my work does not directly intersect with the main body of Scott's work, he was instrumental in getting the cAG riboswitch story off the ground, and - more importantly - proving that cAG was a natural signaling molecule in Geobacter. A primary burden in the first few months of my project to find the synthase was that we had no reproducible evidence of this signaling molecule in Geobacter, and I am forever grateful to Scott for both putting up with my constant doubts and for getting the proof I so desperately wanted. Will let you know if and when the October Project ever finishes.

Colleen was the first person to develop an RNA-Based Fluorescent sensor (RBFs) for cyclic di-GMP that worked in bacteria. This was followed up over the course of two and a half years by the development of RBFs for the other two bacterial CDNs. My thesis project was inspired directly by her work as a mechanism by which we could prove that RBFs were mature enough to be used in discovery. And, of course, there would be no cAG synthase to find had she not helped to craft the cAG riboswitch discovery along with (Dr.) Steve Wilson. I am so grateful to you, Colleen, for letting me do that first gel extraction in grad school, that set my trajectory back to the world of biochemistry.

Thank you (Dr.) Tania Gonzalez. Tania took the splicing cassette that was first discovered by Ming and adapted by Scott into a generalizable method to regulate tight gene expression in plants. But more than that, you were the older student I needed to vent to in the early morning when I was frustrated with what it took to publish. You were the person who introduced me to the most amazing drink (after possibly boba), the Horchata. And most importantly, you were gentle in reminding me of ways that I could grow further in seeing and experiencing other perspectives. I am so grateful to you for the time and patience you showed with me. I am also grateful to your undergraduate Bao Nguyen - Bao, you are going to be an amazing doctor someday. Thank you for believing in me - you were right, I am braver than I thought I was.

Thank you (Dr.) Cindy Wang, my bay-nemy. Of everyone I have known, you have come the closest to “figuring out” how to do graduate school right. The balance between work and life, how not to get too invested while remaining pleasantly optimistic, and most importantly the knowledge that no day in lab is wasted as long as a master mix is involved. I learned from you that efficiency was not a crime, it was a way to make time for other things that mattered. From my second to my fourth years you were my compass for sanity, my mentor in how to live, and the daily reminder that I needed that none of this was about me. And don’t worry - the RBF will not be forgotten :) Thank you also to your two undergraduates, Wanda Thi and Cindy (the Younger) Lam - not only were you two instrumental in making YhjH a thing that the lab used more often (which was very important for developing our phosphodiesterase screen!), as work study students you went above and beyond in both keeping us well-stocked, and the lab super clean.

Soon-to-be Dr. Yichi Su - I will always remember how you remain to this day the only human being who has asked our adviser to “push you harder,” against the advice of two older graduate students being given mere minutes before that happened. I will always remember our late night Ke$ha dance sessions that were interrupted by a certain Cindy Wang waltzing in at midnight and asking “why are you still here?” You are, without a doubt, one of the most talented scientists I have gotten to work with. Remember that your biggest critic will always be yourself. Also to your undergraduate Xavier Aguilar-Enriquez. You honor me by coming to me which Yichi is not around. You are a talented youngster who will do well wherever you go - I cannot wait to see what grad school has in store for you! Also, since I’m a Dr. I can officially weigh in on this and win - Zachary’s is better… except maybe for that Little Star meatball pizza.

Thank you Todd Wright - my partner in crime on Team HyPr. I am eternally grateful that we added you to both the group and to my project when we did - you kept me sane through much of the process of publishing a paper, doing research, writing a thesis… ok pretty much all of grad school starting at third year to the end. I believe in my introduction to your GRS I once joked that you were “boring,” but I see often that you are measured with your emotions, something that I could learn a bit more of. I can think of no one more capable or deserving of being the next to lead the HyPr saga, than a graduate student who found both the first small molecule activator for a HyPr, the first validated phosphodiesterase for cAG in a

deltaproteobacterium, and who is brave enough to continue the *Myxococcus* story. To your undergrad Lucy Jiang I also give thanks - you have been a wonderful baymate this year, and I look forward to choreographing more cardio “dance” with you!

Johnny Truong - I will always remember the first time I met you at my first Chem 200 office hours - you came to ask a question about thermodynamics… which quickly devolved into asking me questions about the Hammond Lab. Ever since you have joined the group, our lab has in many ways benefitted from your role as “lab mom” - there always seems to be new snacks on the table, and you’re a major reason that the place has remained sanitary enough for humans to consume food anywhere near the group room table. You’ve had a challenging project trying to find a ribozyme in a world where there might not be one - but from what I’ve seen the professionalism by which you approach your interpersonal interactions (which I can and should continue to learn from) will serve you well in the future, no matter what you choose to pursue. Thank you for being part of the lab family, thank you personally for being part of QC, and thank you for all the HotS advice!

Rebekah Kitto - Operation JELLO wouldn’t have had its most important rule (we should also eat food multiple points during the operation) had you not helped me come up with it. I am so grateful that you joined the lab. More importantly, I am so grateful that you taught me how to volley - we really should do that more often :) Despite the challenges of having to balance projects both up and down the hill (and both on Earth an in outer space!), you have matured into an amazing scientist and colleague. Thank you also for being willing to take the blame for eating all the food, when we all know that I eat just as much!

Andrew Dippel - Not going to lie - the lab pretty much started rioting when Rebekah told us which group you were going to join. When I started in the Hammond Lab, Cindy’s project with GFRNP was starting to wind down - I never thought someone would bring fluorescent proteins back into the lab. Then you came along - in many ways beyond taking over the fluorescent protein side of the lab, you have become the next Cindy in lab - calm, measured, efficient. You have crafted a beautiful story on making luminescent sensors for cyclic dinucleotides. Thank you also for your advice on who to see at ACL - who knew that EDM with woodwinds could sound good?! Also thank you to your undergraduate Wyatt Anderson - our first grandmaster Overwatch player in the lab.

Thank you Jade Sales-Lee: you were one of my first confidantes in the lab, and a constant reminder to practice more self care than I was administering to myself in my first year (can someone say baking is an amazing form of stress relief?). I am also thankful that you ended up pretty close by - sorry I ended up disappointing your hopes of having even more friends on your floor in Genentech Hall, but also it sounds like you have a story that’s going to be finishing in the next few months, so I’m not as sorry as I would be if you had multiple years left. Thank you for supporting me through the difficult decision of finding a postdoc lab. Thank you also for your amazing younger sister, Quin, who was a stellar roommate for eight months, three of which had me with a broken foot. Thank you for taking are of me during that time - it’s a reminder that the simple things (like helping a person in crutches transport food to a table) are the most important.

To the first and last undergraduates that I got to train - Ryan Muller and James Park. Mentoring both of you has been a delight in my scientific career, and has served as a daily reminder that becoming an academic scientist is what I want to do. Ryan is now a PhD candidate in his own right here at Berkeley, and has never been too busy to talk science with me… even at 3 am. James - thank you for seeing me through the most stressful period of my PhD - I am sorry for all of the balls I have dropped in being a good mentor I could be. That being said, you never

ceased to surprise me by rising to the occasion. You're going to do amazing science things one day, because you already are a fantastic scientist.

Thank you to Susan Meux, the lab administrator. There are many (emphasis on many) times that I have dropped the ball. Susan has always always always come to my rescue. Whether it was getting filing fee to work out, making sure that the reagents I needed came in on time, or simply dropping off baked goods for the lab on holidays, I will always be grateful that Susan had our backs when we didn't even know our backs needed having!

Finally, Ming. I consider myself incredibly lucky to have found my way to your research group, and even more lucky to have had the opportunity to see your lab grow and blossom into an established research program. I know that I am likely the strongest personality that you have mentored - I thank you for your endless patience in that regard. I started in your lab thinking that the most important metric a scientist could be judged on is how cool their results are. I was wrong: you have taught me that the most important metric is character. That wasn't the only way I was wrong - I think I'm currently on the losing side of our many scientific bets over the years, although I suspect that measured over the past year my win/loss ratio is closer to 1:1 (which probably means I should graduate or something). Thank you for trusting me with discovering a new signaling pathway in a super cool organism. More important - thank you for being willing to go into the arena with me. Thank you for giving me access to amazing collaborators throughout the world - from my parents' home state of Minnesota (Professor Daniel Bond and Dr. Chi Ho Chan), to Texas (Dr. Beiyan Nan), to the United Kingdom (Professor Elizabeth Sockett and Dr. Rebecca Lowry), to our own backyard (Professor Jennifer Doudna and Dr. Philip Kranzusch). You put most professors at Berkeley to shame with your willingness to go up to bat for your students.

It has been the honor of a lifetime to call myself your student.

Chapter One

Specificity and Activation of Dinucleotide Cyclases

Abstract

Organisms are constantly bombarded with a variety of changing environmental conditions, including fluctuations in nutrient levels, oxidative stress, surfaces, and the presence of other organisms that play roles as both friends and foes. To adapt to these changing conditions, organisms have sensory systems that allow rapid responses to environmental cues. One major strategy that has been developed for intracellular signaling in response to extracellular changes is the use of second messengers. Second messengers are characterized by a control module where, in response to a palette of primary signals, synthase and hydrolase networks control the intracellular levels of a soluble small molecule, which can bind to various effectors, leading to a physiological response.

Cyclic dinucleotides (CDNs) form an important class of second messenger in bacteria, regulating a variety of cellular processes including division, cell wall homeostasis, and pathogenesis. For over thirty years, only one CDN, cyclic di-GMP (cdiG) was known, and many components of its signaling network in bacteria have been described. However, the past decade has seen the number of CDNs quadruple, with the discovery of cyclic di-AMP (cdiA) in 2008, as well as two flavors of cyclic AMP-GMP, characterized by the phosphate linkage - 3',3'-cyclic AMP-GMP (cAG) in bacteria, and 2',3'-cyclic AMP-GMP (2',3'-cAG) in mammalian cells. While the synthase, hydrolases, and effectors for cdiA are known, no cAG network has been fully mapped in bacteria.

In this chapter, we introduce the known CDN synthases, the diguanylate cyclase harboring the catalytic GGDEF domain, the diadenylate cyclase (DAC), the first bacterial cAG synthase DncV, and the mammalian 2',3'-cAG synthase cGAS. We then focus on their mechanistic determinants, with a focus on how they form their specific products. Finally, we highlight the structural basis for regulation of synthase activity, as well as the known primary signals for dinucleotide cyclase activation.

Introduction to Cyclic Dinucleotides

Cyclic dinucleotides (CDNs) are small, soluble signaling molecules that are produced as an intracellular response to a primary - typically extracellular - signal. These signaling molecules are central to pathways typically characterized by three parts that regulate and respond to the messenger (Figure 1). Synthase enzymes respond to input signals by producing a specific CDN, which accumulates in the cell. When present, the CDN binds to various effector molecules, thus producing an output response, such as activating a transcription factor or enzymatic activity. The transient nature of the signal requires that, in response to changing environmental conditions, the signal is broken down by phosphodiesterases that are either constitutively active or regulated by a different signal. This basic theme can give rise to multiple variations that allow for more complex signaling systems (Romling et al., 2013).

From 1987-2007, only one CDN was described in the literature, cyclic di-GMP (cdiG, Figure 2) (Romling et al., 2013; Ross et al., 1987). In this intervening time, cdiG has been revealed as a major force in bacterial physiology. Broadly, this CDN controls, the transition from the motile, planktonic stage of bacterial lifestyle to sessile, biofilm-forming, community states (Jenal et al., 2017; Romling et al., 2013). For example, cdiG-regulated phenotypes are involved in bacterial pathogenesis, antibiotic resistance, the production of fermented food products (such as cheese, kombucha, and sourdoughs), and play roles in both environmental bioremediation and contamination (Coda et al., 2014; De Roos and De Vuyst, 2017; Edwards and Kjellerup, 2013; Wolfe et al., 2014). In addition, the protein family implicated in cdiG synthesis, the GGDEF domain, has been well-characterized, allowing prediction for cdiG signaling occurring in a bacterium assuming its genome has been sequenced (Ryjenkov et al., 2005; Seshasayee et al., 2010).

Research completed during the past decade has begun to challenge the sovereignty of cdiG as the sole CDN important to bacterial lifestyles. Two new cyclic CDNs, cyclic di-AMP (cdiA) and 3',3'-cyclic AMP-GMP (cAG), were first described in 2008 and 2012, respectively (Figure 2) (Davies et al., 2012; Witte et al., 2008). Additionally, the CDN world experienced a dual paradigm shift with the discovery of a non-canonical CDN native in the mammalian innate immune response, 2',3'-cAG (2',3'-cAG) (Ablasser et al., 2013; Diner et al., 2013; Gao et al., 2013; Sun et al., 2013; Wu et al., 2013).

In the intervening time, a number of studies have elucidated multiple components and physiological implications of both the cdiA and 2',3'-cAG pathways. Cyclic di-AMP plays, broadly, multiple roles in cell wall homeostasis, ion channel conductance, and sporulation (Corrigan and Grundling, 2013). Similarly, 2',3'-cAG has been implicated in both viral and bacterial infections, and risen to prominence as a possible target in cancer immunotherapy (Hornung et al., 2014). In contrast, knowledge of 3',3'-cAG signaling has lagged behind, as cAG has been discovered in organisms for which no synthase has been described (Kellenberger et al., 2015b). The presence of multiple nucleotide second messengers in bacteria leads to many important questions: What techniques have been used to discover CDN synthases, and how prevalent are these synthases in the phylogenetic tree? How do these synthases select the appropriate

substrates to produce the correct CDN product? Finally, what are the primary signals for CDN synthesis, and how do these synthases respond to them?

I - Discovery and Phylogeny of CDN Synthases

I.A - The GGDEF Domain- Making cdiG

The GGDEF domain, originally named DUF1, is responsible for synthesis of cyclic di-GMP (Ausmees et al., 2001; Ryjenkov et al., 2005). Cyclic di-GMP was discovered by Benziman and coworkers in the bacterium *Gluconacetobacter xylinus*, where they biochemically isolated the small molecule as an activator of cellulose synthase, and purified protein fractions containing diguanylate cyclase (DGC) activity (Ross et al., 1987). However, proteomics had not yet reached the stage required to localize the domain that contained this enzymatic activity. Thus, unlike other CDNs, cdiG was a known bacterial second messenger before its synthase was known. Further genetic characterization revealed that this diguanylate cyclase (DGC) activity was found in genes containing both GGDEF and EAL domains, characterized by conserved regions containing these amino acid motifs (Tal et al., 1998). All three characterized DGCs and all three characterized cdiG-specific PDEs from *G. xylinus* were hybrid enzymes containing both GGDEF and EAL domains, preventing full assignment of activity to a specific domain. Further bioinformatics work demonstrated that the GGDEF domain was homologous to the adenylate cyclase domain, and that the GGDEF was the only common element between DGCs characterized in the period from 1987 to 2001 (Ausmees et al., 2001; Pei and Grishin, 2001). Finally, Ryjenkov and coworkers purified six GGDEF domain-containing proteins from diverse phylogenetic branches and demonstrated definitively that all six purified proteins possessed robust DGC activity, and that two isolated GGDEF domains had low yet detectable DGC activity (Ryjenkov et al., 2005). With this, further characterization of GGDEF domains often assumes that only DGC activity is possible, with enzymes being assayed against only GTP for activity (Hobley et al., 2012; Skotnicka et al., 2015).

With the validation of these phylogenetically diverse GGDEFs, it became clear that the presence of cyclic di-GMP could be correlated with the presence of one or more GGDEF domains in an organism. GGDEF domains are primarily found in bacteria, where they are present across many taxa. Of the 1113 representative complete bacterial genomes available from the RefSeq database by the end of 2011, 728 (65%) contain at least one GGDEF domain (Chou and Galperin, 2016; Romling et al., 2013). The domain is only found in one archaeal species. In addition, cdiG appears absent in Bacteroidetes and Chlamydiae groups, which may be due to these species' specific environmental requirements. More recent bioinformatics data has borne out this expected broad distribution - 2394 species from the 4321 curated by Pfam have GGDEF domains (~55%) (Finn et al., 2016). While no members of the candidate phyla radiation (Hug et al., 2016a) have GGDEFs as determined by BLAST searches, a number of candidate phyla, including the Rokubacteria (Hug et al., 2016b), Wallbacteria, Schekmanbacteria, and Eisenbacteria (Anantharaman et al., 2016) contain multiple GGDEF proteins, suggesting that cdiG signaling is also active in many of these mysterious, yet prevalent, microbes. Finally, a small subset of eukaryotes, the slime molds, contain GGDEF

domains, which have DGC activity and are important to cellular differentiation (Chen and Schaap, 2012).

Organisms with GGDEF domains vary in the number of genes encoding GGDEF domain in their genomes (Seshasayee et al., 2010). In particular, some genera of intracellular pathogens (e.g. *Borellia* and *Rickettsia*) have only one gene encoding a GGDEF protein (Figure 3). In contrast, many actinomycetes and proteobacterial classes have an expanded arsenal of up to 102 GGDEF domain-containing proteins (in Actinoplanes sp. SE50) (Seshasayee et al., 2010). Importantly, these numbers do not appear to correlate with genome size.

I.B - The DAC Domain - Making cdiA

In contrast to the GGDEF domain, the diadenylate cyclase (DAC) domain, was assigned function concurrent with the discovery of its product, cyclic di-AMP (cdiA, Figure 2). In *Bacillus subtilis*, DisA acts in a DNA-damage checkpoint during sporulation (Bejerano-Sagie et al., 2006). Previous work demonstrated that it “scans” DNA for double-stranded breaks (DSBs) (Bejerano-Sagie et al., 2006). Structural characterization of the protein revealed cdiA present in the active site of protein crystals (Witte et al., 2008). Nucleotide-free enzyme preparations readily converted ATP to cdiA, and this activity rapidly dropped in response to incubation with branched DNA, which is an expected product of stalled DNA replication induced by DSBs. Further characterization has demonstrated a wide phylogeny of bacteria containing DAC enzymes (Corrigan and Grundling, 2013).

In comparison to the GGDEF domain, the DAC domain is less prevalent. 1712 bacterial species (39%) from the Pfam database contain at least one DAC domain (also known as the DisA_N domain) (Finn et al., 2016). A wide variety of DAC domains are present in Archaea, with 74 (of 235) species containing at least one DAC domain. The DAC from *Methanocaldococcus jannaschii* has been validated for DAC activity, as the first demonstration of CDN signaling in this domain of life (Kellenberger et al., 2015a). A few eukaryotic species, specifically choanoflagellates, harbor a DAC domain; however further validation is required to determine if cdiA exists in these organisms.

The broad expansion of GGDEFs in bacteria is not observed with the DAC domain - most species contain only one or two DAC domains (2792 sequences across 2325 species). Few proteobacteria have a DAC protein (87 of 1907 species total, compared to 1107 species containing at least one GGDEF). Thus, a fundamental difference between cdiA signaling and cdiG signaling is that, in many cases, fewer proteins - and by proxy input signals - are involved in regulating the CDN pool in the cell.

I.C - DncV - The First Hybrid CDN

Similar to the discovery of the DAC domain, 3’,3’-cAG (cAG, Figure 2) was discovered concurrently in bacteria with the first characterized cAG synthase, DncV, previously known as VC0179 (Davies et al., 2012). VC0179 was originally identified as being regulated by the virulence transcription factor VC0177, another gene on the *V. cholerae* 7th pandemic genomic island, through use of a transposon knockout library. Disruption of VC0179 caused a defect in the ability for the pathogen to colonize mouse intestines

(Davies et al., 2012). Homology analysis suggested that VC0179 was a member of the nucleotidyltransferase superfamily, which led to *in vitro* experiments demonstrating that it forms cdiG or cdiA products when assayed against GTP or ATP, respectively. Incubation with both GTP and ATP, however, led to robust consumption of both nucleotides to form the heterodimeric product. Further analysis of cell extracts from WT and Δ*dncV V. cholerae* strains showed that DncV primarily produced cAG *in vivo*, and triggers an unrelated physiological response to that expected from cdiG (Davies et al., 2012).

DncV homologues are the rarest of the CDN synthases in bacteria; a simple BLAST search gives only 308 homologues; indeed, these homologues are spread randomly primarily throughout the proteobacterial phylogeny - most representatives are in the gamma- and beta-proteobacterial phyla, however 5 *Bacillus* species contain a DncV homologue. We therefore expect that many of these examples are due to horizontal gene transfer, and not due to conserved cAG signaling in these organisms. However, the presence of specific cAG phosphodiesterases outside of the Vibrio pathogenicity island, combined with the discovery of riboswitches that specifically bind to and recognize cAG in bacteria for which no DncV homologue exists, suggest yet undiscovered cAG signaling pathways (Gao et al., 2015; Kellenberger et al., 2015b; Nelson et al., 2015).

I.D - cGAS - CDN Signaling in Mammalian Cells

The discovery of the mammalian cAG (2',3'-cAG, Figure 2) synthase and characterization of its activity occurred months before its product was correctly assigned. Previously, it was known that mammalian cells initiate a rapid immune response to DNA located in the cytosol, usually from infection; however, the mechanism by which this cytosolic DNA was sensed remained unknown (Barbalat et al., 2011). Initial work suggested that the elusive cytosolic DNA sensor produced 3',3'-cAG (Sun et al., 2013; Wu et al., 2013). A set of three orthogonal fractionation schemes led to only three potential candidate synthase enzymes (Sun et al., 2013). Bioinformatics analysis showed that one of the three previously uncharacterized proteins belonged to the nucleotidyltransferase superfamily. Subsequent biochemical analysis demonstrated that this enzyme, cyclic GMP-AMP synthase (cGAS) was capable of producing a cyclic dinucleotide that was originally characterized as 3',3'-cAG (Sun et al., 2013; Wu et al., 2013). However, multiple reports called this discovery into question: the cGAS product is not degraded into GMP and AMP by RNAse T1, while 3',3'-cAG is (Ablasser et al., 2013). Furthermore, some alleles of STING, the cytosolic sensor of the cGAS product, were insensitive to cAG, but remained active against the cGAS product (Diner et al., 2013). A combination of X-ray crystallographic analyses of the cGAS protein, thin-layer chromatography using radiolabeled NTP analogues, and NMR and tandem MS/MS structural characterization of the cGAS product conclusively demonstrated that this product contained a unique 2',5'-phosphodiester linkage (Ablasser et al., 2013; Diner et al., 2013; Gao et al., 2013). This linkage is similar to the 2',5'-linear oligoadenylate nucleotide produced by the close structural homologue of cGAS, oligoadenylate synthase (OAS) (Hornung et al., 2014).

The Mab21 domain, which is the family associated with cGAS activity, is widespread throughout the eukaryotic domain, in particular in metazoans (Kranzusch et al., 2013; Wu et al., 2014). Of 288 metazoan species, 223 contain at least one Mab21 domain (Finn et al., 2016). Analogous to the GGDEF domain in proteobacteria, most metazoans contain a high number of Mab21 domains, averaging 5.7 per genome, with humans possessing 14. As only one of these Mab21 domains have been enzymatically characterized, it remains to be seen what role these other proteins play in mammalian signaling.

II - Structural Aspects of Nucleotide-based Second Messenger Enzymes

Synthesis of almost all nucleotide-based second messengers require phosphodiester bond formation, typically between the 3' hydroxyl group and the alpha-phosphate of a 5' nucleotide triphosphate. A conserved catalytic fold acting on a conserved ribonucleotide triphosphate substrate core appears to have diverged into many different subfamilies capable of synthesis of diverse second messengers (Sinha and Sprang, 2006). Indeed, most bacterial nucleotide synthases are part of the nucleotidyl-transferase superfamily, which consists of a conserved beta sheet surrounded by alpha helices (Sinha and Sprang, 2006). This central beta sheet typically scaffolds acidic residues that coordinate one or more magnesium ions that activate the triphosphate backbone. We discuss here the structural determinants that control this conserved catalytic mechanism in each second messenger synthase, as well as the specific residues that are involved in substrate discrimination.

II.A - General Features of Dinucleotide Cyclases

All nucleotidyltransferase enzymes must overcome three challenges to synthesize the proper second messenger. First, the nucleobase must be recognized in a specific manner. Second, the triphosphate region must be recognized and activated as an electrophile for nucleophilic attack. Third, the incipient hydroxyl group must be activated.

Purine recognition typically occurs through two modes: direct hydrogen-bonding interactions with the nucleobase "edges," and pi-pi and cation-pi stacking interactions. Given that hydrogen bond strengths vary between 1-6 kcal mol^{-1} (Nick Pace et al., 2014), we expect that a judicious choice of edge interactions with either the Watson-Crick (W-C) or Hoogsteen faces would be strong enough to distinguish (i.e. selectivity >1000-fold, or ~4.2 kcal mol^{-1}) between adenine and guanine residues. In many cases, adenine is recognized through a hydrogen bond donor (such as serine) interacting with the W-C face nitrogen, while a hydrogen bond acceptor such as aspartate or glutamate is used to interact with guanine (Chan et al., 2004; Gao et al., 2013; Kranzusch et al., 2014; Witte et al., 2008). These interactions are by no means limited to the amino acid side chains, as backbone amides have been observed to act as either H-bond donors or acceptors (Witte et al., 2008). Furthermore, steric occlusion can allow preferential selection of adenine by blocking the exocyclic amine group in guanine from binding. Pi stacking of the nitrogen base also plays a crucial role in recognition - both cation-pi and pi-pi interactions abound in synthase nucleotide binding pockets.

Catalysis of the nucleotidyltransferase reaction requires activation of both the nucleophilic hydroxyl group and the electrophilic alpha phosphate. The nucleophilic group is typically activated through deprotonation - in nucleotidyltransferases this is accomplished through acidic aspartate or glutamate residues (Sinha and Sprang, 2006) (Kranzusch et al., 2014; Witte et al., 2008). In contrast, activation of the alpha phosphate occurs through multiple mechanisms. Positively charged residues are typically arrayed around the phosphates (Chan et al., 2004). In all cases, however, there is a conserved magnesium cation coordinated by multiple acidic aspartate residues - indeed, in many nucleotidyltransferases (such as cGAS and OAS), formation of this catalytic array that coordinates the Mg^{2+} cation is an essential component to activation (Civril et al., 2013; Donovan et al., 2013).

Although one might assume that a final mode of recognition involves discrimination between ribonucleotides and deoxyribonucleotides, there is little biochemical evidence for evolution of specificity in this domain. Diguanylate cyclases have been shown to readily accept both deoxynucleotide triphosphates as well as 9-ß-D-arabinofuranosyl-guanine-5'-triphosphate (Shchokolova et al., 2015). Indeed, the concentration difference between nucleotide triphosphate and deoxynucleotide triphosphate suggests lack of an evolutionary pressure: GTP and ATP typically exist in 100-fold excess above their 2'-deoxy counterparts in bacterial cells (Bennett et al., 2009).

II.B - Structural Aspects of Characterized Dinucleotide Cyclases

The GGDEF and DAC domains possess well characterized C_2-symmetric, homodimeric active sites that synthesize their homodimeric products (Chan et al., 2004; Witte et al., 2008). With highly conserved nucleobase, phosphate, and in the case of the DAC domain, hydroxyl group contacting residues, their catalytic mechanisms and origins of substrate selectivity are well-understood.

In the case of the GGDEF domain, the monomeric fold with topology βααββαβαβ generates an array of contacts that positions the GTP substrate above the signature "GGDEF" motif for which the domain is named (Chan et al., 2004). This motif permits little flexibility for the enzyme to remain active - characterized active enzymes have a [G/A/S]G[D/E]EF motif (Romling et al., 2013). The first two residues of this motif sit directly below the guanosine base, which explains the requirement for small (or nonexistent) side chains at these positions (Hunter et al., 2014). The first acidic residue, either an aspartate or glutamate, participates with another acidic side chain on β1 to coordinate a magnesium ion, which is further coordinated by the substrate alpha phosphate, activating it for nucleophilic attack. While in some crystal structures the second acidic residue, a strictly conserved glutamate, is coordinated to another magnesium ion, this is not commonly observed in GGDEF crystal structures, and may be a crystallization artifact (Schirmer, 2016; Tarnawski et al., 2015). Thus, the role for this conserved residue has not yet been elucidated.

The guanine nucleobase is recognized through two specific contacts on both the Watson-Crick and sugar faces (Chan et al., 2004). In particular, an aspartate residue located on the second alpha helix in the GGDEF domain contacts N_1, and an asparagine residue located on the first α helix in the domain contacting N_7 and the

exocyclic amine group in hydrogen bonding interactions, which are expected to grant exquisite selectivity for the G nucleobase over A. In some crystal structures, there is a third contact between a tyrosine residue preceding the GGDEF motif and the exocyclic carbonyl of the guanosine base along the Hoogsteen face (Gourinchas et al., 2017). Furthermore, an absolutely conserved histidine residue centered over the G base may interact in a cation-π-stacking interaction (Schirmer, 2016).

While it is clear how the electrophilic “acceptor” nucleotide is activated for nucleophilic attack, it remains unknown how the nucleophilic portion is activated. Nucleophilic activation is required, as many isolated GGDEF domains do not possess GTPase activity, which would be observed in an aqueous environment if no nucleophilic activating residue were required. While no general base for 3’-hydroxy-group activation has been observed, this is likely due to the lack of active structures that have been obtained. It is possible that simple proximity effects allow for catalysis, however it is more likely that a cross-dimer general base exists; indeed, the requirement for dimerization to occur for activity to be measured lends further support to this hypothesis. We posit that a number of absolutely conserved residues in the GGDEF domain that have no known function (Schirmer, 2016) may act in this manner or in other cross-dimer substrate interactions.

Diadenylate cyclases contain a conserved DAC fold (also known as the DisA_N domain) similar to the other nucleotidyl transferases, in that it again contains a conserved beta sheet buttressed by a number of helices, with topology αβαβαββαβββ (Witte et al., 2008). As in other dinucleotide cyclases, a conserved set of residues near the active site, specifically DGA and EXGXRHR motifs, are observed. Similar to the GGDEF domain, the DAC active site consists of two half-active sites that form a C2-symmetric catalytic site.

ATP base contacts in the DAC domain occur primarily through hydrogen bonding interactions that allow for high specificity (Witte et al., 2008). The adenosine base is specifically recognized through two contacts: the carboxy group and amide nitrogen in the backbone of I94 (*Bacillus subtilis* DisA numbering) forms Watson-Crick-face hydrogen bonds with the adenine base, and T111 forms a final H-bond with the exocyclic nitrogen along the Hoogsteen face. Unlike diguanylate cyclases, there is also recognition of the ribose ring, in the form of a hydrogen bond between the backbone nitrogen of G76 and the 2’-OH.

In contrast to the homodimeric dinucleotide cyclases, the two characterized heterodimeric dinucleotide cyclases, DncV and cGAS, possess monomeric active sites that have to perform two separate catalytic steps to form a cyclic dinucleotide (Gao et al., 2013; Kranzusch et al., 2014). These monomeric active sites have discrete binding pockets for the two nucleotide triphosphates. Substrates are oriented to produce an initial linear dinucleotide; this intermediate must then be reoriented in the active site for the cyclization reaction to occur. Both enzymes differ in the order and linkage of product formed, in the two steps. Mechanistic studies utilizing non-hydrolyzable phosphate substrate analogs demonstrated that cGAS initially uses a donor ATP molecule to form a phosphodiester bond with the 2’-OH of a GTP acceptor (Gao et al., 2013). In contrast, DncV uses a donor GTP molecule to form a phosphodiester bond with the 3’-OH of an

ATP acceptor (Kranzusch et al., 2014). In both cases, reorientation of the linear substrate in the active site leads to a second 3',5' phosphodiester bond formation to form the cyclic dinucleotide product.

One question is how these two enzymes catalyze different phosphodiester linkages. When one examines the initial linear product formed by both cGAS and DncV, alternation between elongated and contracted side chains orient the nucleotides in the proper orientation to form either the 2',5'- or 3',5'- linkage (Kranzusch et al., 2014). This remodeling permitted engineering efforts to generate a "evolutionarily ancient" cGAS mutant that synthesizes 3',3'-cAG. In support of this evolutionary hypothesis, evidence using the response of various STING alleles to cGAS homologues suggests that the evolutionarily "ancient" version of cGAS, from the sea anemone *Nematostella vectensis*, produces a canonical cAG (Kranzusch et al., 2015). Despite reverse-engineering efforts, forward engineering efforts were unable to generate a DncV analogue capable of 2',5'-cAG synthase activity.

III - Regulation of Dinucleotide Cyclases

Second messengers, by definition, are designed to respond to a primary signal. Two major questions come out of this - how have bacteria evolved conserved strategies for dinucleotide cyclase activation for divergent signals, and what signals activate dinucleotide cyclases? In this section, we discuss the structural aspects of dinucleotide cyclase activation and the signals that have been discovered for specific dinucleotide cyclases. In particular, we highlight a conserved strategy involving reconstitution of the active site upon sensing of the primary signal.

III.A - Regulation of cdiG Synthesis

Diguanylate cyclases are regulated through a number of conserved structural mechanisms, centering primarily upon the obligate dimer nature of the active site. For the DGC reaction to occur, both GGDEF domains must be oriented in a position to bring both GTP substrates in the proper orientation for cyclic product formation (Chan et al., 2004). Indeed, expression and *in vitro* activity assays of GGDEF domains isolated from regulatory domains results in catalytically inactive proteins (De et al., 2009; Ryjenkov et al., 2005). This activity can be restored at the high concentrations (>100 μM) required for crystallization studies, as product has been observed in some isolated GGDEF domain X-ray crystal structures (Deepthi et al., 2014). Because GGDEF dimerization is required for catalysis, and the dimerized domains must be oriented properly for catalysis, two common themes in DGC regulation involves activation by dimerization, and activation by dimer reorientation.

Preventing dimer formation until activation occurs is one common mechanism of GGDEF control. Many GGDEF enzymes are located at the C-terminus of canonical response receiver regulator (Rec) domains, which dimerize upon phosphorylation of a conserved aspartate residue by a histidine kinase (Capra and Laub, 2012). This imposes multiple constraints upon DGC activity: First, the exquisite selectivity observed between histidine kinases and their target Rec domains ensures that activation will occur only in response to a desirable signal. Second, Rec domains possess their own

intrinsic autophosphatase activity, setting an intrinsic time limit to protein activation (Gao et al., 2007).

Some GGDEF domains form obligate dimers even without activator present; two well-studied examples are the zinc-repressed diguanylate cyclase DgcZ from *E. coli* and the red light-regulated diguanylate cyclase PadC from *Idiomarina* sp. A28L (Gourinchas et al., 2017; Zahringer et al., 2013). In each of these cases, the regulatory domain acts to orient the two GGDEF domains for productive reaction only when repression is relieved or activation occurs, presumably by moving the two active sites close enough for reaction. This conformational switch is often communicated through the switching of a homodimeric parallel coiled coil domain between two different registers (De et al., 2009; Gourinchas et al., 2017). Indeed, stabilization of either register with hydrophobic mutations in the a/d residues leads to the corresponding change in activity: stabilization of the inactive form abolishes activity, while stabilization of the active register generates a constitutively active protein (Gourinchas et al., 2017). The wide-ranging presence of the coiled coil domain has allowed the facile generation of constitutively active chimeric DGC proteins - fusion of the well-known GCN4 coiled coil to the GGDEF domain of WspR, for example, generates a synthetic diguanylate cyclase (De et al., 2009). This conserved activation strategy has permitted the generation of artificial chimeric light-regulated GGDEF protein with different photoactivation ratios (Ryu and Gomelsky, 2014).

Another form of regulation not based in coordination of the coiled-coils is found in membrane-bound GGDEFs like YfiN, an oxidative stress-responsive DGC in E. coli (Kim and Harshey, 2016). In this case, the GGDEF domain interacts with a putative helical linker located between the inter membrane HAMP domain and the GGDEF. Rearrangement of an extra-membrane PAS domain caused by degradation of YfiR, a general redox stress response factor, causes rearrangement of the HAMP domain, unlocking the GGDEF domain from the helical linker and allowing for activation (Giardina et al., 2013; Kim and Harshey, 2016).

GGDEF domains also possess at least two forms of “product inhibition by domain immobilization” (Chan et al., 2004). At least two conserved allosteric inhibition sites per monomer have been observed in both biochemical and x-ray crystal structures (Chan et al., 2004; Christen et al., 2006; Wassmann et al., 2007). Binding of product to these conserved I-sites forces the two active sites to orient facing away from each other, preventing enzymatic activity. In the case of both inhibitory sites, cyclic di-GMP binds to each I-site in a dimeric fashion, where two cyclic di-GMP molecules are stacked upon each other (Chan et al., 2004; Wassmann et al., 2007). The binding mode controlling specificity for cdiG is closely conserved between the two regulatory sites: a network of three arginines are involved in Hoogsteen interactions with three of the guanine bases in the cdiG dimer, and an aspartate residue contacts the W-C face of one of these three bases. Importantly, one base is not recognized by any specific hydrogen-bonding contacts, suggesting that multimers of cdiG and cAG could also bind readily to this allosteric site. Additionally, in GGDEF domains that do not possess a conserved I-site, many exhibit product inhibition through binding to the active site (Yang et al., 2011). Thus, the K_D for binding to the various active and inhibitory sites for GGDEFs likely

create an upper limit for cdiG concentration in cells, likely in the μM regime (Whitney et al., 2012).

One question in the field of bacterial signaling is, what signals activate GGDEF enzymes for cdiG synthesis? The magnitude of this question can be appreciated by analyzing the domain organizations of various GGDEF proteins. As annotated by the Pfam database, there are 3,011 specific domain architectures from 43,327 known GGDEF sequences (Finn et al., 2016). Furthermore, approximately 25% (12,157) of these sequences contain a single GGDEF domain with no sensory domains from a predicted family. In GGDEFs that contain a single sensor domain, the most common input domain is the response receiver regulator (Rec) domain (1,535 sequences) - and thousands of sequences contain one of the several subfamilies of PAS and GAF domains, which are general small-molecule sensors that bind amino acids, nucleotides, and in concert with cofactors can act as sensors of light, oxygen, and redox potential (Henry and Crosson, 2011). Less common are more complex domain architectures where multiple GAF and PAS domains form one regulatory region, suggesting more complex signaling decisions required for activation of cdiG synthesis.

III.A.1 - Rec domain GGDEFs

The Rec domain is found in approximately 20% of all GGDEF domain-containing proteins, and is the downstream actuator in bacterial two-component signaling systems. In these systems, a membrane-bound histidine kinase (HisK) senses an extracellular signal through an N-terminal sensory domain (Capra and Laub, 2012). In response to the allosteric changes upon ligand binding, the C-terminal kinase domain is activated, leading to autophosphorylation at a conserved histidine residue. This then permits phosphotransfer to a conserved aspartate residue on the N-terminus of a cognate response receiver regulator (Rec) domain, the conformational change then leading to activation of a C-terminal output domain to modulate enzymatic or cellular function.

Because the Rec domain is not the sensor of the primary signal, two additional problems come with determining activation signals for Rec-GGDEF proteins: what is the histidine kinase responsible for phosphorylating the Rec domain, and what signal activates the cognate HisK? The fact that many two component systems exist in cognate pairs located adjacent to one another - and usually on the same operon - on the genome, makes identification of most HK-RR pairs rapid and facile, and allowed the rapid computational identification of key residues involved in cognate interactions (although some genetically adjacent HK/RR pairs are not cognate pairs) (Skerker et al., 2008).

More difficult is the identification of a so-called “orphan” HK or RR to its cognate receptor or kinase, respectively. Despite advances in computational prediction of two component system interactions only 1 orphan kinase-rec pair has been assigned and experimentally proven from these efforts (Black et al., 2015). Successful discovery of orphan pairs is still primarily pursued through genetic or biochemical methods. Genetic methods often involve assaying individual knockouts or transposon knockout libraries to determine which kinases/response regulators phenocopy the orphan protein (Deutschbauer et al., 2011; Petters et al., 2012). In contrast, if the species in question

has a limited number of sensor histidine kinases, it is often tenable to perform in vitro phosphotransfer profiling with a set of candidate response regulators (or histidine kinases) to determine the appropriate cognate(s) of interest (Laub et al., 2007). Thus, a high-throughput, simple method to determine not only interacting pairs for orphan kinases but also if a Rec domain is part of a more complex phosphorelay pathway, remains a key problem in the field of bacterial signaling.

Despite decades of research, only three primary signals for Rec-GGDEF enzymes are known. The DGC WspR from *Pseudomonas* species is the cognate RR for the HK WspE, which is activated in response to surface sensing, although the exact mechanism for this sensory pathway remains unknown (Hickman et al., 2005). The cell division-regulating Rec-GGDEF PleD from *Caulobacter crescentus* is regulated by intracellular location due to the asymmetric distribution of two histidine kinases, DivJ and PleC (Paul et al., 2004). *Caulobacter* species are alphaproteobacteria that are models for asymmetric cell division, in which one daughter cell remains attached to a surface through a “stalk” structure, whereas the other becomes a “swarmer” cell with a typical bacterial flagellum (Lori et al., 2015). In this case, DGC activity is activated on the stalked cell side by the kinase DivJ, while it is repressed in the daughter swarmer cell by the kinase PleC, which acts as a phosphatase for phosphorylated PleD (Christen et al., 2010). Finally, the *Silicibacter* sp. TrichCH4B diguanylate cyclase SiliDGC is the target of the SiliHK that is regulated by nitric oxide (Rao et al., 2015). This nitric oxide synthase is activated by a the presence of an algal symbiont *Trichodesmium erythraeum*, however the precise mechanism for nitric oxide synthase activation remains unknown.

III.A.2 - Other Small-Molecule Recognition Domains

A number of other small molecule recognition domains are found at the N-terminus of GGDEF proteins. In cytosolic proteins the small recognition domain is typically a PAS or GAF domain. These proteins bind a wide array of small molecules, including amino acids, nucleotides, and phytochromes (Henry and Crosson, 2011). Despite the wide variety of PAS-GGDEF and GAF-GGDEF fusions, only 3 signals for PAS-GGDEF and GAF-GGDEF activation have been well-characterized: the only small-molecule-sensing GAF domain is the cyclic AMP-sensing GGDEF from *Leptospira interrogans*, (da Costa Vasconcelos et al., 2017). Additionally, there are both blue light- and red light-sensing GGDEF domains from various cyanobacteria (Blain-Hartung et al., 2017; Enomoto et al., 2015; Enomoto et al., 2014; Gourinchas et al., 2017; Tarutina et al., 2006). In these cases, most photoreceptors bind to bilin chromophores, which are derived from tetrapyrroles (Enomoto et al., 2015; Gourinchas et al., 2017) Because these domains sense an input orthogonal to most GGDEFs from most species, they have been adapted multiple times for optogenetic manipulation of cdiG levels (Ryu et al., 2017; Ryu and Gomelsky, 2014)

Two other examples of a small molecule-binding domain regulating GGDEF function are the zinc-binding CZB domain and the oxygen-binding globin domain (Tuckerman et al., 2009; Zahringer et al., 2013). The CZB-GGDEF domain organization appears rare and restricted to DGCs primarily in *Enterobacteria*, with the only studied member being DgcZ from *E. coli*, which is negatively regulated by zinc-binding. (Zahringer et al., 2013).

DgcZ is a driver of poly-N-acetyl-glucosamine synthesis, an essential exopolysaccharide for surface attachment, in *E. coli*; upon stimulation with zinc, biofilm formation is suppressed (Lacanna et al., 2016; Zahringer et al., 2013). While pleiotropic effects could be the source of this phenotype, RNA-based fluorescent biosensors have demonstrated an increase in cdiG in a DgcZ dependent manner in response to removal of zinc from the growth media (Yeo et al., 2017). The globin-GGDEF domain is more prevalent, being observed throughout various proteobacterial phyla, and binds oxygen with the assistance of a heme cofactor (Burns et al., 2014; Burns et al., 2016; Tuckerman et al., 2009).

Membrane-bound GGDEFs are often able to sense extracellular primary signals through periplasmic sensory domains, such as the periplasmic binding substrate-binding (PBPb), Cache, and CHASE domains (Koestler and Waters, 2014; Mills et al., 2015; Xu et al., 2017). These domains are generic small molecule binding domains that have diverged to sense many different signals - the only known signals that regulate GGDEFs by these domains are the charged amino acids lysine and arginine, the sugars glucose and N-acetyl-glucosamine (GlcNAc), and various bile acids. However, the mechanisms by which these small molecule-sensing domains function differ: in the case of the diguanylate cyclase CdgH from *Vibrio cholerae*, two periplasmic PBPb domains directly binds arginine/lysine/ornithine and glutamine/histidine, respectively (Xu et al., 2017). In *Salmonella typhimurium* STM1987, the Cache domain likely acts as a hub to integrate signals from periplasmic metabolite-sensing proteins such as the arginine binding protein ArtI (Mills et al., 2015). Similarly, genetic data suggest that the PBPb and CHASE-GGDEF proteins VC1067 and VC1376 from *Vibrio cholerae* are regulated by bile acids present in humans (Koestler and Waters, 2014). How a third DGC, VC1372, is regulated by these small molecules remains unknown, as it contains no obvious domains involved in signaling (Koestler and Waters, 2014).

Unknown protein domains or protein-protein interactions can also contribute to DGC activity. In *E. coli*, cellulose production is regulated by the diguanylate cyclase DgcQ sensing the concentration of various pyrimidine biosynthetic components (Rossi et al., 2017). It has no obvious sensory domains, however, the cytosolic portion binds robustly to both UTP and N-carbamoyl-aspartate (a biosynthetic intermediate of UTP synthesis). UTP activates the protein, while N-carbamoyl-aspartate inactivates the protein, and primary components of pyrimidine biosynthesis interact directly with DgcQ, likely adding a further level of control (Rossi et al., 2017). Also in E. coli, binding of the cdiG phosphodiesterase YciR interacts directly with the DGC YdaM (also known as DgcM), inhibiting its activity (Lindenberg et al., 2013). In this case, YciR acts as a “trigger enzyme” - upon binding of cdiG and hydrolysis, it releases YdaM, activating its DGC activity for curli fimbriae production (Lindenberg et al., 2013). It remains to be seen how common these PDE-DGC interactions are in the regulation of cdiG levels, but preliminary work in E. coli suggests that signaling hubs are an integral part of cdiG metabolism in bacteria (Sarenko et al., 2017).

We note that, of known DGC activation signals, most of these activation signals have been discovered and validated in the past five years, with the three exceptions being oxygen (globin-coupled sensor), cellular localization (PleD), and surfaces (WspR). As

methods improve for high-throughput evaluation of cyclic di-GMP levels in cells, we expect that this rate of primary signal discovery will improve concurrently. In particular, we point out that there are now both RNA- and protein-based fluorescent sensors are now powerful enough to detect changes in cdiG levels in cells stemming from the activities of single enzymes (Christen et al., 2010; Kellenberger et al., 2013; Wang et al., 2016).

III.B - Regulation of DAC Activity -

In contrast to the vast body of literature regarding GGDEF domain activation, corresponding studies on DAC activation have been limited primarily to the initial structural work on DisA, the DNA damage-checkpoint protein that regulates entry into sporulation (Witte et al., 2008). An octamer in solution, DisA forms two C4-symmetric tetramers that dimerize in a head-to-head conformation using the DAC domain. With this, four C-terminal HhH domains are splayed outwards in a symmetric conformation, which is similar to the structure of a Holliday junction or otherwise stalled replication fork (Witte et al., 2008). While gel shift assays demonstrate binding of DisA to branched nucleic acid species, and activity assays demonstrate that these branched DNA intermediates decrease activity, the structural basis for how DNA-binding impacts the catalytic domain remains unknown (Witte et al., 2008). It is expected that other DisA homologues function in this way. However, little is known about the structural basis for activation of other diadenylate cyclases, for instance how the membrane-associated DacA from *Listeria monocytogenes* responds to osmotic stress, which can occur via the introduction of oligopeptides (Whiteley et al., 2017).

There are some reports of protein interactions modulating DAC activity. GlmM, the phosphoglucosamine mutase in *B. subtilis*, binds to and acts as a negative regulator of CdaA, the DAC in this species that does not have any DNA-binding motifs (Gundlach et al., 2015). This interaction is similarly conserved in *Lactobacillus lactis* (Zhu et al., 2016). Similar inhibitory interactions between CdaR, a membrane protein with no known function, and CdaA in various gram positives have been reported (Mehne et al., 2013). Finally, DisA itself has a negative regulator in the recombination protein RadA (Zhang and He, 2013). However, in all of these cases, as little is known about the structural implications of binding, it is unknown how these protein interactions regulate the activity of the DAC domain.

III.C - Activating Dinucleotide Cyclases

OAS and cGAS are core modulators of mammalian innate immunity, acting as cytosolic surveillance systems for double stranded RNA and DNA, respectively (Hornung et al., 2014). As such, they share homologous structural features that allow for binding to double-stranded nucleic acids, in particular a conserved positively-charged surface that can contact the phosphate backbones of their respective substrates with little regard to sequence specificity. The OAS backbone preferentially binds A-form RNA, by a combination of a larger positively charged surface, and a set of key residues required for interaction with the 2'-OH groups in the RNA recognition substrate (Donovan et al., 2013). In contrast, cGAS specifically recognizes B-form DNA substrates through a

conserved zinc loop that blocks A-form nucleic acids from entering the binding cleft (Civril et al., 2013).

The mechanism for OAS activation by RNA recognition gives a paradigm by which these nucleotidyltransferases function (Donovan et al., 2013). Binding of the nucleic acid activator closes the substrate-binding cleft, causing the formation of the final ß sheet in the active site, and moves the final aspartate in the catalytic triad 4 Å to permit magnesium ion binding (Donovan et al., 2013). This activation paradigm is conserved in cGAS (Civril et al., 2013). A central spine helix connecting the N- and C-terminal lobes, as in OAS, is broken upon RNA binding. This, again, leads to motion of the final acidic residue required for active site formation.

Structural work suggested that cGAS would be robustly activated by 20 nt-length DNA (Civril et al., 2013). Both *in vivo* and *in vitro* work presented an interesting conundrum: transfected dsDNA must either have G-overhangs or be ~45 nt in length (or longer) for activation (Andreeva et al., 2017; Herzner et al., 2015). The first hints towards a solution came with the discovery of a second DNA binding site in cGAS, allowing the formation of a higher-order DNA_2:$cGAS_2$ complex (Li et al., 2013). The final advance has come with a recent X-ray crystal structure and corresponding kinetics assays demonstrating that cGAS forms higher-order ladder-like structures for full activation *in vivo* (Andreeva et al., 2017).

While the eukaryotic dinucleotide cyclases have been extensively characterized in the structural requirements for activation by their respective substrates, the bacterial cAG synthase from *Vibrio cholerae*, DncV, is related structurally to cGAS and OAS (Davies et al., 2012). However, unlike these two modulators of the innate immune system, it is constitutively active. Crystallographic and biochemical analysis revealed that this protein is repressed by folate derivatives (Zhu et al., 2014). However, the solved X-ray crystal structure does not differ significantly from the apo enzyme, so it remains unknown how these molecules decrease activity, or their physiological role in *Vibrio* infection (Kranzusch et al., 2014; Zhu et al., 2014).

Outlook

At the outset of my thesis studies, a new riboswitch effector had been discovered for cAG in *Geobacter* species implicated in regulating extracellular electron transfer (Kellenberger et al., 2015b; Nelson et al., 2015; Ren et al., 2015). However, no DncV homologue was found in the genome, raising the question of “what is the cAG synthase?” We hypothesized that, much like the riboswitch being structurally related to and likely evolved from a previously-described cdiG riboswitch, that the cAG synthase had evolved from cdiG synthase machinery. Herein, I describe the discovery and characterization of a cAG signaling pathway that exists naturally in deltaproteobacteria that acts orthogonal to the cdiG signaling system. Chapter 2 breaks the paradigm that GGDEFs are associated solely with cdiG signaling by demonstrating that a GGDEF protein from *Geobacter sulfurreducens* is capable of synthesis of cAG, and the subsequent initial biochemical characterization of this enzyme. Chapter 3 then focuses on a detailed mechanistic analysis of this hybrid product producing (HyPr) GGDEF, and explains how it is able to produce primarily the heterodimeric product under presumably

physiological conditions. In Chapter 4, I obtained the first X-ray crystal structure of a HyPr GGDEF to understand further the structural basis for HyPr activity.

While cdiG has attained dominance in the realm of bacterial signaling, the discovery and characterization of related but orthogonal second messenger pathways has broken the paradigm that the enzymes involved in its synthesis, breakdown, and recognition are involved in only one signaling system. My work sets the stage for the discovery and development of new, orthogonal second messenger systems in bacteria.

Figures

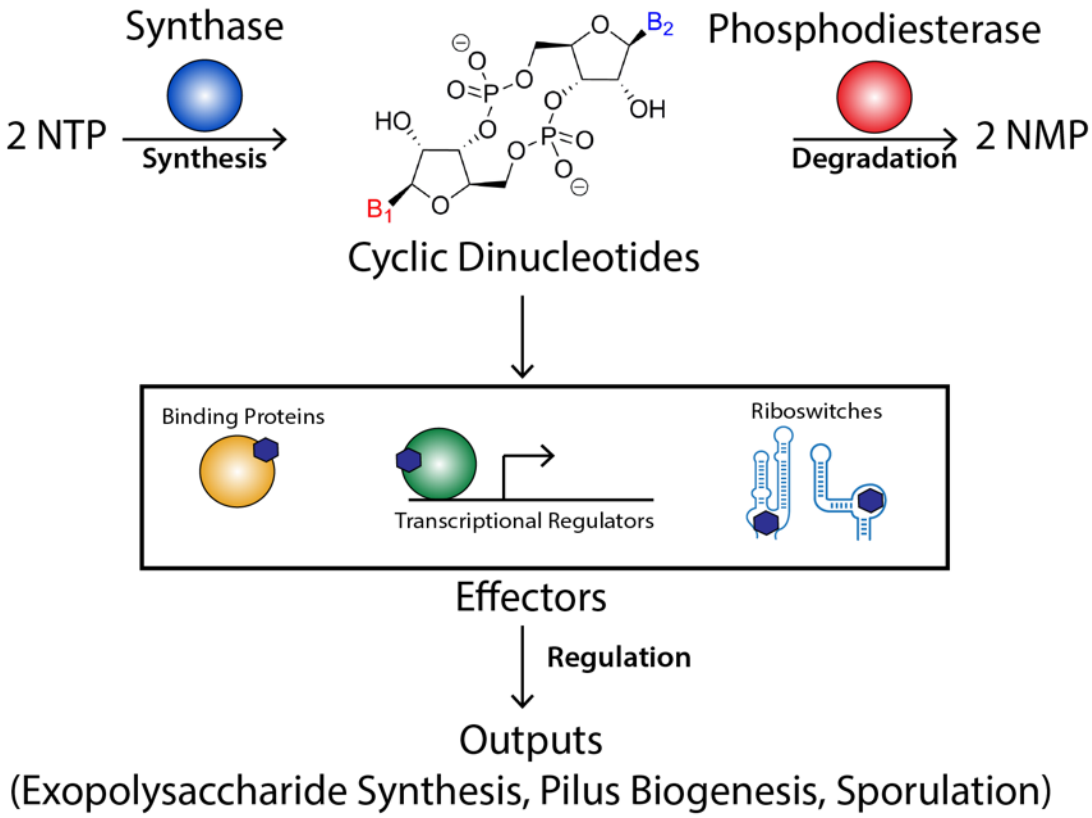


Figure 1 - Cyclic Dinucleotide Signaling Modules. For a functional second-messenger system to work, three things are required: a synthase that generates the second messenger under appropriate activation conditions, a hydrolase (in the case of CDNs, a phosphodiesterase) to degrade the molecule under orthogonal conditions, and a suite of effectors that respond to the presence of the second messenger to cause further outputs. Cyclic dinucleotides regulate diverse outputs, including biofilm and pilus formation, cell wall maintenance, and sporulation.

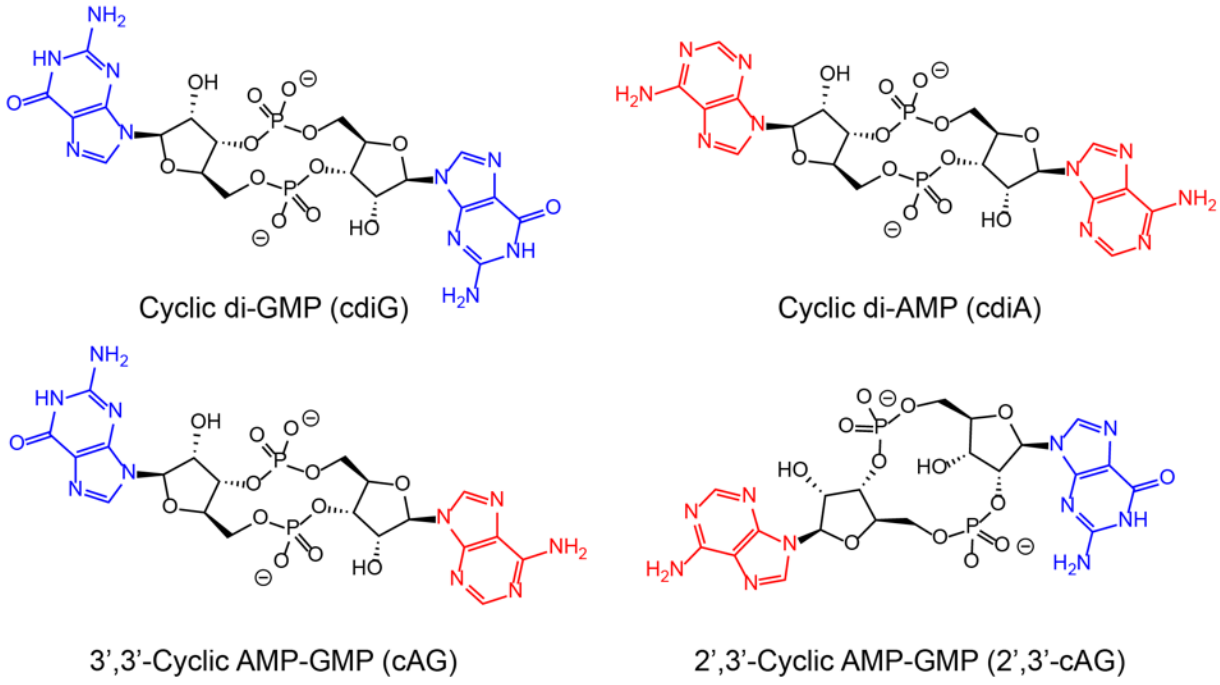


Figure 2 - Characterized Cyclic Dinucleotides. The four cyclic dinucleotides currently known in the literature: cyclic di-GMP, cyclic di-AMP, and 3',3'-cyclic AMP-GMP are found primarily in prokaryotes. In contrast, the non-canonical 2',3'-cyclic AMP-GMP is found as a stimulator of the mammalian innate immune response.

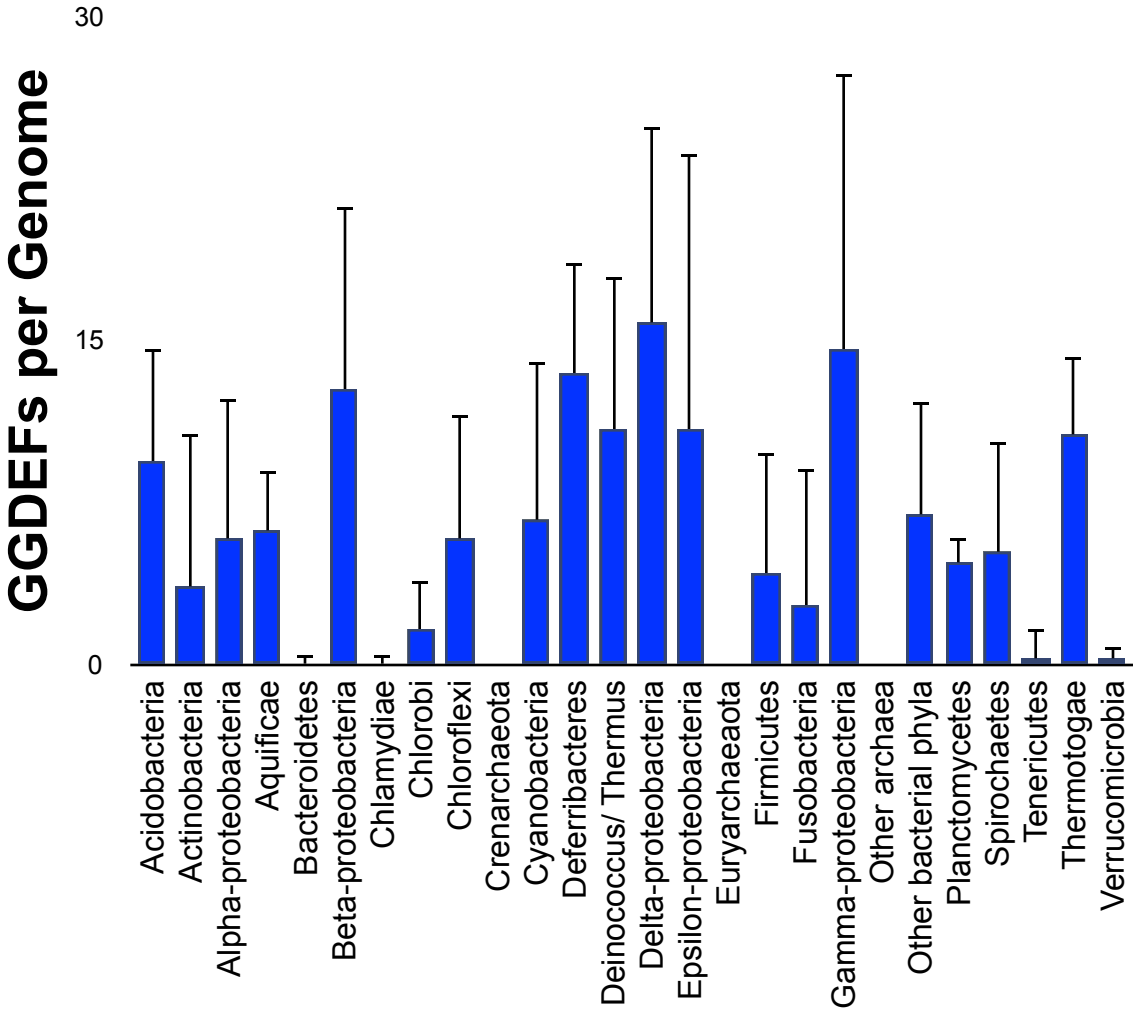


Figure 3 - Phylogeny of GGDEF proteins. Shown are phylum-wide distributions of GGDEFs per genome.

Bibliography

Ablasser, A., Goldeck, M., Cavlar, T., Deimling, T., Witte, G., Rohl, I., Hopfner, K.P., Ludwig, J., and Hornung, V. (2013). cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. Nature *498*, 380-384.

Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U.*, et al.* (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun *7*, 13219.

Andreeva, L., Hiller, B., Kostrewa, D., Lassig, C., de Oliveira Mann, C.C., Jan Drexler, D., Maiser, A., Gaidt, M., Leonhardt, H., Hornung, V.*, et al.* (2017). cGAS senses long and HMGB/TFAM-bound U-turn DNA by forming protein-DNA ladders. Nature *549*, 394-398.

Ausmees, N., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M., and Lindberg, M. (2001). Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. FEMS Microbiol Lett *204*, 163-167.

Barbalat, R., Ewald, S.E., Mouchess, M.L., and Barton, G.M. (2011). Nucleic acid recognition by the innate immune system. Annu Rev Immunol *29*, 185-214.

Bejerano-Sagie, M., Oppenheimer-Shaanan, Y., Berlatzky, I., Rouvinski, A., Meyerovich, M., and Ben-Yehuda, S. (2006). A checkpoint protein that scans the chromosome for damage at the start of sporulation in Bacillus subtilis. Cell *125*, 679-690.

Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., and Rabinowitz, J.D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat Chem Biol *5*, 593-599.

Black, W.P., Wang, L., Davis, M.Y., and Yang, Z. (2015). The orphan response regulator EpsW is a substrate of the DifE kinase and it regulates exopolysaccharide in Myxococcus xanthus. Sci Rep *5*, 17831.

Blain-Hartung, M., Rockwell, N.C., and Lagarias, J.C. (2017). Light-Regulated Synthesis of Cyclic-di-GMP by a Bidomain Construct of the Cyanobacteriochrome Tlr0924 (SesA) without Stable Dimerization. Biochemistry *56*, 6145-6154.

Burns, J.L., Deer, D.D., and Weinert, E.E. (2014). Oligomeric state affects oxygen dissociation and diguanylate cyclase activity of globin coupled sensors. Mol Biosyst *10*, 2823-2826.

Burns, J.L., Rivera, S., Deer, D.D., Joynt, S.C., Dvorak, D., and Weinert, E.E. (2016). Oxygen and Bis(3',5')-cyclic Dimeric Guanosine Monophosphate Binding Control Oligomerization State Equilibria of Diguanylate Cyclase-Containing Globin Coupled Sensors. Biochemistry *55*, 6642-6651.

Capra, E.J., and Laub, M.T. (2012). Evolution of two-component signal transduction systems. Annu Rev Microbiol *66*, 325-347.

Chan, C., Paul, R., Samoray, D., Amiot, N.C., Giese, B., Jenal, U., and Schirmer, T. (2004). Structural basis of activity and allosteric control of diguanylate cyclase. Proc Natl Acad Sci U S A *101*, 17084-17089.

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

Chou, S.H., and Galperin, M.Y. (2016). Diversity of Cyclic Di-GMP-Binding Proteins and Mechanisms. J Bacteriol *198*, 32-46.

Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., Meuwly, M., and Jenal, U. (2006). Allosteric control of cyclic di-GMP signaling. J Biol Chem *281*, 32015-32024.

Christen, M., Kulasekara, H.D., Christen, B., Kulasekara, B.R., Hoffman, L.R., and Miller, S.I. (2010). Asymmetrical distribution of the second messenger c-di-GMP upon bacterial cell division. Science *328*, 1295-1297.

Civril, F., Deimling, T., de Oliveira Mann, C.C., Ablasser, A., Moldt, M., Witte, G., Hornung, V., and Hopfner, K.P. (2013). Structural mechanism of cytosolic DNA sensing by cGAS. Nature *498*, 332-337.

Coda, R., Cagno, R.D., Gobbetti, M., and Rizzello, C.G. (2014). Sourdough lactic acid bacteria: exploration of non-wheat cereal-based fermentation. Food Microbiol *37*, 51-58.

Corrigan, R.M., and Grundling, A. (2013). Cyclic di-AMP: another second messenger enters the fray. Nat Rev Microbiol *11*, 513-524.

da Costa Vasconcelos, F.N., Maciel, N.K., Favaro, D.C., de Oliveira, L.C., Barbosa, A.S., Salinas, R.K., de Souza, R.F., Farah, C.S., and Guzzo, C.R. (2017). Structural and Enzymatic Characterization of a cAMP-Dependent Diguanylate Cyclase from Pathogenic Leptospira Species. J Mol Biol *429*, 2337-2352.

Davies, B.W., Bogard, R.W., Young, T.S., and Mekalanos, J.J. (2012). Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for V. cholerae virulence. Cell *149*, 358-370.

De, N., Navarro, M.V., Raghavan, R.V., and Sondermann, H. (2009). Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. J Mol Biol *393*, 619-633.

De Roos, J., and De Vuyst, L. (2017). Acetic acid bacteria in fermented foods and beverages. Curr Opin Biotechnol *49*, 115-119.

Deepthi, A., Liew, C.W., Liang, Z.X., Swaminathan, K., and Lescar, J. (2014). Structure of a diguanylate cyclase from Thermotoga maritima: insights into activation, feedback inhibition and thermostability. PLoS One *9*, e110912.

Deutschbauer, A., Price, M.N., Wetmore, K.M., Shao, W., Baumohl, J.K., Xu, Z., Nguyen, M., Tamse, R., Davis, R.W., and Arkin, A.P. (2011). Evidence-based

annotation of gene function in Shewanella oneidensis MR-1 using genome-wide fitness profiling across 121 conditions. PLoS Genet *7*, e1002385.

Diner, E.J., Burdette, D.L., Wilson, S.C., Monroe, K.M., Kellenberger, C.A., Hyodo, M., Hayakawa, Y., Hammond, M.C., and Vance, R.E. (2013). The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. Cell Rep *3*, 1355-1361.

Donovan, J., Dufner, M., and Korennykh, A. (2013). Structural basis for cytosolic double-stranded RNA surveillance by human oligoadenylate synthetase 1. Proc Natl Acad Sci U S A *110*, 1652-1657.

Edwards, S.J., and Kjellerup, B.V. (2013). Applications of biofilms in bioremediation and biotransformation of persistent organic pollutants, pharmaceuticals/personal care products, and heavy metals. Appl Microbiol Biotechnol *97*, 9909-9921.

Enomoto, G., Ni Ni, W., Narikawa, R., and Ikeuchi, M. (2015). Three cyanobacteriochromes work together to form a light color-sensitive input system for c-di-GMP signaling of cell aggregation. Proc Natl Acad Sci U S A *112*, 8082-8087.

Enomoto, G., Nomura, R., Shimada, T., Ni Ni, W., Narikawa, R., and Ikeuchi, M. (2014). Cyanobacteriochrome SesA is a diguanylate cyclase that induces cell aggregation in Thermosynechococcus. J Biol Chem *289*, 24801-24809.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A.*, et al.* (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res *44*, D279-285.

Gao, J., Tao, J., Liang, W., Zhao, M., Du, X., Cui, S., Duan, H., Kan, B., Su, X., and Jiang, Z. (2015). Identification and characterization of phosphodiesterases that specifically degrade 3'3'-cyclic GMP-AMP. Cell Res *25*, 539-550.

Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B.L., Zillinger, T., Serganov, A.A., Liu, Y., Jones, R.A., Hartmann, G.*, 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*, 1094-1107.

Gao, R., Mack, T.R., and Stock, A.M. (2007). Bacterial response regulators: versatile regulatory strategies from common domains. Trends Biochem Sci *32*, 225-234.

Giardina, G., Paiardini, A., Fernicola, S., Franceschini, S., Rinaldo, S., Stelitano, V., and Cutruzzola, F. (2013). Investigating the allosteric regulation of YfiN from Pseudomonas aeruginosa: clues from the structure of the catalytic domain. PLoS One *8*, e81324.

Gourinchas, G., Etzl, S., Gobl, C., Vide, U., Madl, T., and Winkler, A. (2017). Long-range allosteric signaling in red light-regulated diguanylyl cyclases. Sci Adv *3*, e1602498.

Gundlach, J., Mehne, F.M., Herzberg, C., Kampf, J., Valerius, O., Kaever, V., and Stulke, J. (2015). An Essential Poison: Synthesis and Degradation of Cyclic Di-AMP in Bacillus subtilis. J Bacteriol *197*, 3265-3274.

Henry, J.T., and Crosson, S. (2011). Ligand-binding PAS domains in a genomic, cellular, and structural context. Annu Rev Microbiol *65*, 261-286.

Herzner, A.M., Hagmann, C.A., Goldeck, M., Wolter, S., Kubler, K., Wittmann, S., Gramberg, T., Andreeva, L., Hopfner, K.P., Mertens, C.*, et al.* (2015). Sequence-specific activation of the DNA sensor cGAS by Y-form DNA structures as found in primary HIV-1 cDNA. Nat Immunol *16*, 1025-1033.

Hickman, J.W., Tifrea, D.F., and Harwood, C.S. (2005). A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. Proc Natl Acad Sci U S A *102*, 14422-14427.

Hobley, L., Fung, R.K., Lambert, C., Harris, M.A., Dabhi, J.M., King, S.S., Basford, S.M., Uchida, K., Till, R., Ahmad, R.*, et al.* (2012). Discrete cyclic di-GMP-dependent control of bacterial predation versus axenic growth in Bdellovibrio bacteriovorus. PLoS Pathog *8*, e1002493.

Hornung, V., Hartmann, R., Ablasser, A., and Hopfner, K.P. (2014). OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids. Nat Rev Immunol *14*, 521-528.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K.*, et al.* (2016a). A new view of the tree of life. Nat Microbiol *1*, 16048.

Hug, L.A., Thomas, B.C., Sharon, I., Brown, C.T., Sharma, R., Hettich, R.L., Wilkins, M.J., Williams, K.H., Singh, A., and Banfield, J.F. (2016b). Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. Environ Microbiol *18*, 159-173.

Hunter, J.L., Severin, G.B., Koestler, B.J., and Waters, C.M. (2014). The Vibrio cholerae diguanylate cyclase VCA0965 has an AGDEF active site and synthesizes cyclic di-GMP. BMC Microbiol *14*, 22.

Jenal, U., Reinders, A., and Lori, C. (2017). Cyclic di-GMP: second messenger extraordinaire. Nat Rev Microbiol *15*, 271-284.

Kellenberger, C.A., Chen, C., Whiteley, A.T., Portnoy, D.A., and Hammond, M.C. (2015a). RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messenger Cyclic di-AMP. J Am Chem Soc *137*, 6432-6435.

Kellenberger, C.A., Wilson, S.C., Hickey, S.F., Gonzalez, T.L., Su, Y., Hallberg, Z.F., Brewer, T.F., Iavarone, A.T., Carlson, H.K., Hsieh, Y.F.*, et al.* (2015b). GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. Proc Natl Acad Sci U S A *112*, 5383-5388.

Kellenberger, C.A., Wilson, S.C., Sales-Lee, J., and Hammond, M.C. (2013). RNA-based fluorescent biosensors for live cell imaging of second messengers cyclic di-GMP and cyclic AMP-GMP. J Am Chem Soc *135*, 4906-4909.

Kim, H.K., and Harshey, R.M. (2016). A Diguanylate Cyclase Acts as a Cell Division Inhibitor in a Two-Step Response to Reductive and Envelope Stresses. MBio *7*.

Koestler, B.J., and Waters, C.M. (2014). Bile acids and bicarbonate inversely regulate intracellular cyclic di-GMP in Vibrio cholerae. Infect Immun *82*, 3002-3014.

Kranzusch, P.J., Lee, A.S., Berger, J.M., and Doudna, J.A. (2013). Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. Cell Rep *3*, 1362-1368.

Kranzusch, P.J., Lee, A.S.Y., Wilson, S.C., Solovykh, M.S., Vance, R.E., Berger, J.M., and Doudna, J.A. (2014). Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. Cell *158*, 1011-1021.

Kranzusch, P.J., Wilson, S.C., Lee, A.S., Berger, J.M., Doudna, J.A., and Vance, R.E. (2015). Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2',3' cGAMP Signaling. Mol Cell *59*, 891-903.

Lacanna, E., Bigosch, C., Kaever, V., Boehm, A., and Becker, A. (2016). Evidence for Escherichia coli Diguanylate Cyclase DgcZ Interlinking Surface Sensing and Adhesion via Multiple Regulatory Routes. J Bacteriol *198*, 2524-2535.

Laub, M.T., Biondi, E.G., and Skerker, J.M. (2007). Phosphotransfer profiling: systematic mapping of two-component signal transduction pathways and phosphorelays. Methods Enzymol *423*, 531-548.

Li, X., Shu, C., Yi, G., Chaton, C.T., Shelton, C.L., Diao, J., Zuo, X., Kao, C.C., Herr, A.B., and Li, P. (2013). Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. Immunity *39*, 1019-1031.

Lindenberg, S., Klauck, G., Pesavento, C., Klauck, E., and Hengge, R. (2013). The EAL domain protein YciR acts as a trigger enzyme in a c-di-GMP signalling cascade in E. coli biofilm control. EMBO J *32*, 2001-2014.

Lori, C., Ozaki, S., Steiner, S., Bohm, R., Abel, S., Dubey, B.N., Schirmer, T., Hiller, S., and Jenal, U. (2015). Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication. Nature *523*, 236-239.

Mehne, F.M., Gunka, K., Eilers, H., Herzberg, C., Kaever, V., and Stulke, J. (2013). Cyclic di-AMP homeostasis in bacillus subtilis: both lack and high level accumulation of the nucleotide are detrimental for cell growth. J Biol Chem *288*, 2004-2017.

Mills, E., Petersen, E., Kulasekara, B.R., and Miller, S.I. (2015). A direct screen for c-di-GMP modulators reveals a Salmonella Typhimurium periplasmic L-arginine-sensing pathway. Sci Signal *8*, ra57.

Nelson, J.W., Sudarsan, N., Phillips, G.E., Stav, S., Lunse, C.E., McCown, P.J., and Breaker, R.R. (2015). Control of bacterial exoelectrogenesis by c-AMP-GMP. Proc Natl Acad Sci U S A *112*, 5389-5394.

Nick Pace, C., Scholtz, J.M., and Grimsley, G.R. (2014). Forces stabilizing proteins. FEBS Lett *588*, 2177-2184.

Paul, R., Weiser, S., Amiot, N.C., Chan, C., Schirmer, T., Giese, B., and Jenal, U. (2004). Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. Genes Dev *18*, 715-727.

Pei, J., and Grishin, N.V. (2001). GGDEF domain is homologous to adenylyl cyclase. Proteins *42*, 210-216.

Petters, T., Zhang, X., Nesper, J., Treuner-Lange, A., Gomez-Santos, N., Hoppert, M., Jenal, U., and Sogaard-Andersen, L. (2012). The orphan histidine protein kinase SgmT is a c-di-GMP receptor and regulates composition of the extracellular matrix together with the orphan DNA binding response regulator DigR in Myxococcus xanthus. Mol Microbiol *84*, 147-165.

Rao, M., Smith, B.C., and Marletta, M.A. (2015). Nitric Oxide Mediates Biofilm Formation and Symbiosis in Silicibacter sp. Strain TrichCH4B. MBio *6*, e00206-00215.

Ren, A., Wang, X.C., Kellenberger, C.A., Rajashankar, K.R., Jones, R.A., Hammond, M.C., and Patel, D.J. (2015). Structural basis for molecular discrimination by a 3',3'-cGAMP sensing riboswitch. Cell Rep *11*, 1-12.

Romling, U., Galperin, M.Y., and Gomelsky, M. (2013). Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. Microbiol Mol Biol Rev *77*, 1-52.

Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., Mayer, R., Braun, S., de Vroom, E., van der Marel, G.A., van Boom, J.H.*, et al.* (1987). Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. Nature *325*, 279-281.

Rossi, E., Motta, S., Aliverti, A., Cossu, F., Gourlay, L., Mauri, P., and Landini, P. (2017). Cellulose production is coupled to sensing of the pyrimidine biosynthetic pathway via c-di-GMP production by the DgcQ protein of Escherichia coli. Environ Microbiol *19*, 4551-4563.

Ryjenkov, D.A., Tarutina, M., Moskvin, O.V., and Gomelsky, M. (2005). Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain. J Bacteriol *187*, 1792-1798.

Ryu, M.H., Fomicheva, A., Moskvin, O.V., and Gomelsky, M. (2017). Optogenetic Module for Dichromatic Control of c-di-GMP Signaling. J Bacteriol *199*.

Ryu, M.H., and Gomelsky, M. (2014). Near-infrared light responsive synthetic c-di-GMP module for optogenetic applications. ACS Synth Biol *3*, 802-810.

Sarenko, O., Klauck, G., Wilke, F.M., Pfiffer, V., Richter, A.M., Herbst, S., Kaever, V., and Hengge, R. (2017). More than Enzymes That Make or Break Cyclic Di-GMP-Local Signaling in the Interactome of GGDEF/EAL Domain Proteins of Escherichia coli. MBio *8*.

Schirmer, T. (2016). C-di-GMP Synthesis: Structural Aspects of Evolution, Catalysis and Regulation. J Mol Biol *428*, 3683-3701.

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

Shchokolova, A.S., Rymko, A.N., Kvach, S.V., Shabunya, P.S., Fatykhava, S.A., and Zinchenko, A.I. (2015). Enzymatic synthesis of 2'-ara and 2'-deoxy analogues of c-di-GMP. Nucleosides Nucleotides Nucleic Acids *34*, 416-423.

Sinha, S.C., and Sprang, S.R. (2006). Structures, mechanism, regulation and evolution of class III nucleotidyl cyclases. Rev Physiol Biochem Pharmacol *157*, 105-140.

Skerker, J.M., Perchuk, B.S., Siryaporn, A., Lubin, E.A., Ashenberg, O., Goulian, M., and Laub, M.T. (2008). Rewiring the specificity of two-component signal transduction systems. Cell *133*, 1043-1054.

Skotnicka, D., Petters, T., Heering, J., Hoppert, M., Kaever, V., and Sogaard-Andersen, L. (2015). Cyclic Di-GMP Regulates Type IV Pilus-Dependent Motility in Myxococcus xanthus. J Bacteriol *198*, 77-90.

Sun, L., Wu, J., Du, F., Chen, X., and Chen, Z.J. (2013). Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. Science *339*, 786-791.

Tal, R., Wong, H.C., Calhoon, R., Gelfand, D., Fear, A.L., Volman, G., Mayer, R., Ross, P., Amikam, D., Weinhouse, H.*, et al.* (1998). Three cdg operons control cellular turnover of cyclic di-GMP in Acetobacter xylinum: genetic organization and occurrence of conserved domains in isoenzymes. J Bacteriol *180*, 4416-4425.

Tarnawski, M., Barends, T.R., and Schlichting, I. (2015). Structural analysis of an oxygen-regulated diguanylate cyclase. Acta Crystallogr D Biol Crystallogr *71*, 2158-2177.

Tarutina, M., Ryjenkov, D.A., and Gomelsky, M. (2006). An unorthodox bacteriophytochrome from Rhodobacter sphaeroides involved in turnover of the second messenger c-di-GMP. J Biol Chem *281*, 34751-34758.

Tuckerman, J.R., Gonzalez, G., Sousa, E.H., Wan, X., Saito, J.A., Alam, M., and Gilles-Gonzalez, M.A. (2009). An oxygen-sensing diguanylate cyclase and phosphodiesterase couple for c-di-GMP control. Biochemistry *48*, 9764-9774.

Wang, X.C., Wilson, S.C., and Hammond, M.C. (2016). Next-generation RNA-based fluorescent biosensors enable anaerobic detection of cyclic di-GMP. Nucleic Acids Res *44*, e139.

Wassmann, P., Chan, C., Paul, R., Beck, A., Heerklotz, H., Jenal, U., and Schirmer, T. (2007). Structure of BeF3- -modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. Structure *15*, 915-927.

Whiteley, A.T., Garelis, N.E., Peterson, B.N., Choi, P.H., Tong, L., Woodward, J.J., and Portnoy, D.A. (2017). c-di-AMP modulates Listeria monocytogenes central metabolism to regulate growth, antibiotic resistance and osmoregulation. Mol Microbiol *104*, 212-233.

Whitney, J.C., Colvin, K.M., Marmont, L.S., Robinson, H., Parsek, M.R., and Howell, P.L. (2012). Structure of the cytoplasmic region of PelD, a degenerate diguanylate cyclase receptor that regulates exopolysaccharide production in Pseudomonas aeruginosa. J Biol Chem *287*, 23582-23593.

Witte, G., Hartung, S., Buttner, K., and Hopfner, K.P. (2008). Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. Mol Cell *30*, 167-178.

Wolfe, B.E., Button, J.E., Santarelli, M., and Dutton, R.J. (2014). Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. Cell *158*, 422-433.

Wu, J., Sun, L., Chen, X., Du, F., Shi, H., Chen, C., and Chen, Z.J. (2013). Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. Science *339*, 826-830.

Wu, X., Wu, F.H., Wang, X., Wang, L., Siedow, J.N., Zhang, W., and Pei, Z.M. (2014). Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. Nucleic Acids Res *42*, 8243-8257.

Xu, M., Wang, Y.Z., Yang, X.A., Jiang, T., and Xie, W. (2017). Structural studies of the periplasmic portion of the diguanylate cyclase CdgH from Vibrio cholerae. Sci Rep *7*, 1861.

Yang, C.Y., Chin, K.H., Chuah, M.L., Liang, Z.X., Wang, A.H., and Chou, S.H. (2011). The structure and inhibition of a GGDEF diguanylate cyclase complexed with (c-di-GMP)(2) at the active site. Acta Crystallogr D Biol Crystallogr *67*, 997-1008.

Yeo, J., Dippel, A.B., Wang, X.C., and Hammond, M.C. (2017). In Vivo Biochemistry: Single-Cell Dynamics of Cyclic Di-GMP in Escherichia coli in Response to Zinc Overload. Biochemistry.

Zahringer, F., Lacanna, E., Jenal, U., Schirmer, T., and Boehm, A. (2013). Structure and signaling mechanism of a zinc-sensory diguanylate cyclase. Structure *21*, 1149-1157.

Zhang, L., and He, Z.G. (2013). Radiation-sensitive gene A (RadA) targets DisA, DNA integrity scanning protein A, to negatively affect cyclic Di-AMP synthesis activity in Mycobacterium smegmatis. J Biol Chem *288*, 22426-22436.

Zhu, D., Wang, L., Shang, G., Liu, X., Zhu, J., Lu, D., Wang, L., Kan, B., Zhang, J.R., and Xiang, Y. (2014). Structural biochemistry of a Vibrio cholerae dinucleotide cyclase reveals cyclase activity regulation by folates. Mol Cell *55*, 931-937.

Zhu, Y., Pham, T.H., Nhiep, T.H., Vu, N.M., Marcellin, E., Chakrabortti, A., Wang, Y., Waanders, J., Lo, R., Huston, W.M.*, et al.* (2016). Cyclic-di-AMP synthesis by the diadenylate cyclase CdaA is modulated by the peptidoglycan biosynthesis enzyme GlmM in Lactococcus lactis. Mol Microbiol *99*, 1015-1027.

Chapter Two

Hybrid Promiscuous (HyPr) GGDEF enzymes produce cyclic AMP-GMP (3', 3'-cGAMP)

This chapter has been published in: Hallberg, Z.F**.** Wang, X.C., Wright, T.A., Nan, B., Ad, O., Yeo, J., Hammond, M.C. *Proc Natl Acad Sci* 2016, *113*, 1790.

Abstract

Over 30 years ago, GGDEF domain-containing enzymes were shown to be diguanylate cyclases that produce cyclic di-GMP (cdiG), a second messenger that modulates the key bacterial lifestyle transition from a motile to sessile, biofilm-forming state. Since then, the ubiquity of genes encoding GGDEF proteins in bacterial genomes has established the dominance of cdiG signaling in bacteria. However, the observation that proteobacteria encode a large number of GGDEF proteins, nearing 1% of coding sequences in some cases, raises the question why bacteria need so many GGDEF enzymes. In this study, we reveal that a subfamily of GGDEF enzymes synthesizes the asymmetric signaling molecule cyclic AMP-GMP (cAG or 3', 3'-cGAMP). This discovery is unexpected because GGDEF enzymes function as symmetric homodimers, with each monomer binding to one substrate NTP. Detailed analysis of the enzyme from *Geobacter sulfurreducens* showed it is a dinucleotide cyclase capable of switching the major cyclic dinucleotide (CDN) produced based on ATP to GTP ratios. We then establish through bioinformatics and activity assays that hybrid CDN-producing and promiscuous substrate-binding (HyPr) GGDEF enzymes are found in *Geobacter* and other deltaproteobacteria. Finally, we validated the predictive power of our analysis by showing that cAG is present in surface grown *Myxococcus xanthus*. This study reveals that GGDEF enzymes make alternative cyclic dinucleotides to cdiG and expands the role of this widely distributed enzyme family to include regulation of cAG signaling.

Results/Discussion

Four cyclic dinucleotides (CDNs) have been discovered to date and are found primarily in bacteria, although recent examples in eukaryotes have further raised the profile of this class of signaling molecules (Chen and Schaap, 2012; Corrigan and Grundling, 2013; Danilchanka and Mekalanos, 2013; Hengge, 2009; Romling et al., 2013). Cyclic di-GMP (cdiG), which has been studied for over 30 years, is the most well characterized CDN and is primarily associated with the transition from the motile planktonic to sessile, biofilm-forming bacterial lifestyles (Hengge, 2009; Romling et al., 2013; Ross et al., 1987). It was first discovered by Benziman as a regulator of cellulose synthase in *Gluconacetobacter xylinus* (Ross et al., 1987). Synthesis of cdiG was later associated with the GGDEF domain (Ausmees et al., 2001; Ryjenkov et al., 2005), which is found in 75% of bacterial species (Seshasayee et al., 2010). On the other hand, cyclic AMP-GMP (cAG or 3',3'-cGAMP) was first discovered as the product of the enzyme DncV in the El Tor strain of *Vibrio cholerae* (Davies et al., 2012). It is a regulator of *V. cholerae* motility and intestinal colonization in mammalian hosts. Other CDNs include cyclic di-AMP (cdiA), which is involved in bacterial cell wall homeostasis and sporulation (Corrigan and Grundling, 2013; Witte et al., 2008), and 2',3'-cyclic AMP-GMP, which is involved in the mammalian innate immune response (Ablasser et al., 2013; Diner et al., 2013; Gao et al., 2013a; Sun et al., 2013; Wu et al., 2013).

We and others recently discovered that *Geobacter*, a genus of deltaproteobacteria, use cAG-sensing riboswitches to regulate genes associated with extracellular electron transfer (Kellenberger et al., 2015; Nelson et al., 2015), an extraordinary activity that involves bacterial colonization on metal oxide surfaces (Reguera et al., 2005). We further showed that *Geobacter sulfurreducens* produces cAG (Kellenberger et al., 2015), but the synthase enzyme remained a mystery. *Geobacter* genomes have no homologues to the cAG synthases DncV or cGAS, which harbor oligoadenylate synthase (OAS)-like domains and produce structurally distinct isomers of cAG (3', 3'-cGAMP and 2', 3'-cGAMP, respectively) (Ablasser et al., 2013; Diner et al., 2013; Gao et al., 2013a; Kranzusch et al., 2013). However, the cAG-sensing riboswitches in *Geobacter* gained function via adapting the ligand binding pocket of GEMM-I riboswitches, which typically bind cdiG (Ren et al., 2015). Thus, we considered that cAG signaling may have evolved in *Geobacter* by co-opting components from the cdiG signaling pathway.

The *Geobacter sulfurreducens* genome encodes 29 GGDEF domain-containing enzymes that are assigned as diguanylate cyclases (Figure 1). These enzymes comprise nearly 1% of the coding sequences. In contrast, there are two predicted DAC domains, which are assigned as diadenylate cyclases. While the existence of GGDEF domains in a genome is considered sufficient proof for cdiG signaling and has been used to establish the presence of cdiG signaling in over 75% of 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 cAG 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 cdiG or cAG (Figure 2A) (Kellenberger et al., 2015; Kellenberger et al., 2013). 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 cdiG biosensor and little to no turn-on with the cAG biosensor, in line with the results for WspR, a well-characterized diguanylate cyclase (De et al., 2008; Hickman et al., 2005). 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 cAG biosensor and little turn-on with the cdiG biosensor. Similar results were observed for DncV, the cAG 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 cAG synthase GSU1658, a candidate diguanylate cyclase GSU1656, or empty plasmid. LC-MS and MS/MS data showed that *E. coli* do not inherently produce cAG, but overexpression of GSU1658 leads to high cAG levels (Fig. 2B, 3). Furthermore, this activity requires an active GGDEF domain, as no cAG was observed with the GLDEF mutant of GSU1658. In contrast, overexpression of GSU1656 leads to high cdiG levels, but no cAG. Together with the *in vivo* screening results, these data reveal GSU1658 as a GGDEF enzyme with unprecedented cAG synthase activity.

The cAG synthase DncV has a monomeric protein fold and a single active site with individual binding sites for ATP and GTP (Kranzusch et al., 2014). Diguanylate cyclases, on the other hand, possess a C2-symmetric active site formed by the homodimeric association of two GGDEF domains, with one GTP bound per monomer (Chan et al., 2004; De et al., 2008). To demonstrate that GSU1658 self-associates, we co-expressed C-terminal His- and HA-tagged monomers and showed that pulldowns with the His-tag isolated the HA-tagged monomer (Figure 4). However, we could not measure a relevant dimerization constant because the wild-type protein is isolated as an auto-inhibited form with cyclic dinucleotides tightly bound to an inhibitory binding site (I-site) in the GGDEF domain (Chan et al., 2004; Christen et al., 2006) (Figures 5, 6). This form is inactive and appears to be mostly monomeric by size-exclusion chromatography (Figure 5), which is similar to observations for PleD, a classic diguanylate cyclase whose domain architecture resembles that of GSU1658 (Chan et al., 2004; Wassmann et al., 2007). Both enzymes contain a response regulator receiver (Rec) domain before the GGDEF domain (Aldridge et al., 2003), as well as an I-site (Chan et al., 2004; Christen et al., 2006). We identified and confirmed the I-site in GSU1658 by purifying the I-site mutant R393A, which did not co-elute with bound cyclic dinucleotides (Figure 5, 6). Unfortunately, the I-site mutant was largely insoluble, even with addition of an MBP tag, as it forms higher order aggregates at low micromolar concentrations *in vitro*, as observed by size exclusion chromatography and isothermal calorimetry experiments (Figures 4, 5). This aggregation phenomenon may be due to misfolding of the R393A variant under *in vitro* conditions, or may be naturally involved in synthase activity, as

oligomerization of the diguanylate cyclase WspR has been observed in *Pseudomonas aeruginosa* to activate enzyme activity (Huangyutitham et al., 2013). In the latter case, WspR oligomerization is controlled by the phosphorylation status of the Rec domain, as opposed to changes in protein expression level. Prior transcriptional profiling of *G. sulfurreducens* grown under diverse conditions (biofilm, electrode, fumarate/acetate, fumarate/ferric citrate) showed that GSU1658 is constitutively expressed (Qiu et al., 2010), so activation of enzyme dimerization or oligomerization likely occurs through the Rec domain as well.

Absence of product bound at the I-site is expected to relieve auto-inhibition of the enzyme. To assay GSU1658 activity *in vitro*, we incubated the enzyme with ATP and GTP doped with trace radiolabeled ATP or GTP and identified the CDNs produced via thin layer chromatography. This assay allows newly synthesized CDNs to be distinguished from pre-bound ones. All GGDEF enzymes were analyzed at micromolar concentrations to favor active dimer or oligomer formation. As expected, the CDN-bound WT GSU1658 was less active *in vitro* than the R393A I-site mutant (Figure 7A). Unexpectedly, however, both the WT and R393A mutant produced significant levels of cdiA and cdiG in addition to cAG. This finding contradicts the cell lysate data in which cAG was the predominant product. Closer inspection of the biosensor and cell lysate results for WT GSU1658, however, did reveal an increase in cdiG signal, albeit at lower levels than observed *in vitro*. We realized this product distribution can be explained by the homodimeric structure of the enzyme, if each of the two identical substrate-binding sites are able to recognize both ATP and GTP. Because R393A and WT GSU1658 had similar cell lysate profiles, we concluded that the I-site mutation has no effect on product ratios (Figure 8). Since we do not know the native oligomeric state of activated WT GSU1658, we cannot rule out that oligomerization affects product distribution.

The discrepancy between *in vivo* and *in vitro* results is likely due to two major factors. First, because cdiG is found naturally in *E. coli*, cdiG-specific phosphodiesterases exist for its breakdown (Paul et al., 2010). The presence of these phosphodiesterases, and the absence of cdiA- or cAG-specific phosphodiesterases, could have led to the lower observed levels of cdiG in the cell lysate. The enzyme TBD1265, for instance, is a canonical EAL domain-containing phosphodiesterase that is 33-fold more selective for cdiG over cAG, with no activity towards cdiA hydrolysis (Shanahan et al., 2013). Furthermore, of the 28 HD-GYP and EAL proteins in *V. cholera* El Tor, only three HD-GYP enzymes showed cleavage activity for cAG (Gao et al., 2015).

Second, the *in vitro* assays were carried out with 1:1 ATP to GTP, but in cells, ATP is usually found in excess relative to GTP. To examine this, we incubated R393A GSU1658 with different ratios of ATP to GTP. As we increased ATP relative to GTP, the product ratio skewed toward cAG and cdiA relative to cdiG (Figures 7B, C, 9). The heterodimeric product, cAG, appeared to be the major product at ATP to GTP ratios between 3:1 and 5:1, in accordance with the measured physiological ranges for enterobacteria (between 2:1 and 3:1) (Bochner and Ames, 1982; Buckstein et al., 2008). Furthermore, it should be noted that published ratios are for total cellular ATP and GTP, whereas the pool of free GTP is likely lower. Taken together, these results reveal that we

have discovered a GGDEF enzyme with dinucleotide synthase activity that produces different CDNs depending on the ratio of ATP to GTP.

We analyzed several features of GSU1658 by mutational analysis. Phosphorylation of the Rec domain is known to activate homodimer formation of canonical GGDEF enzymes such as PleD (Chan et al., 2004; Wassmann et al., 2007), however treatment of GSU1658 with BeF_3, a known phosphomimic compound used to study Rec domains (37), resulted in protein precipitation. Phosphorylation site knock-out (D52A) and mimic (D52E) mutations of GSU1658 in the I-site mutant background had little to no effect on overall enzyme activity (Figure 10A), and the D52E mutation had little to no effect on product ratios (Figure 11). These mutations may not adequately recapitulate the phosphorylated state of the Rec domain in GSU1658, nor the interactions with the cognate histidine kinase, which has not yet been identified.

Sequence alignment and analysis of the x-ray crystal structure of PleD (26) 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 (Figures 10B, 12). 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 13). Thus, we analyzed mutations at this position in the substrate binding pocket of GSU1658 in the I-site mutant background.

Interestingly, the S347A mutant maintained dinucleotide cyclase activity (Figures 10A, 14), so the Ser side chain is not strictly necessary for recognition of ATP or GTP. However, the S347D mutant only makes cdiG, so the Asp side chain restores specificity for GTP. The majority (99%) of annotated GGDEF enzymes harbor an Asp at this position, as do the canonical diguanylate cyclases PleD and WspR(Yan et al., 1999).

Other natural variant residues identified by bioinformatics analysis were analyzed by thin layer chromatography or cell lysate experiments (Figures 10A, 14, 15). In support of our model for side chain-nucleobase interactions, the S347N mutant makes cdiG and the S347T mutant makes all three bacterial CDNs (Figures 10A, 14, 15). These data confirm that this position strongly influences recognition of ATP or GTP. However, the protein background also appears important, as the inverse D344S mutant of PleD was inactive (Figure 16), rather than making all three CDNs as predicted. Likewise, the inverse D-to-S mutations of several *Geobacter* diguanylate cyclases (GSU1400, GSU2313, GSU2534) resulted in enzyme inactivation, while the GSU3350 mutant retained diguanylate cyclase activity (Figure 17). The inability of these inverse mutations to generate a cAG synthase may be due to Ser having a shorter side chain than Asp, so these diguanylate cyclases may not bring the residue in close enough proximity to engage with the substrate. The binding pocket of GSU1658 appears to be more plastic, as it remains active while accommodating both larger (S347D) and smaller (S347A) residues.

We further found that replacing the full GGDEF domain of PleD with the one from GSU1658 generates a functional dinucleotide cyclase (Figure 16). However, the ratio of cdiG to cAG produced by the protein chimera is higher than for GSU1658, which suggests that the difference in conformation between the chimeric and natural homodimers may influence product ratios. Thus, we have identified that a natural amino

acid variation in GGDEF domains can change the specificity of the enzyme, and showed that features outside of the GGDEF domain also affect product ratios.

Whereas previous GGDEF enzymes were uniformly assigned as diguanylate cyclases (DGCs) if they have the active motif [G/A/S]G[D/E]E[F/Y] (Romling 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. Thus, we propose that this newfound sub-family be called HyPr (hybrid, promiscuous) GGDEF enzymes, and that GSU1658 be renamed HyPrA.

To survey this newfound HyPr sub-family, 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 1). 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 12) and have riboswitch effectors that regulate genes in response to cAG (Kellenberger et al., 2015; Nelson et al., 2015). A second HyPr is predicted in *G. sulfurreducens*, but GSU1937 was not active in the biosensor-based screen (Figure 2A). This protein has a predicted transmembrane HAMP domain, so it likely requires membrane insertion for activity and was inactive when expressed heterologously.

Intriguingly, bacteria that do not harbor cAG-selective riboswitches also appear to encode candidate HyPr enzymes in their genomes (Table 1). For example, *Myxococcus xanthus*, a social bacterium that serves as a model for studying group behavior (Nan and Zusman, 2011) and cell-cell communication (Vassallo et al., 2015), harbors one gene homologous to HyPrA (MXAN_4463) and another gene with a GAF sensory domain preceding the HyPr GGDEF catalytic domain (MXAN_2643, renamed HyPrB). Bd0367 (also called DgcA) from *Bdellovibrio bacteriovorus*, an intracellular pathogen of other bacteria described as a “living antibiotic”, has a similar Rec-GGDEF architecture to HyPrA and has been shown to regulate gliding motility and prey escape, a phenotype distinct from those controlled by the other three GGDEF enzymes (Hobley et al., 2012).

These three and other candidate HyPr enzymes were cloned into *E. coli* and analyzed by cell extraction followed by LC-MS. In all tested cases that expressed well (Figure 18), we observed the production of cAG, although to varying levels (Figure 19). In contrast, overexpression of PleD or WspR, two well-characterized diguanylate cyclase GGDEFs, led to production of cdiG only. The cAG to cdiG ratios are different between HyPr enzymes, which reinforces that the Ser specificity residue is not the sole determinant of product ratios. The majority of the HyPr enzymes we tested preferentially make cAG under *in vivo* conditions in which ATP is in excess. Furthermore, we recapitulated that I-site mutants of the *M. xanthus* and *B. bacteriovorus* HyPrA enzymes produce cAG *in vitro* in a similar fashion to GSU1658 at 1:1 ATP to GTP ratios (Figure 18).

Our biochemical validation of cAG synthase activity for diverse HyPr GGDEFs strongly predicted that cAG was an endogenous signaling molecule in these other organisms. We initially found cdiG but not cAG in cell extracts of wild-type *M. xanthus* cultured in

solution. However, based on the involvement of cAG signaling in processes related to surface sensing (intestinal colonization for *V. cholerae* and extracellular electron transfer for *Geobacter*), we hypothesized that cAG was similarly associated with *M. xanthus* growth on solid surfaces. Lysis conditions first required optimization to account for higher EPS content in surface-grown cells, which otherwise impeded successful extraction of all cyclic dinucleotides. This method was used to analyze cyclic dinucleotide content of *M. xanthus* grown in solution versus on 1.5% agar, and revealed that cAG is produced at higher levels upon surface growth (Figures 19, 20). Positive identification of cAG was confirmed by multiple methods, including comparison of HRMS and tandem MS/MS spectra and S1 nuclease digestion profile to synthetic standards (Figures 20-22).

These results provide the first evidence for cAG signaling in *Myxococcus*, and are in line with the conservation of HyPrA and HyPrB across all *Myxococcus* species. In particular, we showed that cAG levels are modulated by solution versus solid growth conditions. This correlates with a proposed role in surface sensing, which provides an advanced starting point for future phenotypic analysis and efforts to discover other components in the putative cAG signaling pathway. Taken together, our results reveal that HyPr activity is more widespread in bacteria than the distribution of cAG riboswitches, expanding the potential scope of cAG signaling (Figure 23).

In searching for the enzyme responsible for cAG synthesis in *Geobacter*, we discovered a sub-family of homodimeric GGDEF proteins that have the capacity to produce all three bacterial cyclic dinucleotides, which we term HyPr GGDEFs. An efficient and selective *in vivo* fluorescent biosensor assay enabled the discovery of this novel activity, highlighting RNA-based biosensors as an attractive technology platform for enzyme screening. Alternative screening strategies such as fractionation (Sun et al., 2013; Wu et al., 2013), *in vitro* enzyme screening (Corrigan et al., 2013), or a phenotype-based assay in the native organism (Lori et al., 2015) are relatively more time- and resource-intensive. The *in vivo* biosensor assay is particularly useful for signal transduction enzymes that are activated by dimerization, because this can be mimicked by overexpression in a heterologous host.

Unlike the other cAG synthases, DncV and cGAS, which have individual binding sites for ATP and GTP, HyPr GGDEFs have two symmetrical binding sites that accommodate either ATP or GTP as substrates. Notably, both DncV and cGAS display weak activity for synthesizing cdiG or cdiA *in vitro* as well, especially in the presence of only one type of substrate (Ablasser et al., 2013; Davies et al., 2012; Kranzusch et al., 2014). However, these enzymes preferentially make cAG (3',3' or 2',3', respectively) at equimolar concentrations of NTPs, whereas the ability of HyPrA to make cAG is tuned by the relative excess of ATP to GTP under physiological conditions.

One outstanding question is whether HyPr GGDEFs are responsible for cAG signaling only, or could the ability of these enzymes to produce all three cyclic dinucleotides be modulated and exploited by bacteria that use them as endogenous second messengers for distinct signaling pathways. While we currently cannot exclude the latter possibility, the former possibility is more straightforward and consistent with current models for

cdiG and cdiA signaling pathways, which are composed of separate sets of enzymes and effectors. In *Geobacter*, we propose that HyPrA (GSU1658) and perhaps GSU1937, which also harbors the D-to-S variation, act as cAG synthases *in vivo*, and activation of these signaling enzymes turn on genes involved in extracellular electron transfer through cAG-specific riboswitches (Figure 23) (Kellenberger et al., 2015). The discovery of HyPrA brings us closer to elucidating a primary signal that regulates this unique metabolic activity, as this enzyme harbors a Rec domain and thus is presumably part of a two-component system. In general, further work is needed to elucidate the activation mechanisms for HyPr GGDEFs and to address whether HyPr GGDEFs have promiscuous functions.

Through analysis of sequence determinants for HyPrA activity and validation of additional HyPr GGDEFs, we predicted that these enzymes are more broadly distributed than cAG riboswitches, which have been the only identified sensors for cAG to date (Figure 23). HyPr GGDEFs appear highly conserved in the bacterial orders Desulfuromonadales (includes *Geobacter*), Myxococcales, and Bdellovibrionales. Additionally, a number of bacteria in the Acidobacteria and Deferribacter phyla also contain candidate HyPr GGDEFs. Based on our demonstration of endogenous cAG production in wild-type *Myxococcus xanthus*, we expect that cAG signaling also is present in these other bacteria due to HyPr GGDEF activity, although other components remain to be identified. Another HyPr GGDEF that we showed *in vitro* is capable of producing cAG is Bd0367 from *Bdellovibrio bacteriovorus* (Figure 4). Prior studies highlighted the conundrum for how Bd0367, which was called DgcA, could give rise to a distinct phenotype from the other three GGDEF enzymes, given that the small size of the bacteria “make c-di-GMP spillover unavoidable” (Hobley et al., 2012). Based on our discovery of HyPr activity, one plausible explanation is that this enzyme is a HyPrA and synthesizes a different signal than the other GGDEFs. Interestingly, gliding motility in *B. bacteriovorus* is associated with type IV pili, whose assembly is predicted to be controlled by cAG riboswitches in *Geobacter* and which is known to play a role in surface-associated twitching motility in *Myxococcus (Nan and Zusman, 2011)*, suggesting a unified connection between cAG and type IV pili-mediated processes.

Following our original hypothesis that components of the cdiG signaling pathway were co-opted for cAG signaling, we predict that variant PilZ domains and potentially novel effectors exist in these cAG-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 delta proteobacteria (Seshasayee et al., 2010), and at least one variant HD-GYP domain has been shown to degrade cAG (Gao et al., 2015). Besides synthesizing cAG, we also showed that GGDEF domains can make cdiA, an activity that had been speculated (Nelson et al., 2013), 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.

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 S3). Cyclic dinucleotide standards were purchased from Axxorra (Farmingdale, NY) or enzymatically synthesized. DFHBI was chemically synthesized following literature protocols (Paige et al., 2011).

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 from reference (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.

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 ºC 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 cdiG biosensor, the significance test was between candidate GGDEF signal and pCOLA signal. The cAG biosensor is ~100-fold selective for cAG over cdiG, but some fluorescence above background is still observed for cdiG synthases. Thus, for the cAG 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 ºC for 4 h. Cells were harvested by centrifugation at 4,700 rpm for 15 min at 4 ºC, and pellets were stored at -80 ºC.

Cell Extraction of *E. coli*

Cyclic dinucleotides 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 1260 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 m 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 cyclic dinucleotides in extracts were found to always elute in the order of cdiG (7.3±0.3 min), cAG (7.6±0.3), and cdiA (7.9±0.4 min). Due to slight variability in retention times, the assignment of cyclic dinucleotide 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 cyclic dinucleotides (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 cyclic dinucleotide was present, but was not the most abundant ion peak observed, even with the standards. This observation suggests that the relative ionization of cyclic dinucleotides is low under these conditions, and furthermore the cyclic dinucleotides 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 cyclic dinucleotides.

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 et al., 2015). High-resolution mass spectrometry (HRMS) and tandem mass spectrometry (MS/MS) measurements of collected fractions were performed as previously described (Kellenberger 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.

Overexpression and Purification of Dinucleotide Cyclases

Full-length proteins with N-terminal His-6-MBP tags encoded in a pET16-derived plasmid were overexpressed in *E. coli* BL21 (DE3) star cells harboring a pRARE2 human tRNA plasmid and were grown in LB/carb/chlor for 10 h after induction at OD_{600} ~ 0.7 with 1 mM IPTG. Cells were lysed by sonication in a lysis buffer containing 25 mM Tris-HCl (pH 8.2), 500 mM NaCl, 20 mM imidazole, and 5 mM beta-mercaptoethanol. Clarified lysate was bound to Ni-NTA agarose (QIAGEN), and resin was washed with lysis buffer prior to elution with lysis buffer supplemented with 500 mM imidazole. Proteins were dialyzed overnight at 4 ºC against buffer containing 25 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5% (v/v) glycerol, and 1 mM DTT. Protein purified in this way was concentrated to ~5-10 mg/mL, flash frozen in liquid nitrogen, and stored at -80 ºC. Protein with C-terminal His-6x tags encoded in pET24a were overexpressed and purified similarly, with the cells grown in LB/kan/chlor.

Detection of Homodimerization

After inoculating fresh LB cultures from overnight starter cultures of *E. coli* BL21 (DE3) Star cells containing the pRARE2 plasmid (Invitrogen), C-terminal His-tagged enzymes in pET24a or pET28a, and C-terminal HA-tagged enzymes in pET22b, cells were grown aerobically to an OD_{600} ~ 0.3, then induced with 1 mM IPTG at 28 ºC for 4 h. Cells were harvested by centrifugation at 4,700 rpm for 15 min at 4 ºC, and pellets were stored at -80 ºC. The cells were resuspended in lysis buffer (25 mM Tris-HCl (pH 8.2), 500 mM NaCl, 20 mM imidazole, and 5 mM beta-mercaptoethanol), and lysed using a Biospec MiniBeadBeater-16. Lysates were cleared by centrifugation at 4 ºC at 13,000 rpm for 45 min. Clarified lysates were suspended in Ni-NTA resin (QIAGEN). The resin was washed five times with lysis buffer. The proteins were eluted with lysis buffer supplemented with 500 mM imidazole. The epitope-tagged HA protein was detected with HRP-conjugated BMG-3F10 anti-HA rat antibody (Roche, 1:1000). Validation information for the antibody is available on the manufacturer website.

Isothermal Titration Calorimetry

Samples of MBP-tagged GSU1658 R393A were first dialyzed into buffer containing 50 mM Tris-HCl [pH 7.5], 100 mM NaCl, and 1 mM TCEP, and then concentrated using a 5 mL Amicon Ultra MWCO 10 kDa concentrator (Millipore). Protein samples and buffer were then loaded onto a Microcal Auto-iTC200 isothermal titration calorimeter (Malvern, Worcestershire, UK). Details of the dissociation experiment and subsequent analysis have been described previously (Velazquez-Campoy et al., 2004; Wassmann et al., 2007). In short, 3.1 µL samples of MBP-GSU1658 R393A (at concentrations of 111 µM and 119 µM) were injected into the cell filled with buffer (400 µL volume) at 4 min intervals. The measurement was obtained at 28 ºC. Both measurements of each state were analyzed with Origen software (OrigenLab) to obtain ΔH and K_D values for the dissociation reaction. An additional fit parameter was also used upon data evaluation to

eliminate constant background heat produced by technical effects and dilution of titrant. However, in all cases, best fit curves corresponded to enthalpic changes in excess of 100 kcal/mol, likely a result of aggregate dissociation instead of dimer dissociation.

Size Exclusion Chromatography

GSU1658 was monitored by size-exclusion chromatography (SEC) by using a Superdex 200 HiLoad 26/60 column (GE Healthcare). Dialysis buffer (100 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM DTT, 5% glycerol) was used as running buffer for the protein samples. Runs were performed on an ÄKTA*purifier* (GE Healthcare) FPLC system at a flow rate of 0.5 ml/min.

***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 (Kranzusch et al., 2013). For full procedure, see SI Appendix. 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 ºC 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 ºC for 30 min to digest the unincorporated NTPs. Reactions were terminated by heating to 95 ºC 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 (Spangler et al., 2010). After termination of the reactions by heating to 95 ºC 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 cyclic dinucleotides in extracts were found to always elute in the order of cdiG (8.7 ± 0.3 min), cAG (10.6 ± 0.3 min), and cdiA (11.0 ± 0.5 min). Due to slight variability in retention times, the assignment of cyclic dinucleotide 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 cyclic dinucleotides (8 to 12 min).

***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 ºC 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 cyclic dinucleotides in extracts were found to always elute in the order of cdiG (8.7 ± 0.3 min), cAG (10.6 ± 0.3 min), and cdiA (11.0 ± 0.5 min). Due to slight variability in retention times, the assignment of cyclic dinucleotide 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 cyclic dinucleotides (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.

Growth of *Myxococcus xanthus*

Wild type (DZ2) *M. xanthus* was cultured at 32 ºC in liquid CYE medium (Campos et al., 1978) to an OD_{600} ~0.6. For liquid grown cultures, cells were pelleted and frozen at -80 ºC until use.

For cultures grown on agar, liquid culture (25 mL at OD_{600} ~0.6) was poured onto a CYE plate (150 mm diameter, 1.5% (w/v) agar), and incubated for 24 h at 32 ºC to allow the cells to settle onto the agar surface. Excess liquid culture was then discarded. The cells attached to the agar surface were incubated at 32 ºC for another 24 h before being harvested using a cell scraper.

Cell Extraction of *Myxococcus xanthus*

To fragment the exopolysaccharide matrix for analysis, a frozen cell pellet (for details on cell growth, see SI Appendix) was weighed, thawed, and resuspended in water (2 mL) supplemented with DNase I (20 mg, Sigma) and Lysozyme (20 mg, Sigma). The cell solution was incubated at 37 ºC with agitation for 6 h. Cells were lysed and extracted by addition of acetonitrile (4 mL) and methanol (4 mL). After centrifugation for 45 min at 10,000 rpm, the supernatant was carefully removed and stored at -20 ºC. Supernatant was evaporated to dryness by rotary evaporation, and the dried material was suspended in 400 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 ºC.

LC-MS analysis of *M. xanthus* cell extracts was performed using an Agilent 6530 Accurate-Mass Q-TOF LC-MS with an Agilent 1290 Infinity UHPLC. This instrumentation is located in the laboratory of Professor Michelle Chang at UC Berkeley. Samples were separated on a Poroshell 120 SB-Aq column (50 mm length x 2.1 mm

internal diameter, 2.7 μm particle size, Agilent) at a flow rate of 0.6 mL/min. For analysis of cell extracts, a linear elution program of 0 to 20% B over 4 min with an initial hold at 0% B for the first 0.2 min was used. Solvent A was H2O + 0.1% formic acid and solvent B was MeCN. MS data were collected from 0.9 to 2.4 min.

Extract samples were analyzed by MS in the positive ion mode using the range of m/z = 50 to 1100 or 1700. MS/MS measurements were performed with a fragmentation voltage of 150 V and a collision energy of 20 V.

Acknowledgements

The authors would like to acknowledge Anthony T. Iavarone from the QB3 / Chemistry Mass Spectrometry Facility for assistance with mass spectrometry experiments and Yichi Su for helpful discussion of the manuscript. This work was supported in part by NIH grant DP2 OD008677 (M.C.H.) and NSF graduate fellowship (Z.F.H.). M.C.H. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund.

Figures

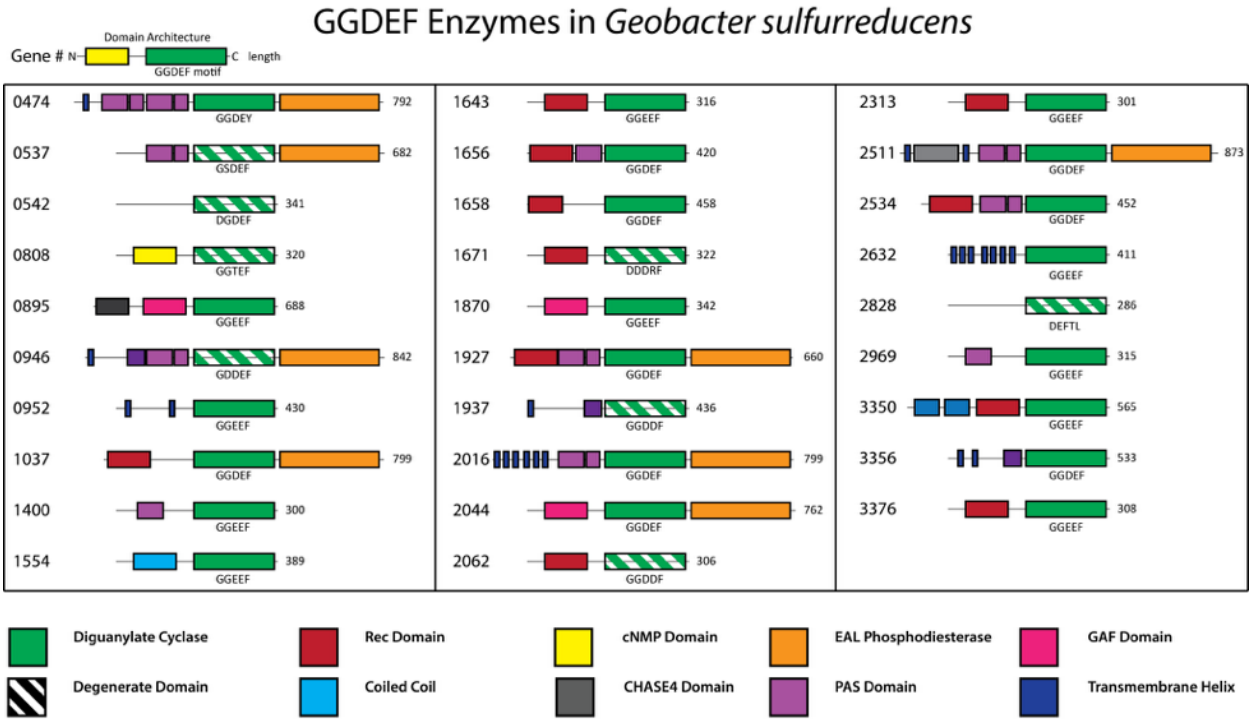


Figure 1 – Domain Architecture of GGDEF-Domain Containing Proteins in *Geobacter sulfurreducens* PCA. Proteins tested for cdiG- and cAG-synthase activity are shown. REC, response receiver regulator domain found in two-component regulatory systems; cNMP, cyclic nucleotide monophosphate binding domain; EAL, cdiG-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.

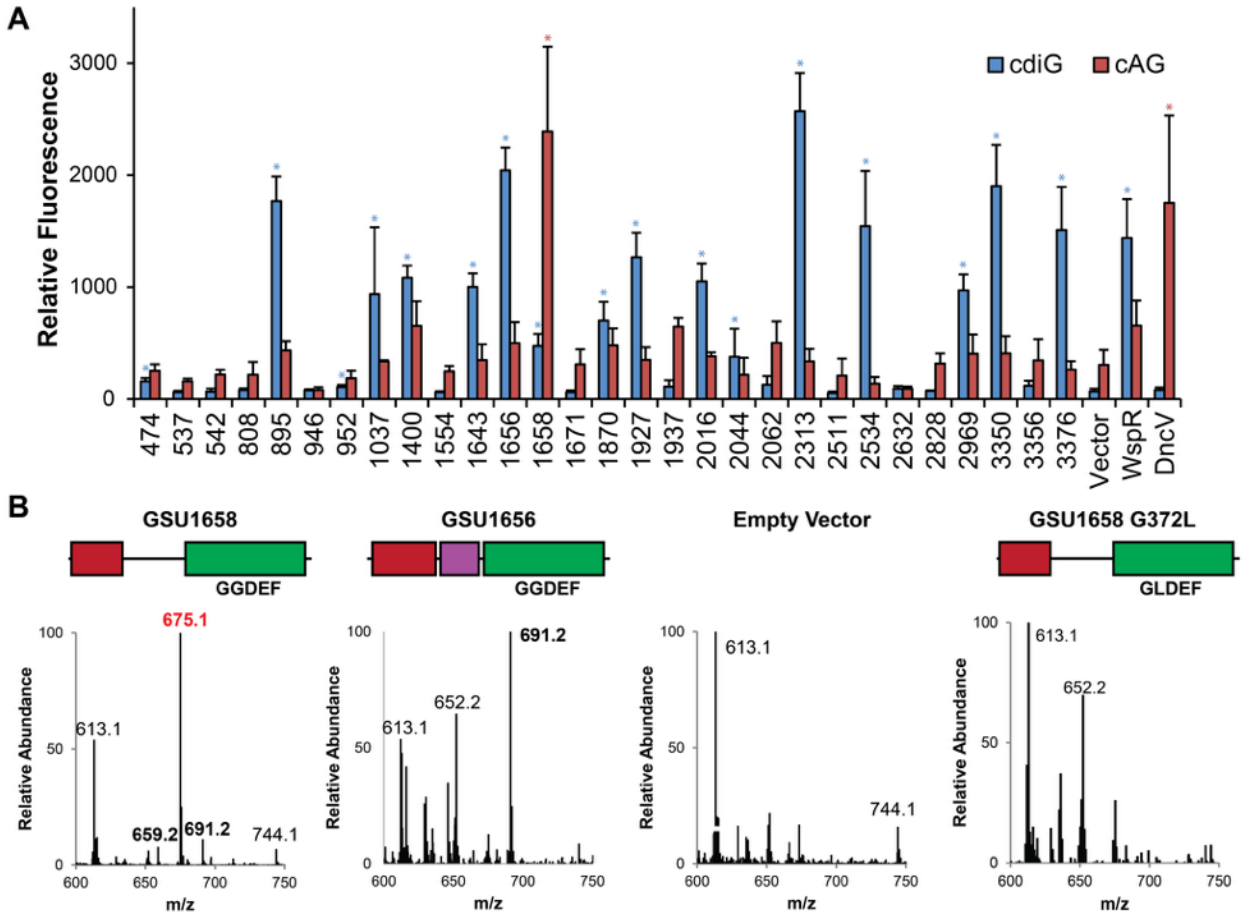


Figure 2 - *In vivo* Fluorescent Biosensor Screen of 29 *Geobacter* GGDEF Genes Reveals a cAG Synthase.

(A) Average fluorescence measured by flow cytometry (n=3, 10,000 cells per run) of *E. coli* BL21 (DE3) Star cells co-expressing the cdiG-selective biosensor Dp-Spinach2 (blue) or cAG-selective biosensor GM0970-p1-4delA-Spinach (red) along with GGDEF domain proteins from *Geobacter sulfurreducens* PCA, diguanylate cyclase WspR, cAG 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 cdiG sensor; above WspR for the cAG sensor).

(B) LC/MS analysis of *E. coli* cell extracts overexpressing constructs shown or empty vector, see Fig. S1 for protein domain color scheme. Shown are the MS spectra from integrating the retention time region containing all three cyclic dinucleotides. Expected masses are for cdiG (m/z = 691), cAG (m/z = 675), and cdiA (m/z = 659).

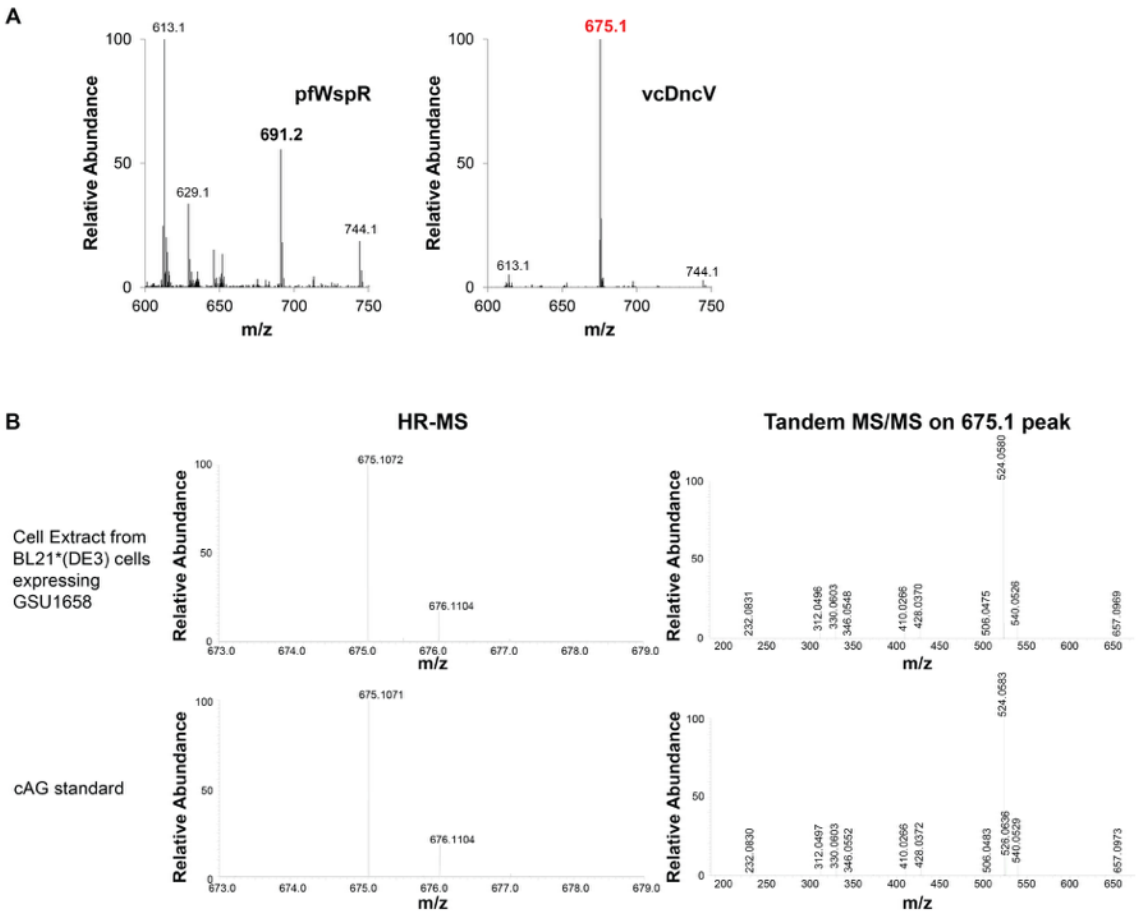
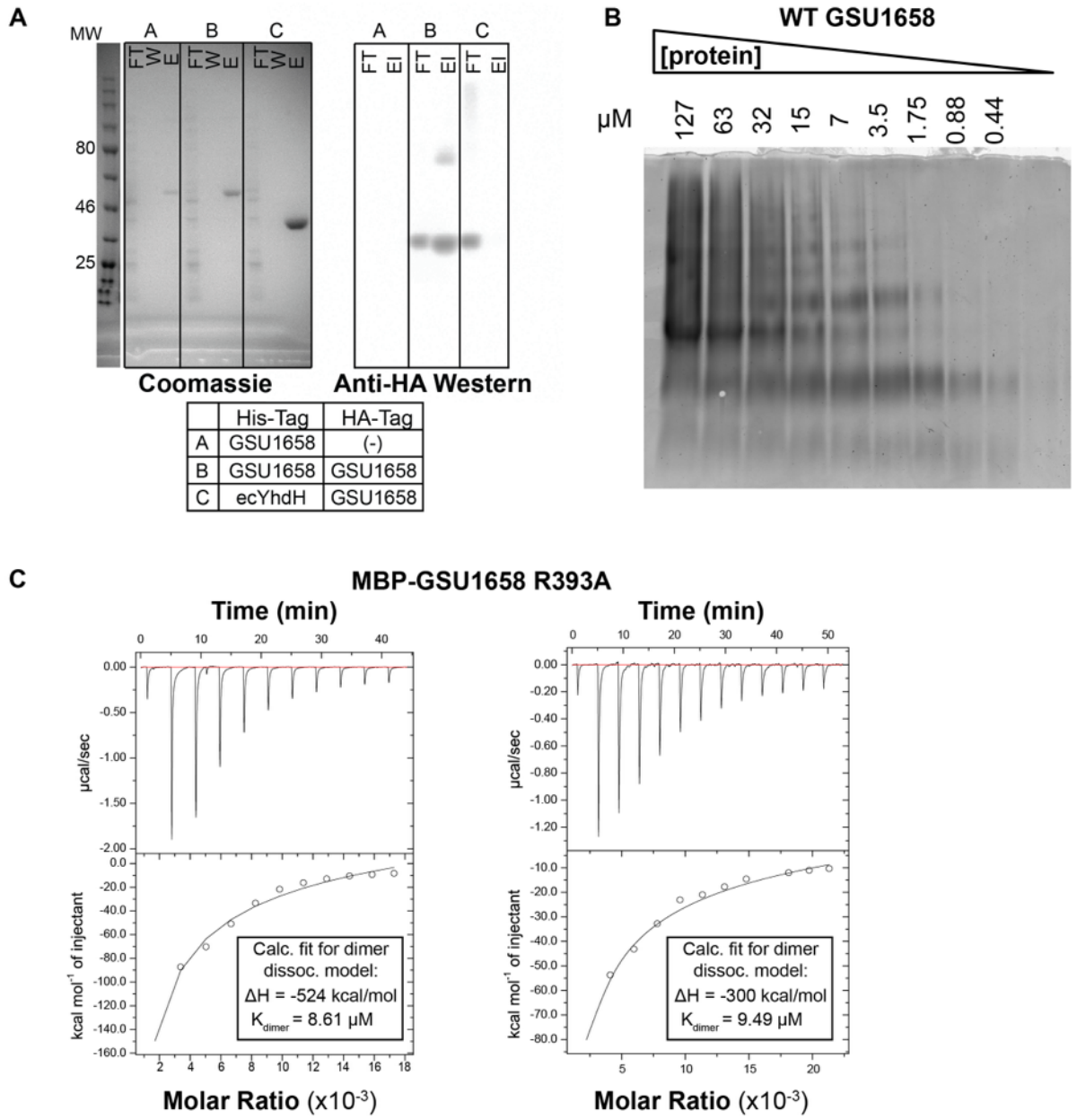


Figure 3 – Cell Extraction of Enzyme Standards and GSU1658. (A) HPLC-MS of lysates from BL21 star (DE3) cells expressing control enzymes. Left: WspR, a cdiG (m/z = 691.1) synthase; right: DncV, a cAG (m/z = 675.2) synthase. **(B)** Mass spectrometry analysis of lysate from BL21 star (DE3) cells expressing GSU1658 or synthetic 3',3'-cAG standard. Left: High-resolution mass-spectrometry; Right: Tandem MS/MS of the 675.1 peak observed.



	His-Tag	HA-Tag
A	GSU1658	(-)
B	GSU1658	GSU1658
C	ecYhdH	GSU1658

Figure 4 – Analysis of GSU1658 Dimerization. (**A**) Pull-down of differentially tagged GSU1658 constructs. His-tagged WT GSU1658 or BL21-derived YhdH was co-expressed with a plasmid control containing no enzyme or HA-tagged WT GSU1658, and the cell lysate purified by Ni-NTA affinity chromatography. Samples were then immunoblotted against the HA epitope. (**B**) Native PAGE analysis of GSU1658-6xHis conformers. Samples were visualized with SYPRO®-Orange (Life Technologies) staining. (**C**) Dilution ITC of MBP-GSU1658 R393A mutants. GSU1658 at 111 and 119 μM were diluted into buffer and the heat change measured over multiple rounds of injection. Enthalpy change and K_{dimer} values were obtained by fitting heat changes to a dimer dissociation model; the high enthalpy values suggest this model is incorrect.

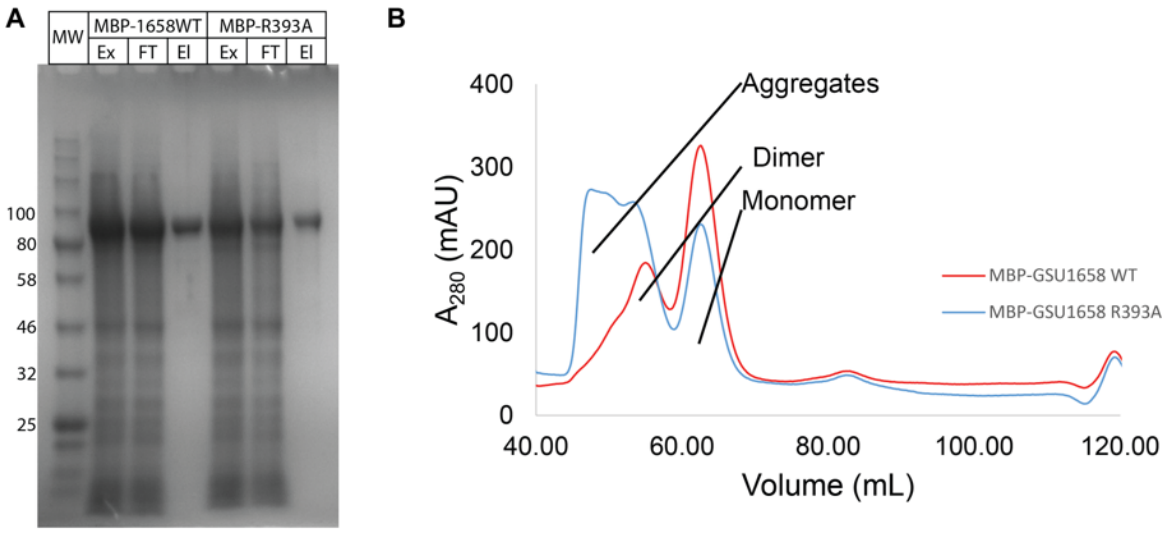


Figure 5 – Purification of MBP-tagged GSU1658. (**A**) SDS-PAGE gel analysis of fractions from the purification of MBP-tagged WT- and R393A-GSU1658. Gels were stained with GelCode Blue (Thermo Scientific). Ex, extract; FT, flow-through; El, elution. (B) Size-exclusion chromatography of MBP-tagged variants of GSU1658. MBP-GSU1658 (WT or R393A, 50 µM, 1 mL) was analyzed by SEC. Shown is the A_{280} trace starting at the void volume of ~40 mL.

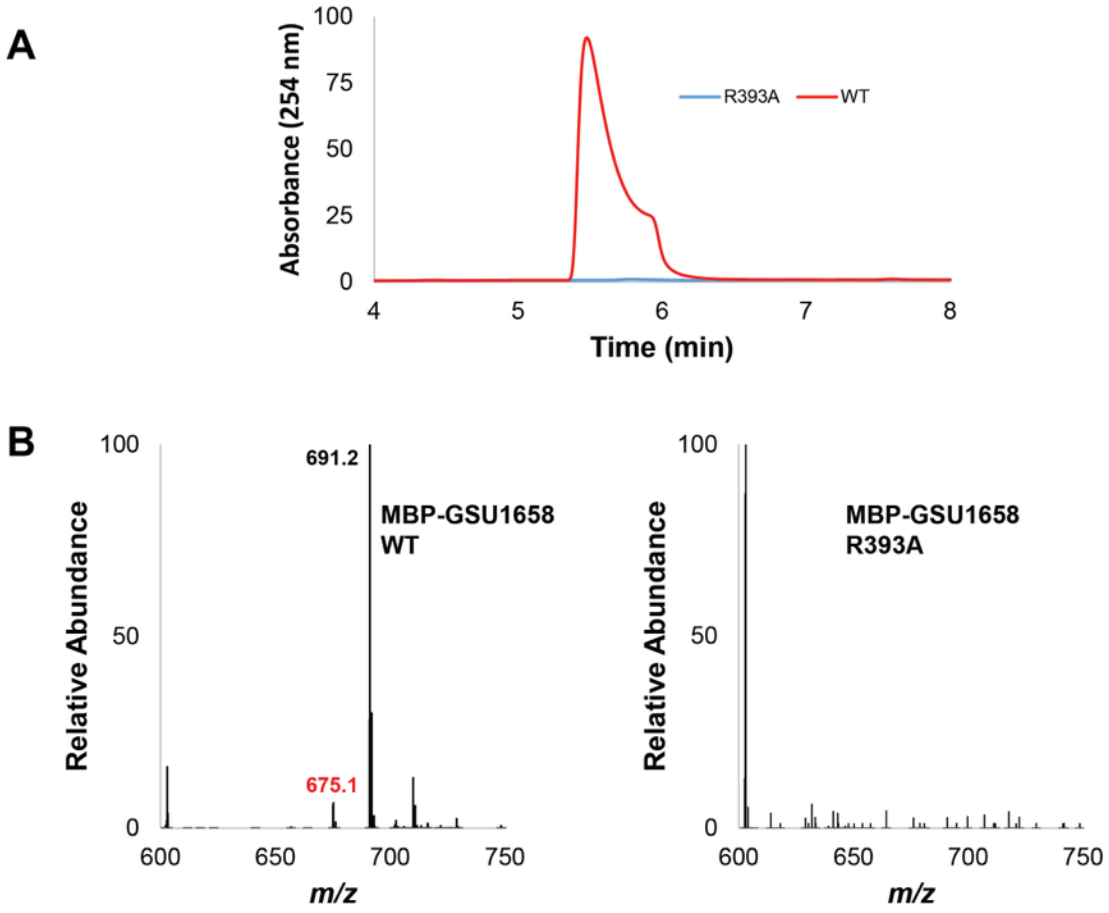


Figure 6 – R393A Mutation of GSU1658 Ablates CDN Binding to I-site. (**A**) Overlay of UV-Vis spectra and (**B**) MS spectra from HPLC-MS analysis of nucleotides bound to MBP-tagged GSU1658 constructs. The MS spectra are from integrating the retention time region containing all three cyclic dinucleotides (6 to 8 min). Expected masses are for cdiG (m/z = 691) and cAG (m/z = 675).

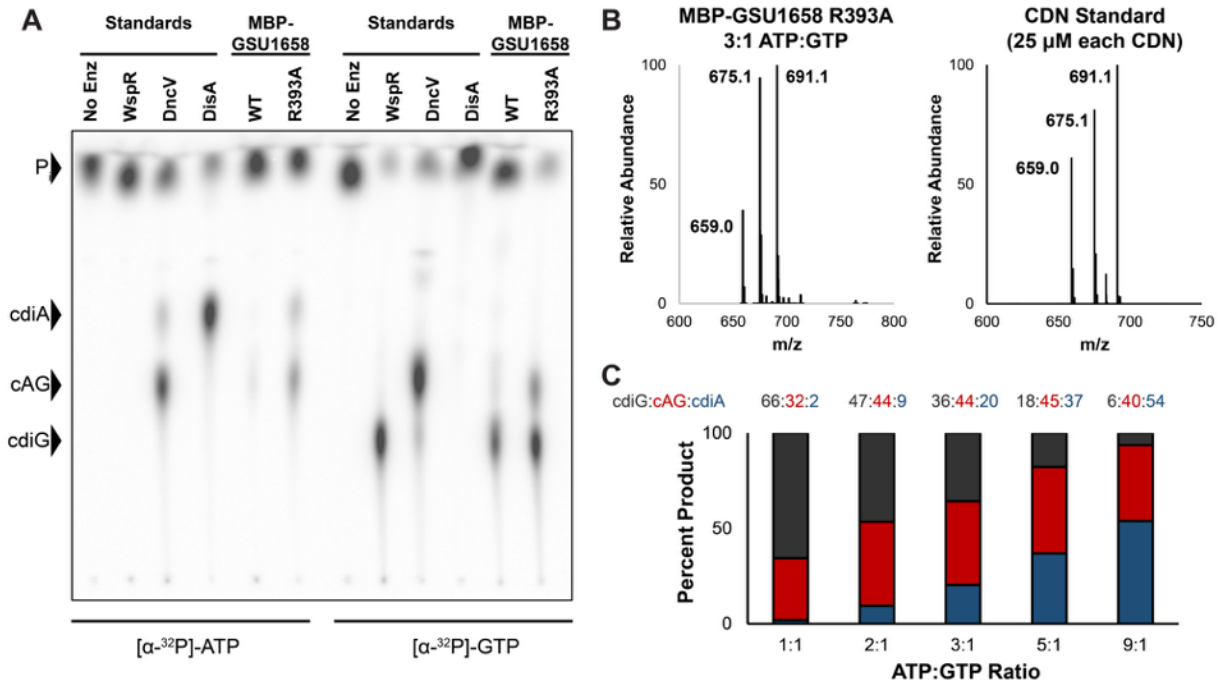


Figure 7 - GSU1658 produces different cyclic dinucleotides depending on ATP to GTP ratios.

(A) Cellulose thin layer chromatography analysis 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 ATP or α-^{32}P-labeled GTP. Prior to loading, reactions were quenched with alkaline phosphatase to digest unreacted nucleotides, resulting in production of inorganic phosphate (P_i). Residue R393 is located in the putative I-site.

(B) Representative LC/MS analysis of an enzymatic assay performed with MBP-GSU1658 R393A at 3:1 ratio of ATP to GTP in comparison to a standard containing all three product CDNs at equal concentrations. Shown is the MS spectrum from integrating the retention time region containing all three cyclic dinucleotides. Expected masses are for cdiG (m/z = 691), cAG (m/z = 675), and cdiA (m/z = 659).

(C) Analysis of product ratios for MBP-GSU1658 R393A at different ATP to GTP ratios based on LC/MS analysis and comparison to CDN standard to account for different ionization efficiencies. Average values of 2 replicate runs are shown.

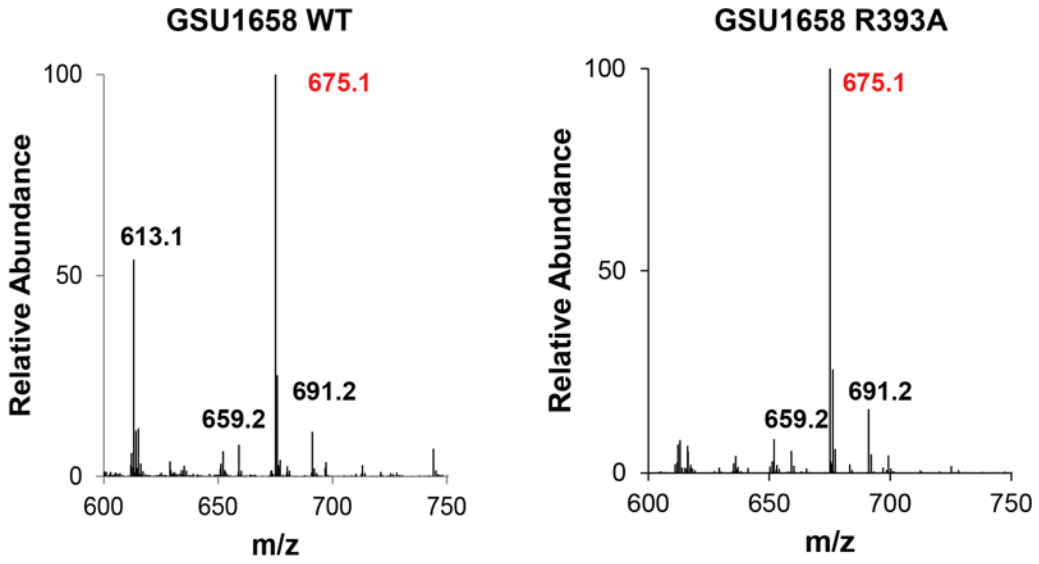


Figure 8 – HPLC-MS Analysis of Lysate from Cells Expressing the GSU1658 Wild-type and R393A Mutant. The MS spectra shown integrates the retention time region containing all three cyclic dinucleotides (6 to 8 min). Expected masses are for cdiG (m/z = 691), cAG (m/z = 675), and cdiA (m/z = 659).

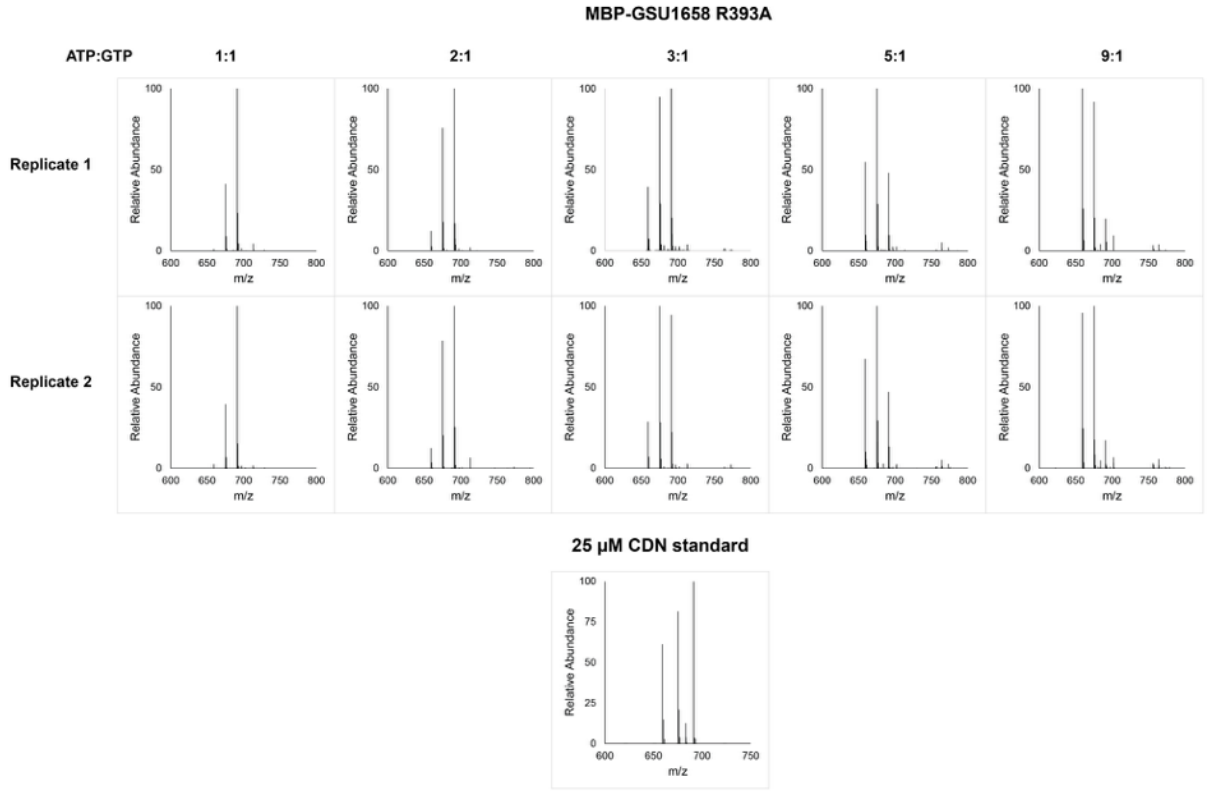


Figure 9 – LC-MS Analysis of CDN Product Distribution for MBP-tagged GSU1658 R393A Mutant. LC-MS analysis of *in vitro* enzyme reactions with varying ratios of ATP to GTP. Shown is the MS spectra from integrating the retention time region containing all three cyclic dinucleotides (6 to 8 min). Expected masses are for cdiG (m/z = 691), cAG (m/z = 675), and cdiA (m/z = 659).

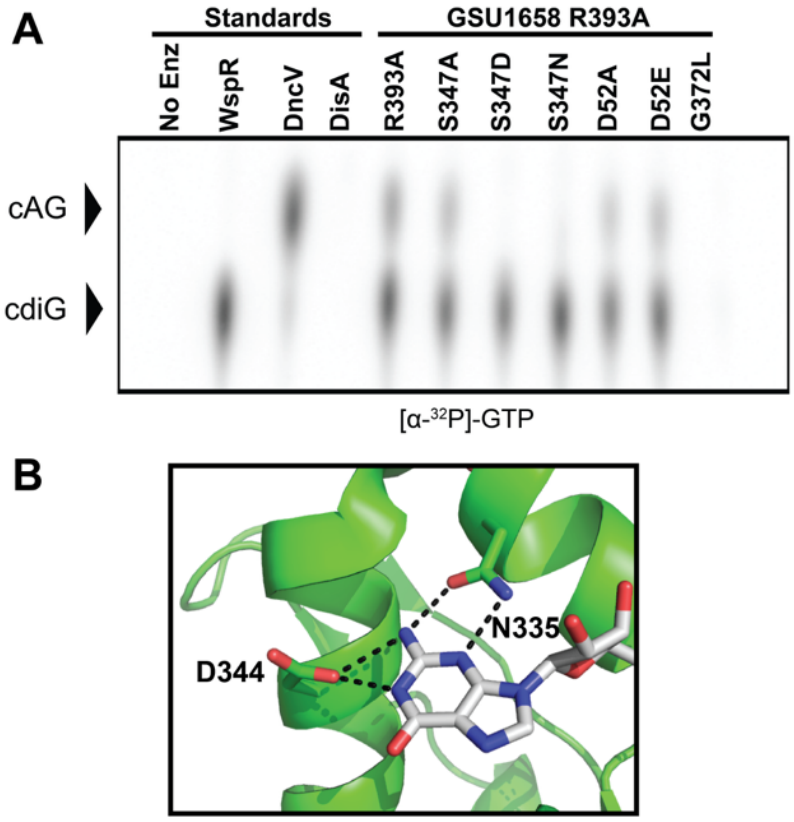


Figure 10 - Identification of Specificity Position in HyPr GGDEF Active Site.

(A) Cellulose thin layer chromatography 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 (Full TLC and results for α-^{32}P-labeled ATP are shown in Figure 14). Prior to loading, reactions were quenched with alkaline phosphatase to digest unreacted nucleotides, resulting in production of inorganic phosphate (Pi). 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, 22). Hydrogen bonding contacts between the guanine base and key protein residues are shown as dotted lines.

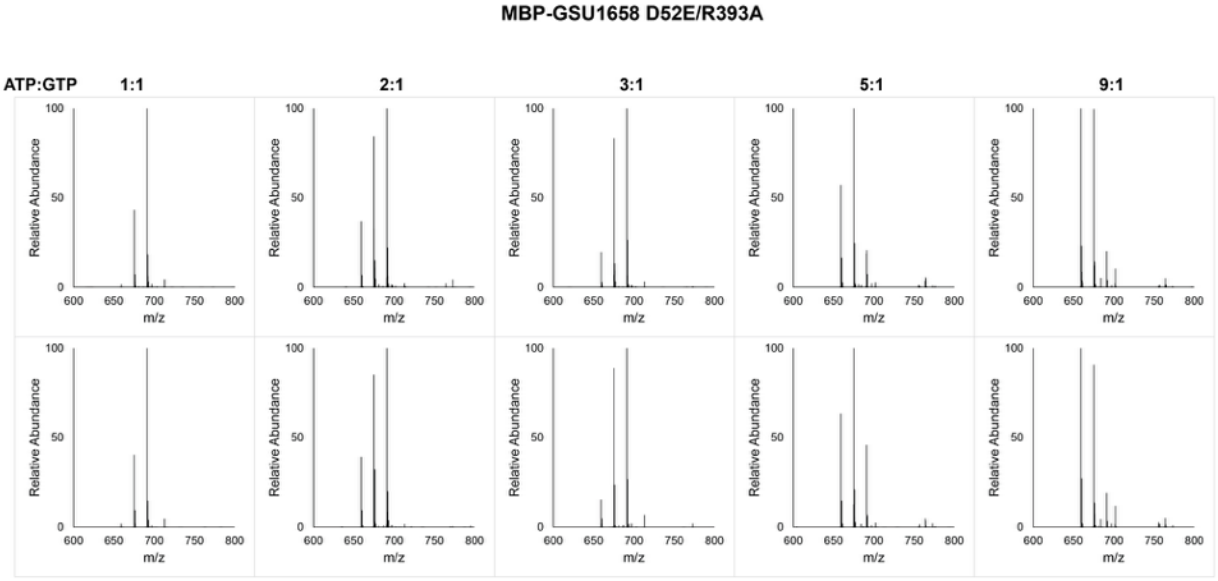


Figure 11 – LC-MS Analysis of CDN Product Distribution for MBP-tagged GSU1658 D52E/R393A Mutant. LC-MS analysis of *in vitro* enzyme reactions with varying ratios of ATP to GTP. Shown is the MS spectra from integrating the retention time region containing all three cyclic dinucleotides (6 to 8 min). Expected masses are for cdiG (m/z = 691), cAG (m/z = 675), and cdiA (m/z = 659).

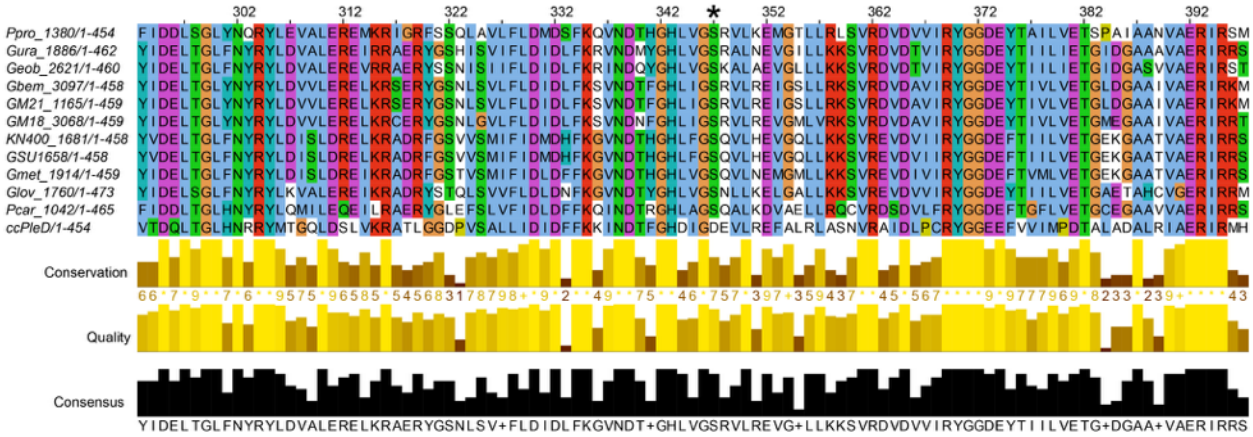


Figure 12 - Alignment of Representative *Geobacter* GGDEF Domains. Sequence alignment of the GGDEF domain of PleD, a canonical diguanylate cyclase, with all HyPrA GGDEF domains in sequenced *Geobacter* and *Pelobacter* species. The encoding gene is conserved and found in the same genomic location, 5' 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).

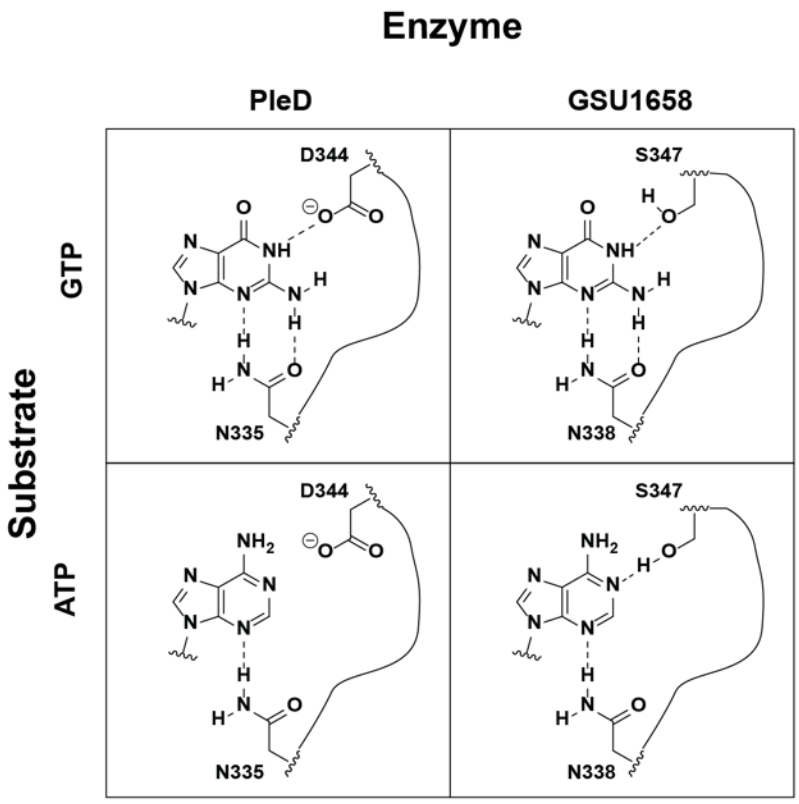


Figure 13 - Proposed Model for Purine Nucleotide Binding by PleD Versus GSU1658.

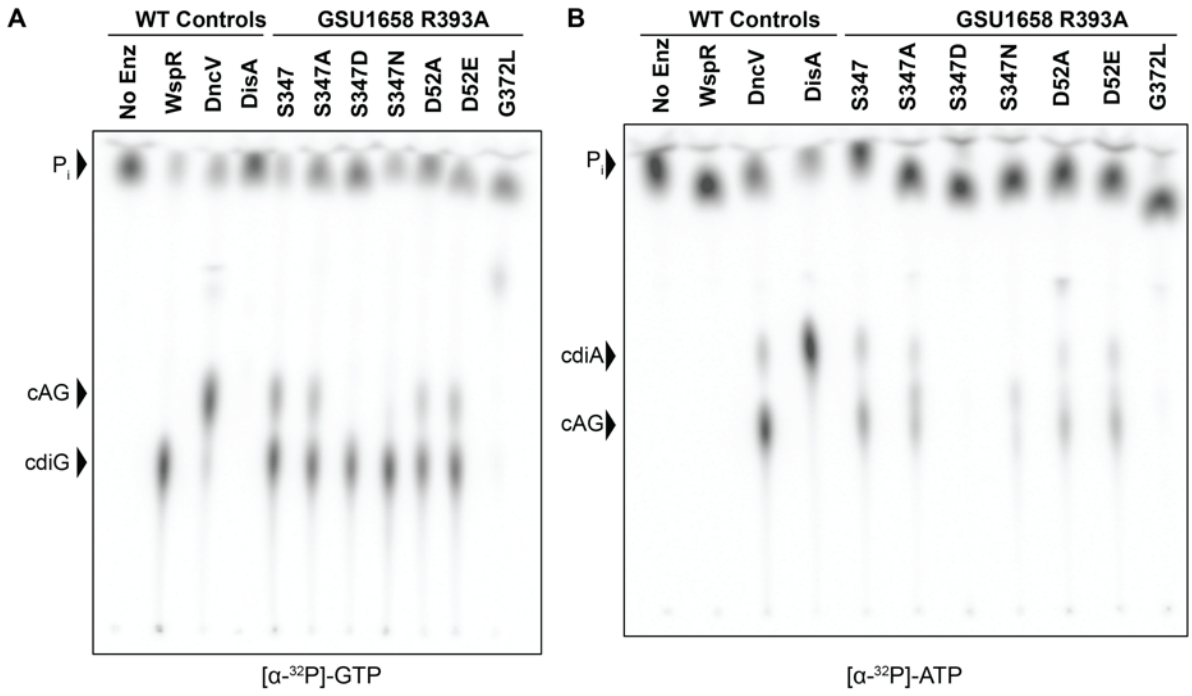


Figure 14 - Activity Assay of GSU1658 Mutants with Radiolabeled NTPs. Cellulose thin layer chromatography of radiolabeled products from enzymatic reactions with 1:1 ATP to GTP substrates in excess and doped with trace amounts of α-^{32}P-labeled (**A**) GTP or (**B**) ATP. (**A**) is the full TLC plate for the inset shown in Figure 3. The positions of the cyclic dinucleoide products (cAG, cdiG, cdiA) and inorganic phosphate (P_i) are marked.

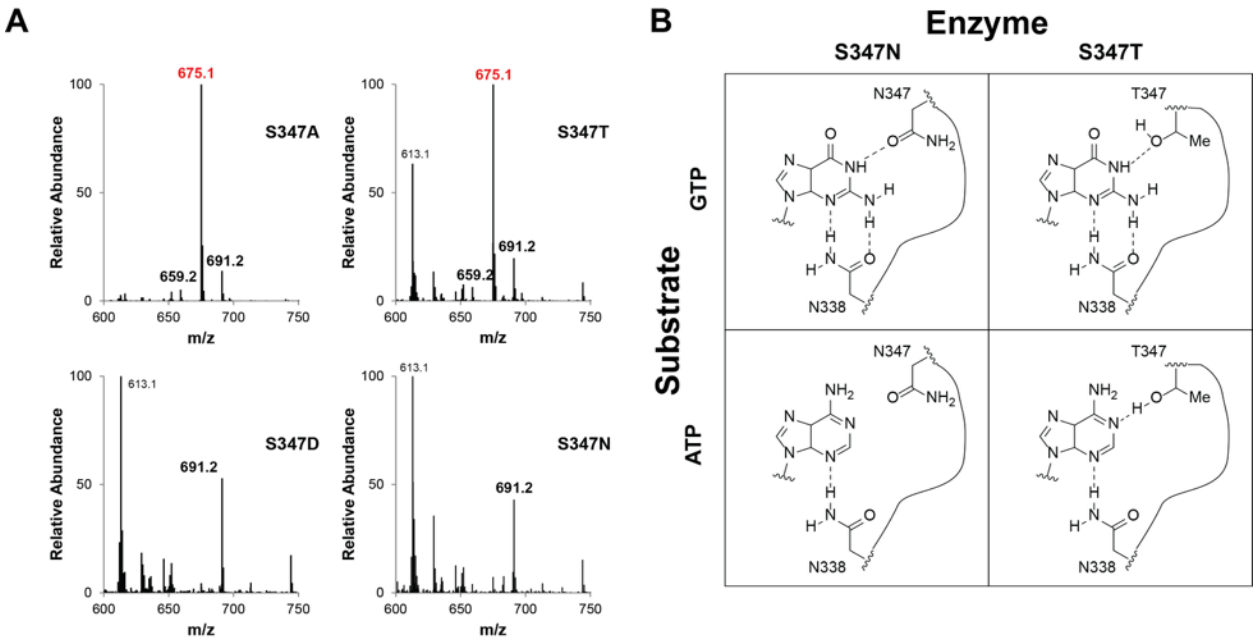


Figure 15 - Analysis of Specificity Residue Mutations in GSU1658. (**A**) LC-MS analysis of extracts from cells expressing GSU1658 single mutants in pET24a with C-terminal 6x-His tags. (**B**) Proposed model for purine nucleotide binding by S347N and S347T GSU1658 mutants.

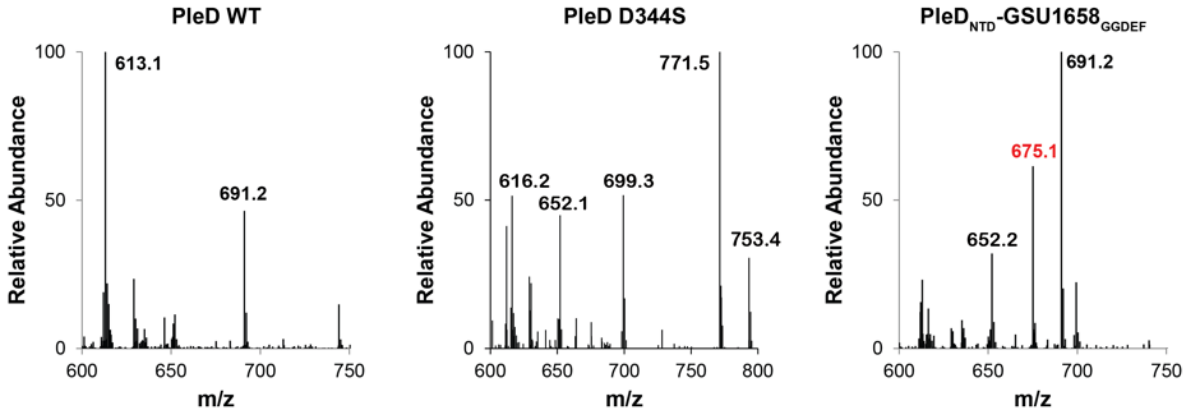


Figure 16 – HPLC-MS Analysis of PleD and Associated Mutants. LC/MS analysis of *E. coli* cell extracts overexpressing PleD variants as shown. PleD$_{NTD}$-GSU1658$_{GGDEF}$ is a fusion between residues 1-293 of PleD, and residues 297-458 of GSU1658. Shown is the MS spectra from integrating the retention time region containing all three cyclic dinucleotides (6 to 8 min). Expected masses are for cdiG (m/z = 691) and cAG (m/z = 675).

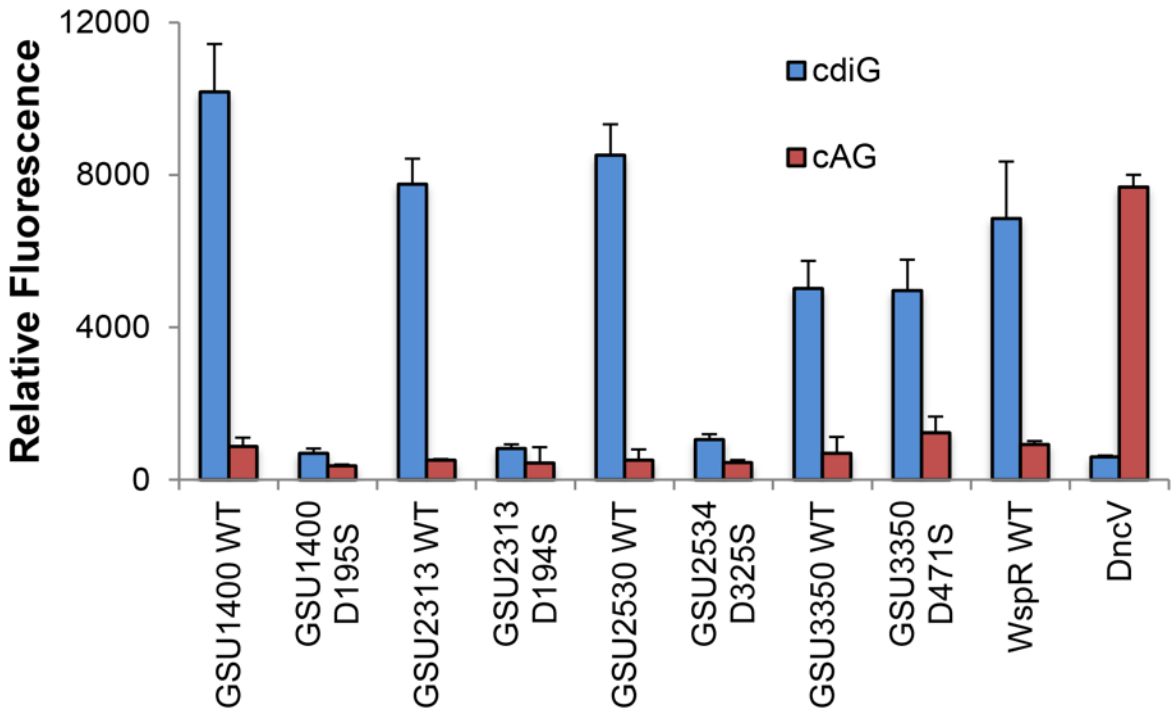


Figure 17 - Analysis of D-to-S Mutations of Several Diguanylate Cyclases from *Geobacter*. Average fluorescence measured by flow cytometry (n=3, 10,000 cells per run) of E. coli BL21 (DE3) Star cells co-expressing the cdiG-selective biosensor DP17-Spinach2 (blue) or cAG-selective biosensor GM0970-p1-4delA-Spinach (red) along wild-type or selectivity site D-to-S mutants of validated diguanylate cyclases from *Geobacter sulfurreducens* PCA.

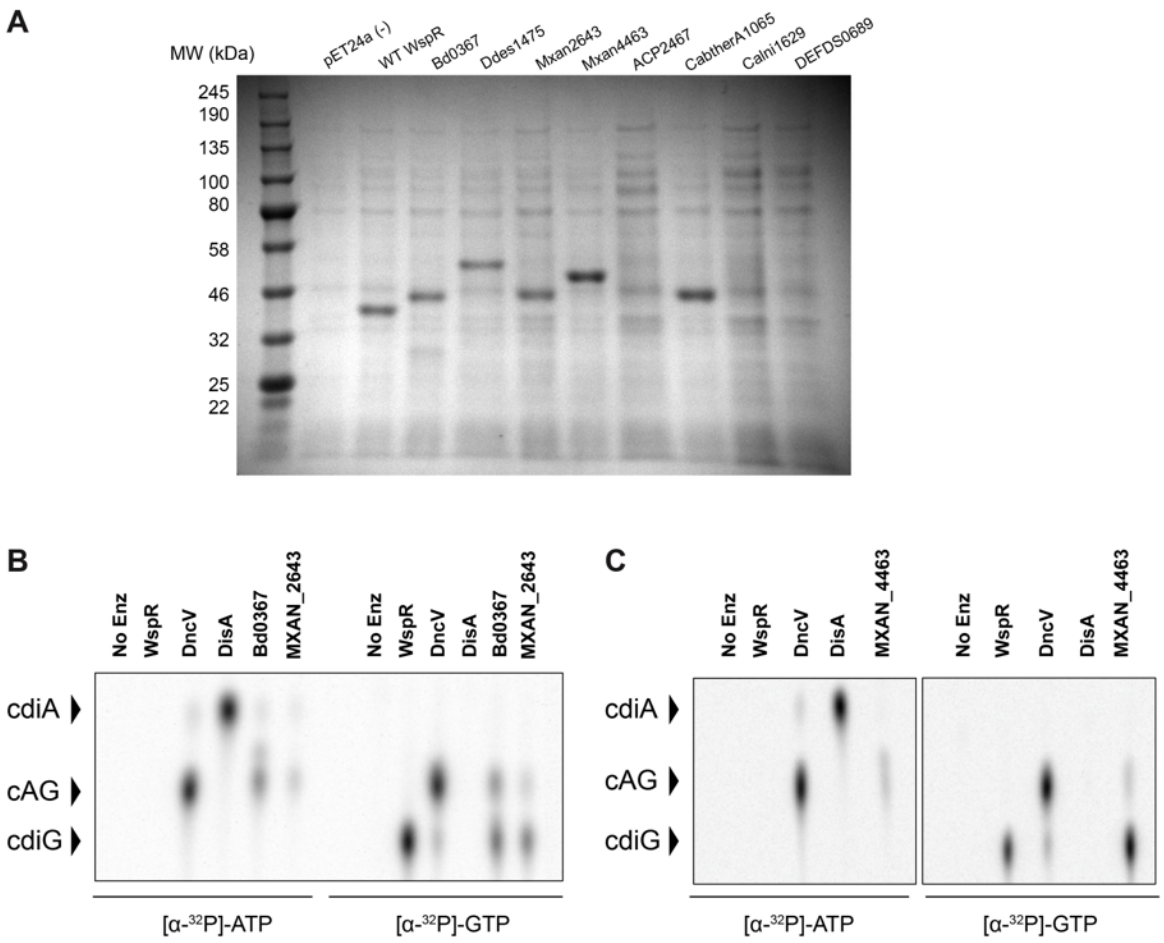


Figure 18 - Activity Assay for Additional HyPr GGDEF Domains. (A) SDS-PAGE gel analysis of lysates from cells expressing C-terminal 6x-His constructs in pET24a, or (for ACP2467, Calni1629, and DEFDS0689) N-terminal 6x-His-MBP constructs. Gel was stained with GelCode Blue (Thermo Scientific). (**B**) Cellulose thin layer chromatography showing cyclic dinucleotide 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 in (B), with Wild-type C-terminal 6x-His tagged MXAN_4463.

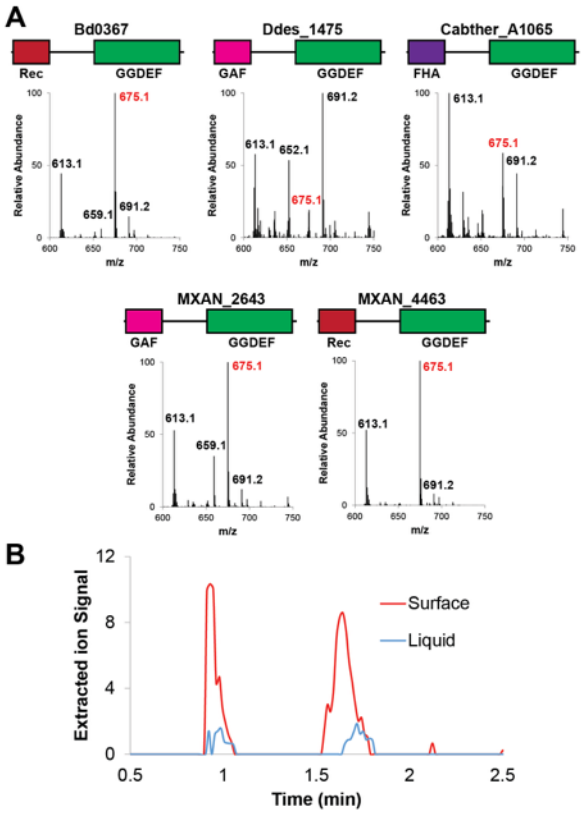


Figure 19 - Validation of HyPr Activity in Select Deltaproteobacteria and Acidobacteria.

(A) LC/MS analysis of *E. coli* cell extracts overexpressing candidate HyPr enzymes, see Fig. S1 for protein domain color scheme and Fig. S12 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 cyclic dinucleotides. Expected masses are for cdiG (m/z = 691), cAG (m/z = 675), and cdiA (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 cAG (m/z=675.1072; ppm<10 cutoff) normalized to the weight of extracted cells. A second biological replicate is shown in Figure 20.

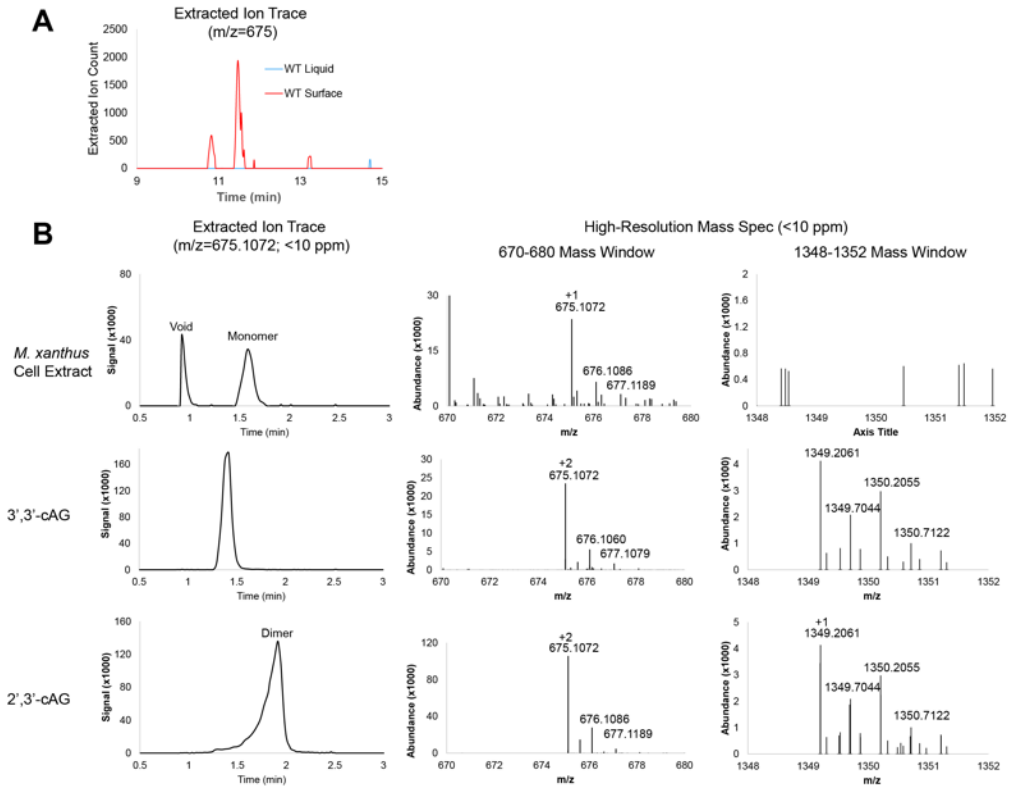


Figure 20 - Mass Spectrometry Analysis of *Myxococcus xanthus* Cell Eextracts. (**A**) Mass spectrometry analysis of second independent biological replicate of *M. xanthus* cell extracts from surface- or liquid-grown WT samples. Shown is the extracted ion trace for cAG (m/z=675) normalized to the weight of extracted cells. The instrumentation and LCMS protocol used in this case was the same as described for *E. coli* cell extract. (**B**) Mass spectrometry analysis of lysate from surface-grown WT *M. xanthus* strain DZ2 or synthetic 3',3'-cAG and 2',3'-cAG standards. Left: Extracted ion trace for cAG (m/z=675.1072). Right: High-resolution mass spectra of the monomer (m/z=675.1072) and dimer (m/z=1349.2064)

Using the same instrumentation as used for analysis of *E. coli* cell extracts, we detect cAG only in WT *M. xanthus* grown on agar surface, not for liquid culture samples.

In the extract from surface-grown *M. xanthus*, we observe two peaks, one at the void volume (1 min), and the other which is intermediate between the retention time of the 3', 3' and 2',3' cAG standards. However, the high-resolution mass spectrum shows that it matches the chemical formula for cAG. The slight discrepancy in retention times compared to synthetic standards is because both synthetic standards appear to elute as dimers, as we observe the dimer mass (1349 mass, +1 charge) as well as the 675 mass as a +2-charged species. In contrast, the cAG present in cell extracts elutes as the monomer (675 mass, +1 charge). We further validated that the peak is 3',3'-cAG by tandem MS/MS, as seen in Figure S16.

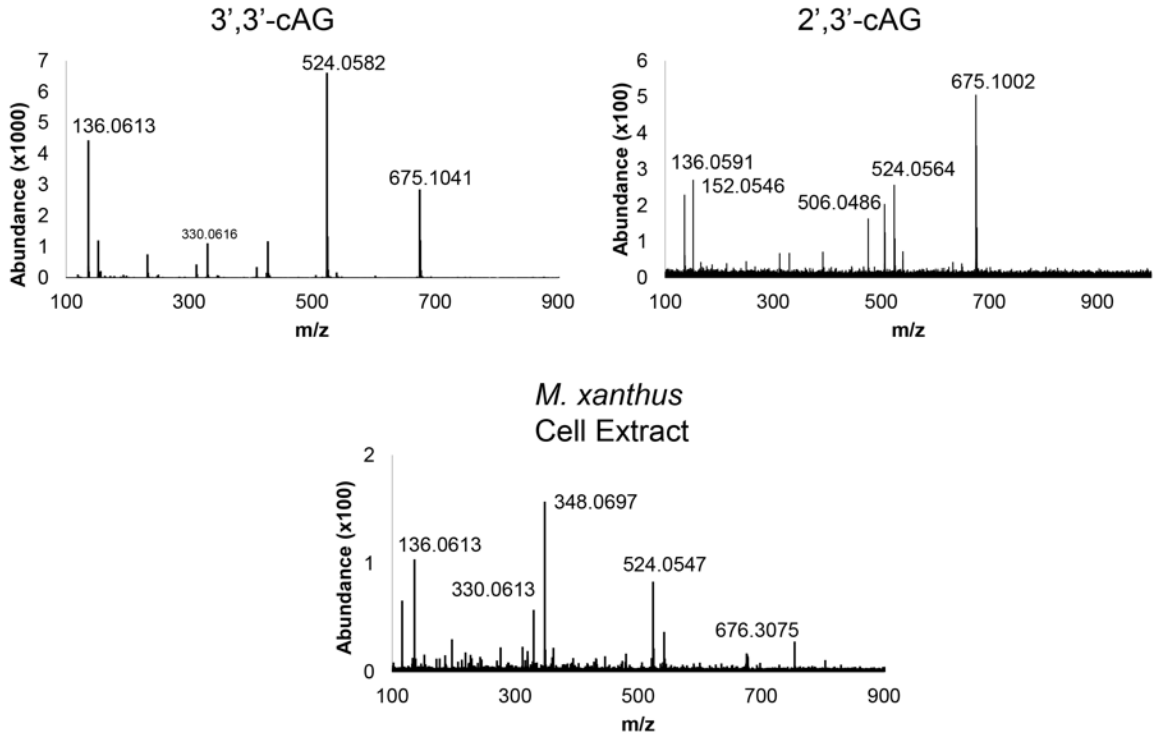


Figure 21 - Tandem MS/MS of Parent Ion (m/z=675.1072) in Cyclic Dinucleotide Standards and *Myxococcus xanthus* Cell Extracts.

Tandem MS/MS analysis of the parent ion shows the cAG present in cell extracts has key fragmentations of 136, 330, and 524, which correspond to the 3',3'-cAG synthetic standard . In contrast, the key fragmentation of m/z=506 corresponding to 2',3'-cAG is not observed in the cell extract sample. A related linear dinucleotide containing a 2',3'-cyclic phosphate has been known to have fragmentation masses of m/z=152 and 540 (Gao et al., 2013b), which we do not observe in the cell extract sample. There is one additional peak in the extract tandem MS/MS spectrum corresponding to AMP, which may be attributed to possible fragmentation differences between the monomeric and dimeric forms of the cyclic dinucleotide.

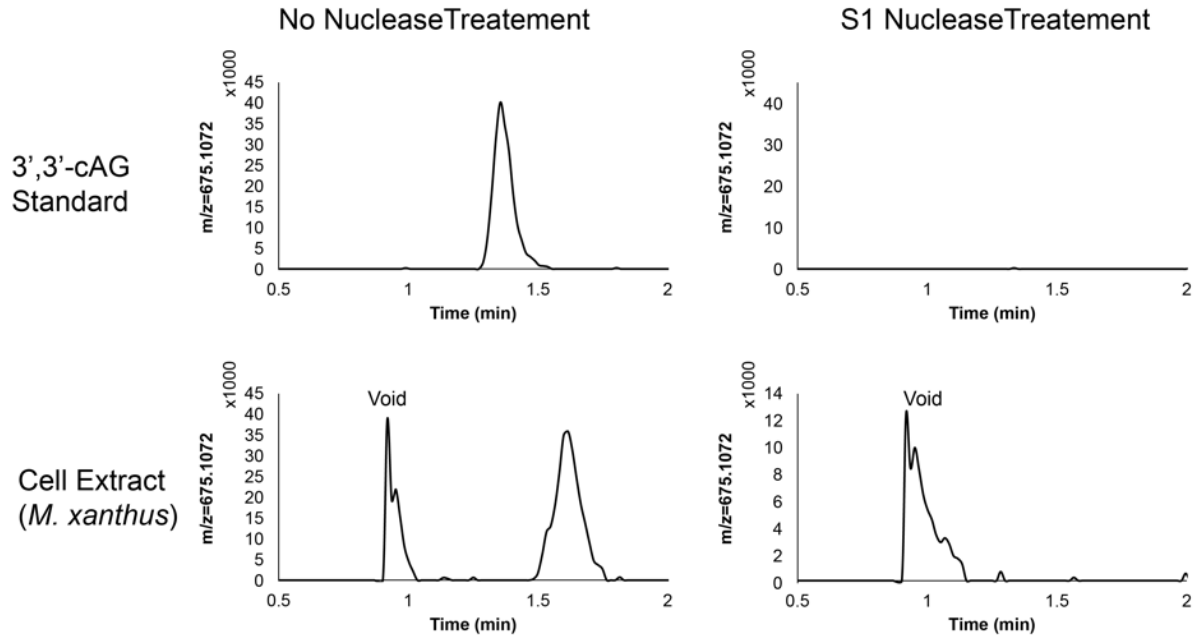


Figure 22 - Effect of Nuclease Treatment on *Myxococcus xanthus* Cell Extracts. Extracted ion trace for m/z=675.1075 with a <10 ppm cutoff of 3',3'-cAG standard (Top) or *M. xanthus* strain DZ2 extracts (bottom) untreated (Left) or treated with S1 Nuclease (Right).

S1 nuclease cleaves 3'-5' phosphodiester bonds, and is shown to cleave 3',3'-cAG. As shown, S1 nuclease is able to cleave both m/z=675 peaks occurring at the void volume and at ~1.6 min in the *M. xanthus* cell extract, further supporting our assignment for the compound as 3',3'-cAG.

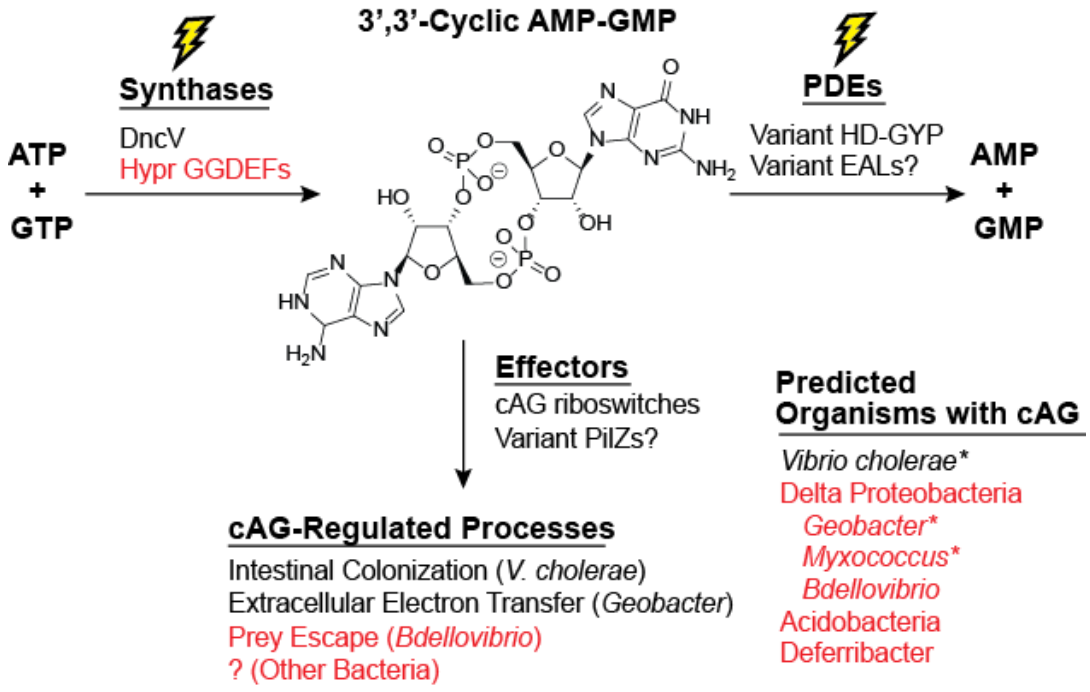


Fig. 23 - Expanded Understanding of the Function and Evolution of cAG Signaling. Levels of the second messenger cAG are regulated by the activity of cAG synthases and phosphodiesterases (PDEs) in response to primary environmental signals. Effectors that bind cAG then propagate downstream effects on bacterial physiology. In *Vibrio cholerae*, DncV serves as the synthase and variant HD-GYP domains as the phosphodiesterases, but no effectors are known. We have shown that diverse delta proteobacteria and acidobacteria contain HyPr GGDEF enzymes that can act as cAG synthases and control various processes through cAG riboswitches (*Geobacter*) and other unidentified effectors. An asterisk indicates organisms confirmed to have endogenous cAG. Text in red indicates new information put forth in this paper.

Table 1 - 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 / MP5AC-TX9)*	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 STI-AU_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 BIF-PSEUDO_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 / MP5AC-TX9)*	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_pU-UH2392p0067	*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 STI-AU_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 tume-faciens (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 2 - 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	ATGGAACGGATTCTCGTTGTCGAAGATGACCGTTTTTTTCGTCAGATGTATGTTGATCTCCTGAAAGAG- GAGGGATACGAGGTCGATACCGTGGCATCGGGCACCGAGGGGTTGAAGCGGCTTGAGAAGCAAGAATAC- CACCTCGTCATTACCGACCTGGTCATGCCCGGAATGAGCGGTATCGAGGTGTTGTCCCGCGTCAAGCAG- AAAGCTCCGAACGTCGATGTCATCCTCGTCACCGGTCACGCCAACCTCGAATCGGCCGTCTATGCCCTC- AAGAATGGTGCCCGCGATTATATTCTCAAACCGTTCAACCATGATGAATTCAAGCACACCGTGGCACTTTG- CTTTGAGCAGCGGAGGCTTATCAACGAAAACTACGAGCTCAAGGAGCTGCTGAATCTTTTTCAAGTTGGG- CAGAACATAGCCAACTGTATCGACTTGGAACGGCTCTCTGCGGTTGTGGTCGATGCTTTCTGCAAGGAG- GTCGGAGTTTCACGCGCTATCGGCCTCTTTCCCGAAAAGAGCGAACCCCACGCCCTCAAGGAGCTGAG- GGGGCTTGAGCCTGAAGTTGCAGCCGCTCTTGCCGAAAAAGCTCTTACCCTTTGCAGTGACGCCGCGG- AGACGGCAGGGGGCTTTCGACGGCTCGACGGTTCCCATTTTTCCGATGGTCTCCTGCGAACTGCGGGGA- TTAATGGCGCCCTTGTGGTTAGCATCCGCCAGCGTACGCTCCTGCAGGGAGTGCTTCTGCTGGTCAATG- ACCAGGGCAAGCCGTTCCCTGCCGTGTTCAAACATAAAAGCATCCAGTTTTTGCTGGAGCAGGCATCGC- TTGCCTTCGACAACGCCCTGCGTTACTCCAGCGCCCGCGACATGCTCTATGTTGACGAACTCACGGGAC- TCTTCAACTACCGTTACCTTGACATCTCGCTGGACCGGGAGTTGAAGCGGGCTGACCGATTCGGCTCGG- TAGTTTCCATGATCTTCATCGACATGGACCACTTCAAGGGAGTCAACGACACCCACGGCCATCTTTTTGG- GAGCCAGGTCCTCCATGAAGTAGGTCAATTGCTCAAGAAGTCGGTCCGTGAGGTCGATGTAATCATTCGC- TACGGTGGCGACGAGTTCACCATAATTCTGGTGGAAACCGGTGAAAAGGGCGCTGCAACCGTGGCTGA- AAGGATTCGTCGCTCCATCGAGGACCACCACTTTCTGGCCTCTGAAGGGCTCGATGTCCGGCTCACCGC- AAGTCTCGGCTACGCCTGTTATCCCCTTGACACCCAGTCCAAAATGGAACTTCTCGAACTGGCGGACAA- AGCCATGTATAGGGGCAAGGAAGAGGGCAAAAACCGTGTATTCCGGGCAACGGCAATCCGTTGA
Mxan2643	Q1D911	ATGAATCCCGCGGACCTCCTGTCGGCCATGAAGCGGACAGTGGAGCAGTTGGCCGCCTTCAATGAGATG- GCGAAGGCCCTGACGTCCACGCTCGAGCTCCGCGAGGTGCTGGCGCTGGTGATGCAGAAG- GTCAGCAGCCTGCTGCTGCCTCGCAACTGGTCGCTCATCCTCCAGGACGAGCGCACCGGAAAGCTCT- ACTTCGAAATCGCGGTGGGTGACGGCGCGGACGTGCTCAAGGGCCTCCAGCTCAACCCGGGCGAGG- GCATTGCCGGCGCCGTCTTCACGTCCGGCGCGGCGCGGCTCGTCCATGACGTGGGTGGGGACCCCA- GCTTCTCGCCACGCTTCGATGAAGCCTCCGCCTTCCACACCCGCTCCATCCTCGCGGTGCCGCTGCT- GGCCCGGGGCCGGGTCCTGGGCATCATCGAACTGGTGAACGGGCCCATGGACCCCCCCTTCACCAA- CGAGGACCTCACCATTCTCACCGCCATCGCGGACTACGCGGCCATCGCGATTGAGAACGCGCGCAAC- TTCCGGCGGGTGCAGGAGTTGACGATTACGGACGAGCACACCGGCTGCTACAACGCCCGGCACCTG- CGCGCCTTGCTGGACCAGGAGGTGAAGCGCTCGGAGCGCTTCAGCCACCCGCTGTCGCTCGTCTTC- CTGGACCTGGACCACTTCAAGAGCATCAACGACACCCATGGGCACCTGGTGGGTAGCGCCACCTTGA- AGGAAGTGGGGGACCTGCTGATGACCCTGGGCCGGCAGAACCTGGACGCCGTCTTCCGCTACGGCG- GCGACGAGTTCGCCATGTTGCTGGTGGAGACGGACCCGGAGGGCGCGGCCGTCATCGGCCAGCGCG- TCTGCGAGGCCTTTCGGGGGCGGGGCTTCCTCCTGGAGCAGGGCCTGGACGTGCGCCTCACCGCCA- GCGTGGGCGTGGCCACCTACCCGGACCATGCCTCGTCCGCGCTGGACCTCATCCGCGCGGCGGACT- TCGCCATGTACGCGGCCAAGGCCCGGGGCCGGGACGCGCTCTGCATCGCCGAGCCCATTGCTCCGA- ACGGCGGCACAGGCTCCCACGAGTTCCCGGAGCGGTAG
Mxan4463	Q1D3Y9	ATGGCGCGAATCCTCCTCGTCGACGACGAAAAGATCGCCCGCACCCTGTACGGCGACTACCTCACCGC- CGTGGGACACGCCGTCACGGCGGTGGGCACGCTACAAGAGGCAAAGGAAGCACTCGCAGGCGAC- CGTTTCGACGCGGTGGTGACGGACCTCATCCTCCCCGGTGGTGACGGCATGGAGGTCCTGCGGCACG- TGCGGGAACATCACCCGGGCGTGGAGGTGGTGGTCATCACTGGCCTGGAGAAGGTGGACCCCGCCGT- GCGCGCCATCAAGAGCGGCGCCGCGGAGTACCTCGTCAAGCCGGTGGCCCCGGAGGCCCTGCAGCA- CGCCGTGCGCCGAGCGCTCACCACGCGCGACCTGATGCAGGAGAACGCGTCGCTGCGCCGCCATGTG- GCCATGTTGGAGGCGGGGCAACGCATCGCCACCACCCTGGACCGCGAGAAGCTGGCCTCGGCCACCG- CCAGCGCGCTGCAGAGCATGGCCTCCGCCAGCGCCGTGGTCCTGCTGGAGCGCGACTCTGCCTTCGC- GCTGCGGCGCCACGGCACCAGCGGCCTGTCCACCGCGCTGGAAGAGCCGCTCATCGCCGAGCTCATC- GAACGCCTGACGAACGAACGCGGTCCGCGCGAGCTGGACGGCATGGACGCGCCCTTTCCTCGCGCAA- TCTCCTTCCCCGCGCTGGAGGGTGACGCCGTGCTGGGACACGCGGTGCTCTTCTTCGGCGGCACGGG- CGCGGAGTGGGCGGGCGAGACGGCCAGCTTCCTGGTTCGCAACTGGGCGCTCGCGCTGCGCAACCT- CGGCCGCTTCGCCGCGGTGGAGGACCTGGCGTACGTCGACGACCTCACGCGCCTGTTCAACACCCGC- TACCTGCACCTGGTGGTGGACCGCGAGGTCCAGGACGCGCTCCAGTCACAGCGCACCTTCAGCCTGCT- GTTCCTGGACCTGGACCACTTCAAGTCCATCAACGATACCCATGGCCACCTCGTGGGCTCCAAGGTGCT- GGTGGAGGCGGCGCGCGTGGTGAAGGGCTGCGTGAGAGACCACGACGTCGTCGCGCGCTACGGCGG- AGACGAATACGTGGTGGTGCTGCGCAACACCGACTCCGGCGGCGCGCTCAAGGTGGCCGAGCGCATC- CGACGCACCATGGAGACGCACAACTTCCTGGCGCGCGAAGGCCTGTCGCTCAAGCTCACCACGTGTAT- CGGCGTGGCCAGCTTCCCCGAGCACGCCCAGGACAAGGCCACGCTGTTGGACCTGTCGGACCGGGCC- ATGTACCGCGGCAAGCGGGGCTCGCGGAACGTCGTCTACATGGCGGCGAAGGACCTGGAGGCCCCAC- CGGCCGAGCGCCGGCAGGCCCACTCCGCGTCCTGA

Ddes1475	B9J0V0	ATGCTGAACAAGTCAAGCATCATACCAGAACATATACAGCTCGAATCACACGACCCTGTCTGGGAGTGGCA- CACGACATCCGACAGGCTTTTCATGAGCGTAGGCGCCCTTGCCCAGCTACGCATGGACGGCAAACCGC- CGCGCAGCATGAAAGATTATCTGGAGCACTGCCCCCTCGAAAGCCTGGCTCCCCTTCTTGAATCTATGG- AAAAAGCGCTCAACGGCTCCCACGGGCCGCACCTTGAAGTGTTTTATCCTTTTGACAGTTTTCTGGTAC- GGTCTCAGATACTGGTCTTGCGGCGCGACGTTTTCGGTCGCGGAACCCTGGTAACAGGCTGCAACGT- GGCTATGGACAGACAAAGGCTTGCACCCACTGCTGCCGCCGCCCCCGTGCCGGCACCCCAGCCCCC- GCCCCGAAGCCTGGCCGAAGCCGCCGTTCCTTCCACGGCCCGCAGCGACGCCAGCCGCCTCATGCT- GGCCCTCAACGCCGCCAGCGATGGCCTGTGGGACTGGGACCCCAGCACCAATGCCATTTATTTCAGTC- CCCGCTACCTCGACATGCTTGGCTACACCAGCGAAGAATTCCCCCCCCCTGTCCACATCATGGACCAGC- AAGGTACACCCCGACGATTACGACAACATCGTTCCCATGCAGATTGAATTCATCAACAACCCCAAAATG- GGCGACAGCTTTGAATGCACCTACCGGATGCAGCGCGGCGACGGCACCTGGGCATGGATTCTCAGCC- GGGGCTATGTGACTCACCGCGACGCAAGCGGCAAGGCCATCCGCGTTGTGGGCCTGCACACAGACGT- CAGCGCGAGCCAGGGCGACAGGGCACGGCTTGAAGAGCTGGTGCGTAACGACGCCCTTACCGGGCT- GCGCAGCCGCACCTATTATGGAATGACTGTCGACAAGCTGGAACAGCAGCAGATGCGGCCCGTCAGCA- TCATCATCGCCGACATGGACGGACTCAAAATGGTCAACGACCATGTAGGCCATACCGAAGGCAGCGAA- ATGCTCTGCCAGGCAGCCATCATACTGCGCGGCAGTCTCAATGCCACCGACTGTATCGCCCGCATGGG- CGGCGATGAATTTGCCGCCATTGTGCCGGGTTGCGCCAAGGAAGACCTTGAGGCGCTCATCCAGCGG- GTCAGAGACGCTTTTGATGCCTATAATGCCGACCCGGACCATGTGCCGACACACATGTCTGTGGGCGG- AGCATGCGCTGACGACATGAACACCACCCTGGCCCAGGCCCTGTCGGAGGCGGATCGCAACATGCTG- GCCGTCAAGCACGAAAGCAGCCCAAAGTGGCGCCTGCGCATAAAAAACTGGATAGAAAACCGGACCG- GCAAAACAATTCAGCTTGAAGACAGCCGCTACAGGATGTCCCCCACGCACGACGACTCTTGA
Bd0367 WT *	Q6MQU2	ATGTCGCGCGCCGAAGTGACGCTCGTATGTAAAATGAGCTTTGAAGTATCGCCGAAGCAACCAAAGAGC- CGCCGTATCCTGGTTATCGACGACGATAAGGACTCATTAGAAATTTTATTGGAACCCCTGCGCTGGGAAGGT- TATGACGCGCGTGGCGTGACTACCGAAGCGGAGGCGCATAAATTAATCGAGTCATGGATTCCGCATATC- GTGATCCTGGATTGGATGGCCCCGTCAATGGCCGGCCTGCGCGTTCTGAAATCCGTACGCGAACGCCT- GAGTCATGTCTCGTGTGTCTTTGTATCGGAAAATTCTTCCACAGAGGCTATTATTGAGGCTTTGGATTCG- GGCGCCGACGATTATATTGTAAAGCCATTCGTGCCATTAGAGTTGTTAGCACGCATCCGCTCTCAACTGC- GCATCCGCGATCTGCACGAGCAGCTGCTGTTTGCCAACGAAAAATTAAAGGAACTGGTTGATACCGACG- ATTTAACCGGTTTATATAATATGCGTAGCTTATACCAGCGTCTGGATTTTGAAATGGAACGTGGCCGCCGC- TTCCACCGCGACGTGTGCGTGGTCATGATGGACATGGACTATTTCAAAACCGTGAATGATGGACACGAC- CACTTATTCGGGAGTTATGTGCTGAGCGAAGTTGGTAAAATCATTCGCGCCAACACTCGTAACATCGATA- TCCCGGCACGTTATGGGGGGGATGAGTTTCTGATGGTCCTGACCGAAACTAATCATGCGGGCGCTATGT- ATTTTTGCGAGCGCCTGCGCGAAAATATTGAAAAAACAACCTTTCGTAACGGCGAGGACAGCATGAAAT- TGACAGCCTCACTGGGCTTTGCGATCACCATCCCCGGCGAAAACATCAGCGCGCGTGAACTGGTTCG- CCGCGCCGACCACGCTCTGTATCAGGCAAAACGCGCCGGGCGCAACCAGGTGGCGCATTACAAACCG- GAGAGCGCGCCCGTAGTTGAGATCAAGTCGGCAGTGCACAAACGCCGTAAAGCCGCCGGTTAA
Cabther-A_1065 WT *	G2LH77	ATGAACCTTAAACTGGGCGCCATCTTACGTCCGGTTAATAGCCTCAACCAAACACAGAAACTGCAAGC- CAACCCACTTGCCCGCCCGCGTCAAGCGCGTCCGGCACTGGTGCACATGCGTGGAGATTATCTGGGCT- CAAGCTTTCGTATTGAACATGCCATTACGCGCATTGGACGCGGATCAGACGCAGAGTTACGTTTAGAAAAT- GATGACGAAGCAAGTCGCTTACACGCCCGTATTGAGCGCCTGGAAACACCTACAGGGCATTTCCAATATT- GGTTGACCGATCTGCGCTCTACCAACGGGACCCAACTGAATGGTATTCCGCTGGTGCCGGGCGAGGCA- GTGTTGCTGCATGATGGCGATAAATTTAGTATCGGCCGTCATATCCTCAAGTTCACTTTTTTAGACGATATTG- ATGAGGAGTTTCATCGTCGTATCACCGAACTCATCACTCATGATGACTTAACCGGTCTGCTGAACCGCAAA- TCGTTCATCCTGGAAATGCAGCGTGAGATGGCCCGTAGTAACCGCTACGGTCACCCATTTGGCCTGCTGA- TGATGGATATTGATCATTTTAAGCGTGTCAATGATACCTATGGTCACCTGGTTGGTTCTCAGGTATTACGCG- AGGTGGCTACCGTTATCCGCGAAACACTGCGTGACTCTGACATTGCAGGTCGTTATGGTGGAGAAGAATA- TATTGCCCTCTTACCAGAAACCGATCGCCTGCGCGCACACGAAGCGGCCGAGCGCATTCGTCAAGCAAT- CGAACGCACCCCGTTCACAGCAAGCCTCAACGTGCCGCACCACAAGTTACGTCTGACCATTAGTATTGGG- ATCGCGAGTTATCCGGGGGACGCAGCCCAAATTAATGATCTGATCCAGCGCGCGGATGAAGCGATGTATG- AAGCAAAACGCCGCGGTCGTAATCTGGTGCAGACGACGGGCCAATCGGCGGCCAATCGCGCGACCCCC- CCTTCACTTCCATTGCCGCCGCCGTCTGGAGATGACAGTCCCACGGAGCATCTGACCGTGGAGCA- GCCGCAGCCTGTCAAACCTTAG
DEFD-S_0689 R248A *	D3PC46	ATGTATGAAAGCCTGAAACGCAACATCTTCGTCATTCTGACAAGCATTCTCCTTATTTATGAAACCTATAA- CAAAACAAATGAGAATTTGCTGCTGCTGACTTCTTTACTGCTCACGTGCTATATTGCTGCAACGTTGAT- CAAAAAGGTTGAACTGGATGAAGTACTGTTTGCTCTGTTCGTGATTTTAATCGGTTATCTGAGCATCGC- AAACCGTGAGTTCATTTATTTTCAAATTCTCGCGATTACATTTTTGGTATTTGATTCGAAGTTCTACGTGA- TTAAAGTGATCCTGGCGATTCTGTTGATTCTCTTCGATTTATTCTACTTGAACATTTCGATCCTTTCCACG- TTTAGCCTGATGATTCTCTATTCCTTATTCTTTTCTATCTTCATCAAATTGCTGATTGATCGTTTGGAAGAA- GAAATCGACGAACTGTCCATTACGGACGACCTCACGGGTCTGCTGAACCAAAAAGGATTCCTGAAAAA- GTTTGAGGAAGAATATTATCGTAGCGTTCGCTACAAGAAAAATTTTACCGTTATCATGTTGGATAGCGAT- GATCTGAAGAAAGTTAATGACACTTATGGGCACAAATACGGGACCAAAGTTATTCTGTTCATCGCGGAT- GAAATTAAGAAGAACATTCGCCGTACCGACTTTGCTTGCCGCTACGGCGGTGACGAGTTTATGATCTG- TCTGGTTGAAACACCTATCAACAACGGCAAAATTTTCGCGGAAGCGCTGAAAAACAACATTGCAATGA- AACCGGTATTTACCGATAAAGGCCGTGGTTTCAATGTGACAGTGTCGGTGGGGGTTGTTGGTTATCCG- CACACAAGCGAAAAGTCGTTCGAGCTGCTTGATCTGGTTGACAAAGCGCTGTACGAAGCAAAAAACA- AAGGCAAAAATCGCGTTGAGATCCTGACCAAAAATTCTTCCTTA

Calni_1629 R268A *	E4TFG3	ATGATTGATAACAAGATTAAACACTTTCAGTATAAGATTGCAGAAGTCTTATATCTCTTTGCGTTTACTATCAT- CATCGCCTCCATTAAATTTCTTGATTCTACCAGTAACAAAGCAAACTACGCGATCTTAGTGTTTTTCCT- GATCTTCATCATCCTGAAATTTAGCATTGACGATAATCTGTTCAGCTCAAAAATCCTGTCTTATTATCT- GCTGTTCCAGTCACAAAACATTTTTGCCGCGTTCATTAACGGGTCAACTCCAAATCTGTTAATCTTTA- TGTCGATGCTGGGCTTGCTGATTTTTAGCATTGTGCTGTATGATAAAAAATATCTCGTCGTCCACTTT- ATTGTGACCGGTATCCTTGCGTTCGTTTTCTTCACAACATTTGATTCGAAAGAGAGTTTTGTGTTCTT- CATCTCACTTCCTTTCATCTTTATTATTTCGTTAAACTTTAATAAGATCTACATTACTACCCGTAACCTG- ATCACCGAGTTATCTATTACTGATGAGATGACCGGCCTCCTCAACCAGAGCGGGTTTATGAAGAAG- ATCGAAGAAGAATTTTATCGCAGCCAGCGTTACCAGAAAACGTTTTCAGTTCTGATGATCGATTCAGA- CAATTTAAAACTGATTAATGATACTTATGGCCATAAATATGGGAGCATCGTCATCAAATCCATTGCGGA- AGTCATTAAGACTAACATTCGCCGTACAGACTTCGCGGCGCGCTATGGCGGAGATGAATTCATTCT- GTGCCTGGTGGAAACTGATTTAGACGGGGCTCTGGAAGTGGCAGAGGCGATCCGTAAGCAGTTC- GAGCTGAAAAGCTTCTTTACCAAAGATGAGAAGAAGTTCACAATCACGATTAGTATCGGAGTAAGCA- ACTATCCTAAAAGCGGCGATTCTCTGATGGATGTGATTGAACTTGCGGACAAGGCCATGTACCATAG- CAAGAACAGCGGTAAGAACAAGACGAGTTTCCTGCTGAAGAAC
ACP_2467 WT *	C1F1G0	ATGGACGCCCACACTATCGTCAGTCTGCCGCCCACTTGGAACCAAGGGATGTCTGCCGAAGCGCG- CAATCAGAACTGGAAGGATTTGGTGGTCTTCCATAACTTAGCACGCGCTCTGACCTCCTCCCTG- GAGCTTGATTCGGTGCTGCATGCAATCATGGAACAGATGCGTCAATTCTTCGAACCGGAGACC- TGGTCGTTGCTTATCCTGGATGAAACAACCCAGGAATTGTATTACGCGGTTGCAGTCGGACAG- TCCGAAGCGGCTCTGCGTAATGTGCGTGTGCCGCTGGGAGAAGGCATGGCGGGTTGGGTGG- CCCAACATGGCGAGTCCCTCATCGTGCCGGATCTGGAACAAGATCCGCGCTTCGCCGCGAC- CTCGGATGCCCGCACCCCAATGCGTAGCGCGATCTGCATGCCACTGCTCTCACGCCAACGC- ACCCTGGGCGTGATTCAACTGTTTAACTGCCGCCTGGAAAGCATGACCGAATACACCATTAGCT- TCCTGCATATCCTGTGCGACTATGCGGCGATTGCAATCGAAAATGCACGTGCAGTGGAGAAAA- TCCAGGCCCTGACGATTACGGATGACTGTACCGGCTTATACAACCAACGTCATCTCCAGCAGA- AGATCGAAGAAGAGGTCACCCGTGCTCGTCGTCACCACCATCCATTCTCAGTCATCTTTCTGG- ATCTTGACCATTTCAAACAAATCAATGACCAACACGGGCACTTAATCGGGAGCCGCCTTCTGG- CGGGTATTGGCCAGTGCCTCCGCCTGCACATTCGCCCGGGAGACCATGCCTTTCGCTATGGT- GGCGATGAATTTATCTTACTGCTTCCAGAAACCACAAAAGCGGAAGCCGAGCAGATTGCGCG- CAACCTGCGTCAAAAACTGCGTAGCCATGTCTTCGAGATGGGCAGCGATCTCCGTTTGCAGG- TTTCGGCCTCATTCGGTGTCGCCAGTTTTCCGGAGGATGGCCGTACGGGCCATCAGATCATT- CGTATGGCCGATGCAATGATGTATTTGGTGAAAGGCTCTACGCGCGACGACGTGGCAGTTGCG- GACCGTAATACCGAACTGCTCCGCAACTCA

Table 3 - List of primer sequences. 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	CCACCTCGTCATTACCGCACTGGTCATGCCCGGAATG	For Primer to Quickchange GSU1658 D52A
71	CATTCCGGGCATGACCAGTGCGGTAATGACGAGGTGG	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	AGAGAGCTCGAGTCATAAGGAAGAATTTTTGGTCAGG	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	CTTCGGTCACGATATCGGCAGTGAGGTGCTGCGC-GAGTTC	For primer to Quickchange PleD D344S
95	GAACTCGCGCAGCACCTCACTGCCGATATCGTGACC-GAAG	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

Bibliography

Ablasser, A., Goldeck, M., Cavlar, T., Deimling, T., Witte, G., Rohl, I., Hopfner, K.P., Ludwig, J., and Hornung, V. (2013). cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. Nature *498*, 380-384.

Aldridge, P., Paul, R., Goymer, P., Rainey, P., and Jenal, U. (2003). Role of the GGDEF regulator PleD in polar development of Caulobacter crescentus. Mol Microbiol *47*, 1695-1708.

Ausmees, N., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M., and Lindberg, M. (2001). Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. FEMS Microbiol Lett *204*, 163-167.

Bochner, B.R., and Ames, B.N. (1982). Complete analysis of cellular nucleotides by two-dimensional thin layer chromatography. J Biol Chem *257*, 9759-9769.

Buckstein, M.H., He, J., and Rubin, H. (2008). Characterization of nucleotide pools as a function of physiological state in Escherichia coli. J Bacteriol *190*, 718-726.

Campos, J.M., Geisselsoder, J., and Zusman, D.R. (1978). Isolation of bacteriophage MX4, a generalized transducing phage for Myxococcus xanthus. J Mol Biol *119*, 167-178.

Chan, C., Paul, R., Samoray, D., Amiot, N.C., Giese, B., Jenal, U., and Schirmer, T. (2004). Structural basis of activity and allosteric control of diguanylate cyclase. Proc Natl Acad Sci U S A *101*, 17084-17089.

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

Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., Meuwly, M., and Jenal, U. (2006). Allosteric control of cyclic di-GMP signaling. J Biol Chem *281*, 32015-32024.

Corrigan, R.M., Campeotto, I., Jeganathan, T., Roelofs, K.G., Lee, V.T., and Grundling, A. (2013). Systematic identification of conserved bacterial c-di-AMP receptor proteins. Proc Natl Acad Sci U S A *110*, 9084-9089.

Corrigan, R.M., and Grundling, A. (2013). Cyclic di-AMP: another second messenger enters the fray. Nat Rev Microbiol *11*, 513-524.

Danilchanka, O., and Mekalanos, J.J. (2013). Cyclic dinucleotides and the innate immune response. Cell *154*, 962-970.

Davies, B.W., Bogard, R.W., Young, T.S., and Mekalanos, J.J. (2012). Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for V. cholerae virulence. Cell *149*, 358-370.

De, N., Pirruccello, M., Krasteva, P.V., Bae, N., Raghavan, R.V., and Sondermann, H. (2008). Phosphorylation-independent regulation of the diguanylate cyclase WspR. PLoS Biol *6*, e67.

Diner, E.J., Burdette, D.L., Wilson, S.C., Monroe, K.M., Kellenberger, C.A., Hyodo, M., Hayakawa, Y., Hammond, M.C., and Vance, R.E. (2013). The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. Cell Rep *3*, 1355-1361.

Gao, J., Tao, J., Liang, W., Zhao, M., Du, X., Cui, S., Duan, H., Kan, B., Su, X., and Jiang, Z. (2015). Identification and characterization of phosphodiesterases that specifically degrade 3'3'-cyclic GMP-AMP. Cell Res *25*, 539-550.

Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B.L., Zillinger, T., Serganov, A.A., Liu, Y., Jones, R.A., Hartmann, G.*, et al.* (2013a). Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. Cell *153*, 1094-1107.

Gao, X., Mukherjee, S., Matthews, P.M., Hammad, L.A., Kearns, D.B., and Dann, C.E., 3rd (2013b). Functional characterization of core components of the Bacillus subtilis cyclic-di-GMP signaling pathway. J Bacteriol *195*, 4782-4792.

Hengge, R. (2009). Principles of c-di-GMP signalling in bacteria. Nat Rev Microbiol *7*, 263-273.

Hickman, J.W., Tifrea, D.F., and Harwood, C.S. (2005). A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. Proc Natl Acad Sci U S A *102*, 14422-14427.

Hobley, L., Fung, R.K., Lambert, C., Harris, M.A., Dabhi, J.M., King, S.S., Basford, S.M., Uchida, K., Till, R., Ahmad, R.*, et al.* (2012). Discrete cyclic di-GMP-dependent control of bacterial predation versus axenic growth in Bdellovibrio bacteriovorus. PLoS Pathog *8*, e1002493.

Huangyutitham, V., Guvener, Z.T., and Harwood, C.S. (2013). Subcellular clustering of the phosphorylated WspR response regulator protein stimulates its diguanylate cyclase activity. MBio *4*, e00242-00213.

Kellenberger, C.A., Wilson, S.C., Hickey, S.F., Gonzalez, T.L., Su, Y., Hallberg, Z.F., Brewer, T.F., Iavarone, A.T., Carlson, H.K., Hsieh, Y.F.*, et al.* (2015). GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. Proc Natl Acad Sci U S A *112*, 5383-5388.

Kellenberger, C.A., Wilson, S.C., Sales-Lee, J., and Hammond, M.C. (2013). RNA-based fluorescent biosensors for live cell imaging of second messengers cyclic di-GMP and cyclic AMP-GMP. J Am Chem Soc *135*, 4906-4909.

Kranzusch, P.J., Lee, A.S., Berger, J.M., and Doudna, J.A. (2013). Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. Cell Rep *3*, 1362-1368.

Kranzusch, P.J., Lee, A.S., Wilson, S.C., Solovykh, M.S., Vance, R.E., Berger, J.M., and Doudna, J.A. (2014). Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. Cell *158*, 1011-1021.

Lori, C., Ozaki, S., Steiner, S., Bohm, R., Abel, S., Dubey, B.N., Schirmer, T., Hiller, S., and Jenal, U. (2015). Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication. Nature *523*, 236-239.

Nan, B., and Zusman, D.R. (2011). Uncovering the mystery of gliding motility in the myxobacteria. Annu Rev Genet *45*, 21-39.

Nelson, J.W., Sudarsan, N., Furukawa, K., Weinberg, Z., Wang, J.X., and Breaker, R.R. (2013). Riboswitches in eubacteria sense the second messenger c-di-AMP. Nat Chem Biol *9*, 834-839.

Nelson, J.W., Sudarsan, N., Phillips, G.E., Stav, S., Lunse, C.E., McCown, P.J., and Breaker, R.R. (2015). Control of bacterial exoelectrogenesis by c-AMP-GMP. Proc Natl Acad Sci U S A *112*, 5389-5394.

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

Paul, K., Nieto, V., Carlquist, W.C., Blair, D.F., and Harshey, R.M. (2010). The c-di-GMP binding protein YcgR controls flagellar motor direction and speed to affect chemotaxis by a "backstop brake" mechanism. Mol Cell *38*, 128-139.

Qiu, Y., Cho, B.K., Park, Y.S., Lovley, D., Palsson, B.O., and Zengler, K. (2010). Structural and operational complexity of the Geobacter sulfurreducens genome. Genome Res *20*, 1304-1311.

Reguera, G., McCarthy, K.D., Mehta, T., Nicoll, J.S., Tuominen, M.T., and Lovley, D.R. (2005). Extracellular electron transfer via microbial nanowires. Nature *435*, 1098-1101.

Ren, A., Wang, X.C., Kellenberger, C.A., Rajashankar, K.R., Jones, R.A., Hammond, M.C., and Patel, D.J. (2015). Structural basis for molecular discrimination by a 3',3'-cGAMP sensing riboswitch. Cell Rep *11*, 1-12.

Romling, U., Galperin, M.Y., and Gomelsky, M. (2013). Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. Microbiol Mol Biol Rev *77*, 1-52.

Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., Mayer, R., Braun, S., de Vroom, E., van der Marel, G.A., van Boom, J.H.*, et al.* (1987). Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. Nature *325*, 279-281.

Ryjenkov, D.A., Tarutina, M., Moskvin, O.V., and Gomelsky, M. (2005). Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain. J Bacteriol *187*, 1792-1798.

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

Shanahan, C.A., Gaffney, B.L., Jones, R.A., and Strobel, S.A. (2013). Identification of c-di-GMP derivatives resistant to an EAL domain phosphodiesterase. Biochemistry *52*, 365-377.

Spangler, C., Bohm, A., Jenal, U., Seifert, R., and Kaever, V. (2010). A liquid chromatography-coupled tandem mass spectrometry method for quantitation of cyclic di-guanosine monophosphate. J Microbiol Methods *81*, 226-231.

Sun, L., Wu, J., Du, F., Chen, X., and Chen, Z.J. (2013). Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. Science *339*, 786-791.

Vassallo, C., Pathak, D.T., Cao, P., Zuckerman, D.M., Hoiczyk, E., and Wall, D. (2015). Cell rejuvenation and social behaviors promoted by LPS exchange in myxobacteria. Proc Natl Acad Sci U S A *112*, E2939-2946.

Velazquez-Campoy, A., Leavitt, S.A., and Freire, E. (2004). Characterization of protein-protein interactions by isothermal titration calorimetry. Methods Mol Biol *261*, 35-54.

Wassmann, P., Chan, C., Paul, R., Beck, A., Heerklotz, H., Jenal, U., and Schirmer, T. (2007). Structure of BeF3- -modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. Structure *15*, 915-927.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.

Witte, G., Hartung, S., Buttner, K., and Hopfner, K.P. (2008). Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. Mol Cell *30*, 167-178.

Wu, J., Sun, L., Chen, X., Du, F., Shi, H., Chen, C., and Chen, Z.J. (2013). Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. Science *339*, 826-830.

Yan, D., Cho, H.S., Hastings, C.A., Igo, M.M., Lee, S.Y., Pelton, J.G., Stewart, V., Wemmer, D.E., and Kustu, S. (1999). Beryllofluoride mimics phosphorylation of NtrC and other bacterial response regulators. Proc Natl Acad Sci U S A *96*, 14789-14794.

Chapter Three

Kinetic Analysis of Hybrid Promiscuous GGDEFs

Introduction

With the discovery of Hybrid product producing, promiscuous substrate binding (HyPr) GGDEF enzymes, a key paradigm in the cyclic dinucleotide field had been broken - the C_2-symmetric GGDEF domain is capable of forming a heterodimeric product. However, multiple mechanistic questions remained regarding how this enzyme could function *in vivo*. Other cAG synthases have a particular order of catalysis, wherein one phosphodiester bond is formed first preferentially, before the cyclization can occur - do HyPr GGDEFs have this same preference when forming the heterodimeric product? Most important to its function *in vivo* - how is cAG the predominant product produced, when the enzyme, as observed from our initial studies, appears promiscuous?

In this chapter, I apply what is known about conventional dinucleotide cyclase mechanism to determine the order of bond formation and the kinetic parameters of the formation of cAG using the HyPr GGDEF from *Geobacter sulfurreducens*. Building upon previous work using non-hydrolyzable nucleotides, we were able to determine that the HyPr GGDEF enzyme exhibits no preference for order of phosphodiester bond formation. We obtained kinetic parameters for the single substrate cases, where only cdiG and cdiA are produced. We then develop a computational method to determine the dissociation constants for the heterodimeric cases by analyzing the ratios of product produced. Our results support the hypothesis that HyPr GGDEF mechanism exhibits cooperative binding, where the enzyme selects for specific substrates dependent on the identity of the first substrate that binds. Finally, we model HyPr activity under conditions where ATP and GTP levels do not change, demonstrating that under physiologically-relevant ATP:GTP ratios, cAG is the primary product.

Results

Our previous biochemical results showed that, under varying ratios of ATP:GTP, HyPr GGDEFs produce different ratios of cyclic dinucleotide products, suggesting that they act in a promiscuous manner (Hallberg et al., 2016). Furthermore, our results suggested that the enzyme was specifically tuned to the ratio of free ATP to free GTP in *Geobacter* cells so as to produce primarily the heterodimeric product. However, these results were based on endpoint ratios of cyclic dinucleotide products, where most of the nucleotide substrate had been consumed. In contrast, *in vivo* we expect that cellular ATP and GTP concentrations remain constant. Therefore, three major questions remained to be answered: Do HyPr GGDEF enzymes have a preferred order for bond formation, is promiscuity in HyPr GGDEFs imparted at substrate binding steps, and what is the major product produced in an *in vivo* situation?

The two characterized dinucleotide cyclases, DncV and cGAS, operate via a two-step mechanism, where a single linear dinucleotide intermediate is formed, followed by cyclization. Importantly, DncV and cGAS possess discrete, unsymmetrical binding sites for both NTP substrates, and so differ in catalytic pathway (Figure 1A): whereas DncV initially produces pppA[3',5']pG, utilizing ATP as the donor molecule and GTP as the acceptor, cGAS produces pppG[2',5']pA. In contrast, HyPr GGDEF enzymes possess what is expected to be a C_2-symmetric active site, which suggests that they would be able to utilize both ATP or GTP as donor or acceptor (Chan et al., 2004; Gourinchas et

al., 2017). However upon binding of both ATP and GTP, a non-symmetric active site would be produced, which could lead to a preferred order of catalysis.

To test if HyPr GGDEFs act in a promiscuous or specific manner during catalysis, we utilized non-hydrolyzable NTP analogues in concert with radiolabeled NTPs to probe the products the enzyme was able to form. In this experiment, the non-hydrolyzable analog can only serve in the donor role of the nucleotidyltransferase reaction. Therefore, cGAS would only produce a linear dinucleotide with non-hydrolyzable GTP and hydrolyzable ATP, while DncV would produce a linear dinucleotide in the opposite case. In contrast to these specific dinucleotide cyclases, the HyPr GGDEF DncG produces both linear heterodimeric dinucleotide products (Figure 1B). Upon further analysis using radiolabeled and non-hydrolyzable nucleotides of the same identity, we observe that GSU1658 (labeled here as DncG for Dinucleotide cyclase in *Geobacter*) can produce all four linear purine-containing dinucleotides (Figure 1B). We conclude that HyPr GGDEFs have no preferred order in formation of the linear intermediate.

For a HyPr GGDEF enzyme in an active, dimeric conformation, two discrete NTP-binding steps occur in two equivalent enzyme binding sites, followed by catalysis and product release to form the cyclic dinucleotide. There is one case where an enzyme binds to two GTP molecules or two ATP molecules, but two different orientations in which enzyme is bound to both ATP and GTP. To test this model, we used it to calculate product ratios for the “neutral scenario,” where k_{cat} and K_D values for ATP and GTP are equal and substrate concentrations are equal. As expected, the model resulted in a product ratio of 1:2:1 for cdiG/cAG/cdiA.

Importantly, because both ATP and GTP are accepted as substrates for these enzymes, we expect that there are different dissociation constants for these different nucleotides. We were intrigued by the possibility of cooperative effects - how does initial binding of GTP or ATP influence the second binding event? For both non-cooperative ($K_1 = K_2$) and cooperative possibilities using a single substrate (ATP or GTP), there is an exact mathematical solution - using the same assumptions as Michaelis-Menten kinetics - for the rate of product formation (Oliveira et al., 2015). We purified enzyme constructs containing R-to-A mutations in the autoinhibitory site to circumvent one potential mode of product inhibition, and with an N-terminal 6x-His-MBP tag for purification and to increase solubility (Hallberg et al., 2016). Using an enzyme-coupled assay for pyrophosphate detection, we measured the initial rates of pyrophosphate formation (as surrogate for CDN formation) under varying concentrations of ATP and GTP (Burns et al., 2014). As expected, DncG was active with both ATP and GTP (Figure 3).

For both individual substrates, we performed least-squares minimization of the initial rates to solutions of both noncooperative and cooperative mechanisms to determine binding constants K_{G1}, K_{G2}, K_{A1}, K_{A2} and catalytic rate constants ($k_{cat,diG}$ and $k_{cat,diA}$) for the respective enzymes (Table 1). As predicted from a mechanism that consists of two substrate binding steps, fits of our initial rates to Michaelis-Menten parameters were suboptimal compared to a multistep regime.

Our analysis reveals that HyPr GGDEFs function similarly with ATP and GTP as substrates at the catalytic steps, with the primary differences being observed in

substrate binding. The rate-determining step, which we expect to be the formation of both phosphodiester bonds and is given by the k_{cat} values, is effectively the same for both substrates. In contrast, assuming a non-cooperative mechanism, the initial dissociation constant for ATP is effectively 4-fold higher for ATP than that of GTP. In a cooperative mechanism, this difference between dissociation constants is increased to 9-fold. Assuming a cooperative mechanism, the effects on substrate binding differ: initial binding of ATP results in a 6-fold increase in affinity for the second ATP binding event, whereas GTP binding doubles the affinity for the second GTP binding event.

Understanding the kinetics of a HyPr GGDEF in a context containing both purine nucleotide triphosphates is more complex (Figure 2). Our single substrate fits allowed determination of all major kinetic parameters except three: the two secondary binding events that would lead to production of the heterodimeric product (i.e. $K_{A|G}$, the equilibrium constant for ATP binding given that GTP occupies the other site, and $K_{G|A}$), as well as the catalytic rate constant for heterodimer formation. As no direct numerical solution to this kinetic scheme exists, we chose to model the reaction pathway using Python and compare fits between modeled product ratios to experimentally determined product ratios under varying ratios of ATP:GTP obtained under the same conditions (Figure 4). In our modeling, we assumed that $k_{cat,cAG}$ was equal to the other catalytic rate constants (k_{cat} = 0.03 sec^{-1}), as the k_{cat} values for the homodimeric CDNs were similar. We varied the effective equilibrium constants of the two unknown values between 1-100 μM in 1 μM increments, and generated a least squares 'fit' between the calculated ratios of CDNs and the observed ratios (Figure 5C). A contour map of this least squares fit shows a local minimum of $K_{A|G}$ = 86 μM and $K_{G|A}$ = 1 μM. The error for the computational fit is more sensitive to $K_{G|A}$ values than $K_{A|G}$ values; in the lowest contour the $K_{A|G}$ values range from 79-100 μM, whereas the $K_{G|A}$ remains at the minimum of 1 μM. The computation fit utilizing our cooperative model more accurately replicate our previous experimental results under different ratios of ATP and GTP than fits using a non-cooperative model (Figure 5A, 5B). Thus, we conclude that GGDEF cooperativity is an essential component of HyPr GGDEF mechanism.

Another consequence of the in vitro modeling was that we observed the experimental product ratios are skewed by substrate depletion. However, bacteria and other living organisms maintain homeostasis of their NTP levels, and therefore the product ratios observed may not be characteristic of the actual CDN ratios produced in *Geobacter*. Performing these simulations under conditions where the concentrations of ATP and GTP do not change reveals that cAG is the predominant product produced (>50%) above ratios of 2.3:1 ATP:GTP (Figure 5D). Thus, while additional intracellular factors could contribute further to dinucleotide cyclase specificity, HyPr enzymes are uniquely tuned to produce primarily cAG in a physiological context.

Discussion

A primary question of our investigation has been: how does a homodimeric active site produce primarily a heterodimeric product *in vivo*? DncV and cGAS produce heterodimeric cyclic dinucleotides through a "simple" answer - there are discrete binding sites selective for the donor and acceptor nucleotide triphosphates, so the enzyme can synthesizes only one linear intermediate. This linear intermediate then gives rise to only

one cyclic product, as the enzyme reorients it in the active site to form the second phosphodiester bond. However, this strategy does not occur in DncG, because HyPr enzymes have a promiscuous active site that can bind either ATP or GTP in each nucleotide binding site. The symmetric nature of the active site permits the enzyme to produce all four possible isomers of 3',5'-linear dinucleotide triphosphate.

Our kinetics investigations demonstrate that, in contrast to the paradigm exemplified by cGAS and DncV, Nature can make a symmetric active site preferentially synthesize a non-symmetric product by cooperative substrate binding. Previous work has demonstrated cooperative binding of GTP in a diguanylate cyclase from *Xanthomonas*, but it is unknown if this is a general paradigm for this enzyme family, and how large of an effect it might play in HyPr enzymes (Oliveira et al., 2015). Our results demonstrate that cooperative binding extends to HyPr enzymes, in both homodimeric and heterodimeric product-generating cases. The cell extract results obtained from HyPr GGDEF enzymes from *Bdellovibrio* and *Myxococcus* species, where cAG predominates above the other two CDNs, suggests that cooperative substrate binding is a common paradigm in the HyPr subclass, and likely GGDEF enzymes as a whole.

Utilizing our kinetics simulations to determine the full kinetic parameters of DncG gives further support to this hypothesis: we are able to achieve reasonable fits to our experimental ratio data by varying only the effective dissociation constants for the various substrates. However, this does not preclude the possibility that subtle difference in the catalytic rate constants may occur. For instance, even a two-fold increase in the rate constant for cAG formation above that observed for cdiA and cdiG would increase the amount of AG formed compared to the other two CDNS from a ratio of 26:62:12 cdiG:cAG:cdiA to 16:77:7 at a 5:1 ratio of ATP:GTP.

DncG has a canonical response regulator receiver domain, which is likely the target of a histidine kinase for phosphorylation (Capra and Laub, 2012). Given precedent from other Rec-GGDEF proteins, we expect that phosphorylation is required for activation, and that activation will change the activity by up to 50-fold (Huangyutitham et al., 2013; Paul et al., 2007). We expect that this change in activity is due almost exclusively to a change in the "effective" catalytic rate constant: we determined kcat values by assuming vmax = kcat*[E]total, but not all of our enzyme is in the active state. We have observed that GSU1658 at μM concentrations is present as both inactive dimer and dimer , which we assume to be "active-like" (Hallberg et al., 2016). Thus, increasing the amount of protein in the active state will give a higher value for k_{cat}. In further support of this assumption, the effective dissociation constants of constitutively active GGDEF enzymes for GTP remain similar to other measurements for these unactivated proteins (Blain-Hartung et al., 2017; Burns et al., 2016; da Costa Vasconcelos et al., 2017; Oliveira et al., 2015; Wassmann et al., 2007). Indeed, our kinetic data recapitulate the effective binding and catalytic rate constants of I-site mutants of unactivated PleD, the Rec-GGDEF from *Caulobacter crescentus*, suggesting further that this protein requires phosphorylation to become active (Wassmann et al., 2007). Therefore, it remains possible that phosphorylation of one or both Rec domains, or interaction with another partner protein, could change the effective catalytic rate constants for production of cAG compared to the homodimeric CDNs by a different amount.

Materials and Methods

General Reagents and Oligonucleotides

All oligonucleotides were purchased from Elim Biopharmaceuticals (Hayward, CA) or IDT (Coralville, IA). Plasmids for expression of MBP-tagged GSU1658 were reported previously (Hallberg et al., 2016). Plasmids encoding MBP-tagged cGAS and DncV were obtained from the laboratory of Jennifer A. Doudna at UC Berkeley (Kranzusch et al., 2013; Kranzusch et al., 2014).

Overexpression and Purification of Dinucleotide Cyclase Enzymes

Full-length proteins with N-terminal His-6-MBP tags encoded in a pET16-derived plasmid were overexpressed in E. coli BL21 (DE3) star cells harboring a pRARE2 human tRNA plasmid and were grown in LB/carb/chlor for 10 h after induction at OD600 ~ 0.7 with 1 mM IPTG. Cells were lysed by sonication in a lysis buffer containing 25 mM Tris-HCl (pH 8.2), 500 mM NaCl, 20 mM imidazole, and 5 mM beta-mercaptoethanol. Clarified lysate was bound to Ni-NTA agarose (QIAGEN), and resin was washed with lysis buffer prior to elution with lysis buffer supplemented with 500 mM imidazole. Proteins were dialyzed overnight at 4 ºC against buffer containing 20 mM HEPES-KOH (pH 7.5), 250 mM KCl, 1 mM TCEP, and 5% (v/v) glycerol. Protein purified in this way was concentrated to ~5-10 mg/mL, flash frozen in liquid nitrogen, and stored at -80 ºC. Protein with C-terminal His-6x tags encoded in pET24a were overexpressed and purified similarly, with the cells grown in LB/kan/chlor.

In Vitro Activity Assay of Dinucleotide Cyclases using Radiolabeled NTPs

In vitro activity assays were performed as previously described, with the following modifications (Kranzusch et al., 2014). Enzyme (10 μM) was incubated in a solution of 50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 10 mM $MgCl_2$, and 5 mM dithiothreitol 100 μM NTP (and/or nonhydrolyzable analog) and ~0.1 μCi radiolabeled [α-^{32}P]-ATP or [α-^{32}P]-GTP (Perkin Elmer) as indicated. Reactions were incubated at 28 ºC for 1 h.

Following incubation, the reaction was treated with 20 units of Calf Intestinal Alkaline Phosphatase (NEB) at 28 ºC for 30 min to digest the unincorporated NTPs. Reactions were terminated by heating to 95 ºC 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 radio labeled products were detected using a Phosphor-image screen (GE Healthcare) and a Typhoon scanner (GE Healthcare).

In Vitro Activity Assay of Dinucleotide Cyclases using Pyrophosphatase Assay

In vitro activity assays were performed as previously described, with slight modifications (Burns et al., 2014). The EnzChek pyrophosphate kit (Life Technologies) was used according to the manufacturer's instructions with the exception that the buffer was supplemented with KCl to a final concentration of 100 mM and $MgCl_2$ to a final concentration of 10 mM, and the reactions were initiated with addition of ATP or GTP.

Assays were performed in triplicated in 96 well plates containing 1 μM protein and varying NTP concentrations (0-10 mM). Plates were monitored using a SpectraMax® i3x plate reader (Molecular Devices) and SoftMax Pro 6.5.1 software (Molecular Devices). Subsequent analyses to determine enzymatic rates were performed using the Excel Solver package (Microsoft).

In Vitro Activity Assay of Dinucleotide Cyclases using LC-MS

Activity assays were performed as described previously (Burhenne and Kaever, 2013; Hallberg et al., 2016). LC-MS analysis of enzyme reactions was performed using an Agilent 1260 Quadrupole LC-MS with an Agilent 1260 Infinity liquid chromatograph equipped with a diode array detector. Sample volumes of 10 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 enzyme reactions, a elution program consisting of 0% B for 5 minutes, followed by linear elution of 0 to 10% B over 1.5 minutes, isocratic elution at 10% B for 2 minutes, linear elution of 10 to 30% B over 2.5 minutes, linear elution from 30 to 0% B over 10 seconds, and isocratic elution of 0% B for 4 minutes, 50 seconds was used. Solvent A was 10 mM ammonium acetate/0.1% acetic acid and solvent B was HPLC-grade methanol.
Under these conditions, the cyclic dinucleotides in extracts were found to always elute in the order of cdiG (10.23±0.02 min), cAG (10.56±0.02 min), and cdiA (11.09±0.05 min). Due to slight variability in retention times, the assignment of cyclic dinucleotide identity was made through analysis of the mass spectra. Additionally, we note that, while cdiG and cAG run as clearly defined peaks even at high concentrations, cdiA exhibits a fronted peak shape at and above 25 μM concentration. Samples were analyzed by MS in the positive ion mode using the range of m/z = 150 to 1000.

Computational-Based Modeling of Dinucleotide Cyclase Activity

A Python program (version 3.6.1) was developed using the NumPy ODEInt function to model the kinetics of HyPr GGDEFs. One example program is included below. In particular, to generate Kd values, we assume a set k_{on} value of 1 $\mu M^{-2}sec^{-1}$, and vary the k_{off} value. These values for on and off rate are significantly higher than the catalytic rate constant such that the steady-state approximation is obeyed.

To determine the error from our experimental data, we summed the least squares difference between predicted and experimental data across four ATP:GTP ratios: 1:1, 3:1, 5:1, and 9:1.

For *in vivo* simulations, we set $dG/dt = dA/dt = 0$. Our data was obtained by using initial values of [Enzyme] = 1 μM, [GTP] = 100 μM, and varied [ATP] accordingly, however we observe that the results hold absolutely for higher initial concentrations of NTP, provided that the ratios remain constant.

Python (v.3.6.1) Program for GGDEF Simulations

```
import numpy as np
def ggdef(y, t, k1A, kn1A, k2A, kn2A, k1G, kn1G, k2G, kn2G, kAG, knAG, kGA, knGA, kcat):
  E, EGn, EnG, EGG, EnA, EAn, EAA, EAG, EGA, G, A, cdiG, cAG, cdiA = y

  dEdt = kn1G*EnG - k1G*E*G + kn1G*EGn - k1G*E*G + kn1A*EAn - k1A*E*A +
kn1A*EnA - k1A*E*A + kcat*EGG + kcat*EGA + kcat*EAA + kcat*EAG
  dEnGdt = k1G*E*G - kn1G*EnG - k2G*EnG*G + kn2G*EGG + knAG*EAG - kAG*EnG*A
  dEGndt = k1G*E*G - kn1G*EGn + kn2G*EGG - k2G*EGn*G + knAG*EGA - kAG*EGn*A
  dEnAdt = k1A*E*A - kn1A*EnA + knGA*EGA - kGA*EnA*G + kn2A*EAA - k2A*EnA*A
  dEAndt = k1A*E*A - kn1A*EAn + kn2A*EAA - k2A*EAn*A + knGA*EAG - kGA*EAn*G
  dEGGdt = k2G*EnG*G - kn2G*EGG + k2G*EGn*G - kn2G*EGG - kcat*EGG
  dEGAdt = kAG*EGn*A - knAG*EGA + kGA*EnA*G - knGA*EGA - kcat*EGA
  dEAGdt = kAG*EnG*A - knAG*EAG + kGA*EAn*G - knGA*EAG - kcat*EAG
  dEAAdt = k2A*EnA*A - kn2A*EAA + k2A*EAn*A - kn2A*EAA - kcat*EAA
  dcdiAdt = kcat*EAA
  dcdiGdt = kcat*EGG
  dcAGdt = kcat*EAG + kcat*EGA
  dAdt = -k1A*E*A + kn1A*EAn - k1A*E*A + kn1A*EnA - kAG*EGn*A + knAG*EGA -
k2A*EnA*A + kn2A*EAA - k2A*EAn*A + kn2A*EAA - kAG*EnG*A + knAG*EAG
  dGdt = -k1G*E*G + kn1G*EnG - k1G*E*G + kn1G*EGn - k2G*EnG*G + kn2G*EGG -
k2G*EGn*G + kn2G*EGG - kGA*EnA*G + knGA*EGA - kGA*EAn*G + knGA*EAG
  #dAdt = 0
  #dGdt = 0
  dydt = [dEdt, dEGndt, dEnGdt, dEGGdt, dEnAdt, dEAndt, dEAAdt, dEAGdt, dEGAdt, dGdt,
dAdt, dcdiGdt, dcAGdt, dcdiAdt]
  return dydt

k1A = 1
kn1A = 80
k2A = 1
kn2A = 80
k1G = 1
kn1G = 25
k2G = 1
kn2G = 25
kAG = 1
knAG = 80
kGA = 1
knGA = 25
kcat = 0.03

y0 = [1, 0, 0, 0 ,0 , 0, 0, 0, 0, 0, 100, 200, 0, 0, 0]
```

```
t = np.linspace(0,3600,3600)

from scipy.integrate import odeint
sol = odeint(ggdef, y0, t, args =(k1A, kn1A, k2A, kn2A, k1G, kn1G, k2G, kn2G, kAG, knAG,
kGA, knGA, kcat))

def totalCDN(x,y,z):
  sol = odeint(ggdef, [1, 0, 0, 0, 0, 0, 0, 0, 0, 100, z, 0, 0, 0], t, args =(k1A, kn1A, k2A, kn2A,
k1G, kn1G, k2G, kn2G, kAG, x, kGA, y, kcat))
  return(sol[3599,11]+sol[3599,12]+sol[3599,13])

def ratdiG(x,y,z):
  sol = odeint(ggdef, [1, 0, 0, 0, 0, 0, 0, 0, 0, 100, z, 0, 0, 0], t, args =(k1A, kn1A, k2A, kn2A,
k1G, kn1G, k2G, kn2G, kAG, x, kGA, y, kcat))
  return(100*sol[3599,11]/totalCDN(x,y,z))

def ratAG(x,y,z):
  sol = odeint(ggdef, [1, 0, 0, 0, 0, 0, 0, 0, 0, 100, z, 0, 0, 0], t, args =(k1A, kn1A, k2A, kn2A,
k1G, kn1G, k2G, kn2G, kAG, x, kGA, y, kcat))
  return(100*sol[3599,12]/totalCDN(x,y,z))

def ratdiA(x,y,z):
  sol = odeint(ggdef, [1, 0, 0, 0, 0, 0, 0, 0, 0, 100, z, 0, 0, 0], t, args =(k1A, kn1A, k2A, kn2A,
k1G, kn1G, k2G, kn2G, kAG, x, kGA, y, kcat))
  return(100*sol[3599,13]/totalCDN(x,y,z))
```

Python (v.3.6.1) Program to optimize least squares fit:

```
import numpy as np

diGmat = np.full((100,100,10), 0, dtype=float)
AGmat = np.full((100,100,10),0, dtype=float)
diAmat = np.full((100,100,10),0, dtype=float)

print(diGmat)

for r in [1,3,5,9]:
  for y in range(0,100):
    for z in range(0,100):
      diGmat[y,z,r] = diGmat[y,z,r] + ratdiG(y,z,r*100)
      AGmat[y,z,r] = AGmat[y,z,r] + ratAG(y,z,r*100)
      diAmat[y,z,r] = diAmat[y,z,r] + ratdiA(y,z,r*100)

print(diGmat)

diGexp = np.full((100,100,10), 0, dtype=float)
AGexp = np.full((100,100,10), 0, dtype=float)
diAexp = np.full((100,100,10), 0, dtype=float)
for y in range(0,100):
  for z in range(0,100):
    diGexp[y,z,1] = 66
    diGexp[y,z,3] = 36
    diGexp[y,z,5] = 18
    diGexp[y,z,9] = 6
    AGexp[y,z,1] = 32
    AGexp[y,z,3] = 44
    AGexp[y,z,5] = 45
    AGexp[y,z,9] = 40
    diAexp[y,z,1] = 2
    diAexp[y,z,3] = 20
    diAexp[y,z,5] = 37
    diAexp[y,z,9] = 54

final = np.full((100,100), 0, dtype=float)

for y in range(0,100):
  for z in range(0,100):
    for m in range(0,10):
      final[y,z] = final[y,z] + (diGexp[y,z,m]-diGmat[y,z,m])**2 + (AGexp[y,z,m]-
AGmat[y,z,m])**2 + (diAexp[y,z,m]-diAmat[y,z,m])**2
```

Figures

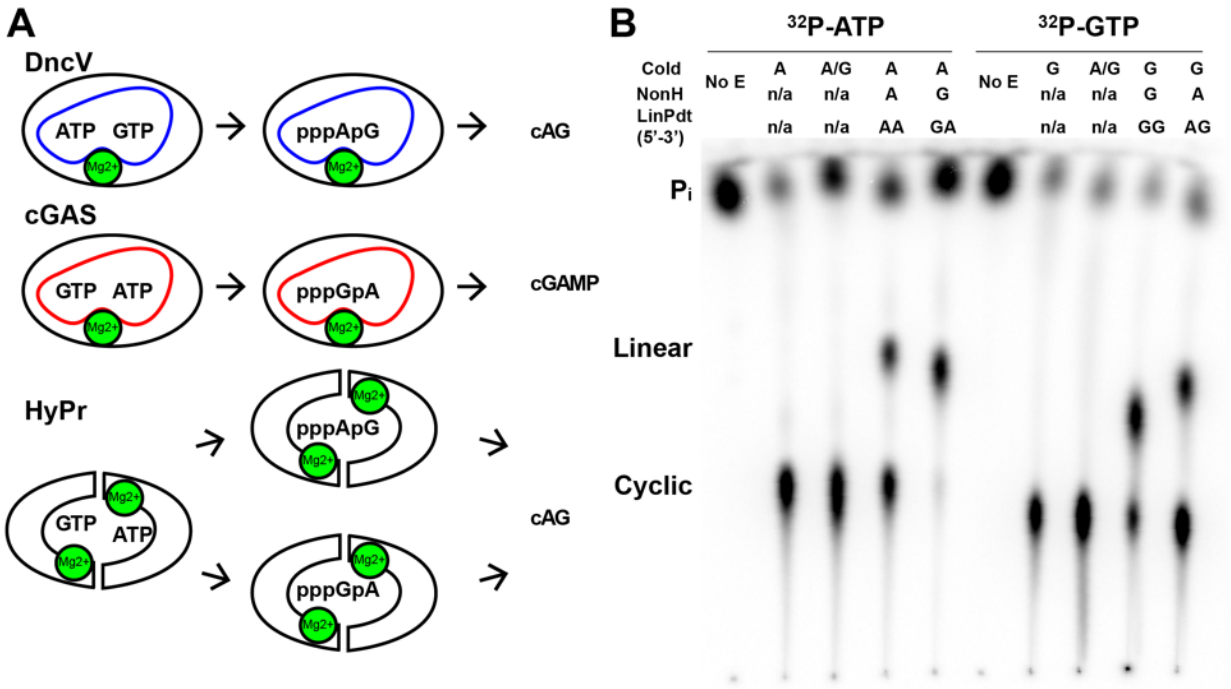


Figure 1 - Hybrid Promiscuous GGDEF Eenzymes Can Produce the Heterodimer Product 3',3'-cAG in Any Order.

(A) Schematic for possible reaction pathways to form the hybrid cyclic dinucleotide cAG. DncV catalyzes product formation using the ATP 3'-OH first to from pppApG. cGAS catalyzes product formation through initial attack by the GTP 3'-OH to form pppGpA, followed by cyclization. HyPr GGDEF enzymes can catalyze product formation through either order.

(B) Cellulose TLC analysis of radiolabeled products from enzymatic reactions with NTP substrates and nonhydrolyzable analogues doped with trace amounts of α-^{32}P-labeled ATP or α-^{32}P-labeled GTP. Before loading, reactions were quenched with alkaline phosphatase to digest unreacted nucleotides, resulting in production of inorganic phosphate (P_i). Linear products are listed as NN, where the full chemical structure of the non-hydrolyzable nucleotide expected after alkaline phosphatase treatment is p(CH_2)pNpN.

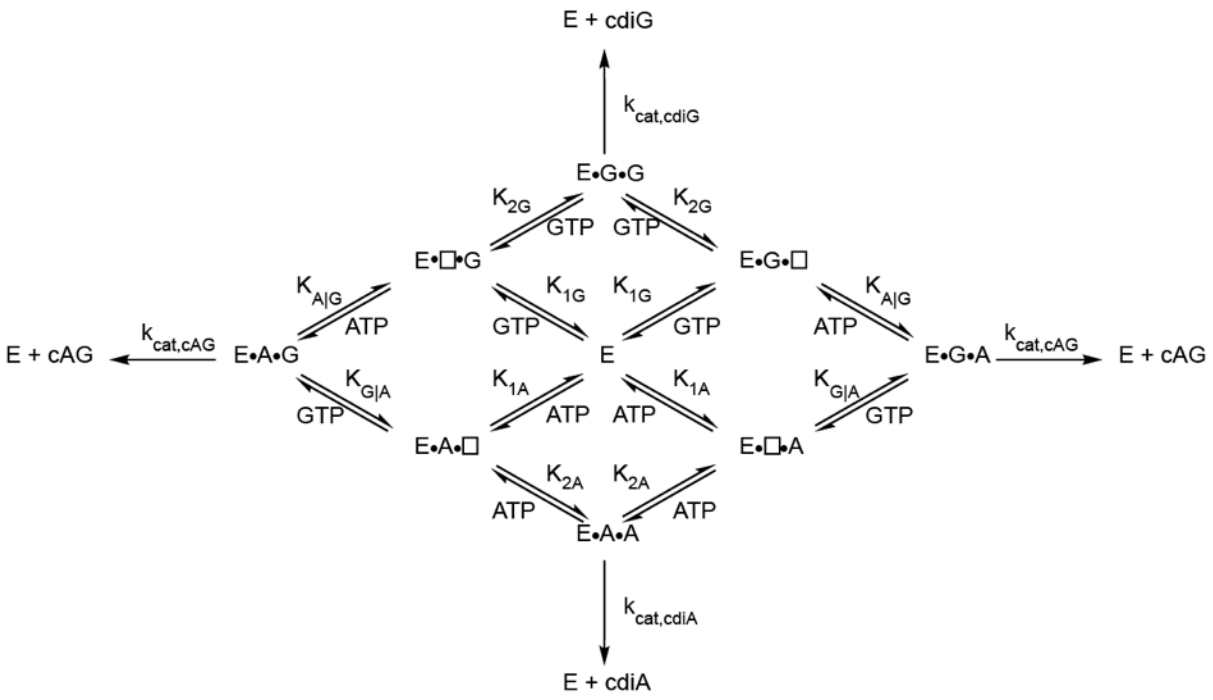


Figure 2 - Mechanistic Scheme for HyPr GGDEFs.

“E” represents a GSU1658 homodimer in the ‘active’ conformation. “E-N□” and “E-□N” represents a GSU1658 homodimer with a NTP bound in the first or second half-active sites, respectively. “E-NN” represents a GSU1658 homodimer with two NTPs bound. We assume that the two half-active sites are equivalent. K_{N1} represents the affinity for the first NTP binding event, and K_{N2} represents the affinity for the second NTP binding event in the homodimeric case. $K_{X|Y}$ represents the affinity for the second NTP binding event in which binding of X is conditioned upon the fact that Y is already bound. Thus, $K_{G|A}$ is the binding affinity for a GTP in the second half active site given an enzyme that already has ATP bound.

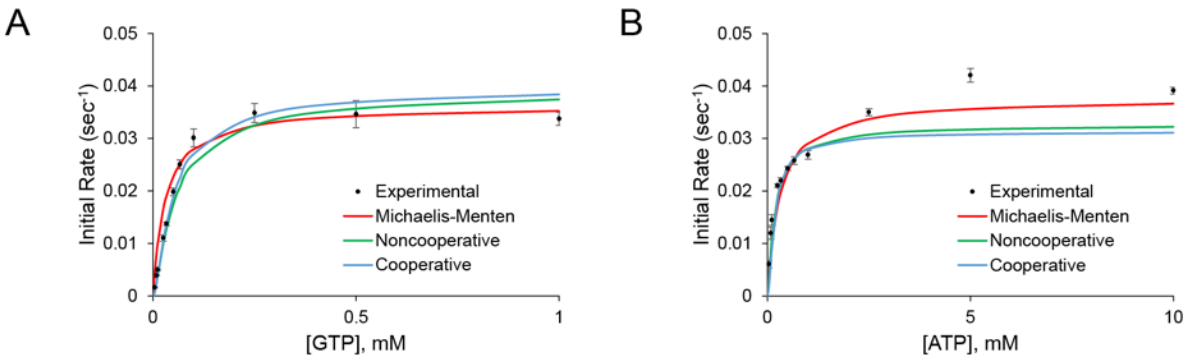


Figure 3- Steady-State Kinetic Measurements of MBP-tagged GSU1658 R393A Diguanylate Cyclase and Diadenylate Cyclase Activity. Initial rate measurements of product formation using (a) ATP or (b) GTP as a substrate.

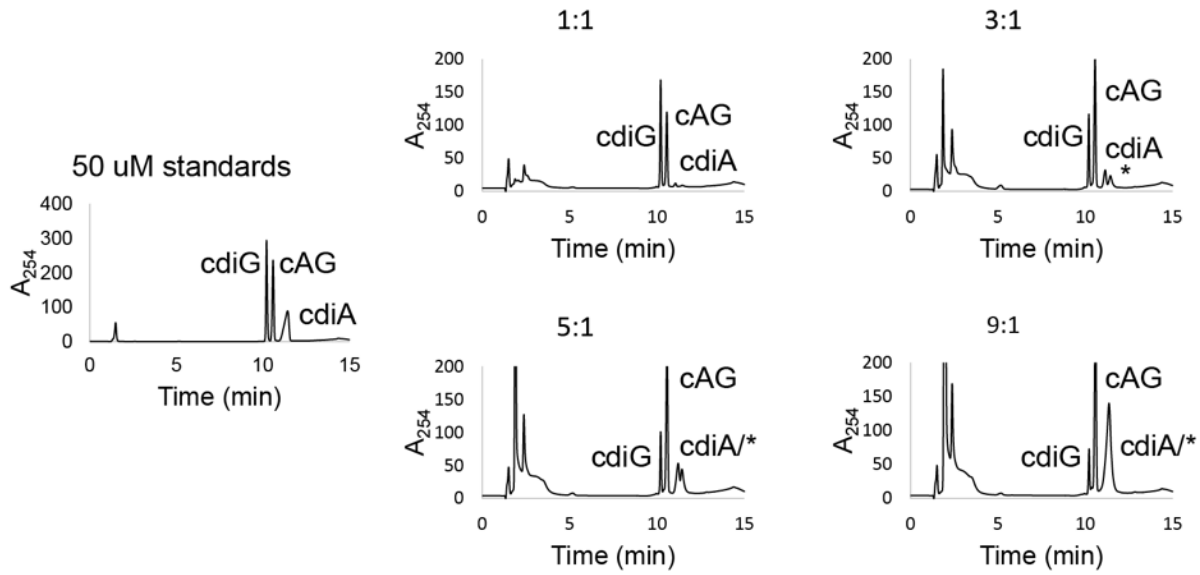


Figure 4 - LC-MS Based Product Ratio Analysis of MBP-tagged GSU1658 R393A. HPLC-MS analysis of endpoint reactions of a 6xHis-MBP-tagged, I-site knockout (R393A) mutant of GSU1658 (1 μM) under varying ATP:GTP ratios, with [GTP] kept constant at 100 μM. Ratios of ATP:GTP are shown above each trace. * represents an additional enzymatic product that absorbs at 254 nm and has a retention time that overlaps with the cdiA peak.

Under these conditions, the cyclic dinucleotides in extracts were found to always elute in the order of cdiG (10.23±0.02 min), cAG (10.56±0.02 min), and cdiA (11.09±0.05 min).

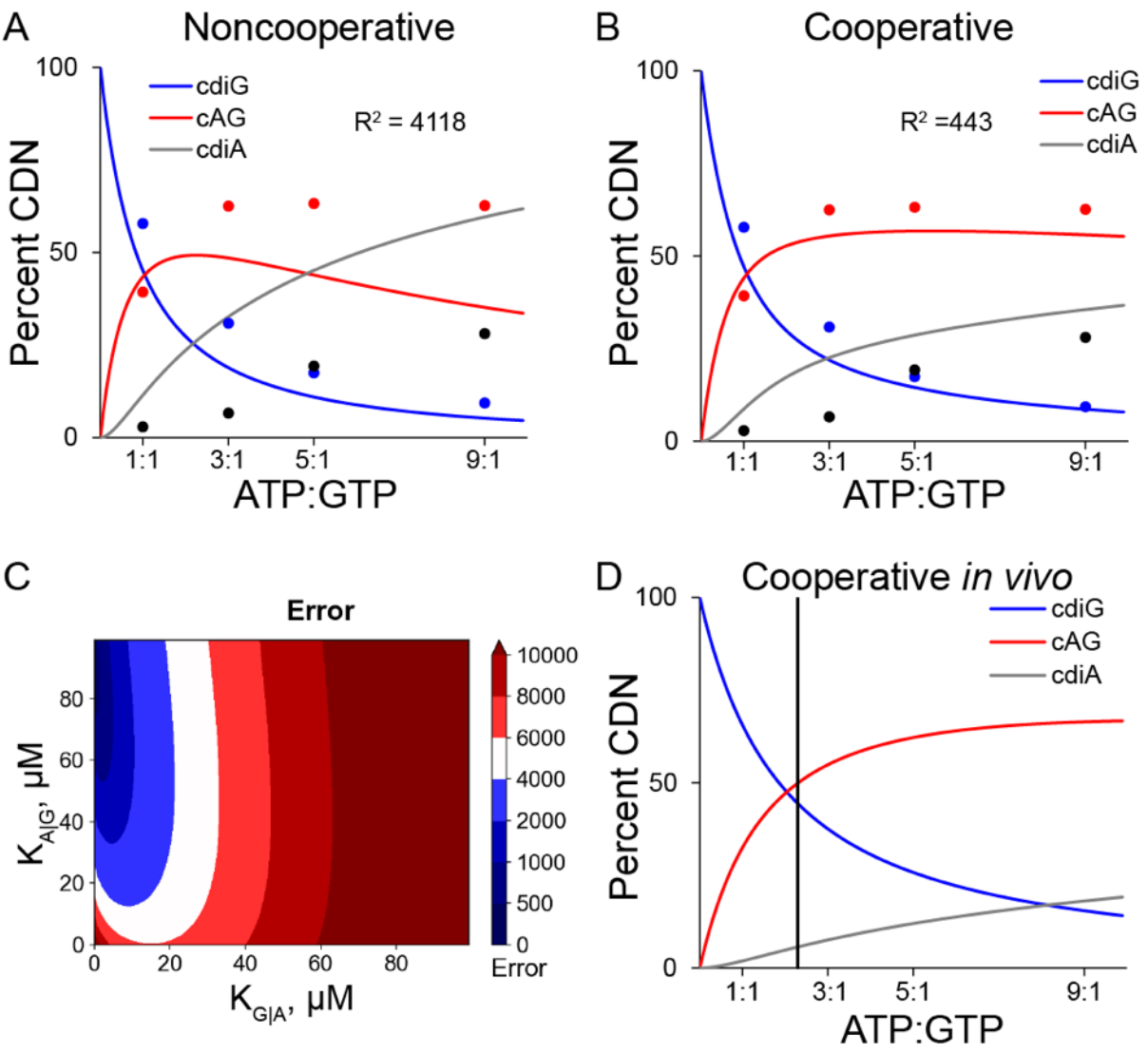


Figure 5 - Python Based Modeling of HyPr GGDEF Activity.

(A) Analysis of product ratios for MBP-GSU1658 R393A at different ATP to GTP ratios using Python modeling. Curves represent ratios obtained by fitting to a non-cooperative mechanism. Data points are an average of n=3 experiments.

(B) As in (A), with best-fit cooperative modeling of $K_{G|A}$=1 μM, and $K_{A|G}$=86 μM.

(C) Least squares error obtained by varying $K_{G|A}$ and $K_{A|G}$ between 0-100 μM and subtracting the obtained ratios of cdiG:cAG:cdiA against the experimental values previously obtained. A minimum is observed at $K_{G|A}$ = 1 μM and $K_{A|G}$ = 86 μM.

(D) Modeled ratios of cdiG:cAG:cdiA produced under “biological” conditions (dATPdt = dGTPdt = 0) at varying ratios of ATP:GTP. The vertical black line represents the ratio (ATP:GTP = 2.3:1) at which cAG becomes the predominant product.

Table 1 - Kinetic Parameters of MBP-GSU1658 R393A for ATP and GTP According to Different Models. For modeling purposes, we assume that $k_{cat,diA} = k_{cat,diG} = k_{cat,AG} = 0.03\ sec^{-1}$.

	Non-cooperative	Cooperative
K_{1A}, μM	80	343
K_{2A}, μM	80	53
$k_{cat,cdiA}$, sec^{-1}	0.04	0.03
K_{1G}, μM	25	39
K_{2G}, μM	25	20
$k_{cat,cdiG}$, sec^{-1}	0.04	0.04
$K_{A\|G}$, μM	80	86
$K_{G\|A}$, μM	25	1
$k_{cat,cAG}$, sec^{-1}	0.04	0.03

Bibliography

Blain-Hartung, M., Rockwell, N.C., and Lagarias, J.C. (2017). Light-Regulated Synthesis of Cyclic-di-GMP by a Bidomain Construct of the Cyanobacteriochrome Tlr0924 (SesA) without Stable Dimerization. Biochemistry *56*, 6145-6154.

Burhenne, H., and Kaever, V. (2013). Quantification of cyclic dinucleotides by reversed-phase LC-MS/MS. Methods Mol Biol *1016*, 27-37.

Burns, J.L., Deer, D.D., and Weinert, E.E. (2014). Oligomeric state affects oxygen dissociation and diguanylate cyclase activity of globin coupled sensors. Mol Biosyst *10*, 2823-2826.

Burns, J.L., Rivera, S., Deer, D.D., Joynt, S.C., Dvorak, D., and Weinert, E.E. (2016). Oxygen and Bis(3',5')-cyclic Dimeric Guanosine Monophosphate Binding Control Oligomerization State Equilibria of Diguanylate Cyclase-Containing Globin Coupled Sensors. Biochemistry *55*, 6642-6651.

Capra, E.J., and Laub, M.T. (2012). Evolution of two-component signal transduction systems. Annu Rev Microbiol *66*, 325-347.

Chan, C., Paul, R., Samoray, D., Amiot, N.C., Giese, B., Jenal, U., and Schirmer, T. (2004). Structural basis of activity and allosteric control of diguanylate cyclase. Proc Natl Acad Sci U S A *101*, 17084-17089.

da Costa Vasconcelos, F.N., Maciel, N.K., Favaro, D.C., de Oliveira, L.C., Barbosa, A.S., Salinas, R.K., de Souza, R.F., Farah, C.S., and Guzzo, C.R. (2017). Structural and Enzymatic Characterization of a cAMP-Dependent Diguanylate Cyclase from Pathogenic Leptospira Species. J Mol Biol *429*, 2337-2352.

Gourinchas, G., Etzl, S., Gobl, C., Vide, U., Madl, T., and Winkler, A. (2017). Long-range allosteric signaling in red light-regulated diguanylyl cyclases. Sci Adv *3*, e1602498.

Hallberg, Z.F., Wang, X.C., Wright, T.A., Nan, B., Ad, O., Yeo, J., and Hammond, M.C. (2016). Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3', 3'-cGAMP). Proc Natl Acad Sci U S A *113*, 1790-1795.

Huangyutitham, V., Guvener, Z.T., and Harwood, C.S. (2013). Subcellular clustering of the phosphorylated WspR response regulator protein stimulates its diguanylate cyclase activity. MBio *4*, e00242-00213.

Kranzusch, P.J., Lee, A.S., Berger, J.M., and Doudna, J.A. (2013). Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. Cell Rep *3*, 1362-1368.

Kranzusch, P.J., Lee, A.S.Y., Wilson, S.C., Solovykh, M.S., Vance, R.E., Berger, J.M., and Doudna, J.A. (2014). Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. Cell *158*, 1011-1021.

Oliveira, M.C., Teixeira, R.D., Andrade, M.O., Pinheiro, G.M., Ramos, C.H., and Farah, C.S. (2015). Cooperative substrate binding by a diguanylate cyclase. J Mol Biol *427*, 415-432.

Paul, R., Abel, S., Wassmann, P., Beck, A., Heerklotz, H., and Jenal, U. (2007). Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization. J Biol Chem *282*, 29170-29177.

Wassmann, P., Chan, C., Paul, R., Beck, A., Heerklotz, H., Jenal, U., and Schirmer, T. (2007). Structure of BeF3- -modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. Structure *15*, 915-927.

Chapter Four

Structural Insights into HyPr GGDEF Function

Introduction

Canonical diguanylate cyclases are members of the nucleotidyltransferase superfamily of enzymes, a diverse class of enzymes that includes adenylate cyclases, the CRISPR polymerase Cas10, the CCA-adding enzyme, and the cytosolic nucleic acid surveillance proteins cyclic GMP-AMP synthase (cGAS) and oligoadenylate synthase (OAS) (Burroughs et al., 2015; Kranzusch et al., 2013; Sinha and Sprang, 2006). These proteins have a conserved beta sheet that contains both acidic residues required for coordination to a divalent metal and activation of the nucleotide substrate, and basic residues that contact the triphosphate backbone (Chan et al., 2004; Donovan et al., 2013; Gao et al., 2013; Kranzusch et al., 2014). Flanking helical regions serve as a scaffold on which nucleobase-recognizing residues are displayed and lead to a wide variety of substrate-binding modes in these different enzymes (Gao et al., 2013; Kranzusch et al., 2014; Osawa et al., 2013).

One open question is how this conserved fold has evolved new diverse functions in catalyzing the formation of nucleotide-based signaling molecules. One intriguing example is seen in our discovery of Hybrid product producing, Promiscuous substrate binding (HyPr) GGDEF proteins. The canonical GGDEF domain possesses diguanylate cyclase (DGC) activity, synthesizing only cyclic di-GMP (cdiG). We have shown that a number of GGDEF domains-containing proteins from deltaproteobacteria can synthesize all three cyclic dinucleotides. While a conserved aspartate-to-serine mutation leading to HyPr activity was found, corresponding mutations in canonical diguanylate cyclases result primarily in inactive enzymes. Furthermore, these HyPr GGDEF enzymes preferentially synthesize the hybrid product by exhibiting cooperative effects for substrate binding, a phenomenon that has been observed in these signaling proteins but for which no structural basis is known.

In this chapter, I present the first solved x-ray crystal structure of a HyPr GGDEF domain, the DncG enzyme from *Geobacter metallireducens*. This x-ray crystal structure with GTP bound provides further clarity into how HyPr GGDEFs select their substrate. First, we demonstrate that a ‘molecular ruler’ between the signature serine residue and the GGDEF motif is important to permit nucleotide binding and conservation of diguanylate cyclase activity. Furthermore, we identify two previously unrecognized cross-dimer interactions involved in substrate selection and catalysis. These results set the stage for understanding the role of cAG in Geobacter and related organisms with HyPr signaling pathways, as well as future rational re-engineering attempts of GGDEFs as signaling enzymes in bacteria.

Results

X-ray Crystal Structure of GGDEF Domain of the *G. metallireducens* HyPr GGDEF Enzyme

Nature employs two major strategies to generate new protein activity: active-site remodeling to accommodate a new substrate, and generation of a new active site (Khersonsky and Tawfik, 2010; Siddiq et al., 2017). In the former, the new function evolves with low negative tradeoff from the original function, leading to a ‘generalist’

enzyme capable of performing either activity. In contrast, generation of a new active site would likely lead to loss of function followed by gain of new function. Our kinetic and biochemical analyses strongly suggest the former strategy, because HyPr GGDEF enzymes function equivalently to diguanylate cyclases when given only GTP as a substrate. To gain further support for this hypothesis, and insight into active site remodeling that must exist in HyPr GGDEFs, we pursued structural characterization and obtained a 1.4 Å resolution crystal structure of an N-terminal T4 Lysozyme fusion of the GGDEF domain (Y_{294}-R_{459}) from Gmet_1914, the HyPr found in *Geobacter metallireducens,* bound to GTP (Figure 1).

Overlay of the HyPr GGDEF domain with the canonical GGDEF domain from a light-activated DGC from *Idiomarina* sp. A28L (Figure 2A) (Gourinchas et al., 2017) shows that the global fold of the HyPr enzyme recapitulates that of the canonical GGDEF. One region behind the active site has been modified from a beta sheet region to a helical/ loop motif; however, this loop region on the opposite side of the two alpha helices supporting the GTP binding site varies considerably between GGDEF structures, and we expect that it is not required for HyPr activity (Chen et al., 2016; Dahlstrom et al., 2015; Deepthi et al., 2014; Yang et al., 2011).

We observed electron density for three bound guanine nucleotides in our structure, two in conserved nucleotide-interacting regions (Figure 2B, 3) (Chan et al., 2004). One guanine nucleotide is found near the canonical allosteric inhibitory site of GGDEF enzymes (Figure 3B) (Chan et al., 2004; Christen et al., 2006). The second nucleotide is bound in the active site, above the GGDEF motif. For this nucleotide, we were unable to find localized electron density for the alpha phosphate (Figure 4). It is likely that, under our crystallization conditions, the GTP nucleotide was hydrolyzed to GMP and pyrophosphate, which both remained coordinated to the active site (Figure 4). For the purpose of clarity, we have chosen to model in the alpha phosphate for this and subsequent sections. A final nucleotide occurs at the T4 Lysozyme-GGDEF interface, presumably acting to stabilize the construct, which may offer a potential explanation for our inability to obtain crystals in the presence of ATP, because the exocyclic ATP amine cannot participate in a hydrogen bonding with the T4 Lysozyme R145 residue, and there is no exocyclic amine at the C2 position to interact with D314 from the HyPr enzyme (Figure 3C). This represents the 5th GGDEF domain with a substrate nucleotide bound in the active site, with three other GGDEF domain structures containing GTPαS (Tarnawski et al., 2015; Wassmann et al., 2007; Zahringer et al., 2013), and one containing GTP (Gourinchas et al., 2017).

Our structural results show strong conservation of the active site between HyPr and canonical GGDEF domains, suggesting a generalist strategy was pursued to evolve HyPr activity (Figure 5). The ribose sits above the GGDEF motif, with the guanosine nucleotide facing towards $\alpha_1\alpha_2$ (the first and second alpha helices with respect to the N-terminus) and the triphosphate moiety splayed around $\alpha_2\alpha_3$. Two specific hydrogen-bonding interactions are made with the nucleobase - a Watson-Crick interaction with S348, and a sugar-face interaction with N339. An additional pi-stacking interaction may exist between the nucleobase and H344, which might explain its role as a conserved residue among GGDEF enzymes that we expect to be active (Figures 2,6). Specific

interactions, analogous to those in canonical GGDEFs, coordinate to the triphosphate backbone. The β and γ phosphates interact with a magnesium ion that is also coordinated to side chain of D331, the backbone carbonyl group of L332, and the side chain of D374 found in the GGDEF motif. Additional interactions occur between the gamma phosphate and K445 and K449.

Watson-Crick Base Contacts - S348

With the global framework of the GGDEF domain intact in HyPr enzymes, we focused our attention on individual substrate-interacting residues to determine their effect on HyPr activity. S348 was the first residue that we found being a signature for HyPr activity (Hallberg et al., 2016). Our original hypothesis has borne out to be true - it forms a Watson-Crick hydrogen bond with the guanosine residue. In contrast, we expect that ATP forms a similar Watson-Crick hydrogen bond with the serine acting as a hydrogen-bond donor. Restoration of the canonical aspartate residue restores full diguanylate cyclase activity.

Our radiolabeled TLC data, together with cell extract data, suggests that the S347D mutation produces only cyclic di-GMP (Hallberg et al., 2016). We chose to characterize this enzyme through endpoint LC-MS based analysis to determine what other products are produced (Figure 7). Indeed, we observe cdiG, along with a side product whose abundance is below the detection limit for characterization by our LC-MS assay.

We further investigated the activity of the S347D mutation in GSU1658 using our RNA-based fluorescent biosensors for cyclic dinucleotides (Wang et al., 2016). *In vivo*, *E. coli* cells coexpressing both and enzyme and a biosensor can be rapidly assessed for synthase activity of the specific CDN. As expected, the S347D mutant shows no response to our cAG biosensor GM790p1-4delA-Spinach1, in agreement with our previous results (Figure 8A). Surprisingly, however, the S347D mutant show minimal response to our cdiG biosensor DP17-Spinach2 (Figure 8A). This result was confusing, as both *in vitro* endpoint analyses and LC-MS analysis of cell extracts with the S347D mutant showed cdiG signal (Figure 7) (Hallberg et al., 2016).

The fact that the S347D mutant *in vitro* produces cdiG at similar levels to that observed by the WT enzyme suggests that this mutant enzyme would have similar kinetic parameters to that of the WT enzyme with respect to GTP, whereas our *in vivo* results suggest that the mutant enzyme is defective for cdiG formation. These seemingly contradictory results were resolved through initial rate analysis of product formation (Figure 9, Table 1). In the case of ATP, as expected, no activity was observed in contrast to the WT enzyme. At low GTP concentrations, the S347D mutant shows a typical kinetic profile that fits well to our kinetic model with little change in GTP dissociation constants from the wild-type enzyme. At higher concentrations (>250 μM) of GTP, however, we observe a linear decrease in enzymatic activity. This suggested to us that product release from the active site was rate-limiting, or that the product itself was now able to act as an inhibitor for the DGC reaction. Our *in vivo* results likely stem from a situation where enzyme produces enough cdiG to cause autoinhibition, preventing cdiG accumulation to levels that can be observed by our biosensors. Taken together, we conclude that the S347 mutation is the primary driver in allowing both ATP and GTP to

act as substrates, and that potential pitfalls - such as autoinhibition, may prevent analysis of enzymatic activity *in vivo* even under over-expression conditions where we expect dimerization - and thus activity - to be forced.

A molecular ruler in the GGDEF motif - D374

Our results suggest that the D-to-S mutation should result in active GGDEFs. However, in four out of five cases, performing the D-to-S mutation on a validated diguanylate cyclase results in an inactive protein (Hallberg et al., 2016). One protein, GSU3350, remained active, but solely as a diguanylate cyclase. One potential explanation comes from the overlay of the canonical GGDEF in complex with GTP with the HyPr GGDEF domain (Figure 8B) (Gourinchas et al., 2017). We observe that the GTP nucleotide is shifted approximately 0.8 angstroms closer to α_2 in the HyPr GGDEF structure. In contrast, the hydrogen bonding distances are approximately equal (2.8 Å in PleD and 2.7 Å in Gmet_1914), which is expected because serine is a shorter side chain than aspartate.

We expect that corresponding mutations or conformational flexibility throughout the GGDEF domain permit this shift in substrate binding by an analogous amount. In the HyPr GGDEF, we indeed observe that the magnesium cation coordinating to the ß and γ phosphates is also shifted by 0.7 Å towards the α_2 helix. This is likely due to the presence of the GG**D**EF motif in Gmet_1914, whereas PleD and WspR contain GG**E**EF motifs, suggesting that this fundamental difference might play into HyPr activity (Chan et al., 2004; De et al., 2009).

These observations suggest a simple ‘molecular ruler’-type model for NTP substrate selection in both HyPr and canonical GGDEFs (Figure 10): if the residue contacting the Watson-Crick face of the nucleotide and magnesium are the appropriate distances apart, then NTPs can bind and catalysis occurs. If, however, the two components are too far apart, substrate binding cannot occur. Therefore, the S347/GG**E**EF residue combination is inactive because the distance is too long for the NTP substrate to make both interactions. In contrast, the D347/GG**D**EF combination should remain active, because the aspartate residue can vary its effective length through conformational flexion.

To investigate the possibility that the GGDEF motif participates in a ‘molecular ruler’ for NTP binding, we performed computational analysis of “active” GGDEFs from the Uniprot database (any GGDEF that has an intact [G/A/S]G[**D/E**]E[F/Y] motif), and compared the motif occurrence in HyPr enzymes (S or T347 as numbered in GSU1658) vs in canonical DGCs (D347). Out of 78,976 sequences that were deemed active (from a total of 119,761 sequences), 33,874 members contain a [S/G/A]G**E**E[F/Y] motif, and 45,102 members contain a [S/G/A]G**D**E[F/Y] motif. While the HyPr enzyme data set is significantly smaller than the canonical data set (91 GGDEFs with a S/T347 and 77,911 with a D347, respectively), we were surprised to find that no HyPr contains a glutamate at this position, whereas canonical diguanylate cyclases alternate between D and E readily (57% D, 43% E - 44,432 D, 33,479 E) (Figures 11, 12).

To lend additional support for the hypothesis that the “GGDEF” motif is required for cdiG synthase activity when an S347 is present, we performed mutational analysis of both this residue and the homologous D226 residue on WspR. Previously, we observed in cell extracts that the WspR D226S mutant possessed no dinucleotide cyclase activity (Hallberg et al., 2016). Mutation of the GGEEF motif of WspR to a GGDEF motif (E370D) maintains catalytic activity when assayed against our RNA-based biosensors *in vivo* (Figure 8A). Surprisingly, however, the double mutant retains diguanylate cyclase activity rather than being promiscuous promiscuous like the HyPr enzyme. Furthermore, we note that in our previous reprogramming efforts, all tested D to S mutations that resulted in inactive proteins have a “GGEEF” motif, while the sole DGC that remained active, GSU3350, has a “GGDEF” motif (Hallberg et al., 2016).

A similar trend is observed when performing corresponding mutations of the HyPr enzyme. Replacing the GG**D**EF motif of GSU1658 with GG**E**EF results in what appears to be an inactive enzyme by *in vivo* biosensor as well as *in vitro* kinetic and endpoint LC-MS analyses (Figure 7, 8A, 9). We expected that the double mutation would restore diguanylate cyclase activity. However, the double mutant displayed no *in vivo* activity against our cdiG biosensor. Kinetic probing and endpoint analyses revealed that the S347D/D373E mutation remained active as a diguanylate cyclase; however, it had an approximately 2-fold reduction in k_{cat} value. We expect that the *in vivo* biosensor results can be explained by the presence of natural phosphodiesterases for cdiG in *E. coli* (Sarenko et al., 2017). If the k_{cat} is not high enough, then the native *E. coli* PDEs are able to prevent an increase in cellular cdiG levels (Sarenko et al., 2017).

Cross-Dimer Contacts and Catalysis - Y304 and E375

Our monomer x-ray crystallographic results demonstrate structural remodeling required to maintain the ability to bind GTP and acquire the ability to bind ATP in HyPr GGDEFs, but offered no further insight into potential sources of cooperativity in these enzymes. We hypothesized that additional residues at the active site interface might be important for substrate selection in GGDEFs. Further support for this hypothesis comes from the fact that GGDEFs must dimerize for NTP catalysis to occur. Furthermore, a number of absolutely conserved residues in the GGDEF domain that have no known function (Schirmer, 2016) may act in a cross-dimer fashion. Multiple structures of GGDEF dimers have been obtained; we chose to analyze the GGDEF dimer from *Idiomarina* sp. A28L, because the C_2 symmetric nature of the structure might come close to recapitulating the expected C_2 symmetric structure of the HyPr GGDEF (Gourinchas et al., 2017).

Superimposition of our structure onto both individual GGDEF domains of the A28L dimer reveals probable residues that may be involved in cross-dimer contact of the nucleobase. In the canonical diguanylate cyclase, we observe an arginine residue (R537) involved in a potential cation-pi base stacking interaction with the guanine residue in the opposing active site (Figure 13A). This arginine residue is conserved in diguanylate cyclases, but has no known function (Schirmer, 2016). In contrast, R537 is replaced with a tyrosine residue (Y304) in the HyPr GGDEF (Figure 13B). While the conformation of the Y304 residue we observe is likely a crystallization artifact of the monomer, simple rotation about the C_α-C_β bond of 170 degrees results in an orientation

that could generate a pi-pi stacking interaction with either adenine or guanine (Figure 14).

We hypothesized that, in canonical diguanylate cyclases, this cross-dimer arginine residue acts as an additional selectivity determinant, forming a cation-π stacking interaction that is antagonistic towards binding of ATP (Figure 15). To investigate this potential cross-dimer residue further, we analyzed the Uniprot database of GGDEF sequences. Limiting our search as before to only “active” sequences (containing an active [G/A/S]G[D/E]E[F/Y] motif, we found that 94% of active GGDEF domains with an aspartate W-C selectivity residue (D344 in PleD) contain a cross-dimer arginine residue, with the majority of mutations observed remaining conservative (H or K) (Figure 13). In contrast, the HyPr GGDEF set contains a broader distribution of cross-dimer interface residues, with only 4% of sequences containing a positively charged residue, and the majority containing either hydrogen bond-donor residues tyrosine, serine, glutamine, and threonine (representing 52% of HyPr sequences), or hydrophobic residues such as alanine or methionine (35%).

In support of the hypothesis that this cross dimer residue is involved in substrate selection, two natural “HyPr” GGDEF sequences with cross-dimer arginine residues, Cabther_A1065 and Ddes1475, produce cdiG competitively to or in excess of cAG, whereas HyPrs with cross-dimer tyrosine, serine, alanine, and glutamine residues produce predominantly cAG (Hallberg et al., 2016). However, as in previous reverse engineering cases, mutations of wild-type WspR cross-dimer arginine residues to any residue observed in HyPr GGDEFs resulted in inactive protein. We conclude that while the cross-dimer tyrosine contact is not necessary to confer HyPr activity, it plays an essential role as an arginine in conferring substrate selectivity in canonical GGDEFs.

Our kinetics analyses previously demonstrated that cooperativity is present in HyPr GGDEF enzymes, however, the origin of this cooperativity remains unknown. We postulate that this cross-dimer residue may play a role in cooperative binding, because it directly communicates binding from one active site to the other monomer. Subjecting GSU1658 Y303R mutant to kinetic analyses with just ATP and GTP revealed similar potential kinetic parameters to the wild-type enzyme, albeit with a slight change in the curve shape for ATP (Figure 9). Intrigued by these results, we performed ratio analysis of the mutant (Figure 16). As expected, cdiG predominated at all ratios tested. While a product with similar retention time to cdiA was also observed, no cdiA was detected by mass spectrometry, even at 9:1 ATP:GTP ratios. Thus, we conclude that the Y303R mutant has altered cooperativity, and that the mutation permits catalysis only when a G molecule is present in at least one binding pocket.

One open problem in GGDEF mechanism remains the identity of the general base presumably required for deprotonation of the 3’ hydroxyl group of the nucleophilic GTP. A potential solution resolves itself upon looking at both the canonical and HyPr superimposition of the GGDEF structures. It appears that the conserved glutamate residue (the fourth residue in GGDEF) is oriented appropriately to engage the incipient hydroxyl group proton from the nucleotide on the opposite face, suggesting that it acts as the general base required to active the nucleophilic component of the cross-dimer substrate (Figure 17A). Indeed, mutation of this residue to glutamine in GSU1658

ablates catalytic activity *in vivo* in a cell extract context (Figure 17B), in line with previous observations demonstrating the requirement for this residue.

Discussion

A primary challenge in understanding nucleotidyltransferase enzymes is determining how specific enzymes bind specific substrates: while the majority of GGDEF enzymes are diguanylate cyclases, reports of GGDEF domains containing cyclic AMP phosphodiesterase and ATP hydrolysis activities are known, in addition to the HyPr subclass (Rao et al., 2010; Sasakura et al., 2002). It is interesting to note that, in many of these A-binding GGDEF proteins, the signature serine for HyPr activity is present, with concomitant loss of the active “GGDEF” motif. Furthermore, the distantly-related Cas10 CRISPR polymerase domain utilizes ATP as a substrate to form cyclic oligoadenylates, and has the same serine residue, but a GGDDF active-site motif (Kazlauskiene et al., 2017; Niewoehner et al., 2017). Therefore, we expect that the serine motif is a general solution for recognition of adenine-containing nucleotides.

Our biochemical and structural results suggest that the GGDEF domain gained HyPr activity through a set of mutations generating a promiscuous binding pocket capable of accepting either ATP or GTP. Our inability to generate HyPr activity in a model canonical GGDEF (WspR) suggests that, in addition to the three residues we have identified, additional variable residues (or combinations of residues) are required to create promiscuous activity in a context-dependent manner. Indeed, while covariation analysis detects the cross-dimer interaction as covarying with the Watson-Crick facing residue, we expect that the use of current mathematical methods to determinate additional residues that may be important in HyPr enzymatic activity would be challenging: the use of covariation to predict protein structure often requires thousands of related sequences to pick up strongly covarying residues (Ovchinnikov et al., 2015), and the HyPr GGDEF dataset consists of less than 100 ‘active’ proteins, of which some do not preferentially produce the heterodimeric product *in vivo* (Hallberg et al., 2016). Furthermore, because the S-to-D mutation is sufficient to abolish HyPr enzyme activity, it is likely that many canonical GGDEF enzymes already possess the requisite enzymatic background to gain HyPr activity upon performing the D-to-S mutation, which would further mask these residues from mathematical analysis.

It has been observed that active diguanylate cyclases can permit either the GGDEF or GGEEF motif, with no discernible rationale for the choice of specific residue used (Schirmer, 2016) However, our results demonstrate a previously unrealized role of this central acidic residue in permitting substrate selection, as GGDEFs with D to S mutations retain activity, while GGEEFs with the same mutation have no catalytic activity. Our studies conclusively demonstrate that the aspartate residue, in combination with the D to S mutation previously identified, acts as a molecular ruler necessary for permitting HyPr activity, in acting to help shift the ligand binding pocket towards the substrate-interacting alpha helix. It remains to be seen what role the D/E mutation has in canonical diguanylate cyclases - it is possible that this ‘ruler' could act to fine-tune both substrate affinity and catalytic activity for the particular role of a diguanylate cyclase.

Biochemical characterization of HyPr and canonical GGDEFs have demonstrated that cooperative substrate-binding is a common paradigm in these enzymes (Oliveira et al., 2015), however it remained unclear what residues could contribute to cross-talk between the two different GGDEF domains could contribute to substrate selection. Our identification of the cross-dimer arginine/tyrosine residue in canonical and HyPr enzymes, respectively, suggests its role as a central player in communicating between the two substrate-binding sites. In one direction, the R to Y mutation on the canonical WspR abolishes activity, whereas the Y to R mutation in the HyPr GSU1658 ablates only cdiA synthase activity. In the former case, it appears that either substrate is unable to bind, or that substrate binding can no longer be relayed to the other active site, preventing either the second substrate-binding event or catalysis. In the latter case, the stark contrast in product ratios observed shows a clear preference for the enzyme to only allow productive cyclization to occur only when at least one GTP molecule is bound. It remains to be seen how other residues across the dimer interface in both HyPr and canonical GGDEF enzymes affect enzyme activity. For example, in at least one active canonical GGDEF enzyme background, GSU3350, we observe a cross-dimer leucine residue at the same position as the highly conserved arginine (Hallberg et al., 2016).

The structural basis for GGDEF enzyme mechanism has focused primarily on contacts between the individual monomeric unit and electrophilic activation of the alpha-phosphate for attack by the incipient hydroxyl group of the ribose sugar. Studies of dimeric GGDEF proteins have often focused on mechanisms of activation, with particular emphasis focused on the linker domain (De et al., 2009; Gourinchas et al., 2017). We demonstrate that the cross-dimer portion of the GGDEF domain that faces the opposite substrate-binding pocket has underappreciated roles both involved in substrate selection and catalysis. In addition to the previously-mentioned cross-dimer arginine/tyrosine residue, we posit that the second acidic residue in the GGDEF signature motif is the general base required for nucleophilic activation. Our results will permit the analysis of HyPr GGDEF signaling in bacteria, and point to additional key residues that may hide the key to finding other signaling molecules that could be produced by this ubiquitous enzyme class. We expect that, with additional structure-activity studies, the versatility of this domain will allow both the discovery and engineering of diverse dinucleotide cyclases capable of accepting more diverse substrates.

Materials and Methods:

General Reagents and Oligonucleotides
All oligonucleotides were purchased from Elim Biopharmaceuticals (Hayward, CA) or IDT (Coralville, IA). Genomic DNA from *G. metallireducens* was obtained from the laboratory of Daniel Bond at UM Twin Cities. The codon-optimized WspR gene was purchased as a gBlock from IDT (Table 1). Cyclic dinucleotide standards were purchased from Axxorra (Farmingdale, NY) or enzymatically synthesized. NTP stocks were purchased from New England Biolabs (Boston, MA).

Molecular Cloning
The T4-Lysozyme(E11Q) gene was inserted upstream of the Gmet_1914$^{294\text{-}459}$ sequence into pET24a.
For C-terminal 6x-His-tagged constructs, gene sequences were inserted between NdeI and XhoI restriction sites of pET24a or pET31b. Standard cloning techniques were used for all other constructs (See Table 3 for Primer list).

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 genes encoding dinucleotide cyclase enzymes in pET24a were inoculated into LB media and grown aerobically to an OD_{600} ~ 0.3. Cultures were then induced with 1 mM IPTG at 28 ºC for 4 h. Cells were harvested by centrifugation at 4,700 rpm for 15 min at 4 ºC, and pellets were stored at -80 ºC.

Cell Extraction of *E. coli*
Cyclic dinucleotides were extracted as described previously (Hallberg et al., 2016).

LC-MS Analysis of *E. coli* Cell Extracts
LC-MS analysis of E. coli cell extracts was performed as described previously (Hallberg et al., 2016).

Overexpression and Purification of Dinucleotide Cyclase Enzymes
Full-length proteins with N-terminal His_6-MBP tags encoded in a pET16-derived plasmid were overexpressed in *E. coli* BL21 (DE3) star cells harboring a pRARE2 human tRNA plasmid and were grown in LB/carb/chlor for 10 h after induction at OD_{600} ~ 0.7 with 1 mM IPTG. Cells were lysed by sonication in a lysis buffer containing 25 mM Tris-HCl (pH 8.2), 500 mM NaCl, 20 mM imidazole, and 5 mM beta-mercaptoethanol. Clarified lysate was bound to Ni-NTA agarose (QIAGEN), and resin was washed with lysis buffer prior to elution with lysis buffer supplemented with 500 mM imidazole.
Proteins were dialyzed overnight at 4 ºC against buffer containing 20 mM HEPES-KOH (pH 7.5), 250 mM KCl, 1 mM TCEP, and 5% (v/v) glycerol. Protein purified in this way was concentrated to ~5-10 mg/mL, flash frozen in liquid nitrogen, and stored at -80 ºC. Protein with C-terminal His-6x tags encoded in pET24a were overexpressed and purified similarly, with the cells grown in LB/kan/chlor.

The catalytic domain of Gmet_1914 (amino acids 294–459) with an N-terminal T4 lysozyme co-crystallization tag was further purified by size-exclusion chromatography on a Superdex 200 16/60 column in gel-filtration buffer (20 mM HEPES-KOH [pH 7.5], 250 mM KCl, 1 mM TCEP, and 5% (v/v) glycerol), and eluted protein was concentrated to 10 mg ml^{-1}. Purified protein was used immediately for crystallography or flash frozen in liquid nitrogen and stored at −80°C for biochemical experiments.

Crystallization and Structure Determination of T4Lys-Gmet_1914

Prior to crystallization, T4Lysozyme-Gmet_$1914^{294–459}$ protein was incubated for 10 min at room-temperature at a concentration of 6.5 mg ml^{-1} in the presence of 10 mM GTP and 10 mM $MgCl_2$. The T4Lysozyme-Gmet_$1914^{294–459}$–GTP complex was crystallized in a hanging-drop vapor diffusion format using the final optimized crystallization conditions of 30 mM HEPES-KOH [pH 7.5], 300 mM Na(OAc), and 26% PEG-4000. Crystals were grown in Easy-Xtal 15-well trays (QIAGEN) in 2 µl hanging drops with a 1:1 (protein:reservoir) ratio over 350 µl of reservoir solution. Crystals required incubation at 18°C for 2–4 days for complete growth, and were then transferred with a nylon loop to a new drop containing reservoir solution supplemented with 10% glycerol as a cryoprotectant and incubated for 30 s before flash-freezing in liquid nitrogen. Native and anomalous data were collected under cryogenic conditions at the Lawrence Berkeley National Laboratory Advanced Light Source (Beamline 8.3.1).

X-ray diffraction data were processed with XDS and AIMLESS (Kabsch, 2010 PMID 20124692) in the monoclinic spacegroup *C* 2. Phase information was determined with a combination of molecular replacement and sulfur single-wavelength anomalous dispersion (SAD). Briefly, iterative sulfur-SAD data sets were collected at ~7,235 eV and merged from independent portions of a large T4Lysozyme-Gmet_$1914^{294–459}$ crystal as previously described (Lee et al., 2016a). A minimal core of T4-Lysozyme (PDB 5JWS) (Lee et al., 2016b) was used a search model for molecular replacement and sub-structure determination. The placed T4-Lysozyme fragment was then used to guide SAD identification of 17 sites with HySS in PHENIX (Adams et al., 2010) corresponding to 12 sulfur atoms in T4Lysozyme-Gmet_$1914^{294–459}$ and 5 solvent ion positions. SOLVE/RESOLVE (Terwilliger, 1999) was used to extend phases to the native T4Lysozyme-Gmet_$1914^{294–459}$ data processed to ~1.35 Å and model building and refinement were completed with Coot (Emsley and Cowtan, 2004) and PHENIX.

In Vitro Activity Assay of Dinucleotide Cyclases using Pyrophosphatase Assay

In vitro activity assays were performed as previously described (Burns et al., 2014), with slight modifications. The EnzChek pyrophosphate kit (Life Technologies) was used according to the manufacturers instructions with the exception that the buffer was supplemented with KCl to a final concentration of 100 mM and $MgCl_2$ to a final concentration of 10 mM, and the reactions were initiated with addition of ATP or GTP.
Assays were performed in triplicate in 96 well plates containing 1 µM protein and varying NTP concentrations (0-10 mM). Plates were monitored using a SpectraMax® i3x plate reader (Molecular Devices) and SoftMax Pro 6.5.1 software (). Subsequent

analyses to determine enzymatic rates were performed using the Excel Solver package (Microsoft).

In Vitro Activity Assay of Dinucleotide Cyclases using LC-MS

Activity assays were performed as described previously (Burhenne and Kaever, 2013; Hallberg et al., 2016). LC-MS analysis of enzyme reactions was performed using an Agilent 1260 Quadrupole LC-MS with an Agilent 1260 Infinity liquid chromatograph equipped with a diode array detector. Sample volumes of 10 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 enzyme reactions, a elution program consisting of 0% B for 5 minutes, followed by linear elution of 0 to 10% B over 1.5 minutes, isocratic elution at 10% B for 2 minutes, linear elution of 10 to 30% B over 2.5 minutes, linear elution from 30 to 0% B over 10 seconds, and isocratic elution of 0% B for 4 minutes, 50 seconds was used. Solvent A was 10 mM ammonium acetate/0.1% acetic acid and solvent B was HPLC-grade methanol.
Under these conditions, the cyclic dinucleotides in extracts were found to always elute in the order of cdiG (10.23±0.02 min), cAG (10.56±0.02 min), and cdiA (11.09±0.05 min). Due to slight variability in retention times, the assignment of cyclic dinucleotide identity was made through analysis of the mass spectra. Additionally, we note that, while cdiG and cAG run as clearly defined peaks even at high concentrations, cdiA exhibits a fronted peak shape at and above 25 µM concentration. Samples were analyzed by MS in the positive ion mode using the range of m/z = 150 to 1000.

Bioinformatic Analysis of GGDEF Variants

A Python-based program was developed to extract alignment data for a library of 139,801 putative GGDEF domain-containing proteins from the Uniprot database (obtained through Pfam, 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

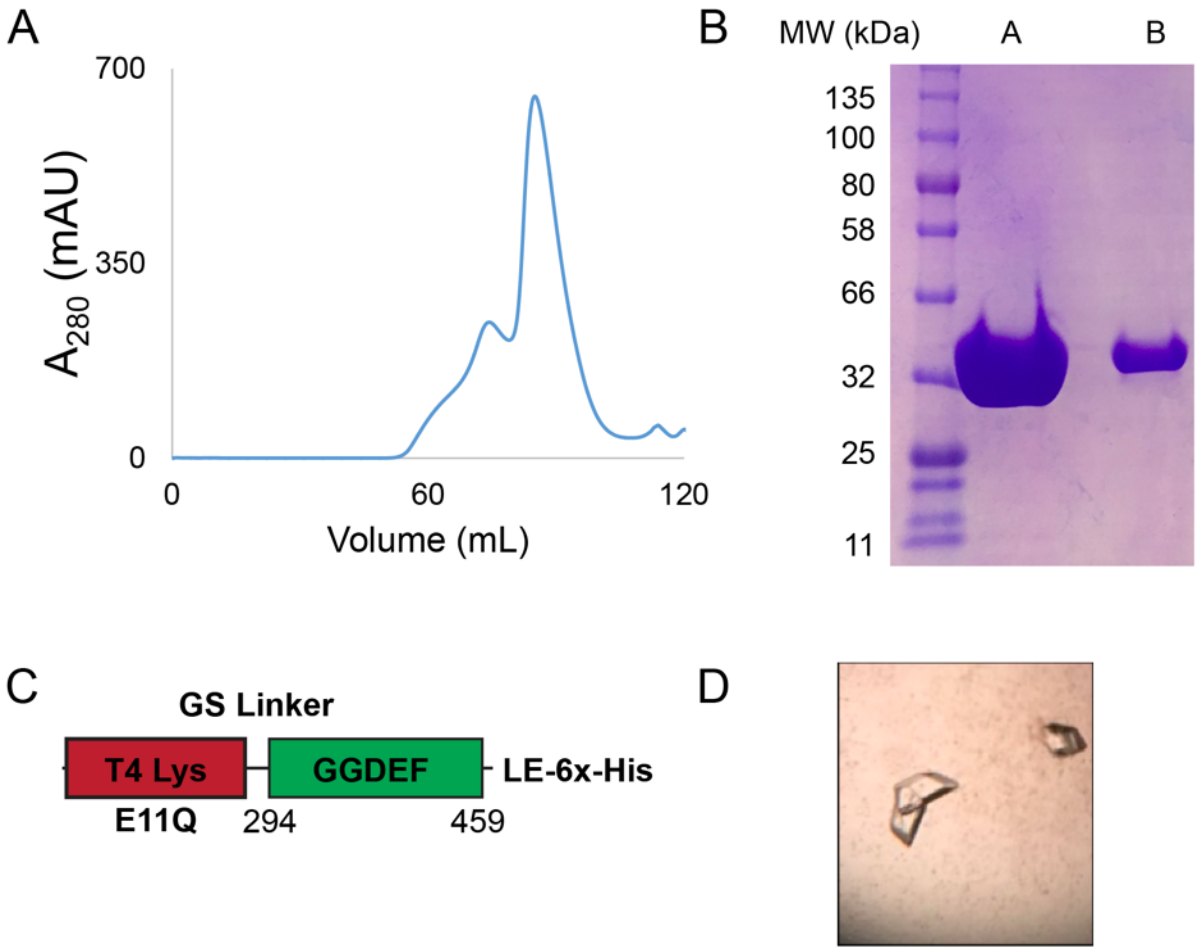


Figure 1 - Purification of T4 Lysozyme-Gmet_1914$_{GGDEF}$ Fusion

(A) Size-exclusion chromatography of T4-Lysozyme-tagged Gmet_1914$_{GGDEF}$. Shown is the A_{280} trace.

(B) SDS-PAGE gel analysis purified T4Lysozyme-Gmet$_{GGDEF}$. Gels were stained with Phastgel® Blue R. A - 60 µg protein sample (as determined by Bradford); B - 15 µg protein.

(C) Schematic of protein construct crystalized. Full sequence given in Table 3.

(D) Photo of T4-Lys-Gmet_1914$_{294-459}$ crystals.

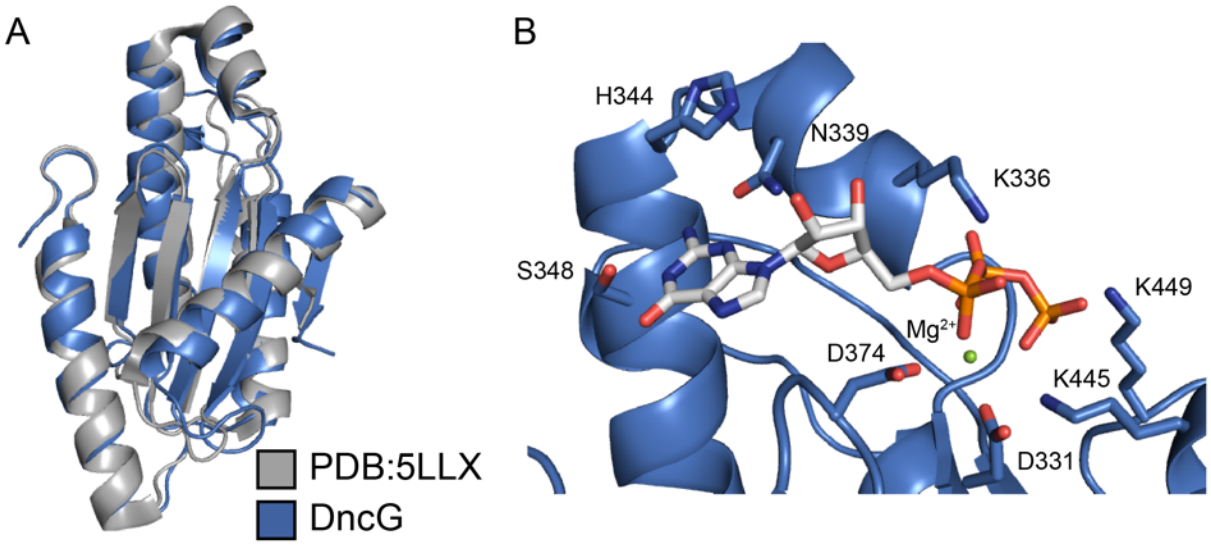


Figure 2 - HyPr GGDEF Structure and GTP binding pocket

(A) Overlay of Gmet_1914 GGDEF (blue) with the GGDEF domain of PleD (Grey, PDB: 2V0N).

(B) Active site of Gmet_1914 GGDEF in complex with GTP. Residues expected to interact with the GTP or Mg2+ cation are labeled. The alpha phosphate electron density was unable to be resolved, but was modeled into the structure.

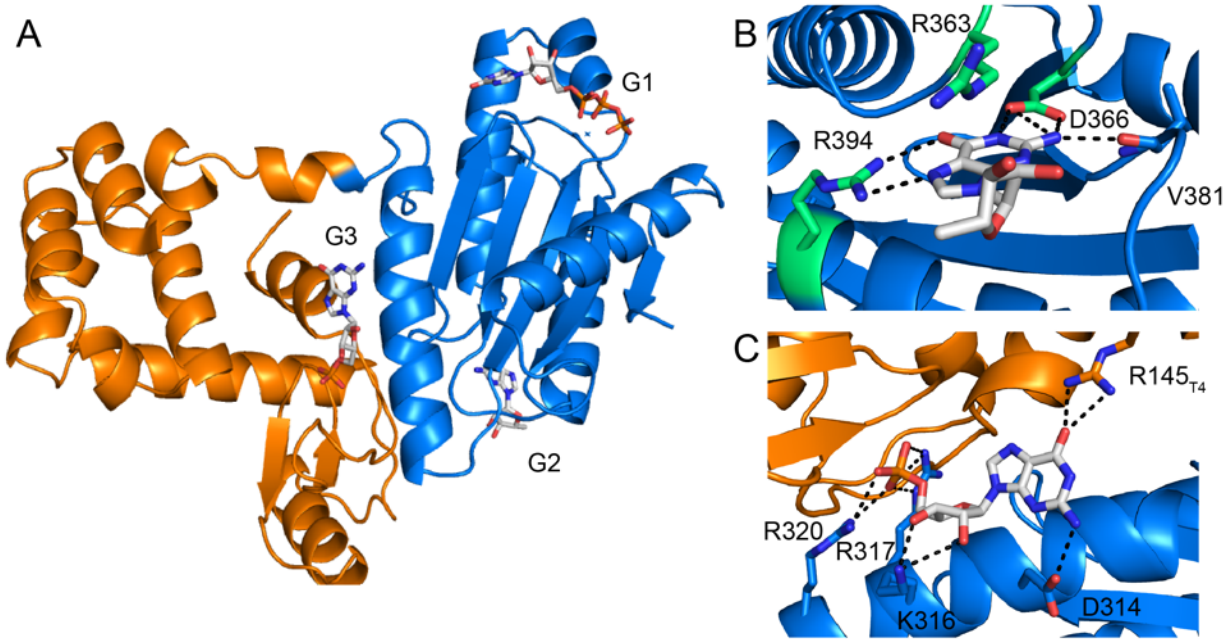


Figure 3 - Guanine Nucleotide Binding in HyPr GGDEF Crystal Structure.

(A) Full structure of the T4-Lysozyme-$Gmet_{GGDEF}$ fusion showing the location of three guanine nucleotides. The T4Lysozyme structure is in orange, with the $Gmet_{GGDEF}$ in blue.

(B) Interactions of second guanine nucleotide near the canonical I-site region of the HyPr GGDEF.

(C) Interactions of third guanine nucleotide at the interface between the T4-Lysozyme and GGDEF domains. The interactions between $R145_{T4}$ and $D314_{Gmet1914}$ are expected to occur only for GTP.

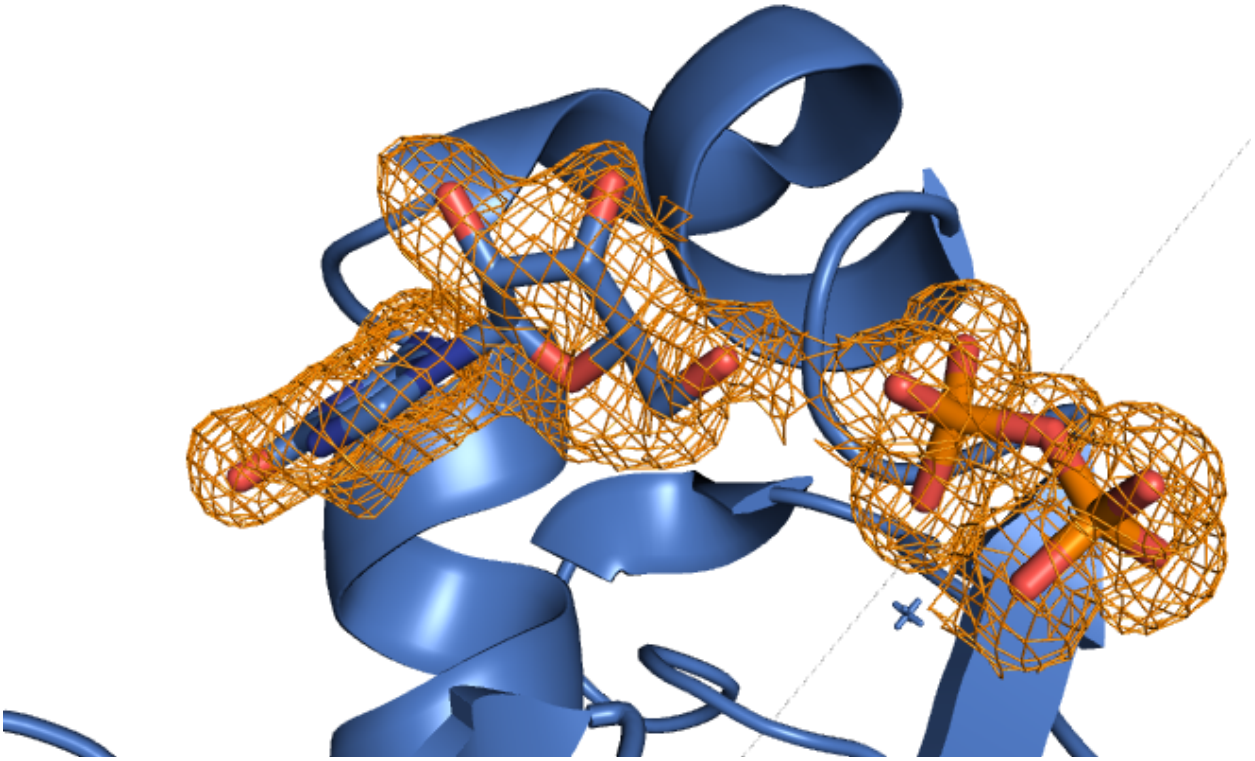


Figure 4 - GTP Binding Pocket Electron Density. An Fo-Fc omit map of electron density contoured at 2.0 σ is shown for the bound guanine nucleotide is colored in orange.

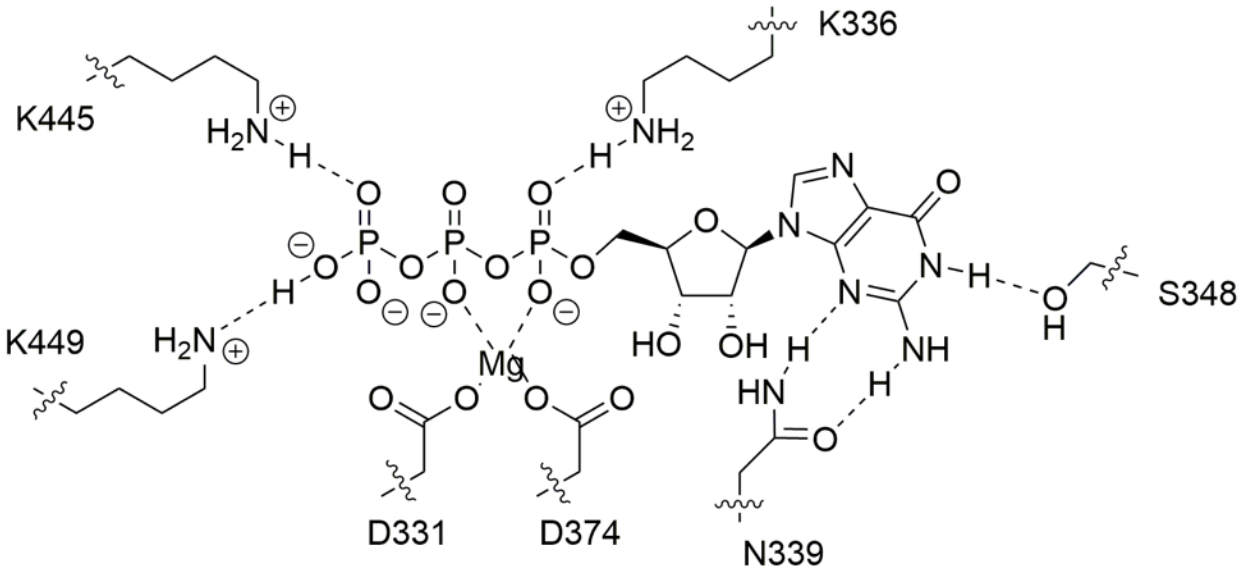


Figure 5- Diagram of GTP Substrate Contacts with Residues from HyPr GGDEF Protein Monomer.

Schematic of the chemical structure of the GTP ligand highlighting hydrogen-bonding interactions between the protein and nitrogen base (N339 and S348), phosphate backbone (K336, K445, K449), and Magnesium cation that interacts with the alpha and beta phosphates (D331 and D374).

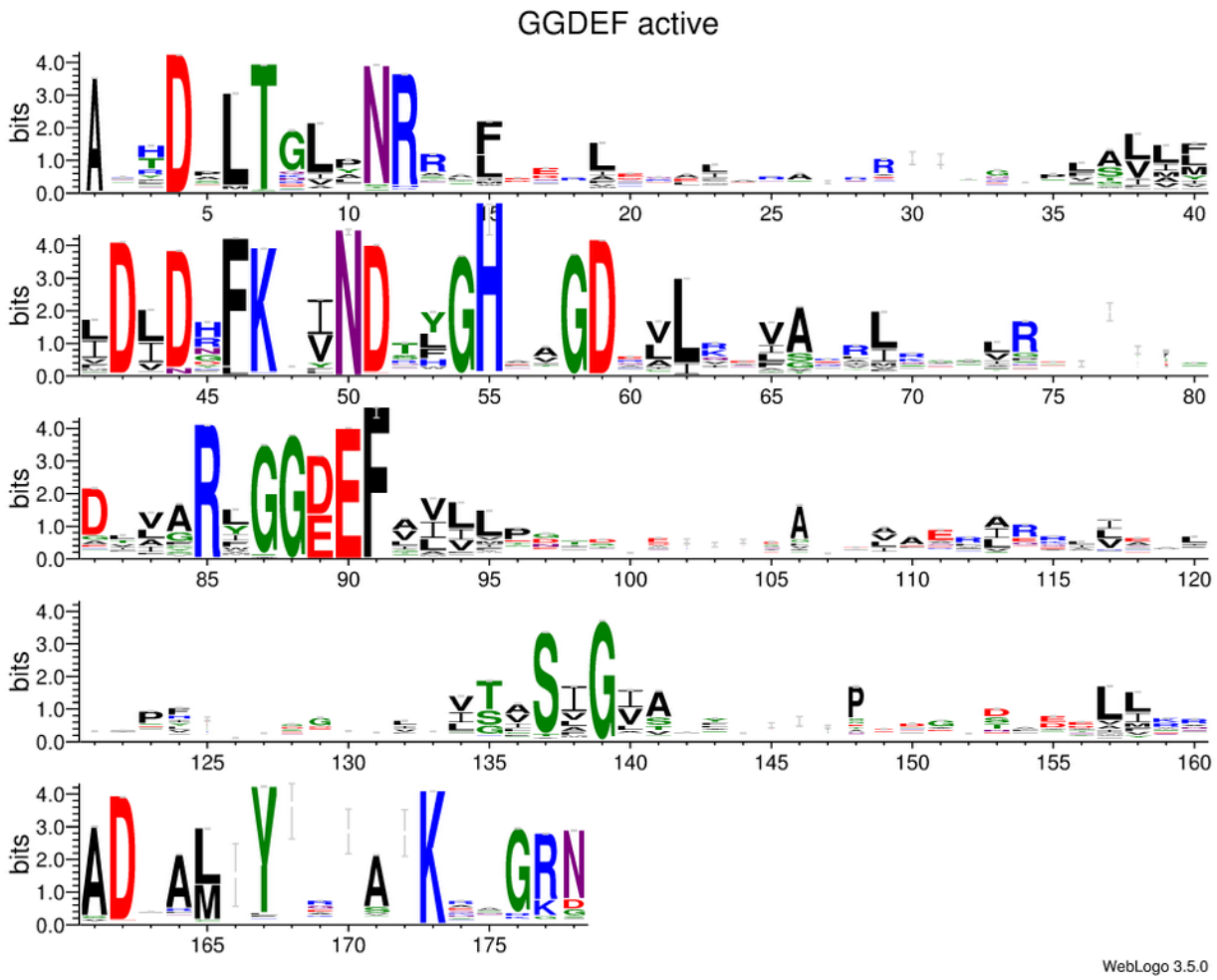


Figure 6 - "Active" GGDEF Consensus. Sequence logo generated from GGDEF domains that contain an intact "[G/A/S]G[D/E]EF" motif. Image generated with WebLogo (Crooks et al., 2004).

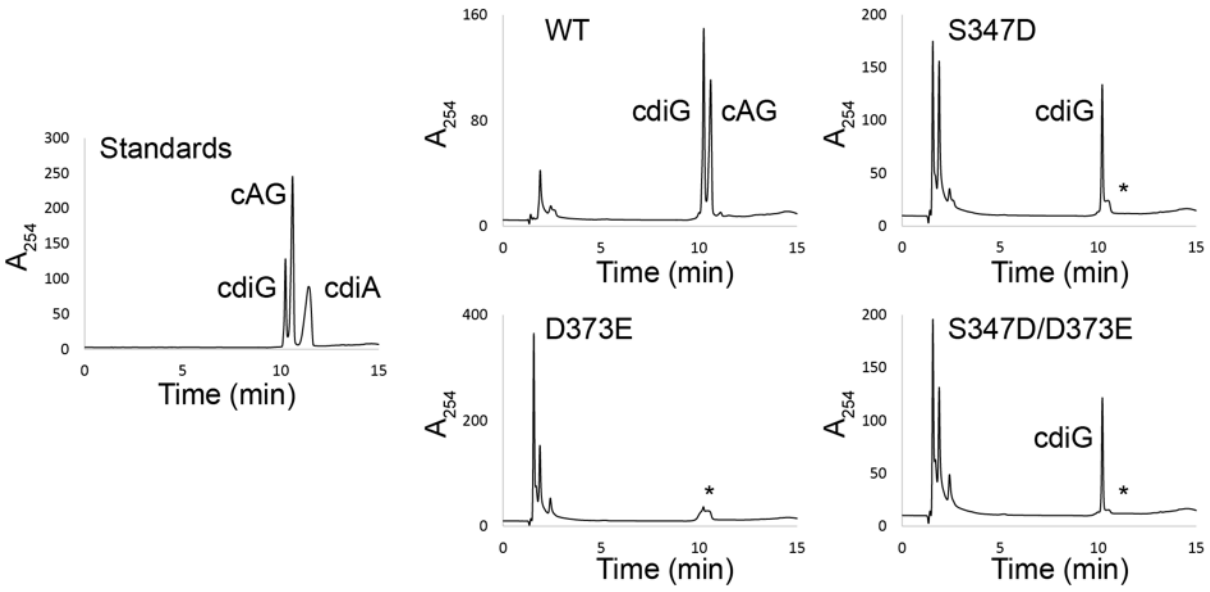


Figure 7 - LC-MS Endpoint Analysis of HyPr Mutant Products. HPLC-MS analysis of endpoint reactions of selected 6xHis-MBP-tagged, I-site knockout (R393A) mutants of GSU1658.

Under these conditions, the cyclic dinucleotides in extracts were found to always elute in the order of cdiG (10.23±0.02 min), cAG (10.56±0.02 min), and cdiA (11.09±0.05 min). * represents products that co-elute at times similar to CDNs but that do not have masses corresponding to CDN products.

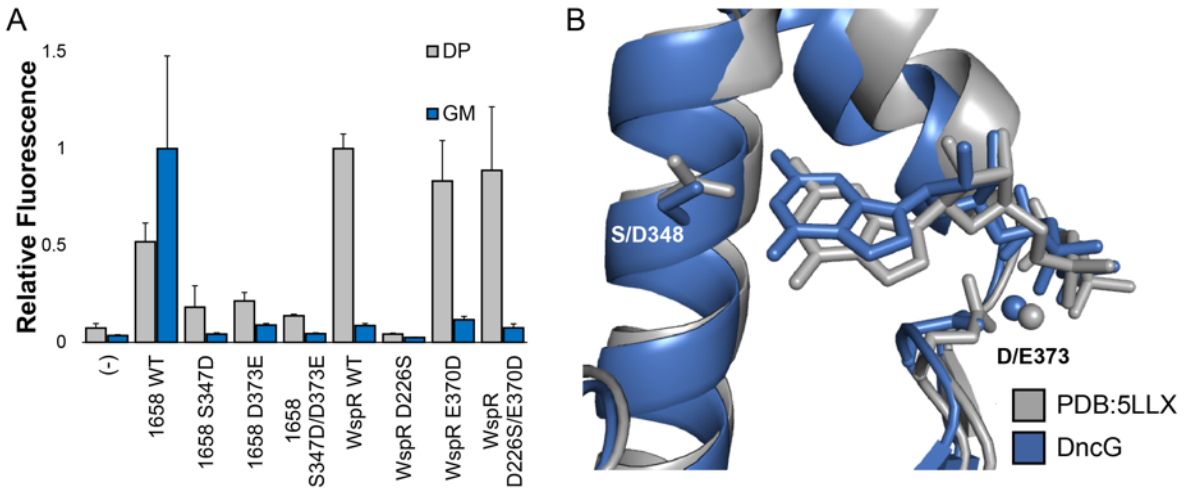


Figure 8 - Molecular Ruler in GGDEF Enzyme Activity

(A) Average fluorescence measured by flow cytometry (n=3, 10,000 cells per run) of E. coli BL21 (DE3) Star cells co-expressing the cdiG-selective biosensor Dp-Spinach2 (grey) or cAG-selective biosensor GM0970-p1-4delA-Spinach (blue) along with WT and mutant versions of GSU1658, codon-optimized WspR, or empty vector.

(B) Superposition of the crystal structures of the Gmet_1914 (blue) and the light-regulated GGDEF from BLAH (grey) (PDB 5LLX) in complex with crystallized guanine nucleotides. Residues interacting with the Watson-Crick face (S/D348 by Gmet_1914 numbering) and the central residue of the “GGDEF” motif (D/E373) are shown as sticks.

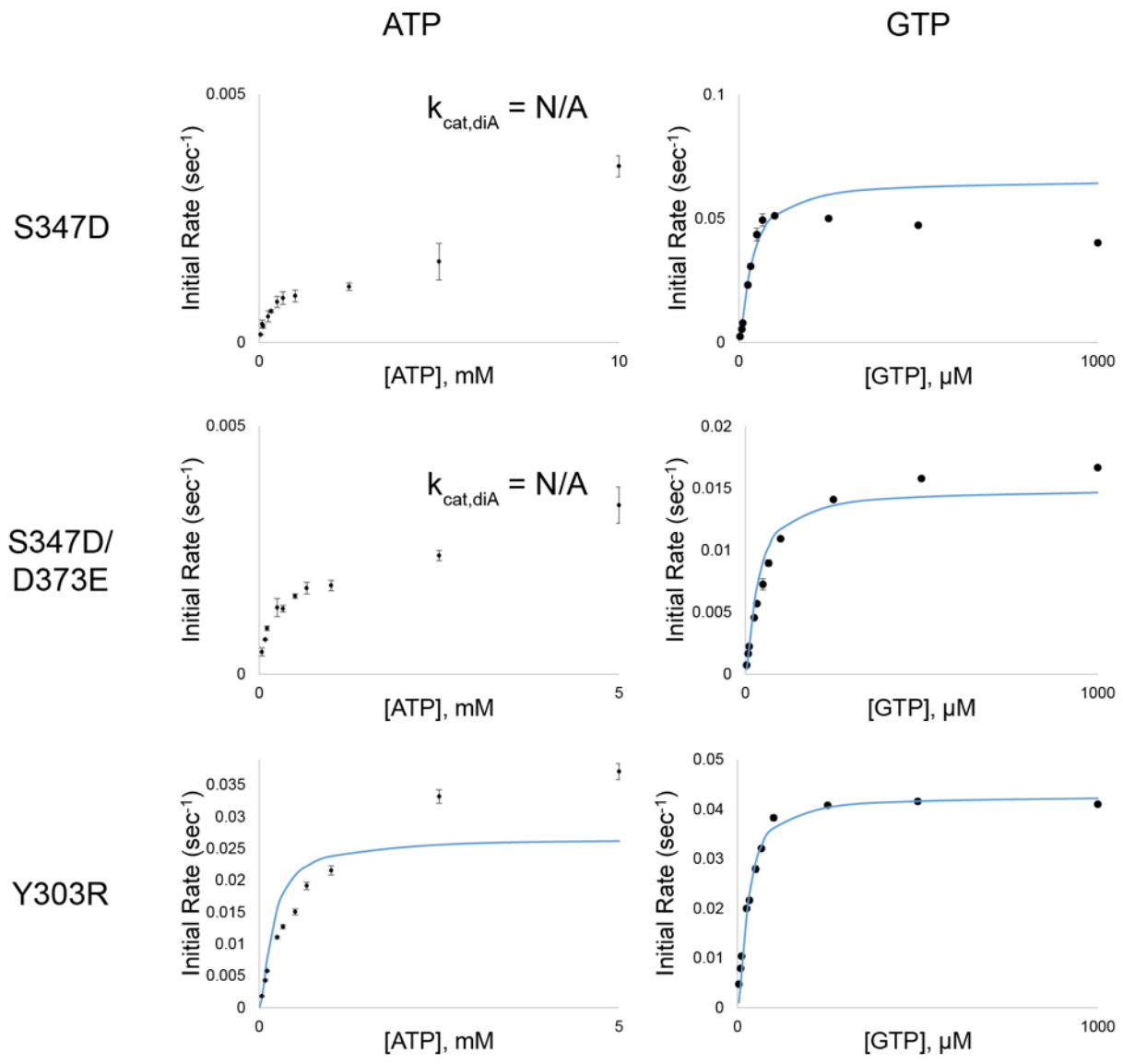


Figure 9 - Initial Rates of Product Formation with single NTP substrates for selected 6xHis-MBP-tagged GSU1658 mutants containing an R393A mutation in the allosteric inhibitory site. K_{cat} and K_D values for cooperative fits are given on Table 1.

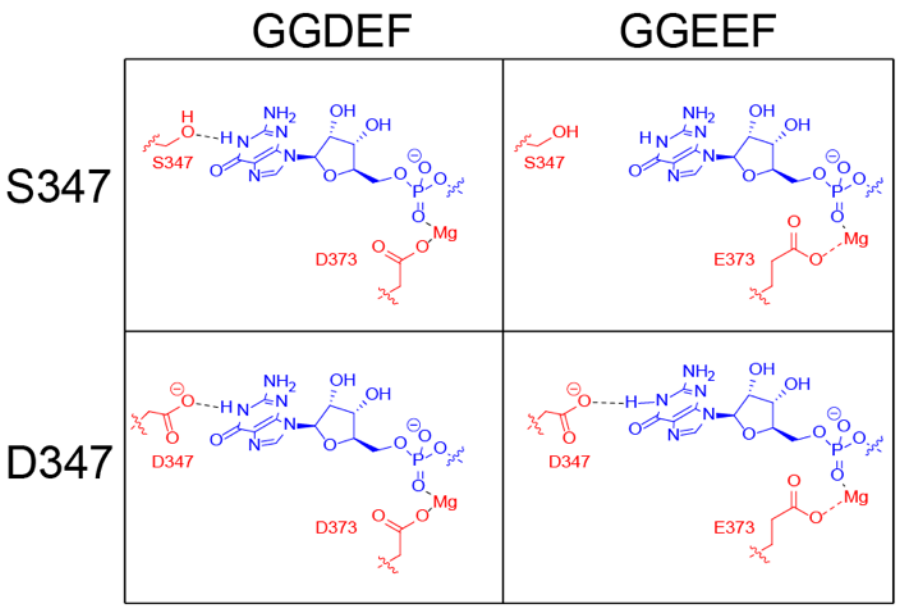


Figure 10 - HyPr GGDEF Active Site Remodeling.

Schematic of the chemical structure of the guanine nucleobase with various Watson-Crick (S/D347) and magnesium (D/E 373) interacting residues.

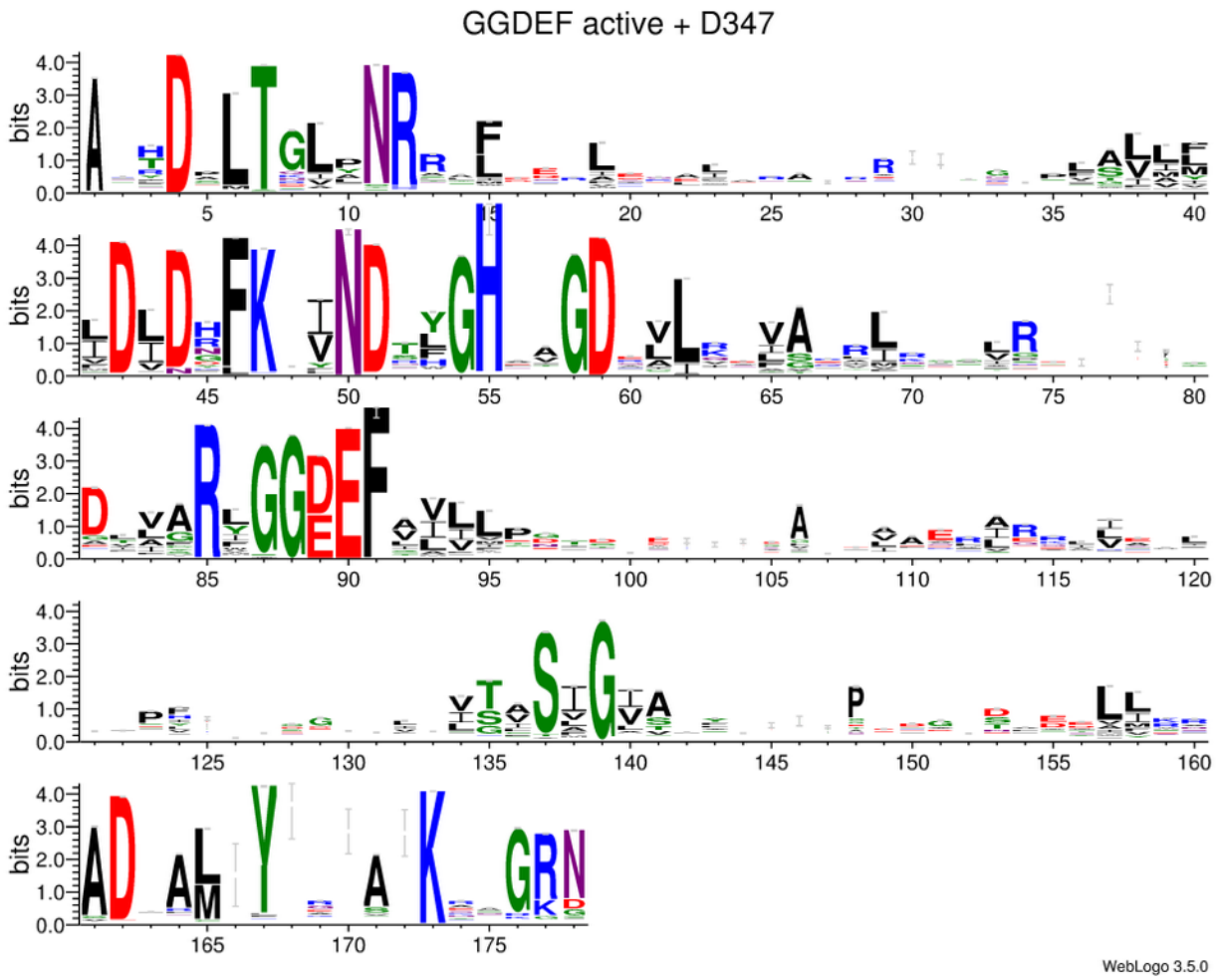


Figure 11- Canonical GGDEF Consensus. Sequence logo generated from GGDEF domains that contain an D347 residue (by GSU1658 numbering). Image generated with WebLogo (Crooks et al., 2004).

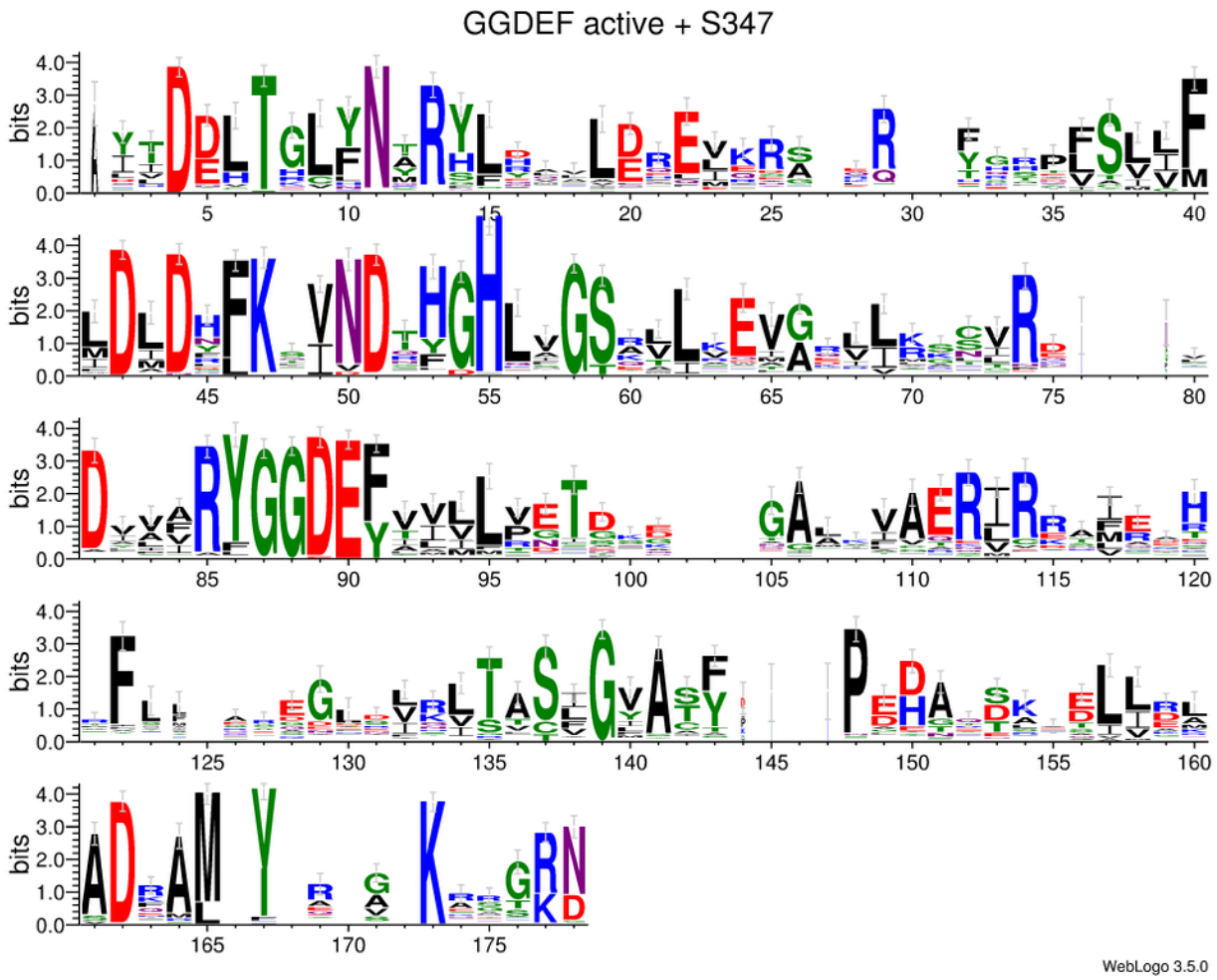


Figure 12 - HyPr GGDEF Consensus. Sequence logo generated from GGDEF domains that contain an S/T347 residue (by GSU1658 numbering). Image generated with WebLogo (Crooks et al., 2004).

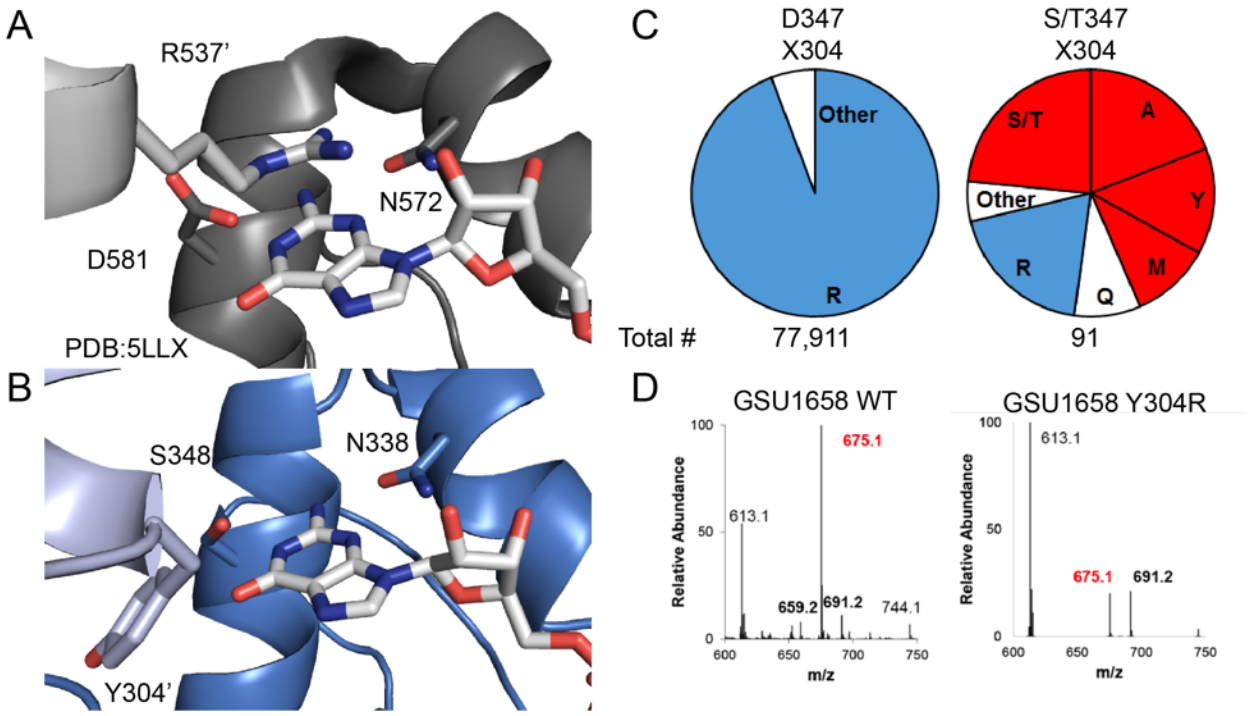


Figure 13 - Cross-Dimer Substrate Interactions in GGDEF Proteins

(A) and (B) Crystal structure of the light-regulated GGDEF dimer from *Idiomarina* sp. A28L (grey, A, PDB 5LLX) and the superimposed Gmet_1914 GGDEF domain (blue, B) in complex with crystallized guanine nucleotides. Residues interacting with the nucleobase (S/D348, N338 and the cross-dimer Y/R304' by Gmet_1914 numbering) are shown as sticks.

(C) Distribution of cross-dimer interacting residue given either canonical (D347) or HyPr (S/T347) residues in predicted active GGDEF proteins.

(D) HPLC-MS analysis of lysate from cells expressing the GSU1658 wild-type and Y303R mutant. The MS spectra shown integrates the retention time region containing all three cyclic dinucleotides (6 to 8 min). Expected masses are for cdiG (m/z = 691), cAG (m/z = 675), and cdiA (m/z = 659).

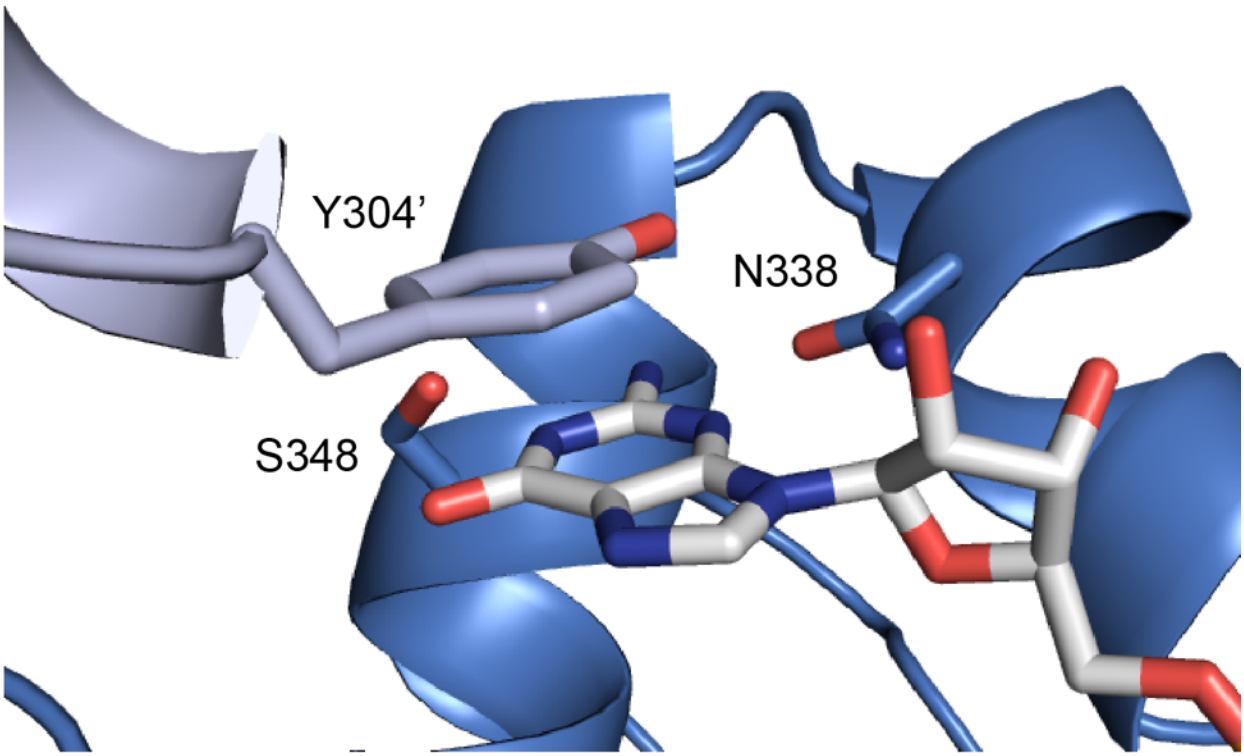


Figure 14 - Proposed Cross-Dimer Substrate Interaction in HyPr GGDEF Proteins.

As in figure 13A, with a 170˚ rotation about the C-Cb bond of Y304'

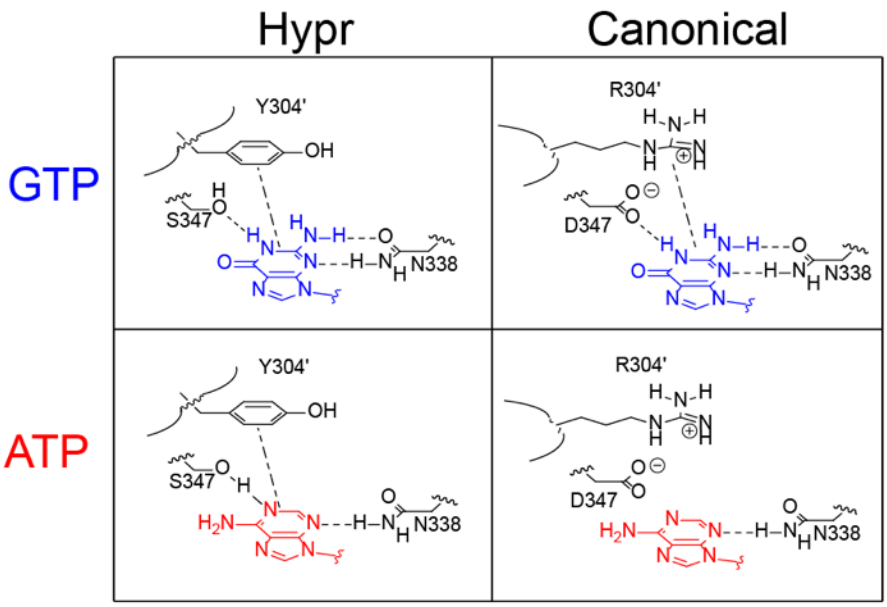


Figure 15 - Cross-dimer Substrate Contacts.

Schematic of the chemical structure of the nucleobases for adenine and guanine in complex with either a HyPr or canonical GGDEF, highlighting confirmed (S/D347 and N338 as numbered in GSU1658) monomer-ligand contacts, as well as a potential (Y/R303') cross-dimer interaction.

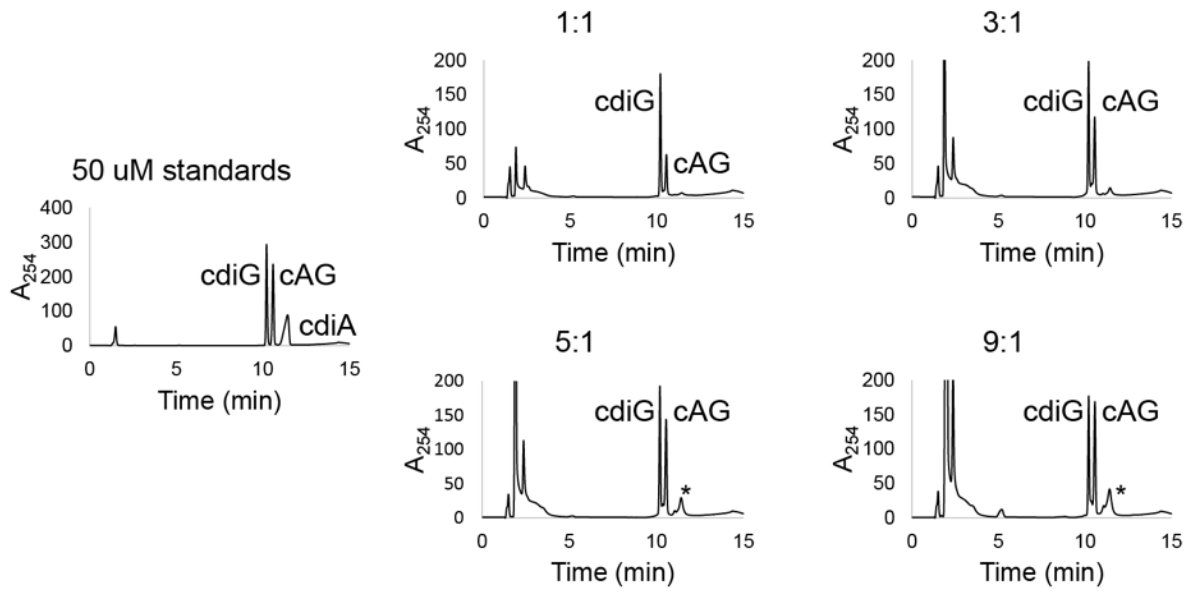


Figure 16 - LC-MS Based Product Ratio Analysis of MBP-tagged GSU1658 R393A/ Y303R. HPLC-MS analysis of endpoint reactions of a 6xHis-MBP-tagged, I-site knockout (R393A), cross-dimer (Y303R) mutant of GSU1658 (1 μM) under varying ATP:GTP ratios, with [GTP] kept constant at 100 μM. Ratios of ATP:GTP are shown above each trace.

Under these conditions, the cyclic dinucleotides in extracts were found to always elute in the order of cdiG (10.23±0.02 min), cAG (10.56±0.02 min), and cdiA (11.09±0.05 min). * represents products that co-elute at times similar to CDNs but that do not have masses corresponding to CDN products.

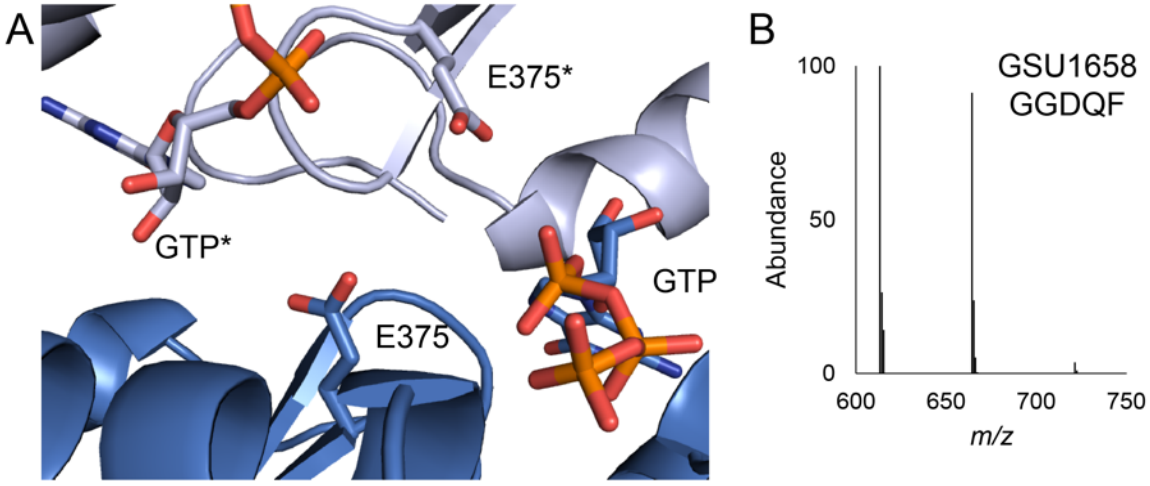


Figure 17 - Cross-Dimer Catalytic Base in GGDEF Proteins

(A) Crystal structure of the Gmet_1914 GGDEF domain (blue, A) superimposed with the the light-regulated GGDEF dimer from *Idiomarina* sp. A28L (grey, B) (PDB 5LLX) in complex with crystallized guanine nucleotides. Potential cross-dimer catalytic residue (E375) and guanine nucleotides are shown as sticks.

(B) HPLC-MS analysis of lysate from cells expressing the GSU1658 E374Q mutant. The MS spectra shown integrates the retention time region containing all three cyclic dinucleotides (6 to 8 min). Expected masses are for cdiG (m/z = 691), cAG (m/z = 675), and cdiA (m/z = 659).

Table 1. Kinetic Parameters of Select GSU1658 R393A Mutants. Kinetic Parameters for D373E could not be determined.

	S347D	S347D/D373E
K_{1G}, μM	38	38
K_{2G}, μM	12	12
$k_{cat,cdiG}$, sec^{-1}	0.07	0.015

Table 2. Codon-optimized WspR WT Sequence (5'→3')

ATGCATAATCCGCATGAATCAAAGACGGACCTGGGAGCTCCACTTGACGGAGCCGT
GATGGTTTTATTAGTGGACGACCAGGCGATGATCGGTGAGGCGGTCCGCCGTTCTC
TGGCTTCTGAAGCGGGCATCGACTTCCATTTTTGCTCCGATCCGCAGCAAGCGGTA
GCGGTAGCCAATCAAATTAAGCCCACGGTTATCCTGCAGGATCTGGTCATGCCTGG
CGTGGATGGGCTGACATTGTTAGCAGCTTATCGCGGAAACCCTGCAACACGCGACA
TTCCGATCATTGTGCTGAGTACCAAGGAGGAACCCACTGTTAAGTCAGCTGCATTT
GCAGCCGGGGCGAATGACTACCTGGTCAAACTTCCAGATGCGATCGAATTAGTTGC
TCGCATCCGCTACCACAGTCGCAGCTACATCGCGCTTCAGCAACGCGATGAAGCCT
ACCGCGCCTTGCGCGAATCCCAGCAGCAGCTTCTTGAAACGAACCTGGTTTTGCA
GCGTCTGATGAACTCCGACGGTTTAACGGGTTTGTCTAATCGCCGTCATTTTGATGA
ATACTTAGAGATGGAATGGCGTCGTAGTTTGCGTGAACAATCTCAGTTGTCATTACTT
ATGATCGACGTCGACTACTTTAAATCGTACAACGATACCTTCGGCCATGTAGCGGGT
GACGAAGCATTACGTCAAGTCGCTGGCGCGATCCGTGAAGGGTGCTCCCGTTCTT
CTGACCTTGCGGCTCGCTATGGTGGAGAGGAGTTTGCAATGGTTCTGCCTGGGAC
ATCACCGGGGGGCGCTCGCCTGTTGGCTGAGAAAGTGCGTCGCACGGTGGAAAG
TTTGCAGATCTCGCATGATCAACCGCGTCCAGGCTCGCATTTAACGGTGTCGATCG
GCGTATCCACCTTGGTTCCTGGAGGTGGAGGCCAGACCTTTCGCGTTTTGATCGAA
ATGGCTGACCAGGCATTATACCAGGCCAAAAATAATGGACGTAATCAGGTGGGATTG
ATGGAACAACCAGTACCTCCGGCACCTGCTGGA

Table 3. Protein Sequence of T4Lysozyme(E11Q)-Gmet1914$_{GGDEF}$ Fusion (N→C).
T4 Lysozyme (E11Q) is given in red, Gmet_1914$_{GGDEF}$ sequence is in green.

MNIFEMLRIDQGLRLKIYKDTEGYYTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITK
DEAEKLFNQDVDAAVRGILRNAKLKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSL
RMLQQKRWDEAAVNLAKSRWYNQTPNRAKRVITTFRTGTWDAYGSYIDELTGLFNYR
YLDISLDREIKRADRFGSTVSMIFIDLDFFKGVNDTHGHLVGSQVLNEMGMLLKKSVRE
VDIVIRYGGDEFTVMLVETGEKGAATVAERIRRSIEGHTFLAAEGFNIRLTASLGYACYPA
DTQSKLELLELADKAMYQGKEQGKNCVFRATAIRLEHHHHHH

Table 4. List of Primer Sequences. Restriction sites are denoted by an underline.

#	Nucleotide Sequence (5'→3')	Purpose
1	GAGA AGATCT TATATCGATGAACTGACCGG	For primer for Gmet_1914 GGDEF into pET16-MBP-T4lys
2	TTGTTAGCAGCgGccgCACTCGAGCTA ACGAATTGCAGTTGCCCTG	Rev primer for Gmet_1914 GGDEF into pET16-MBP-T4lys
3	GAGA CATATG AATATATTTGAAATGTTACGTATAGATCAAGG	For primer for T4Lys-Gmet_1914 GGDEF into pET24a
4	gtggtggtggtggtggtgctcgagACGAATTGCAGTTGCCCTG	Rev primer for T4Lys-Gmet_1914 GGDEF into pET24a
5	GAGACAT ATGCATAATCCGCATGAATCA	For primer for Codon-optimized WspR into pCOLA
6	GAGACTCGAG TCCAGCAGGTGCCGGA	Rev primer for Codon-optimized WspR into pCOLA
7	GAG GAGTTCACCATAATTCTGGTGGAA	Forward primer for round-the-horn mutagenesis of GSU1658 D373E
8	GCCACCGTAGCGAATGATTA	Reverse primer for round-the-horn mutagenesis of GSU1658 D373E
9	CGT CGTTACCTTGACATCTCGCTG	Forward primer for round-the-horn mutagenesis of GSU1658 Y304R
10	GTTGAAGAGTCCCGTGAGTTC	Reverse primer for round-the-horn mutagenesis of GSU1658 Y304R
11	AGT GAAGCATTACGTCAAGTCG	Forward primer for round-the-horn mutagenesis of Codon-optimized WspR D226S
12	ACCCGCTACATGGCCGAAG	Reverse primer for round-the-horn mutagenesis of Codon-optimized WspR D226S
13	GAT GAGTTTGCAATGGTTCTGC	Forward primer for round-the-horn mutagenesis of Codon-optimized WspR E253D
14	TCCACCATAGCGAGCC	Reverse primer for round-the-horn mutagenesis of Codon-optimized WspR E253D

Bibliography

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Burhenne, H., and Kaever, V. (2013). Quantification of cyclic dinucleotides by reversed-phase LC-MS/MS. Methods Mol Biol *1016*, 27-37.

Burns, J.L., Deer, D.D., and Weinert, E.E. (2014). Oligomeric state affects oxygen dissociation and diguanylate cyclase activity of globin coupled sensors. Mol Biosyst *10*, 2823-2826.

Burroughs, A.M., Zhang, D., Schaffer, D.E., Iyer, L.M., and Aravind, L. (2015). Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic Acids Res *43*, 10633-10654.

Chan, C., Paul, R., Samoray, D., Amiot, N.C., Giese, B., Jenal, U., and Schirmer, T. (2004). Structural basis of activity and allosteric control of diguanylate cyclase. Proc Natl Acad Sci U S A *101*, 17084-17089.

Chen, Y., Liu, S., Liu, C., Huang, Y., Chi, K., Su, T., Zhu, D., Peng, J., Xia, Z., He, J.*, et al.* (2016). Dcsbis (PA2771) from Pseudomonas aeruginosa is a highly active diguanylate cyclase with unique activity regulation. Sci Rep *6*, 29499.

Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., Meuwly, M., and Jenal, U. (2006). Allosteric control of cyclic di-GMP signaling. J Biol Chem *281*, 32015-32024.

Dahlstrom, K.M., Giglio, K.M., Collins, A.J., Sondermann, H., and O'Toole, G.A. (2015). Contribution of Physical Interactions to Signaling Specificity between a Diguanylate Cyclase and Its Effector. MBio *6*, e01978-01915.

De, N., Navarro, M.V., Raghavan, R.V., and Sondermann, H. (2009). Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. J Mol Biol *393*, 619-633.

Deepthi, A., Liew, C.W., Liang, Z.X., Swaminathan, K., and Lescar, J. (2014). Structure of a diguanylate cyclase from Thermotoga maritima: insights into activation, feedback inhibition and thermostability. PLoS One *9*, e110912.

Donovan, J., Dufner, M., and Korennykh, A. (2013). Structural basis for cytosolic double-stranded RNA surveillance by human oligoadenylate synthetase 1. Proc Natl Acad Sci U S A *110*, 1652-1657.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B.L., Zillinger, T., Serganov, A.A., Liu, Y., Jones, R.A., Hartmann, G.*, 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*, 1094-1107.

Gourinchas, G., Etzl, S., Gobl, C., Vide, U., Madl, T., and Winkler, A. (2017). Long-range allosteric signaling in red light-regulated diguanylyl cyclases. Sci Adv *3*, e1602498.

Hallberg, Z.F., Wang, X.C., Wright, T.A., Nan, B., Ad, O., Yeo, J., and Hammond, M.C. (2016). Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3', 3'-cGAMP). Proc Natl Acad Sci U S A *113*, 1790-1795.

Kazlauskiene, M., Kostiuk, G., Venclovas, C., Tamulaitis, G., and Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. Science *357*, 605-609.

Khersonsky, O., and Tawfik, D.S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. Annu Rev Biochem *79*, 471-505.

Kranzusch, P.J., Lee, A.S., Berger, J.M., and Doudna, J.A. (2013). Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. Cell Rep *3*, 1362-1368.

Kranzusch, P.J., Lee, A.S.Y., Wilson, S.C., Solovykh, M.S., Vance, R.E., Berger, J.M., and Doudna, J.A. (2014). Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. Cell *158*, 1011-1021.

Lee, A.S., Kranzusch, P.J., Doudna, J.A., and Cate, J.H. (2016a). eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. Nature *536*, 96-99.

Lee, H., Fischer, M., Shoichet, B.K., and Liu, S.Y. (2016b). Hydrogen Bonding of 1,2-Azaborines in the Binding Cavity of T4 Lysozyme Mutants: Structures and Thermodynamics. J Am Chem Soc *138*, 12021-12024.

Niewoehner, O., Garcia-Doval, C., Rostol, J.T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L.A., and Jinek, M. (2017). Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. Nature *548*, 543-548.

Oliveira, M.C., Teixeira, R.D., Andrade, M.O., Pinheiro, G.M., Ramos, C.H., and Farah, C.S. (2015). Cooperative substrate binding by a diguanylate cyclase. J Mol Biol *427*, 415-432.

Osawa, T., Inanaga, H., and Numata, T. (2013). Crystal structure of the Cmr2-Cmr3 subcomplex in the CRISPR-Cas RNA silencing effector complex. J Mol Biol *425*, 3811-3823.

Ovchinnikov, S., Kinch, L., Park, H., Liao, Y., Pei, J., Kim, D.E., Kamisetty, H., Grishin, N.V., and Baker, D. (2015). Large-scale determination of previously unsolved protein structures using evolutionary information. Elife *4*, e09248.

Rao, F., See, R.Y., Zhang, D., Toh, D.C., Ji, Q., and Liang, Z.X. (2010). YybT is a signaling protein that contains a cyclic dinucleotide phosphodiesterase domain and a GGDEF domain with ATPase activity. J Biol Chem *285*, 473-482.

Sarenko, O., Klauck, G., Wilke, F.M., Pfiffer, V., Richter, A.M., Herbst, S., Kaever, V., and Hengge, R. (2017). More than Enzymes That Make or Break Cyclic Di-GMP-Local Signaling in the Interactome of GGDEF/EAL Domain Proteins of Escherichia coli. MBio *8*.

Sasakura, Y., Hirata, S., Sugiyama, S., Suzuki, S., Taguchi, S., Watanabe, M., Matsui, T., Sagami, I., and Shimizu, T. (2002). Characterization of a direct oxygen sensor heme protein from Escherichia coli. Effects of the heme redox states and mutations at the heme-binding site on catalysis and structure. J Biol Chem *277*, 23821-23827.

Schirmer, T. (2016). C-di-GMP Synthesis: Structural Aspects of Evolution, Catalysis and Regulation. J Mol Biol *428*, 3683-3701.

Siddiq, M.A., Hochberg, G.K., and Thornton, J.W. (2017). Evolution of protein specificity: insights from ancestral protein reconstruction. Curr Opin Struct Biol *47*, 113-122.

Sinha, S.C., and Sprang, S.R. (2006). Structures, mechanism, regulation and evolution of class III nucleotidyl cyclases. Rev Physiol Biochem Pharmacol *157*, 105-140.

Tarnawski, M., Barends, T.R., and Schlichting, I. (2015). Structural analysis of an oxygen-regulated diguanylate cyclase. Acta Crystallogr D Biol Crystallogr *71*, 2158-2177.

Terwilliger, T.C. (1999). Reciprocal-space solvent flattening. Acta Crystallogr D Biol Crystallogr *55*, 1863-1871.

Wang, X.C., Wilson, S.C., and Hammond, M.C. (2016). Next-generation RNA-based fluorescent biosensors enable anaerobic detection of cyclic di-GMP. Nucleic Acids Res *44*, e139.

Wassmann, P., Chan, C., Paul, R., Beck, A., Heerklotz, H., Jenal, U., and Schirmer, T. (2007). Structure of BeF3- -modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. Structure *15*, 915-927.

Yang, C.Y., Chin, K.H., Chuah, M.L., Liang, Z.X., Wang, A.H., and Chou, S.H. (2011). The structure and inhibition of a GGDEF diguanylate cyclase complexed with (c-di-GMP)(2) at the active site. Acta Crystallogr D Biol Crystallogr *67*, 997-1008.

Zahringer, F., Lacanna, E., Jenal, U., Schirmer, T., and Boehm, A. (2013). Structure and signaling mechanism of a zinc-sensory diguanylate cyclase. Structure *21*, 1149-1157.

Appendix Chapter

Engineering and in vivo Applications of Riboswitches

This chapter has been published in: Hallberg, Z.F., Su, Y., Kitto, R.Z., Hammond, M.C. *Annu. Rev. Biochem* 2017, *86*, 515.

Abstract

Riboswitches are common gene regulatory units mostly found in bacteria that are capable of altering gene expression in response to a small molecule. These structured RNA elements consist of two modular subunits: an aptamer domain that binds with high specificity and affinity to a target ligand and an expression platform that transduces ligand binding to a gene expression output. Significant progress has been made in engineering novel aptamer domains for new small molecule inducers of gene expression. Modified expression platforms have also been optimized to function when fused with both natural and synthetic aptamer domains. As this field expands, the use of these privileged scaffolds has permitted the development of tools such as RNA-based fluorescent biosensors. In this review, we summarize the methods that have been developed to engineer new riboswitches, and highlight applications of natural and synthetic riboswitches in enzyme and strain engineering, in controlling gene expression and cellular physiology, and in real-time imaging of cellular metabolites and signals.

I. INTRODUCTION

The centrality of RNA in gene regulation is not only due to its role as messenger RNA, but also the regulatory roles of many new classes of non-coding RNAs that have been discovered. Some non-coding RNA classes function in the context of ribonucleoprotein complexes, including microRNAs, which typically target eukaryotic transcripts for translational repression(Wilson and Doudna, 2013), and Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) RNAs, which target phage DNA for cleavage as a bacterial defense mechanism (Wright et al., 2016). Other non-coding RNA classes function independently, without any protein partner. Riboswitches are a remarkable example class, as they are cis-regulatory structured RNA elements capable of controlling expression of downstream genes by direct response to a small molecule ligand. Riboswitches are comprised of a ligand sensing domain, called an aptamer, and a regulatory domain, called the expression platform. Aptamer binding to the ligand compound stabilizes the aptamer structure and causes a conformational change or other activation mechanism in the expression platform domain, which mediates gene regulation.

In the past two decades, over twenty distinct riboswitch-ligand pairs have been discovered (Serganov and Nudler, 2013) that showcase the capability of these natural aptamers for selective, tight molecular recognition of diverse compounds in cells (Figure 1). In addition, the study of natural expression platforms has revealed several different mechanisms for ligand-dependent gene regulation by riboswitches in bacteria, plants, and fungi (Figure 2). These findings reinforce that Nature makes broad use of riboswitches for gene regulation (although mostly in bacteria and conspicuously absent in animals), which has fueled researchers' interest in engineering riboswitch-like systems for their own *in vivo* applications.

In this review, we will present advances in riboswitch engineering and case studies for *in vivo* applications. Riboswitch engineering includes the design of new aptamer domains to recognize xenobiotic compounds, the development of new expression platforms, and the adaptation of riboswitches to make fluorescent biosensors. *In vivo* applications include use of riboswitches for reporting on ligands, regulation of cellular function, control of metabolic flux, and real-time biosensing of ligands.

II. RIBOSWITCH ENGINEERING

II.A. Aptamer Domain Engineering

To achieve efficient and accurate gene regulation, aptamer domains in natural riboswitches must have biologically relevant specificity and binding affinity to respond to their ligands. For example, a representative of the *S*-adenosylhomocysteine (SAH) riboswitch class exhibits a dissociation constant (K_d) of 20 nanomolar for SAH and discriminates more than 1000-fold against *S*-adenosylmethionine (SAM), which differs by one small methyl group and is positively charged (Wang et al., 2008). From bioinformatics, structural and biochemical studies, researchers have demonstrated that ligand-contacting nucleotides in natural riboswitches tend to be highly conserved across the evolutionary tree, while less conserved positions still can play roles in tuning the

thermodynamics and/or kinetics of the binding event. Therefore, mutation of conserved nucleotides can result in changing ligand specificity, typically through altering the hydrogen-bonding pattern of ligand recognition. Here we present some case studies on natural riboswitch variants and corresponding engineering efforts to alter ligand specificity and sensitivity of riboswitch aptamers for gene regulation and biosensing applications.

II.A.1. Purine riboswitches

Purine riboswitches consist of structurally homologous subclasses that sense ligands including guanine, adenine and 2'-deoxyguanosine, and therefore represent a great model for studying the relationship between aptamer sequence and ligand specificity (Figure 1A). Within the purine riboswitch class, guanine riboswitches were first identified. Detailed characterization of the *xpt* guanine riboswitch in *Bacillus subtilis (B. subtilis)* revealed its tight binding affinity for guanine (K_d < 5 nM) and discrimination against many purine analogs, including adenine (Mandal et al., 2003). A database search for sequences with secondary structure and sequence similarity to the *xpt* aptamer motif revealed a broad distribution of guanine riboswitches across the bacterial kingdom. Intriguingly, among these putative guanine riboswitches, only three out of 32 sequences were found to harbor a uracil instead of a cytosine at position 74 (numbering based on the *xpt* riboswitch) (Mandal et al., 2003; Mandal and Breaker, 2004). These U74 variants turned out to preferentially bind adenine instead of guanine (Mandal and Breaker, 2004). Based on this observation, it was hypothesized that the guanine and adenine riboswitches most likely recognized their respective ligands via Watson-Crick base pair to nucleotide 74, and thus this nucleotide determined ligand selectivity (Mandal and Breaker, 2004). This base pairing interaction, along with other important nucleotide-ligand interactions, was later validated by x-ray crystal structures and an NMR study of the purine riboswitch-ligand complexes (Figure 1A) (Batey et al., 2004; Noeske et al., 2005; Serganov et al., 2004).

The single-base determinant of purine riboswitch selectivity offered an important precedent for rational engineering of riboswitch aptamers. However, these early studies also raised a cautionary note that riboswitch function is not just about engineering ligand specificity, but also binding affinity and regulatory function. For example, it was observed *in vitro* that the binding affinity of the C74U *xpt* riboswitch variant for adenine (K_D ~100 nM) is poorer than that of the wild-type riboswitch for guanine (K_D ≤ 5 nM) (Mandal and Breaker, 2004). Additionally, guanine or adenine riboswitches incorporating the single mutation to nucleotide 74 that switched selectivity did not demonstrate the expected gene regulation *in vivo* (Mandal and Breaker, 2004). These observations suggest that other positions, although not systematically different between guanine and adenine riboswitches, contribute to ligand binding affinities and *in vivo* function, and thus multiple mutations are required to engineer a useful riboswitch. This idea was supported by a systematic mutational analysis of nucleotides not directly contacting the ligand, but involved in loop-loop interactions and the three-way junction in the riboswitch structure. These nucleotides were found to still play essential roles in riboswitch function by promoting a binding-competent RNA fold or organizing a secondary structure switch with the expression platform (Gilbert et al., 2007; Tremblay et al., 2011).

Similar principles were again demonstrated when 2'-deoxyguanosine (2'-dG) riboswitches were discovered a few years later as a third subclass of purine riboswitches (Kim et al., 2007). Attempts to mutate the guanine-sensing *xpt* riboswitch to switch specificity toward 2'-deoxyguanosine showed that the pyrimidine at position 51 (numbering based on *xpt* rboswitch) was the principle determinant. However, the x-ray crystal structure showed that other positions, including base pairs proximal to the riboswitch binding pocket and nucleotides participating in loop-loop interactions, were also critical in organizing the tertiary structure to accommodate binding of 2'-dG (Edwards and Batey, 2009), which was later corroborated by another structural study on a wild-type 2'-dG riboswitch (Pikovskaya et al., 2011).

The insights gained about natural purine riboswitches have led to engineered counterparts as orthogonal gene regulatory tools. Structure-guided mutagenesis was performed on an adenine riboswitch by fully randomizing 2 residues (U47 and U51) around the ligand binding pocket to generate 15 candidate riboswitches (Dixon et al., 2010; Vincent et al., 2014). These sequences were subjected to a chemical genetics screen in *E. coli* involving nearly 80 heterocyclic analogues. The screen, along with further rational sequence optimization, resulted in two engineered riboswitches, called M6" and M6C", that are selective for ammeline (Amm) and azacytosine (Azc), respectively (Figure 1A) (Dixon et al., 2010). These unnatural riboswitch-ligand pairs exhibited desirable orthogonality to the parent adenine riboswitch *in vitro*. This orthogonality was further demonstrated in *E. coli* by using the M6 and adenine riboswitches to independently control the expression of two reporter genes with the use of Amm and 2-aminopurine (Dixon et al., 2012). To improve the affinity of the engineered riboswitch-ligand pairs, a broader range of heterocyclic compounds were screened and second-generation ligands (PPDA and PPAO, Figure 1A) with extended aromatic rings were developed for M6" and M6C" (Robinson et al., 2014). These improved riboswitch-ligand pairs enabled construction of both turn-on and turn-off riboswitches for flexible gene regulation and application to orthogonal riboswitch-based control of bacterial cell behavior and physiology (Robinson et al., 2014).

II.A.2. $PreQ_1$ riboswitch

7-aminomethyl-7-deazaguanine ($PreQ_1$) is a queuosine precursor that is important for proper tRNA function in bacteria (Figure 1B). The class I $PreQ_1$ riboswitch was first discovered and characterized in 2007 (Roth et al., 2007), followed by discovery of two other classes later (McCown et al., 2014; Meyer et al., 2008). These riboswitches preferentially bind to $PreQ_1$ and exhibit a slightly poorer affinity for its metabolic precursor 7-cyano-7-deazaguanine ($PreQ_0$) (McCown et al., 2014).

Similar to the purine riboswitch, the $PreQ_1$ riboswitch has been structurally and biochemically characterized (Kang et al., 2009; Klein et al., 2009), and structural analogues of the ligand are synthetically accessible and cell permeable, which makes the system promising for development of additional orthogonal riboswitch-ligand pairs. Structure-guided mutagenesis of nucleotides involved in hydrogen bonding to the ligand (Figure 1B) was performed on a class I $PreQ_1$ riboswitch to generate 7 candidate riboswitches (Robinson et al., 2016; Wu et al., 2015). After first ruling out candidates that responded to endogenous $PreQ_1$/$PreQ_0$, the remaining riboswitches were screened

against 6 synthetic $PreQ_0$ analogues to identify ones that displayed ligand-dependent gene repression in *B. subtilis*. The resulting artificial riboswitch, M1 (C17U), responded only to the diamino analogue of $PreQ_0$ (DPQ_0) and effectively repressed transcription of *mreB*, a gene for cell morphology in *B. subtilis,* upon addition of 2 mM DPQ_0 (Figure 1B).

II.A.3. Flavin mononucleotide (FMN) riboswitches

Roseoflavin is a rare but interesting example of a natural product whose antibacterial activity involves targeting riboswitches, and studying the mechanism of natural orthogonality in this system has led to some lessons for riboswitch aptamer engineering (Figure 1C). FMN riboswitches are found across a broad spectrum of bacteria and regulate a number of essential genes involved in biosynthesis and transport of riboflavin, the non-phosphorylated precursor of FMN. Similarly, roseoflavin is phosphorylated *in vivo* to make roseoflavin mononucleotide (RoFMN), which competes with FMN for riboswitch binding and downregulates FMN biosynthesis, resulting in growth defects and bacterial cell death (Lee et al., 2009; Mansjö and Johansson, 2011; Ott et al., 2009).

Intriguingly, the bacterium that produces roseoflavin, *Streptomyces davawensis (Sd)*, has an FMN riboswitch located upstream of the *ribB* gene, but evolutionary pressure on this producer organism apparently led to mutations that conferred response to FMN but insensitivity to RoFMN *in vivo* (Pedrolli et al., 2012). However, *in vitro* characterization of this FMN riboswitch revealed that its aptamer domain does not selectively discriminate against roseoflavin under equilibrium binding conditions (Lee et al., 2009), and that the complete riboswitch (aptamer plus expression domain) is unable to bind either FMN or RoFMN (Pedrolli et al., 2012). The likely explanation for these observations is that the *ribB* riboswitch is kinetically rather than thermodynamically controlled, as was shown for other FMN riboswitches (Wickiser et al., 2005). In co-transcriptional regulation, kinetically driven riboswitches do not reach thermodynamic equilibrium with the ligand before RNA polymerase continues or terminates transcription. In other words, the requirement is for ligand to bind fast enough to trigger downstream gene regulation, so the key parameter is the rate constant k_{on}, rather than the equilibrium K_d. Therefore, based on the observations, it was proposed that a key nucleotide A61 within the *ribB* riboswitch alters the ligand specificity by slowing down RoFMN binding (Pedrolli et al., 2012).

Kinetic analysis of FMN binding to several natural FMN riboswitch variants revealed that sequence variation in the stem regions, which are only conserved in length and pairing pattern but not sequence composition, resulted in different dissociation rates and affinities (up to 100-fold difference between 8 variants). Based on this observation for natural variants, a wild-type FMN riboswitch was engineered by systematically replacing A-U / G-U pairs with more stable G-C pairs in the stem regions, leading to 10-fold variance in both binding affinity and dissociation rates (Rode et al., 2015). Thus, this case study demonstrates the lesson that instead of focusing on conserved nucleotides for engineering ligand specificity, systematic analysis of stem regions separated from the ligand binding pocket can also tune the binding kinetics and affinity of a riboswitch. Another important lesson from FMN riboswitches is that kinetically driven riboswitches

complicate engineering efforts.

II.A.4. Cyclic di-nucleotide riboswitches

We have already reviewed several cases of engineering inspired by Nature. Conversely, this next case study shows how riboswitch aptamer engineering efforts led to the discovery of novel natural riboswitches. Cyclic di-GMP (cdiG) riboswitches were first identified by bioinformatics analysis as a highly conserved RNA structured element called the GEMM motif (Genes for Environment, Membranes, and Motility) (Weinberg et al., 2007). These GEMM motifs were typically located upstream of genes responsible for synthesis and degradation of cdiG or genes controlled by cdiG, leading to the hypothesis that they are cdiG riboswitches. This was later confirmed by *in vitro* chemical probing showing binding of the Vc2 GEMM-I riboswitch to cdiG with high selectivity and affinity, as well as *in vivo* reporter assays (Sudarsan et al., 2008).

For the GEMM-I riboswitch, structural analysis revealed an asymmetrical mode of recognition for cdiG, as the two guanine bases in the ligand are recognized by a guanosine at position 20 via a Hoogsteen interaction and a cytosine at position 92 via a Watson-Crick interaction, respectively (numbering based on Vc2 riboswitch) (Figure 1D) (Kulshina et al., 2009; Smith et al., 2009). Structure-based mutation of these residues, G20A and/or C92U, generated artificial variants that preferentially bound to cyclic di-AMP (cdiA) or cyclic AMP-GMP (cAG), as expected from the nucleobase interactions (Figure 1D) (Shanahan et al., 2011; Smith et al., 2009). This was an intriguing finding, because around the same time, these other cyclic dinucleotides were in fact discovered as second messengers in bacteria.

Whereas the prior study showed that only C92U conferred binding to cAG, we showed in the context of a riboswitch-Spinach biosensor fusion that the C92U variant did not bind cAG, while the G20A variant was promiscuous for both cdiG and cAG (Kellenberger et al., 2013). This result that A20 could support cAG binding was exciting, because a prior bioinformatics analysis had revealed that while the majority of GEMM-I riboswitch sequences had G20 like Vc2, 23% instead harbored A20 (Smith et al., 2010). By applying our ability to engineer riboswitch-Spinach biosensor fusions, we performed a phylogenetic screen for natural A20 GEMM-I riboswitches with altered ligand specificity, leading to the discovery of a subclass of GEMM-I riboswitches (named GEMM-1b) that selectively respond to cAG (Kellenberger et al., 2015b). Independent research based on the observation of unusual gene associations for a subset of GEMM-I riboswitches also led to the discovery of the same cAG riboswitches in Deltaproteobacteria (Nelson et al., 2015). The identification of GEMM-Ib riboswitches by our group and others presents an inspiring example of riboswitch engineering facilitating biological discovery (Kellenberger et al., 2015b; Kellenberger et al., 2013; Shanahan et al., 2011; Smith et al., 2009).

Like other riboswitch classes, detailed characterization including structural elucidation of natural variants offers valuable lessons for engineering (Ren et al., 2015). In the phylogenetic screen, the majority of A20 sequences from GEMM-I class riboswitches displayed indiscriminate binding to cdiG and cAG. This observation indicated that, similar to position 74 in the purine riboswitch class, residue 20 in GEMM-I riboswitches represents a primary but not sole determinant for ligand specificity. Other residues

located at the peripheral region of the ligand binding pocket also affect ligand recognition. It was found that the first two base pairs that stack above the ligand and residue 20 in GEMM-I riboswitches could affect ligand discrimination (Ren et al., 2015). This result again reinforces that nucleotides directly interacting with the ligand act as primary determinants, but subsequent optimization of peripheral positions might be necessary to generate artificial riboswitches or riboswitch-based biosensors with the desired selectivity.

We recently engineered another class of cdiG riboswitches for novel ligand response. The GEMM-II riboswitch class recognizes the guanines of cdiG via hydrogen bonding, but makes fewer interactions with the ligand phosphate backbone relative to the GEMM-I class (Lee et al., 2010; Shanahan et al., 2011; Smith et al., 2011). Mutation of key residues responsible for hydrogen bonding with cdiG resulted in increased affinity for cAG (Bose et al., 2016). Furthermore, this engineered riboswitch responds to 2', 3'-cGAMP, a non-canonical cyclic dinucleotide with mixed phosphate backbone linkages involved in the mammalian innate immune response (Figure 1D) (Bose et al., 2016).

Together, the above examples demonstrate the potential for riboswitch aptamer domains to be engineered for sensing novel ligands. A general strategy that has been employed is one of semi-rational structure-based design: to mutate conserved nucleotides around the ligand binding pocket, and sometimes to also mutate less conserved positions to fine-tune binding properties. However, one critical assumption made in structure-based design is that point mutations will not destabilize the RNA fold. There are many examples of outlier mutants in the above studies that confirm this assumption does not always hold. Furthermore, these studies also show that regulatory function may require more extensive screening of riboswitch libraries to optimize the dynamic range for gene regulation, in order to tune ligand binding affinities to physiologically relevant ranges or ligand on-rates to complement transcriptional rates for kinetically-driven riboswitches.

II.B. Expression platform engineering

While natural riboswitch aptamer domains are conserved in secondary structure across broad bacterial phyla, the associated expression platforms can vary greatly in both sequence and mode of action between species, and even between examples within the same species (Lee et al., 2010). Therefore, engineering expression platforms requires choice of regulatory mode and fine-tuning of allosteric regulation, which often cannot be done rationally. In part, the difficulty in design arises from having limited information on ligand-free structures of riboswitches, and also having few full riboswitch structures that include the expression platform, as opposed to just the aptamer domain. In this section, we summarize the mechanisms of natural expression platforms, as well as alternative expression platforms that have been invented for use in bacteria and eukaryotes. We then describe the three major approaches that have been taken toward engineering new expression platforms: rational design, screening and selection methods, and computational design.

II.B.1. Mechanisms of natural expression platforms

Natural riboswitches are typically found in the 5' untranslated region of bacterial mRNAs. One primary mechanism for bacterial riboswitch function *in vivo* is co-transcriptional regulation (Figure 2A) (Winkler et al., 2002). Ligand-dependent premature termination of transcription can turn off gene expression by preventing RNA polymerase from transcribing the coding region. Ligand binding typically destabilizes the antiterminator and permits formation of a terminator hairpin that disrupts the active transcription complex and leads to RNA polymerase dissociation from the mRNA. Alternatively, ligand binding may affect rho-dependent transcription termination (Hollands et al., 2012). While most riboswitches turn off genes, the inverse to these mechanisms (e.g. ligand stabilizes the antiterminator) are used by riboswitches that turn on gene expression.

Another common mechanism in bacteria is translational regulation (Figure 2B). Protein translation in bacteria initiates with binding of the 30S ribosome to the ribosome binding site (RBS), so blocking the RBS prevents translation. Ligand binding to a riboswitch aptamer can either influence stem formation to block the RBS, or conversely to reveal the RBS (Winkler et al., 2002). A rarer mechanism involves ligand-induced ribozyme activation (Figure 2C), which only has been observed for the glucosamine-6-phosphate riboswitch class. It was found that ligand binding activates self-cleavage of the GlmS RNA, leading to mRNA degradation (Winkler et al., 2004). GlmS is a natural example of an aptazyme - an aptamer coupled to a ribozyme as the expression platform - that initiates phosphodiester cleavage in response to ligand binding.

So far, the only natural riboswitches found in eukaryotes are thiamine pyrophosphate (TPP) riboswitches in plants and filamentous fungi. Eukaryotic riboswitches regulate gene expression via alternative splicing (Figure 2D). In filamentous fungi and algae, the ligand-free aptamer masks a 5'-splice site, causing an alternative 5'-splice site to be utilized. The ligand-bound state unmasks this splice site, permitting inclusion of the exon located between the two alternative 5'-splice sites. This leads either to inclusion of an upstream open reading frame which is translated instead of the main ORF (in fungi), or inclusion of a premature stop codon (in algae) (Cheah et al., 2007; Croft et al., 2007). In higher plants, the ligand-free form in the 3’ UTR sequesters both 5'- and 3'-splice sites, leading to inclusion of a polyadenylation sequence and a mature RNA transcript. The ligand-bound form unveils both splice sites, leading to removal of the polyadenylation sequence and a transcript that is quickly degraded (Bocobza et al., 2007; Wachter et al., 2007).

II.B.2. Mechanisms of artificial expression platforms

Artificial riboswitches have been designed to regulate gene expression at the translational level in eukaryotes. Protein translation in eukaryotes begins with the 43S initiation complex binding to the 5' 7-methylguanosine cap of the mRNA followed by scanning to the start codon. Insertion of aptamers downstream of the 5' cap, but before the start codon, can block ribosome scanning and translation of the downstream gene (Figure 2E). This mechanism appears to only work in yeast, presumably because the mammalian ribosome is able to efficiently scan through structured RNAs before the start codon (Babendure et al., 2006).

Artificial aptazymes have been used to modulate gene expression in eukaryotes. Aptazymes located in either the 5'- or 3'-untranslated regions can prevent translation. In the former case, aptazyme cleavage leads to loss of the 5'-cap, preventing ribosome initiation (Figure 2F) (Auslander et al., 2010). Aptazyme activity at the 3'-end leads to loss of the poly-A tail, favoring degradation of the mRNA transcript (Figure 2G) (Nyan Win and Smolke, 2007). These aptazymes can act to activate or repress gene expression depending upon whether the small molecule ligand turns the ribozyme on or off (Nyan Win and Smolke, 2007).

Besides these cis-regulatory mechanisms, aptamers and aptazymes can also regulate genes in trans. In principle, any RNA structure that can be unmasked by alternate stem formation or in response to ribozyme activity has the potential to be regulated by small molecules. Aptamer or aptazyme fusion to pre-miRNAs, for instance, permits their processing and subsequent activity only in response to small molecule addition (Beisel et al., 2011; Kumar et al., 2009). This engineering effort resulted in ligand-inducible RNA interference. In two other examples, orthogonal tRNAs for unnatural amino acid incorporation and ribosome function have been successfully regulated by aptazymes (Berschneider et al., 2009; Wieland et al., 2010). We expect that, given the diversity already seen in Nature and in engineered examples, new mechanisms for riboswitch-mediated gene regulation will continue to be discovered and developed.

II.B.3. Rational/Semi-rational design

Careful study of riboswitch mechanisms has allowed the rational design of new riboswitches. In some transcriptionally-regulated riboswitches, expression platforms can be decoupled from the activity of the natural aptamer domain and fused to foreign aptamers, which can yield riboswitches responsive to other ligands. The *Bacillus subtilis metE*, *yitJ*, and *lysC* expression platforms were amenable to exchange with a number of aptamer domains that bind lysine, guanine, FMN, or theophylline to generate riboswitches that work in *E. coli* (Ceres et al., 2013a). Although these fusions exhibited lower dynamic ranges than the parent riboswitches, it was found that the dynamic range of a chimeric theophylline-lysC riboswitch could be restored to match the natural lysC riboswitch through optimization of P1 stem length. A similar aptamer swap design strategy has been used to generate a turn-off switch for cdiG in *B. subtilis* with the magnesium-sensing M-box riboswitch expression platform (Gao et al., 2014).

Natural turn-on riboswitches, such as the purine and cyclic di-GMP riboswitches, often contain significant overlap between the aptamer domain and expression platform, which prevents the generation of a simple, modular platform. To circumvent this issue, “decoupler sequences” were engineered to create portable expression platforms of the *B. subtilis* pbuE adenine and *Dechloromonas aromatica* metH SAH riboswitches (Ceres et al., 2013b). Chimeric turn-on riboswitches were generated for SAM, other purines, and theophylline with higher dynamic ranges than the parent riboswitches (Ceres et al., 2013b).

Translationally-regulated riboswitches have required different approaches to rational design, depending on the organism. It was found in *B. subtilis* that the minimal length of accessible RNA required for stable binding of the ribosome to the RBS is 14 nt (Suess

et al., 2004). Thus, inserting the theophylline aptamer (Jenison et al., 1994) with a transducer element designed to act by a "helix slipping" mechanism (Soukup and Breaker, 1999) resulted in theophylline-induced gene expression via changing the accessible length of the RBS. Additional efforts were based on the hypothesis that inserting an aptamer upstream of the start codon would affect translation. This worked in the case of the theophylline aptamer for *E. coli* and the tetracycline aptamer in *Saccharomyces cerevisiae*. The highest fold-changes in gene expression upon ligand addition were obtained when the aptamer was inserted 8 base pairs from the ribosome binding site or 5 base pairs upstream of the start codon, respectively (Desai and Gallivan, 2004; Hanson et al., 2003).

II.B.4. Screens/Selections

One challenge with rational riboswitch design is that the aptamer must function in the context of the expression platform *in vivo* for gene regulation to occur (Lemay et al., 2011). While highly specific aptamers can be readily generated through *in vitro* systematic evolution of ligands by exponential enrichment (SELEX) (Ellington and Szostak, 1990), SELEX-generated aptamer/expression platform hybrids do not always function *in vivo* (Wittmann and Suess, 2011). As a result, most methods for generating riboswitches requires *in vivo* screening for regulatory function before a desirable riboswitch is obtained.

In a typical screen, the riboswitch library is cloned in front of a reporter gene to generate a signal output for selection. For colorimetric reporter genes such as LacZ, plate-based screening is used (Desai and Gallivan, 2004; Lynch et al., 2007). One improved screening method uses the CheZ gene as a reporter to modulate cell motility, which requires only basic molecular biology reagents and can screen libraries with ~10^5 clones (Topp and Gallivan, 2008b). Employing a green fluorescent protein (GFP) reporter for riboswitch function allows the use of fluorescence-activated cell sorting (FACS) to screen much larger libraries for both basal and activated levels of gene expression (Lynch and Gallivan, 2009). However, these studies also revealed a possible pitfall for *in vivo* screens: individual cells in a population often exhibit heterogeneous expression of the reporter gene, which can complicate analysis and selection. This problem has been overcome by using an additional “noise reporter” system wherein the riboswitch gene is expressed along with a normalization control, such as mCherry (Liang et al., 2012).

While these screening techniques offer significant improvements in throughput, the development of an antibiotic resistance-based selection process would permit analysis of even larger riboswitch libraries. Construct analysis would only require multiple rounds of growth under increasingly stringent selection conditions until a desired level of riboswitch activity was obtained. However, one problem with antibiotic resistance cassettes is that there is no way to select for low cassette expression in the absence of ligand and mutations rendering the riboswitch constitutively active escape the selection process. This problem has been overcome through the development of a dual selection system with the tetracycline resistance cassette, tetA. Under high levels of tetA expression, cells become resistant to tetracycline, but sensitive to Ni^{2+}. In this way, riboswitches can be selected for both turn-on and turn-off function, by selecting for low

expression with Ni^{2+} and high expression with tetracycline (Muranaka et al., 2009a; Muranaka et al., 2009b; Nomura and Yokobayashi, 2007; Sharma et al., 2008).

Similar selection methods have been developed recently for use in *S. cerevisiae* (Klauser et al., 2015). Based on the two-hybrid assay used commonly to study protein interactions, the riboswitch is placed on the 3'-end of the GAL4 transcription factor gene. High Gal4 levels then lead to increased expression of the three reporter genes HIS3, URA3, and LacZ. Under high levels of expression, cells are able to survive on histidine-deficient media. A negative selection can be employed with the addition of 5-fluorooratic acid to the medium: the URA3 gene product, orotidine 5-phosphate decarboxylase, converts 5-fluorooratic acid to the toxic 5-fluorouracil. This system allowed screening of an aptazyme library of 10^6 members. Because only a small percentage of riboswitches that function in yeast can also function in mammalian systems, a similar dual selection method for use in mammalian cells would be highly desirable (Wei et al., 2013; Wittmann and Suess, 2011). Until then, candidate riboswitches will have to be screened individually in mammalian cells (Beilstein et al., 2015).

II.B.5. Computational design

While combinatorial libraries are useful for many ligands, high throughput screening *in vivo* is difficult with ligands that have poor solubility, are cytotoxic, or are impermeable. Thus, drawing on the rich body of literature on the thermodynamics and kinetics of RNA function, some groups have turned to computational modeling to generate functional riboswitches for use *in vivo*. Initial attempts involved the generation of a hypothetical model for riboswitch function based on changes in rate constants; however, while modeling allowed for predictive validation of known riboswitches, it did not allow *de novo* prediction of riboswitch functionality from sequence and thermodynamic constraints alone (Beisel and Smolke, 2009).

One computational approach towards transcriptionally-regulated riboswitches analyzed potential sequences based on three energetic criteria: 1) no secondary structure formation could occur between the aptamer and spacer, 2) the full-length transcript must contain a single hairpin formed from part of the aptamer, linker, and terminator regions, and 3) the mean free energy of the folded sequence must be more stable than the average free energy of a random sequence containing the same number and types of bases (Wachsmuth et al., 2013). Using these criteria, an artificial transcriptionally-regulated riboswitch was designed that possessed similar fold-regulation to that of natural riboswitches in response to theophylline. Further computational studies revealed that two parameters are likely important for transcriptionally-regulated riboswitches: the difference in free energy between the terminator hairpin and the ligand-bound structure, and the kinetic barrier to terminator formation (Wachsmuth et al., 2015).

A more generalizable computational pipeline was developed for translationally-regulated riboswitches (Espah Borujeni et al., 2016). The model for riboswitch activity takes into account the folding energy of the riboswitch combined with the free energy change resulting from ribosome binding to the RBS and the start of translation (Espah Borujeni et al., 2016). Using this computational model, a theophylline-responsive riboswitch was generated that activated gene expression in *E. coli* by 383-fold. More broadly, an

automated riboswitch design system based upon this model generated computationally designed riboswitches that responded *in vivo* to fluoride or dinitrotoluene. This automated pipeline can be used to generate new riboswitches from known RNA aptamers for which ligand-bound structures and binding free energies are known.

II.C. Biosensor engineering

Natural riboswitch aptamers exhibit high affinity and dynamic response to a target ligand, making them privileged scaffolds for biosensor development. A biosensor binds specifically to a biological molecule of interest to generate a signal that can be detected by a non-invasive method such as fluorescence. Biosensors allow for the real time imaging of biologically relevant molecules *in vivo*. General features of RNA-based fluorescent biosensors include: 1) a recognition domain for binding the target ligand, e.g. riboswitch aptamer, which is fused to 2) a transducer module that communicates ligand binding to activate 3) a signaling domain that binds the dye to produce a fluorescent output (Stojanovic and Kolpashchikov, 2004). The modular design of these sensors makes them easily adaptable for sensing a wide range of metabolites. Here we will confine our discussion to riboswitch-based sensors shown to function *in vivo*; for *in vitro* assay applications, we refer readers to reviews on alternative strategies to modulate fluorescence in aptamer-based systems (Navani and Li, 2006; Ouellet, 2016).

The signaling domain for RNA-based biosensors consists of an *in vitro* selected aptamer sequence with high affinity to a profluorescent dye molecule. The first aptamer used as a signaling domain was the malachite green (MG) aptamer, which exhibits >2000-fold fluorescent turn-on when bound to its target fluorophore (Babendure et al., 2003). Fusing the MG aptamer to ligand-binding aptamer domains led to development of theophylline, ATP, and FMN sensors that induced up to 8-fold fluorescent turn-on upon binding to target ligand (Stojanovic and Kolpashchikov, 2004). However, because MG is a non-specific DNA intercalator, these biosensors cannot be used *in vivo*. More recently, the Spinach aptamer was selected to bind 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), a small-molecule mimic of the GFP chromophore (Paige et al., 2011). Similar to MG, the DFHBI molecule is non-fluorescent in dilute solution but exhibits 1000-fold increase in fluorescence quantum yield upon binding and little to no cellular toxicity. Rational modifications to the Spinach scaffold have generated the Spinach2 aptamer, a variant with improved thermostability and enhanced fluorescent signal *in vivo* (Strack et al., 2013). High throughput screening approaches *in vivo* yielded a further advance with the Broccoli aptamer (Filonov et al., 2014).

Spinach, Spinach2, and Broccoli aptamers have been fused to riboswitch aptamers to generate ligand-responsive biosensors for live-cell imaging of target metabolites. The key challenge to biosensor engineering is the design of the transducer module, which must give allosteric regulation of the signaling domain by the recognition domain. An RNA-based biosensor that exhibited 25-fold fluorescence turn-on with SAM was constructed using the SAM-III riboswitch aptamer fused to ten candidate transducer modules, which were stem sequences predicted to have weak thermodynamic stabilities (Paige et al., 2012). The same strategy was applied to generate an ADP biosensor (Paige et al., 2012).

For natural riboswitches, their regulatory mechanisms often involve ligand binding affecting the P1 stem, so we expected that a transducer module involving the natural P1 stem to be effective for allosteric regulation. This P1-P2' strategy (Figure 3A) was used to construct an RNA-based fluorescent biosensor capable of visualizing the signaling molecule cdiG in live bacterial cells (Kellenberger et al., 2013). With the P1-P2' strategy, it was possible to screen a library of riboswitch aptamers fused to Spinach, where the natural riboswitch sequences were identified by bioinformatics analysis of bacterial genomes. The screen resulted in a suite of four biosensors for cdiG that exhibited a broad range of affinities, including one with K_d < 5 nM, and improved fluorescent turn-on, with maximal fluorescence signal that was brighter than the parent Spinach2 aptamer *in vivo* and useful in flow cytometry assays (Wang et al., 2016). The P1-P2' strategy also has been employed to generate biosensors for cyclic di-AMP (Kellenberger et al., 2015a) and 2', 3'-cGAMP (Bose et al., 2016).

Although the conventional riboswitch-Spinach fusion design is effective for appending most riboswitch aptamers to a Spinach aptamer, it is problematic for sequences that contain a terminal pseudoknot structure, such as the one found in the 3' portion of the SAH riboswitch aptamer. To solve this problem and generate a biosensor that exhibits >11-fold fluorescence turn-on in response to SAH, we instead fused the riboswitch through the natural P2 stem to a circularly permuted Spinach2 (cpSpinach2) aptamer (Figure 3B) (Su et al., 2016). The design of cpSpinach2 was necessary in order to keep the biosensor as one contiguous RNA sequence. Similar to the second-generation cdiG biosensors, highly functional SAH biosensors were identified in one round of phylogenetic screening. We also have preliminary data suggesting that this cpSpinach2 strategy is generalizable to other terminal pseudoknot-containing riboswitches.

An alternate strategy to engineering Spinach-based biosensors involves the incorporation of Spinach as an expression platform-like element, separate from the riboswitch aptamer (Figure 3C). Similar to the antiterminator-terminator structure switch in transcriptionally-regulated riboswitches, the ligand-free riboswitch aptamer interacts with part of the Spinach aptamer and blocks DFHBI binding, but upon ligand binding, Spinach is restored to its functional state. These Spinach fusions provide clear fluorescence turn-on for detection of guanine, adenine, TPP and SAM (You et al., 2015). Similar construction principles were used to produce a turn-off glycine biosensor, which assumes the functional Spinach conformation in the absence of ligand (Ketterer et al., 2016).

These different design strategies offer a generalizable approach for the production of fluorescent biosensors to image a variety of small molecule ligands *in vivo*. Our natural stem approach has proven effective for generating sensors that exhibit greater sensitivity and fluorescence turn-on in response to natural targets. This strategy has also proven useful when engineering aptamers selective for new target ligands, as evidenced by the development of our cGAMP and cAG responsive biosensors (Bose et al., 2016; Kellenberger et al., 2015b). The development of new fluorophore-aptamer pairs *in vitro* bodes well for the expansion of RNA-based biosensors into further wavelengths for orthogonal imaging (Dolgosheina et al., 2014; Sato et al., 2015; Sunbul and Jaschke, 2013). The future is bright for RNA-based fluorescent biosensors.

III. IN VIVO APPLICATIONS OF RIBOSWITCHES

The ability of both natural and synthetic riboswitches to respond to levels of a particular small molecule in a concentration-dependent manner has made them valuable for a wide range of applications in living systems, which we review in this section. These include the use of riboswitches in reporter systems for biochemically important ligands, the generation of riboswitches responsive to xenobiotic molecules for control of gene expression and cellular behaviors, and the use of riboswitches in metabolic engineering to optimize small molecule production. We also discuss recent developments in the use of riboswitch-based biosensors for *in vivo* detection of small molecules.

III.A. Riboswitches as Small Molecule Reporters

A common experiment to validate function of newly discovered riboswitches fuses 5'-UTRs containing potential riboswitches to the coding sequence of Beta-galactosidase, and these reporter constructs are used to demonstrate modulation of recombinant protein levels in response to added ligand (Mandal et al., 2004; Nahvi et al., 2002; Winkler et al., 2002). Furthermore, once the riboswitch ligand is known, similar reporter constructs can be used to monitor ligand levels *in vivo*. In this section, we review case studies for how riboswitches have been used to report *in vivo* levels of small molecules for enzyme screening, protein engineering, and strain engineering.

Vitamin B12 is an important corrinoid cofactor involved in C-C bond isomerization reactions; however, much remains unknown about regulation of its levels in the cell. A B12 reporter was constructed by fusing the promoter and 5'-UTR of the *btuB* operon from *E. coli* to the GFP coding sequence, which leads to B12-dependent repression of GFP expression in *E. coli* (Fowler et al., 2010). This reporter was used to validate the corrinoid salvage pathway by analyzing the fluorescence signal from genetic knockout strains upon addition of an exogenous precursor to B12, cobinamide. The reporter-knockout strategy was also applied to evaluate the role of various transporters in maintaining B12 homeostasis (Fowler et al., 2010). In a follow-up study on the $BtuC_2D_2F$ transporter, the same riboswitch reporter enabled rapid *in vivo* screening of transporter activity of over 100 mutants of the substrate binding pocket to identify the amino acids important for specific interactions to B12 (Fowler et al., 2013).

cdiG is the master regulator of the transition between the motile planktonic and sessile biofilm-forming states of bacteria, and multiple reporters and sensors have been developed for its detection. Genomic insertion of a cdiG reporter into *B. subtilis* allowed a high-throughput screen of 37 predicted cdiG metabolizing enzymes from *Clostridium difficile*, a pathogen associated with colitis, to be performed through flow-cytometry analysis (Gao et al., 2014). Similar validation of two membrane-bound diguanylate cyclases from *Bacillus thuringiensis* and *Xanthomonas oryzae* was possible through adaptation of a naturally-occurring triple-tandem GEMM-I riboswitch from *B. thuringiensis* into a fluorescent reporter for use in *E. coli (Zhou et al., 2016)*.

Riboswitch reporters can also be used in screens or selections to engineer or tune enzyme activities. One proof-of-principle study in yeast developed an enzyme capable of specifically removing a methyl group from caffeine to yield theophylline (Michener

and Smolke, 2012). A tandem plate/FACS-based screen of a mutant library constructed from a monooxygenase with low theophylline synthase activity used a theophylline aptazyme fused to a GFP coding sequence to successfully identify variants with higher theophylline synthase activity (Michener and Smolke, 2012). Similarly, a lysine riboswitch controlling a tetracycline resistance cassette facilitated selection of a chimeric aspartate kinase optimized for lysine fermentation (Wang et al., 2015). An analogous riboswitch-resistance cassette was used to select the appropriate promoter from a plasmid-based library to drive production of the TCA cycle enzyme phosphoenolpyruvate carboxylase, leading to a greater than 10-fold increase in lysine titers with little effect on cellular growth rate (Yang et al., 2013). These rapid optimizations of enzyme mutants and expression levels were performed without having to engineer a novel theophylline- or lysine-responsive transcription factor.

Riboswitch reporters can also be used to select for cell strains producing a desired cell-permeable product. To avoid the need to transform strain libraries and to have reporter expression take resources away from production, riboswitch reporters can be used in reporter cells separate from the producer cell to be optimized. To improve riboflavin (vitamin B2) production, an *E. coli* reporter strain was engineered that lacked B2 synthesis machinery but was able to convert riboflavin to flavin mononucleotide (FMN), and carried a *ribB* riboswitch-based aptazyme controlling GFP production (Meyer et al., 2015). Thus, the reporter *E. coli* generate fluorescence in the presence of B2. Utilizing nanoliter reactors, in which individual members of a *Bacillus* production strain library were grown with *E. coli* reporter cells and analyzed by FACS, strains were isolated possessing up to 150% of the B2 synthesis capabilities of the parent industrial strain (Meyer et al., 2015).

III.B. Riboswitches for Conditional Gene Regulation

While riboswitches can be used to understand and engineer biological processes involving natural metabolites, it is often desirable to couple gene expression to addition of a xenobiotic small molecule. Riboswitch-based gene regulation has been utilized in a variety of bacterial and eukaryotic species. This ability to precisely regulate gene expression has made higher-level applications, such as controlling cellular behaviors and developing logic gates, possible.

Riboswitch-based regulation of protein production has been demonstrated in a wide array of bacterial phyla. Modern riboswitches optimized through screens or computational methods have induction levels above 300-fold in *E. coli* (Espah Borujeni et al., 2016). Riboswitches can be incorporated as part of the transgene component, in contrast to other regulatory systems that require both regulatory protein and transgene to be introduced. A set of engineered theophylline-inducible riboswitches function in a wide range of bacterial phyla, including both gram positive and negative species (Reynoso et al., 2012; Topp et al., 2010). Optimized riboswitches activate gene expression 80-fold in *Acinetobacter baumanii*, 70-fold in *B. subtilis*, and 60-fold in *Mycobacterium smegmatis*. Although the optimized system resulted in only an 8.2-fold activation in the pathogen *M. tuberculosis*, it functioned robustly to regulate protein production in a macrophage infection model (Seeliger et al., 2012). Similar systems of theophylline-responsive riboswitches have been demonstrated in streptomyces and

cyanobacterial species (Ma et al., 2014; Nakahira et al., 2013; Ohbayashi et al., 2016; Rudolph et al., 2013).

Riboswitch engineering has also been applied to eukaryotic gene regulation: tetracycline and neomycin riboswitches have been developed that block ribosome scanning in yeast (Hanson et al., 2003; Suess et al., 2003; Weigand et al., 2008) and the tetracycline and theophylline aptamers have been used to regulate alternative splicing, similar to the fungal and plant TPP riboswitches described earlier (Kim et al., 2005; Weigand and Suess, 2007). The development of aptazyme systems based primarily on the theophylline and tetracycline aptamers have added options for control of gene expression (Nyan Win and Smolke, 2007). Other recent reports have demonstrated theophylline riboswitch-based control of genes in plant plastids and in viral replication in mammalian cell hosts (Bell et al., 2015; Ketzer et al., 2012; Ketzer et al., 2014; Verhounig et al., 2010; Wang and White, 2007).

The ability to induce gene expression has been applied to control important cellular behaviors such as cell motility. Control of the CheZ motility protein in *E. coli* through riboswitch-related activity has generated bacteria which preferentially move towards theophylline and AMM (Robinson et al., 2014; Topp and Gallivan, 2008a). Other desirable traits to control include cell proliferation and division. Use of modular theophylline and tetracycline aptazymes controlling different proliferative cytokines in murine and human T-cells led to a doubling of viability in cells exposed to small molecule inducer (Chen et al., 2010). Aptazyme-based regulation of cellular behaviors has also been demonstrated through the modulation of mating efficiency in yeast and control of the cell cycle in the U2-OS cell line (Galloway et al., 2013; Wei and Smolke, 2015).

Another application of riboswitch-based protein regulation is to generate conditional knockouts of essential genes for basic biological studies. Using a tetracycline-controlled expression system, a conditional knockdown system for five essential genes in *S. cerevisiae* was created by varying the promoter sequence to change the expression strength (Kötter et al., 2009). Notably, two of these genes, NOP8 and SEC1, were not successfully knocked down with previously established conditional expression systems. Additionally, riboswitches can enable the selective expression of dominant negative mutations, as has been shown with theophylline-responsive csrA in *E. coli* to demonstrate its role in autoaggregation and cell cycle control (Jin et al., 2009).

Besides simple on/off functionalities that respond to one small molecule, riboswitches can be engineered to engage in more complex modes of regulation through implementation of biochemical logic gates and feedback systems. Riboswitch-based logic gates exist in nature, with tandem riboswitches controlling pathways involving multiple ligands, or allowing more digital control over gene expression (Sudarsan et al., 2006). Similar logic gates have been implemented through the use of synthetic riboswitches and aptazymes (Klauser et al., 2012; Sharma et al., 2008; Win and Smolke, 2008). Additionally, a tandem riboswitch that only generates signal in a tight range of ligand concentration was developed by combining a translationally-controlled ON riboswitch with a less sensitive transcriptionally-controlled OFF TPP riboswitch (Muranaka and Yokobayashi, 2010).

III.C. Riboswitches to Engineer Metabolic Flux

One primary goal of metabolic engineering is to create organisms that produce industrial and fuel compounds at higher yields than can be obtained from a naturally occurring organism. A major associated challenge is diverting biochemical building blocks to produce the desired compound without compromising cell viability. Because riboswitches can regulate gene expression in response to a particular compound, they hold promise as tools for metabolic engineering. As described earlier, riboswitch reporters have been applied to screen or select for mutant enzymes or strains with higher production capacities. In addition, riboswitches have been used to directly control central metabolism enzymes to improve yields.

The lysine riboswitch was used to regulate the flux of carbon-containing precursors through the citric acid (TCA) cycle. Lysine is produced from the TCA intermediate oxaloacetate (OAA), which alternatively is converted to citrate by citrate synthase as part of the TCA cycle. Inserting an *E. coli* lysine-OFF riboswitch upstream of the citrate synthase gene in the industrially relevant lysine fermenting gram-positive bacterium *Corynebacterium glutamicum* led to 63% higher lysine yields (Zhou and Zeng, 2015b). Addition of an engineered lysine-ON switch upstream of the natural lysine transporter *lysE* gene further increased total lysine yield (Zhou and Zeng, 2015a).

III.D. Riboswitch-based Fluorescent Biosensors

Quantitative, real-time detection and tracking of small molecules in living cells has traditionally relied upon the use of protein-based biosensors (Okumoto et al., 2012). Riboswitch-based biosensors present an appealing alternative for several reasons. First, unlike proteins, these RNAs can be synthesized *in vitro* from commercial DNA templates using a single enzyme, T7 RNA polymerase, which accelerates high-throughput screening for biosensor development. Second, riboswitch aptamers are naturally evolved for *in vivo* function with high ligand specificity, affinity, and allosteric regulation in response to ligand. These qualities enable RNA biosensors to produce a high fluorescence turn-on in the presence of their target molecule, with excellent discrimination against other cellular metabolites. Third, Spinach-based biosensors employ a fluorophore dye molecule that enables cellular imaging under anaerobic conditions (Wang et al., 2016). In contrast, protein-based sensors generally require extensive engineering to optimize signal output and affinity, and fluorescent proteins in the GFP family require oxygen for chromophore maturation. In this section, we review the applications of riboswitch-based fluorescent biosensors for monitoring metabolites and small molecule signals in bacteria.

To date, riboswitch-based biosensors have been developed for *in vivo* fluorescent imaging of metabolites SAM, ADP, TPP, and SAH in bacteria (Paige et al., 2012; Su et al., 2016; You et al., 2015). For example, the TPP biosensor was used to monitor TPP biosynthesis in *E. coli* after addition of thiamine, detecting different rates of accumulation within the cell population (You et al., 2015). A more stringent example for *in vivo* metabolite detection is the SAH biosensor, which must detect SAH selectively because the related SAM cofactor has been measured to be in ~300-fold excess in *E. coli* (Halliday et al., 2010). We showed that the SAH biosensor has higher selectivity than commercial monoclonal antibodies, and further demonstrated fluorescence

monitoring of the chemical inhibition of 5'-methylthioadenosine nucleosidase (MTAN), an endogenous enzyme involved in SAH turnover, in live *E. coli* using flow cytometry (Su et al., 2016). Another potential application of SAH biosensors is to detect methyltransferase activity *in vivo*, although natural product inhibition of this enzyme class by SAH is common, and will require a more sensitive biosensor to detect low turnover enzyme activity. These initial applications do illustrate the potential for using riboswitch-based biosensors to study real-time dynamics of specific metabolic pathways in bacteria.

To date, riboswitch-based biosensors also have been developed for *in vivo* fluorescent imaging of small molecule signals cdiG, cdiA, and cAG in bacteria (Kellenberger et al., 2015a; Kellenberger et al., 2015b; Kellenberger et al., 2013; Wang et al., 2016). These biosensors have been applied toward studying the activity of the corresponding signaling enzymes within cells. We employed the cdiA biosensor in a flow cytometry assay to validate the activity of a putative archaeal diadenylate cyclase, providing the first experimental evidence for cdiA as a signaling molecule in this domain of life (Kellenberger et al., 2015a). A high-throughput flow cytometry screen of 29 candidate GGDEF enzymes from *Geobacter sulfurreducens* using our cdiG and cAG biosensors recently revealed a novel subclass of hybrid promiscuous (Hypr) GGDEF enzymes that produce cAG (Hallberg et al., 2016). This discovery breaks the long-standing paradigm that GGDEF enzymes only make cdiG, and the distribution of Hypr GGDEFs shows that cAG signaling is more widespread in bacteria than previously understood. These initial applications demonstrate the power of riboswitch-based biosensors to study *in vivo* enzyme activity and signaling activity.

IV. CONCLUSIONS

Among the diverse roles and functions of RNA, natural riboswitches stand out as small molecule induced regulators of gene expression and privileged scaffolds for future engineering. This has been facilitated by the multiple strengths possessed by riboswitch aptamers, including their modular nature, small genetic footprint, and portability.

Riboswitches can be engineered through their aptamer domains to sense unnatural small molecules for orthogonal gene regulation. Diverse expression platforms have permitted their use in modulating genetic outputs in a broad spectrum of organisms. Beyond traditional roles in gene regulation, the interchangeable nature of the aptamer domain has allowed their import to new signal outputs such as fluorescence. Engineered riboswitches have already demonstrated multiple applications in biological discovery and the development of synthetic gene networks. As this field continues to expand, new engineering efforts must meet the growing demand for more stringent applications.

Riboswitch engineering has been enabled by multiple strategies, including rational structure-based design, computational models, and screens/selections. To meet the rising need, more strategies must be developed to overcome limitations inherent in these methods. The current strategy coupling structure-based design and screening can lead to engineered aptamer domains, but this method is laborious and low-yielding, making higher-throughput selection strategies desirable. SELEX-based methods offer a

possible solution, but *in vitro* activity does not always correlate with *in vivo* function. This implies that selection methods must have an *in vivo* component. Although a general strategy for expression platform engineering is desirable, the best mammalian riboswitches possess lower efficiency in gene regulation than their bacterial counterparts, necessitating the development of selection methods specific to these organisms. In the context of riboswitch-based biosensors, a broadened palette of fluorescent output wavelengths would permit multiplexed imaging of metabolites. We envision that with these new developments, engineered riboswitches will attain a powerful place in the toolbox of synthetic and chemical biologists.

Figures:

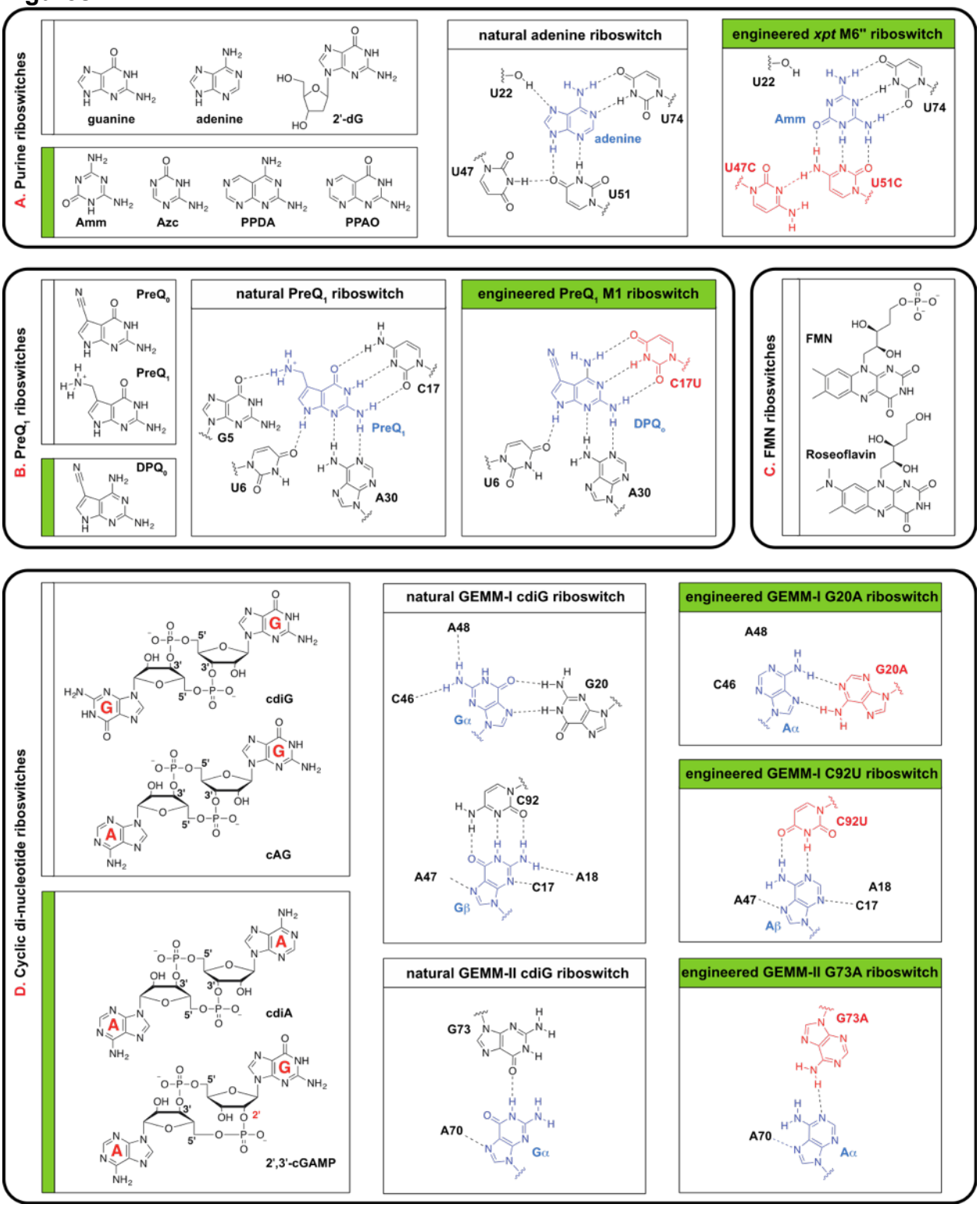

A. Purine riboswitches
guanine
adenine
2'-dG
Amm
Azc
PPDA
PPAO
natural adenine riboswitch
U22
U74
adenine
U47
U51
engineered xpt M6" riboswitch
U22
Amm
U74
U47C
U51C
B. PreQ1 riboswitches
PreQ0
PreQ1
DPQ0
natural PreQ1 riboswitch
G5
C17
PreQ1
U6
A30
engineered PreQ1 M1 riboswitch
C17U
DPQ0
U6
A30
C. FMN riboswitches
FMN
Roseoflavin
D. Cyclic di-nucleotide riboswitches
cdiG
cAG
cdiA
2',3'-cGAMP
natural GEMM-I cdiG riboswitch
A48
C46
G20
C92
A47
C17
A18
engineered GEMM-I G20A riboswitch
A48
C46
G20A
engineered GEMM-I C92U riboswitch
C92U
A47
C17
A18
natural GEMM-II cdiG riboswitch
G73
A70
engineered GEMM-II G73A riboswitch
G73A
A70

Figure 1 - Ligands for Natural and Engineered Aptamer Domains and Their Binding Pocket Hydrogen-bonding Network. White rectangles represent natural ligands for the indicated class of riboswitch, while green rectangles denote artificial ligands which require aptamer domain engineering. (A) Natural purine riboswitches have been found that respond to guanine, adenine, and 2'-deoxyguanosine. Mutants which instead interact with Amm, Azc, PPDA, and PPAO have been engineered. The adenine ligand interacts with the riboswitch via a Watson-Crick interaction with U74, a Hoogsteen interaction with the hydroxyl group from the ribose of U22, and a sugar face interaction with U51 and U47 to impart specificity. In contrast, the engineered apt M6'' riboswitch has C51 and C47 mutations, which allow specific interaction with Amm. (B) Natural PreQ1 riboswitches interact with PreQ0 and PreQ1. A C17U mutation in the riboswitch confers preferential binding to DPQ0. (C) FMN riboswitches naturally bind to both flavin mononucleotide (FMN) and roseoflavin, which acts as an antibacterial compound by targeting FMN riboswitches. Natural riboswitches from species which biosynthesize roseoflavin, however, selectively respond only to FMN. (D) The GEMM-I and GEMM-II riboswitch families are natural receptors for cdiG and cAG, while engineered variants have been generated which bind to cdiA and 2',3'-cGAMP. Engineered and natural GEMM-I riboswitches that bind to cAG have an A20 instead of a G, which interacts with the Hoogsteen face of the A. In addition, cdiA-specific GEMM-I riboswitches have a C92U mutation which interacts with the Watson-Crick face of Abeta. Engineered GEMM-II riboswitches with a G73A mutation can bind cAG and 2',3'-cGAMP, which allows interaction with the endocyclic nitrogen on the Watson-Crick face of the adenine residue.

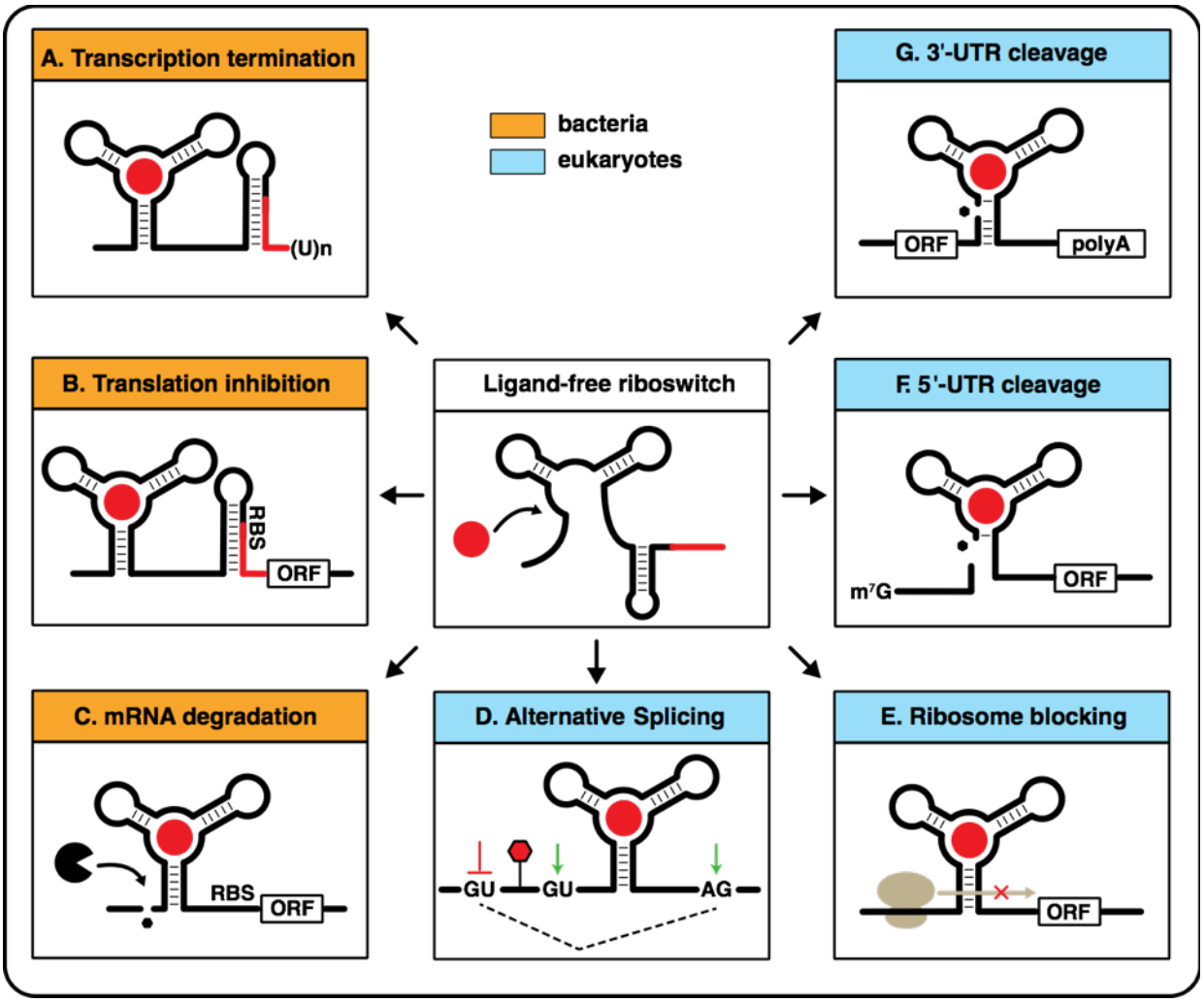


Figure 2 - Mechanisms for Natural and Synthetic Riboswitch Regulation. Orange mechanisms are found in bacteria, while blue mechanisms exist in eukaryotes. (A) Transcriptional attenuation: in the ligand-bound form, a terminator hairpin is formed, causing transcription termination. (B) Translation inhibition: ligand binding causes an alternative structure to form, occluding the ribosome-binding site (RBS), and preventing initiation of translation. (C) mRNA degradation caused by ribozyme activity: ligand-induced self-cleavage by ribozyme activity destabilizes the mRNA, leading to rapid degradation. (D) Ribosome blocking: in yeast, aptamer-ligand interactions can inhibit the ribosome from scanning from the 5' 7-methylguanosine cap to the translation start site, preventing translation. (E) Alternative splicing: ligand binding leads to an alternate mature mRNA forming through changes in splice site selection. Depending on the organism, this leads to addition of a small upstream ORF, inclusion of a premature stop codon, or exclusion of the poly-A tail and mRNA degradation. (F) 5'-UTR cleavage: removal of the 5' 7-methylguanosine cap by aptazyme activity prevents ribosome progression and destabilizes the mRNA. (G) 3'-UTR cleavage: removal of the polyadenosine tail by aptazyme activity leads to rapid mRNA degradation.

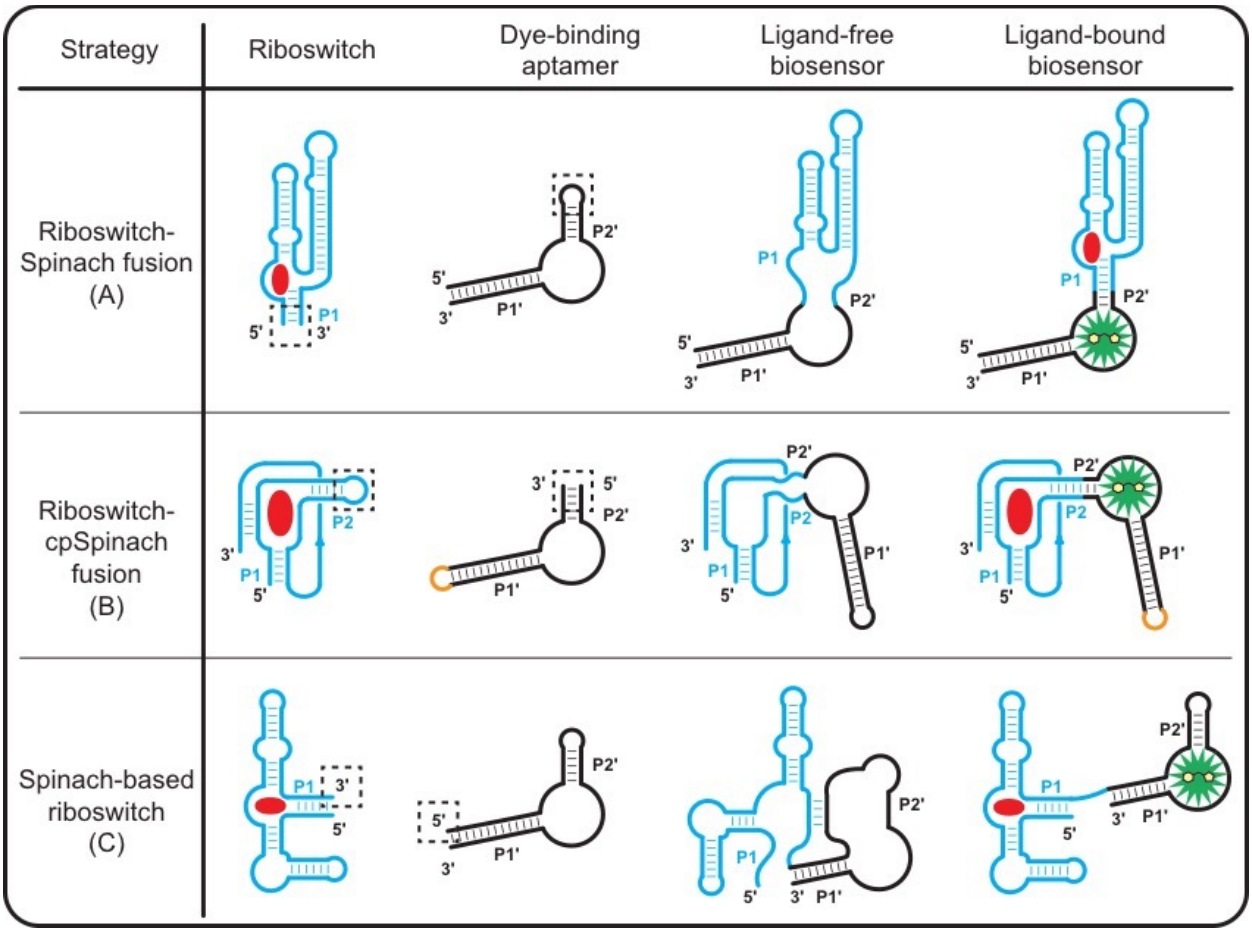


Figure 3 - Structures of the Three Main Classes of Riboswitch-based Biosensors. Riboswitch domains and their corresponding metabolites are depicted in blue and red, while the Spinach aptamer and DFHBI are shown in black and green. (A) The riboswitch-Spinach fusion strategy for biosensor development involves grafting of the P1' stem of a sensor domain to the P2 stem of Spinach. (B) To accommodate pseudoknot-containing riboswitches with non-interacting 5' and 3' ends, circularly permuted Spinach (cpSpinach), which has an engineered P2 open stem, can be fused to the P2' stem of a riboswitch. (C) The final approach creates a Spinach-based riboswitch, linking Spinach to the sensor domain through a transmitter sequence that prevents fluorophore binding in the absence of metabolite. A Spinach riboswitch can be generated by linking the Spinach aptamer to a sensor domain through a transmitter sequence that allows fluorophore binding only in the presence of the metabolite.

Bibliography

Auslander, S., Ketzer, P., and Hartig, J.S. (2010). A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression. Mol Biosyst *6*, 807-814.

Babendure, J.R., Adams, S.R., and Tsien, R.Y. (2003). Aptamers switch on fluorescence of triphenylmethane dyes. J Am Chem Soc *125*, 14716-14717.

Babendure, J.R., Babendure, J.L., Ding, J., and Tsien, R.Y. (2006). Control of mammalian translation by mRNA structure near caps. RNA, 851-861.

Batey, R.T., Gilbert, S.D., and Montange, R.K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. Nature *432*, 411-415.

Beilstein, K., Wittmann, A., Grez, M., and Suess, B. (2015). Conditional Control of Mammalian Gene Expression by Tetracycline-Dependent Hammerhead Ribozymes. ACS Synth Biol *4*, 526-534.

Beisel, C.L., Chen, Y.Y., Culler, S.J., Hoff, K.G., and Smolke, C.D. (2011). Design of small molecule-responsive microRNAs based on structural requirements for Drosha processing. Nucleic Acids Res *39*, 2981-2994.

Beisel, C.L., and Smolke, C.D. (2009). Design principles for riboswitch function. PLoS Comput Biol *5*, e1000363.

Bell, C.L., Yu, D., Smolke, C.D., Geall, A.J., Beard, C.W., and Mason, P.W. (2015). Control of alphavirus-based gene expression using engineered riboswitches. Virology *483*, 302-311.

Berschneider, B., Wieland, M., Rubini, M., and Hartig, J.S. (2009). Small-molecule-dependent regulation of transfer RNA in bacteria. Angew Chem Int Ed *48*, 7564-7567.

Bocobza, S., Adato, A., Mandel, T., Shapira, M., Nudler, E., and Aharoni, A. (2007). Riboswitch-dependent gene regulation and its evolution in the plant kingdom Riboswitch-dependent gene regulation and its evolution in the plant kingdom. Genes Dev, 2874-2879.

Bose, D., Su, Y., Marcus, A., Raulet, D.H., and Hammond, M.C. (2016). An RNA-Based Fluorescent Biosensor for High-Throughput Analysis of the cGAS-cGAMP-STING Pathway. Cell Chem Biol.

Ceres, P., Garst, A.D., Marcano-Velázquez, J.G., and Batey, R.T. (2013a). Modularity of select riboswitch expression platform enables facile engineering of novel genetic regulatory devices. ACS Synth Biol *2*, 463-472.

Ceres, P., Trausch, J.J., and Batey, R.T. (2013b). Engineering modular 'ON' RNA switches using biological components. Nucleic Acids Res *41*, 10449-10461.

Cheah, M.T., Wachter, A., Sudarsan, N., and Breaker, R.R. (2007). Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. Nature *447*, 497-500.

Chen, Y.Y., Jensen, M.C., and Smolke, C.D. (2010). Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. Proc Natl Acad Sci USA *107*, 8531-8536.

Croft, M.T., Moulin, M., Webb, M.E., and Smith, A.G. (2007). Thiamine biosynthesis in algae is regulated by riboswitches. Proc Natl Acad Sci USA *104*, 20770-20775.

Desai, S.K., and Gallivan, J.P. (2004). Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. J Am Chem Soc *126*, 13247-13254.

Dixon, N., Duncan, J.N., Geerlings, T., Dunstan, M.S., McCarthy, J.E., Leys, D., and Micklefield, J. (2010). Reengineering orthogonally selective riboswitches. Proc Natl Acad Sci USA *107*, 2830-2835.

Dixon, N., Robinson, C.J., Geerlings, T., Duncan, J.N., Drummond, S.P., and Micklefield, J. (2012). Orthogonal riboswitches for tuneable coexpression in bacteria. Angew Chem Int Ed *51*, 3620-3624.

Dolgosheina, E.V., Jeng, S.C., Panchapakesan, S.S., Cojocaru, R., Chen, P.S., Wilson, P.D., Hawkins, N., Wiggins, P.A., and Unrau, P.J. (2014). RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking. ACS Chem Biol *9*, 2412-2420.

Edwards, A.L., and Batey, R.T. (2009). A Structural Basis for the Recognition of 2′-Deoxyguanosine by the Purine Riboswitch. J Mol Biol *385*, 938-948.

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

Espah Borujeni, A., Mishler, D.M., Wang, J., Huso, W., and Salis, H.M. (2016). Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. Nucleic Acids Res *44*, 1-13.

Filonov, G.S., Moon, J.D., Svensen, N., and Jaffrey, S.R. (2014). Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. J Am Chem Soc *136*, 16299-16308.

Fowler, C.C., Brown, E.D., and Li, Y. (2010). Using a riboswitch sensor to examine coenzyme B12 metabolism and transport in E. coli. Chem Biol *17*, 756-765.

Fowler, C.C., Sugiman-Marangos, S., Junop, M.S., Brown, E.D., and Li, Y. (2013). Exploring intermolecular interactions of a substrate binding protein using a riboswitch-based sensor. Chem Biol *20*, 1502-1512.

Galloway, K.E., Franco, E., and Smolke, C.D. (2013). Dynamically reshaping signaling networks to program cell fate via genetic controllers. Science *341*, 1235005.

Gao, X., Dong, X., Subramanian, S., Matthews, P.M., Cooper, C.A., Kearns, D.B., and Dann, C.E. (2014). Engineering of Bacillus subtilis strains to allow rapid characteri-

zation of heterologous diguanylate cyclases and phosphodiesterases. Appl Environ Microbiol *80*, 6167-6174.

Gilbert, S.D., Love, C.E., Edwards, A.L., and Batey, R.T. (2007). Mutational Analysis of the Purine Riboswitch Aptamer Domain. Biochemistry *46*, 13297-13309.

Hallberg, Z.F., Wang, X.C., Wright, T.A., Nan, B., Ad, O., Yeo, J., and Hammond, M.C. (2016). Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3', 3'-cGAMP). Proc Natl Acad Sci USA *113*, 1790-1795.

Halliday, N.M., Hardie, K.R., Williams, P., Winzer, K., and Barrett, D.A. (2010). Quantitative liquid chromatography-tandem mass spectrometry profiling of activated methyl cycle metabolites involved in LuxS-dependent quorum sensing in Escherichia coli. Anal Biochem *403*, 20-29.

Hanson, S., Berthelot, K., Fink, B., McCarthy, J.E.G., and Suess, B. (2003). Tetracycline-aptamer-mediated translational regulation in yeast. Mol Microbiol *49*, 1627-1637.

Hollands, K., Proshkin, S., Sklyarova, S., Epshtein, V., Mironov, A., Nudler, E., and Groisman, E.A. (2012). Riboswitch control of Rho-dependent transcription termination. Proc Natl Acad Sci USA *109*, 5376-5381.

Jenison, R.D., Gill, S.C., Pardi, A., and Polisky, B. (1994). High-resolution molecular discrimination by RNA. Science *263*, 1425-1429.

Jin, Y., Watt, R.M., Danchin, A., and Huang, J.D. (2009). Use of a riboswitch-controlled conditional hypomorphic mutation to uncover a role for the essential csrA gene in bacterial autoaggregation. J Biol Chem *284*, 28738-28745.

Kang, M., Peterson, R., and Feigon, J. (2009). Structural Insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. Mol Cell *33*, 784-790.

Kellenberger, C.A., Chen, C., Whiteley, A.T., Portnoy, D.A., and Hammond, M.C. (2015a). RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messenger Cyclic di-AMP. J Am Chem Soc *137*, 6432-6435.

Kellenberger, C.a., Wilson, S.C., Hickey, S.F., Gonzalez, T.L., Su, Y., Hallberg, Z.F., Brewer, T.F., Iavarone, A.T., Carlson, H.K., Hsieh, Y.-F.*, et al.* (2015b). GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. Proc Natl Acad Sci USA *112*, 5383-5388.

Kellenberger, C.A., Wilson, S.C., Sales-Lee, J., and Hammond, M.C. (2013). RNA-based fluorescent biosensors for live cell imaging of second messengers cyclic di-GMP and cyclic AMP-GMP. J Am Chem Soc *135*, 4906-4909.

Ketterer, S., Gladis, L., Kozica, A., and Meier, M. (2016). Engineering and characterization of fluorogenic glycine riboswitches. Nucleic Acids Res *44*, gkw465.

Ketzer, P., Haas, S.F., Engelhardt, S., Hartig, J.S., and Nettelbeck, D.M. (2012). Synthetic riboswitches for external regulation of genes transferred by replication-deficient and oncolytic adenoviruses. Nucleic Acids Res *40*, e167.

Ketzer, P., Kaufmann, J.K., Engelhardt, S., Bossow, S., von Kalle, C., Hartig, J.S., Ungerechts, G., and Nettelbeck, D.M. (2014). Artificial riboswitches for gene expression and replication control of DNA and RNA viruses. Proc Natl Acad Sci USA *111*, E554-562.

Kim, D.-S., Gusti, V., Pillai, S.G., and Gaur, R.K. (2005). An artificial riboswitch for controlling pre-mRNA splicing. RNA *11*, 1667-1677.

Kim, J.N., Roth, A., and Breaker, R.R. (2007). Guanine riboswitch variants from Mesoplasma florum selectively recognize 2'-deoxyguanosine. Proc Natl Acad Sci USA *104*, 16092-16097.

Klauser, B., Atanasov, J., Siewert, L.K., and Hartig, J.S. (2015). Ribozyme-Based Aminoglycoside Switches of Gene Expression Engineered by Genetic Selection in S. cerevisiae. ACS Synth Biol *4*, 516-525.

Klauser, B., Saragliadis, A., Ausländer, S., Wieland, M., Berthold, M.R., and Hartig, J.S. (2012). Post-transcriptional Boolean computation by combining aptazymes controlling mRNA translation initiation and tRNA activation. Mol Biosyst *8*, 2242.

Klein, D.J., Edwards, T.E., and Ferre-D'Amare, A.R. (2009). Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. Nat Struct Mol Biol *16*, 343-344.

Kötter, P., Weigand, J.E., Meyer, B., Entian, K.-D., and Suess, B. (2009). A fast and efficient translational control system for conditional expression of yeast genes. Nucleic Acids Res *37*, e120.

Kulshina, N., Baird, N.J., and Ferre-D'Amare, A.R. (2009). Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch. Nat Struct Mol Biol *16*, 1212-1217.

Kumar, D., An, C.I., and Yokobayashi, Y. (2009). Conditional RNA interference mediated by allosteric ribozyme. J Am Chem Soc *131*, 13906-13907.

Lee, E.R., Baker, J.L., Weinberg, Z., Sudarsan, N., and Breaker, R.R. (2010). An Allosteric Self-Splicing Ribozyme Triggered by a Bacterial Second Messenger. Science *329*, 845-848.

Lee, E.R., Blount, K.F., and Breaker, R.R. (2009). Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. RNA Biol *6*, 187-194.

Lemay, J.-F., Desnoyers, G., Blouin, S., Heppell, B., Bastet, L., St-Pierre, P., Massé, E., and Lafontaine, D.A. (2011). Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. PLoS Genetics *7*, e1001278.

Liang, J.C., Chang, A.L., Kennedy, A.B., and Smolke, C.D. (2012). A high-throughput, quantitative cell-based screen for efficient tailoring of RNA device activity. Nucleic Acids Res *40*, 1-14.

Lynch, S.A., Desai, S.K., Sajja, H.K., and Gallivan, J.P. (2007). A High-Throughput Screen for Synthetic Riboswitches Reveals Mechanistic Insights into Their Function. Chem Biol *14*, 173-184.

Lynch, S.A., and Gallivan, J.P. (2009). A flow cytometry-based screen for synthetic riboswitches. Nucleic Acids Res *37*, 184-192.

Ma, A.T., Schmidt, C.M., and Golden, J.W. (2014). Regulation of gene expression in diverse cyanobacterial species by using theophylline-responsive riboswitches. Appl Environ Microbiol *80*, 6704-6713.

Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C., and Breaker, R.R. (2003). Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. Cell *113*, 577-586.

Mandal, M., and Breaker, R.R. (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. Nat Struct Mol Biol *11*, 29-35.

Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L., and Breaker, R.R. (2004). A glycine-dependent riboswitch that uses cooperative binding to control gene expression. Science *306*, 275-279.

Mansjö, M., and Johansson, J. (2011). The Riboflavin analog Roseoflavin targets an FMN-riboswitch and blocks Listeria monocytogenes growth, but also stimulates virulence gene-expression and infection. RNA Biol *8*, 674-680.

McCown, P.J., Liang, J.J., Weinberg, Z., and Breaker, R.R. (2014). Structural, functional, and taxonomic diversity of three preq1 riboswitch classes. Chem Biol *21*, 880-889.

Meyer, A., Pellaux, R., Potot, S., Becker, K., Hohmann, H.-P., Panke, S., and Held, M. (2015). Optimization of a whole-cell biocatalyst by employing genetically encoded product sensors inside nanolitre reactors. Nat Chem *7*, 673-678.

Meyer, M.M., Roth, A., Chervin, S.M., Garcia, G.A., and Breaker, R.R. (2008). Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria. RNA *14*, 685-695.

Michener, J.K., and Smolke, C.D. (2012). High-throughput enzyme evolution in Saccharomyces cerevisiae using a synthetic RNA switch. Metab Eng *14*, 306-316.

Muranaka, N., Abe, K., and Yokobayashi, Y. (2009a). Mechanism-guided library design and dual genetic selection of synthetic OFF riboswitches. ChemBioChem *10*, 2375-2381.

Muranaka, N., Sharma, V., Nomura, Y., and Yokobayashi, Y. (2009b). An efficient platform for genetic selection and screening of gene switches in Escherichia coli. Nucleic Acids Res *37*, e39.

Muranaka, N., and Yokobayashi, Y. (2010). A synthetic riboswitch with chemical bandpass response. Chem Commun *46*, 6825-6827.

Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L., and Breaker, R.R. (2002). Genetic control by a metabolite binding mRNA. Chem Biol *9*, 1043-1049.

Nakahira, Y., Ogawa, A., Asano, H., Oyama, T., and Tozawa, Y. (2013). Theophylline-dependent riboswitch as a novel genetic tool for strict regulation of protein expression in cyanobacterium synechococcus elongatus PCC 7942. Plant Cell Physiol *54*, 1724-1735.

Navani, N.K., and Li, Y. (2006). Nucleic acid aptamers and enzymes as sensors. Curr Opin Chem Biol *10*, 272-281.

Nelson, J.W., Sudarsan, N., Phillips, G.E., Stav, S., Lunse, C.E., McCown, P.J., and Breaker, R.R. (2015). Control of bacterial exoelectrogenesis by c-AMP-GMP. Proc Natl Acad Sci USA *112*, 5389-5394.

Noeske, J., Richter, C., Grundl, M.A., Nasiri, H.R., Schwalbe, H., and Wöhnert, J. (2005). An intermolecular base triple as the basis of ligand specificity and affinity in the guanine- and adenine-sensing riboswitch RNAs. Proc Natl Acad Sci USA *102*, 1372-1377.

Nomura, Y., and Yokobayashi, Y. (2007). Reengineering a natural riboswitch by dual genetic selection. J Am Chem Soc *129*, 13814-13815.

Nyan Win, M., and Smolke, C.D. (2007). A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc Natl Acad Sci USA *104*, 14283-14288.

Ohbayashi, R., Akai, H., Yoshikawa, H., Hess, W.R., and Watanabe, S. (2016). A tightly inducible riboswitch system in Synechocystis sp. PCC 6803. J Gen Appl Microbiol *159*, 154-159.

Okumoto, S., Jones, A., and Frommer, W.B. (2012). Quantitative imaging with fluorescent biosensors. Annu Rev Plant Biol *63*, 663-706.

Ott, E., Stolz, J., Lehmann, M., and Mack, M. (2009). The RFN riboswitch of Bacillus subtilis is a target for the antibiotic roseoflavin produced by Streptomyces davawensis. RNA Biol *6*, 276-280.

Ouellet, J. (2016). RNA Fluorescence with Light-Up Aptamers. Front Chem *4*, 29.

Paige, J.S., Nguyen-Duc, T., Song, W., and Jaffrey, S.R. (2012). Fluorescence Imaging of Cellular Metabolites with RNA. Science *335*, 1194.

Paige, J.S., Wu, K.Y., and Jaffrey, S.R. (2011). RNA Mimics of Green Fluorescent Protein. Science *333*, 642-646.

Pedrolli, D.B., Matern, A., Wang, J., Ester, M., Siedler, K., Breaker, R., and MacK, M. (2012). A highly specialized flavin mononucleotide riboswitch responds differently to similar ligands and confers roseoflavin resistance to Streptomyces davawensis. Nucleic Acids Res *40*, 8662-8673.

Pikovskaya, O., Polonskaia, A., Patel, D.J., and Serganov, A. (2011). Structural principles of nucleoside selectivity in a 2'-deoxyguanosine riboswitch. Nat Chem Biol *7*, 748-755.

Ren, A., Wang, X.C., Kellenberger, C.A., Rajashankar, K.R., Jones, R.A., Hammond, M.C., and Patel, D.J. (2015). Structural basis for molecular discrimination by a 3',3'-cGAMP sensing riboswitch. Cell Rep *11*, 1-12.

Reynoso, C.M.K., Miller, M.A., Bina, J.E., Gallivan, J.P., and Weiss, D.S. (2012). Riboswitches for intracellular study of genes involved in francisella pathogenesis. mBio *3*.

Robinson, C.J., Medina-Stacey, D., Wu, M.-C., Vincent, H.A., and Micklefield, J. (2016). Rewiring Riboswitches to Create New Genetic Circuits in Bacteria. Methods Enzymol *575*.

Robinson, C.J., Vincent, H.A., Wu, M.C., Lowe, P.T., Dunstan, M.S., Leys, D., and Micklefield, J. (2014). Modular riboswitch toolsets for synthetic genetic control in diverse bacterial species. J Am Chem Soc *136*, 10615-10624.

Rode, A.B., Endoh, T., and Sugimoto, N. (2015). Tuning riboswitch-mediated gene regulation by rational control of aptamer ligand binding properties. Angew Chem Int Ed *54*, 905-909.

Roth, A., Winkler, W.C., Regulski, E.E., Lee, B.W.K., Lim, J., Jona, I., Barrick, J.E., Ritwik, A., Kim, J.N., Welz, R.*, et al.* (2007). A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. Nat Struct Mol Biol *14*, 308-317.

Rudolph, M.M., Vockenhuber, M.P., and Suess, B. (2013). Synthetic riboswitches for the conditional control of gene expression in Streptomyces coelicolor. Microbiology *159*, 1416-1422.

Sato, S., Watanabe, M., Katsuda, Y., Murata, A., Wang, D.O., and Uesugi, M. (2015). Live-cell imaging of endogenous mRNAs with a small molecule. Angew Chem Int Ed *54*, 1855-1858.

Seeliger, J.C., Topp, S., Sogi, K.M., Previti, M.L., Gallivan, J.P., and Bertozzi, C.R. (2012). A riboswitch-based inducible gene expression system for mycobacteria. PLoS ONE *7*, e29266.

Serganov, A., and Nudler, E. (2013). A decade of riboswitches. Cell *152*, 17-24.

Serganov, A., Yuan, Y.R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A.T.n., Hobartner, C., Micura, R., Breaker, R.R., and Patel, D.J. (2004). Structural basis for

discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. Chem Biol *11*, 1729-1741.

Shanahan, C.A., Gaffney, B.L., Jones, R.A., and Strobel, S.A. (2011). Differential analogue binding by two classes of c-di-GMP riboswitches. J Am Chem Soc *133*, 15578-15592.

Sharma, V., Nomura, Y., and Yokobayashi, Y. (2008). Engineering complex riboswitch regulation by dual genetic selection. J Am Chem Soc *130*, 16310-16315.

Smith, K.D., Lipchock, S.V., Ames, T.D., Wang, J., Breaker, R.R., and Strobel, S.A. (2009). Structural basis of ligand binding by a c-di-GMP riboswitch. Nat Struct Mol Biol *16*, 1218-1223.

Smith, K.D., Lipchock, S.V., Livingston, A.L., Shanahan, C.A., and Strobel, S.A. (2010). Structural and biochemical determinants of ligand binding by the c-di-GMP riboswitch. Biochemistry *49*, 7351-7359.

Smith, K.D., Shanahan, C.A., Moore, E.L., Simon, A.C., and Strobel, S.A. (2011). Structural basis of differential ligand recognition by two classes of bis-(3'-5')-cyclic dimeric guanosine monophosphate-binding riboswitches. Proc Natl Acad Sci USA *108*, 7757-7762.

Soukup, G.A., and Breaker, R.R. (1999). Engineering precision RNA molecular switches. Proc Natl Acad Sci USA *96*, 3584-3589.

Stojanovic, M.N., and Kolpashchikov, D.M. (2004). Modular aptameric sensors. J Am Chem Soc *126*, 9266-9270.

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

Su, Y., Hickey, S.F., Keyser, S.G.L., and Hammond, M.C. (2016). In vitro and in vivo enzyme activity screening via RNA-based fluorescent biosensors for S-adenosyl-L-homocysteine (SAH). J Am Chem Soc, jacs.6b01621.

Sudarsan, N., Hammond, M.C., Block, K.F., Welz, R., Barrick, J.E., Roth, A., and Breaker, R.R. (2006). Tandem Riboswitch Architectures Exhibit Complex Gene Control Functions. Science *314*, 300-304.

Sudarsan, N., Lee, E.R., Weinberg, Z., Moy, R.H., Kim, J.N., Link, K.H., and Breaker, R.R. (2008). Riboswitches in eubacteria sense the second messenger cyclic di-GMP. Science *321*, 411-413.

Suess, B., Fink, B., Berens, C., Stentz, R., and Hillen, W. (2004). A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. Nucleic Acids Res *32*, 1610-1614.

Suess, B., Hanson, S., Berens, C., Fink, B., Schroeder, R., and Hillen, W. (2003). Conditional gene expression by controlling translation with tetracycline-binding aptamers. Nucleic Acids Res *31*, 1853-1858.

Sunbul, M., and Jaschke, A. (2013). Contact-mediated quenching for RNA imaging in bacteria with a fluorophore-binding aptamer. Angew Chem Int Ed *52*, 13401-13404.

Topp, S., and Gallivan, J.P. (2008a). Guiding Bacteria with Small Molecules and RNA. J Am Chem Soc *129*, 6807-6811.

Topp, S., and Gallivan, J.P. (2008b). Random walks to synthetic riboswitches - A high-throughput selection based on cell motility. ChemBioChem *9*, 210-213.

Topp, S., Reynoso, C.M.K., Seeliger, J.C., Goldlust, I.S., Desai, S.K., Murat, D., Shen, A., Puri, A.W., Komeili, A., Bertozzi, C.R.*, et al.* (2010). Synthetic riboswitches that induce gene expression in diverse bacterial species. Appl Environ Microbiol *76*, 7881-7884.

Tremblay, R., Lemay, J.F., Blouin, S., Mulhbacher, J., Bonneau, É., Legault, P., Dupont, P., Penedo, J.C., and Lafontaine, D.A. (2011). Constitutive regulatory activity of an evolutionarily excluded riboswitch variant. J Biol Chem *286*, 27406-27415.

Verhounig, A., Karcher, D., and Bock, R. (2010). Inducible gene expression from the plastid genome by a synthetic riboswitch. Proc Natl Acad Sci USA *107*, 6204-6209.

Vincent, H.A., Robinson, C.J., Wu, M.-C., Dixon, N., and Micklefield, J. (2014). Generation of Orthogonally Selective Bacterial Riboswitches by Targeted Mutagenesis and In Vivo Screening. In Methods Mol Biol, pp. 107-129.

Wachsmuth, M., Domin, G., Lorenz, R., Serfling, R., Findeiss, S., Stadler, P.F., and Morl, M. (2015). Design criteria for synthetic riboswitches acting on transcription. RNA Biol *12*, 221-231.

Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P.F., and Mörl, M. (2013). De novo design of a synthetic riboswitch that regulates transcription termination. Nucleic Acids Res *41*, 2541-2551.

Wachter, A., Tunc-Ozdemir, M., Grove, B.C., Green, P.J., Shintani, D.K., and Breaker, R.R. (2007). Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. Plant Cell *19*, 3437-3450.

Wang, J., Gao, D., Yu, X., Li, W., and Qi, Q. (2015). Evolution of a chimeric aspartate kinase for L-lysine production using a synthetic RNA device. Appl Microbiol Biotechnol *99*, 8527-8536.

Wang, J.X., Lee, E.R., Morales, D.R., Lim, J., and Breaker, R.R. (2008). Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. Mol Cell *29*, 691-702.

Wang, S., and White, K.A. (2007). Riboswitching on RNA virus replication. Proc Natl Acad Sci USA *104*, 10406-10411.

Wang, X.C., Wilson, S.C., and Hammond, M.C. (2016). Next-generation RNA-based fluorescent biosensors enable anaerobic detection of cyclic di-GMP. Nucleic Acids Res *44*, e139.

Wei, K.Y., Chen, Y.Y., and Smolke, C.D. (2013). A yeast-based rapid prototype platform for gene control elements in mammalian cells. Biotechnol Bioeng *110*, 1201-1210.

Wei, K.Y., and Smolke, C.D. (2015). Engineering dynamic cell cycle control with synthetic small molecule-responsive RNA devices. J Biol Eng *9*, 21.

Weigand, J.E., Sanchez, M., Gunnesch, E.-B., Zeiher, S., Schroeder, R., and Suess, B. (2008). Screening for engineered neomycin riboswitches that control translation initiation. RNA *14*, 89-97.

Weigand, J.E., and Suess, B. (2007). Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast. Nucleic Acids Res *35*, 4179-4185.

Weinberg, Z., Barrick, J.E., Yao, Z., Roth, A., Kim, J.N., Gore, J., Wang, J.X., Lee, E.R., Block, K.F., Sudarsan, N.*, et al.* (2007). Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. Nucleic Acids Res *35*, 4809-4819.

Wickiser, J.K., Winkler, W.C., Breaker, R.R., and Crothers, D.M. (2005). The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. Mol Cell *18*, 49-60.

Wieland, M., Berschneider, B., Erlacher, M.D., and Hartig, J.S. (2010). Aptazyme-Mediated Regulation of 16S Ribosomal RNA. Chem Biol *17*, 236-242.

Wilson, R.C., and Doudna, J.A. (2013). Molecular mechanisms of RNA interference. Annu Rev Biophys *42*, 217-239.

Win, M.N., and Smolke, C.D. (2008). Higher-order cellular information processing with synthetic RNA devices. Science *322*, 456-460.

Winkler, W., Nahvi, A., and Breaker, R.R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature *419*, 952-956.

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

Wittmann, A., and Suess, B. (2011). Selection of tetracycline inducible self-cleaving ribozymes as synthetic devices for gene regulation in yeast. Mol Biosyst *7*, 2419-2427.

Wright, A.V., Nuñez, J.K., and Doudna, J.A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell *164*, 29-44.

Wu, M.-C., Lowe, P.T., Robinson, C.J., Vincent, H.A., Dixon, N., Leigh, J., and Micklefield, J. (2015). Rational Re-engineering of a Transcriptional Silencing PreQ1 Riboswitch. J Am Chem Soc *137*, 9015-9021.

Yang, J., Seo, S.W., Jang, S., Shin, S.I., Lim, C.H., Roh, T.-Y., and Jung, G.Y. (2013). Synthetic RNA devices to expedite the evolution of metabolite-producing microbes. Nat Commun *4*, 1413.

You, M., Litke, J.L., and Jaffrey, S.R. (2015). Imaging metabolite dynamics in living cells using a Spinach-based riboswitch. Proc Natl Acad Sci USA *112*, E2756-2765.

Zhou, H., Zheng, C., Su, J., Chen, B., Fu, Y., Xie, Y., Tang, Q., Chou, S.-H., and He, J. (2016). Characterization of a natural triple-tandem c-di-GMP riboswitch and application of the riboswitch-based dual-fluorescence reporter. Sci Rep *6*, 20871.

Zhou, L.B., and Zeng, A.P. (2015a). Engineering a Lysine-ON Riboswitch for Metabolic Control of Lysine Production in Corynebacterium glutamicum. ACS Synth Biol *4*, 1335-1340.

Zhou, L.B., and Zeng, A.P. (2015b). Exploring Lysine Riboswitch for Metabolic Flux Control and Improvement of L-Lysine Synthesis in Corynebacterium glutamicum. ACS Synth Biol *4*, 729-734.