



Published in final edited form as:

*Org Biomol Chem.* 2012 December 14; 10(46): 9113–9129. doi:10.1039/c2ob26724a.

## The bacterial second messenger c-di-GMP: Probing interactions with protein and RNA binding partners using cyclic dinucleotide analogs

Carly A. Shanahan<sup>†</sup> and Scott A. Strobel<sup>†,‡,\*</sup>

<sup>†</sup>Department of Chemistry, Yale University, New Haven, Connecticut 06520

<sup>‡</sup>Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520

### Abstract

The ability of bacteria to adapt to a changing environment is essential for their survival. One mechanism used to facilitate behavioral adaptations is the second messenger signaling molecule bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP). c-di-GMP is widespread throughout the bacterial domain and plays a vital role in regulating the transition between the motile planktonic lifestyle and the sessile biofilm forming state. This second messenger also controls the virulence response of pathogenic organisms and is thought to be connected to quorum sensing, the process by which bacteria communicate with each other. The intracellular concentration of c-di-GMP is tightly regulated by the opposing enzymatic activities of diguanylate cyclases and phosphodiesterases, which synthesize and degrade the second messenger, respectively. The change in the intracellular concentration of c-di-GMP is directly sensed by downstream targets of the second messenger, both protein and RNA, which induce the appropriate phenotypic response. This review will summarize our current state of knowledge of c-di-GMP signaling in bacteria with a focus on protein and RNA binding partners of the second messenger. Efforts towards the synthesis of c-di-GMP and its analogs are discussed as well as studies aimed at targeting these macromolecular effectors with chemically synthesized cyclic dinucleotide analogs.

### Second messenger signaling in bacteria

Signal transduction is used by organisms from all domains of life to sense changes in the environment and translate these cues into metabolic, physiological, and behavioral adaptations<sup>1</sup>. These regulatory networks are complex and rely on the coordination of many different interacting components to induce the appropriate biological response. One vital component of many signaling pathways are intracellular small molecules known as second messengers<sup>2,3</sup> (Figure 1). Second messengers relay external stimuli (i.e. the first message) received by cell-surface receptors to macromolecular effectors within the cell. The intracellular concentration of the second messenger is tightly controlled by proteins that synthesize or degrade the small molecule in response to specific extracellular inputs. These changing second messenger levels are sensed by specific effector molecules, which, in response to second messenger binding, act to induce the appropriate physiological output<sup>2,3</sup>. The behavioral responses induced by second messengers are often critical for survival, underscoring the crucial role of these small molecules in signal transduction.

\*To whom correspondence should be addressed. Phone (203) 432-9772. Fax: (203) 432-5767. scott.strobel@yale.edu.

For many years, it has been known that bacteria, and also eukaryotes, utilize the nucleotide based second messengers 3'-5'-cyclic adenosine monophosphate (cAMP) and guanosine tetraphosphate/pentaphosphate ((p)ppGpp) for signaling<sup>3-6,3</sup> (Figure 1). And while the role of the well known eukaryotic second messenger 3'-5'-cyclic guanosine monophosphate (cGMP) in bacterial signaling has been controversial, recent evidence suggests that cGMP may function to regulate prokaryotic development<sup>7</sup>. In bacteria, cAMP primarily regulates catabolite repression, glucose sensing, and the usage of alternative carbon sources<sup>4</sup>, whereas (p)ppGpp is produced in response to cellular stress and starvation conditions<sup>5,6</sup>. Both of these second messengers interact with a single class of transcription factors, the cAMP receptor protein (CRP) for cAMP and RNA polymerase for (p)ppGpp, which allosterically regulates the transcription of target genes to facilitate the appropriate behavioral adaptations<sup>3</sup>. Thus, cAMP and (p)ppGpp control a diverse set of biological processes primarily through transcriptional regulation.

In recent years, the scope of bacterial nucleotide based second messengers has expanded well beyond that of cAMP and (p)ppGpp with the realization that dinucleotide signaling molecules are also prevalent within this domain of life<sup>7</sup>. Bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) was the first of this class to be recognized as a global second messenger<sup>8,9</sup> (Figure 1). Similar to cAMP and (p)ppGpp, c-di-GMP regulates many different biological processes and one of its most crucial functions is to mediate the transition between the motile, planktonic lifestyle and the sessile, biofilm forming state<sup>10-14</sup>. In addition, the related dinucleotide bis-(3'-5')-cyclic dimeric adenosine monophosphate (c-di-AMP) has also gained attention as another potentially global bacterial second messenger involved in controlling DNA damage response, cell wall homeostasis, and membrane stress<sup>15,16</sup> (Figure 1). These recent advances in uncovering new bacterial second messengers have raised important questions related to their overall role in signal transduction, how bacteria regulate these multiple pathways in parallel, the specific mechanisms by which these small molecules induce many different physiological responses and how to design small molecule probes of these pathways to function as chemical tools and novel therapeutics. This review will focus specifically on the c-di-GMP signaling pathway, recent advances in the synthesis of second messenger analogs, and the use of these analogs to target c-di-GMP binding partners.

## **c-di-GMP is a ubiquitous bacterial second messenger that regulates diverse biological processes**

### **The enzymes that synthesize and degrade c-di-GMP**

c-di-GMP was first discovered during the late 1980's in the lab of Moshe Benziman who demonstrated that it allosterically activated the membrane bound enzyme cellulose synthase in *Gluconoacetobacter xylinus*<sup>17</sup>. Furthermore, Benziman and coworkers proposed a model for the regulation of cellulose production in this organism via the enzymatic synthesis and degradation of c-di-GMP based on the ability of purified membrane extracts to metabolize this dinucleotide<sup>17</sup>. However, it wasn't until about a decade later that the specific protein domains responsible for c-di-GMP turnover were identified<sup>18</sup> and found to be widely distributed throughout the bacterial domain<sup>19</sup>. This suggested a much broader role for this small molecule in regulating bacterial behavior and physiology<sup>19,20</sup>. c-di-GMP is now recognized as a global bacterial second messenger signaling molecule that mediates a wide variety of biological processes which extend well beyond its initially characterized role in cellulose production.

c-di-GMP is synthesized inside the cell from two molecules of GTP by GGDEF domain diguanylate cyclases (DGCs)<sup>21,22</sup> (Figure 2). It is degraded primarily by EAL domain

phosphodiesterases (PDEs) to the linear dinucleotide pGpG<sup>23-25</sup> (Figure 2). A second, less common class of c-di-GMP phosphodiesterases has also been identified that display no sequence similarity to the EAL domain PDEs known as the HD-GYP domain PDEs<sup>26,27</sup>. This class of PDEs degrades c-di-GMP to two molecules of GMP through a pGpG intermediate<sup>27</sup>. These c-di-GMP metabolizing enzymes are named for the conserved amino acids in the active site that are critical for catalysis and therefore the ability to regulate the intracellular concentration of the second messenger<sup>3</sup>. GGDEF domain DGCs often contain a secondary site of the sequence RxxD, referred to as the I-site, that functions as a high affinity c-di-GMP binding motif and controls DGC activity through product feedback inhibition<sup>28,29</sup>. GGDEF, EAL and HD-GYP domains are widely found throughout bacterial genomes suggesting that c-di-GMP is utilized as a second messenger by many bacterial species to control behavioral outcomes<sup>19</sup>. In contrast to cAMP and (p)ppGpp which are used by both prokaryotes and eukaryotes for signaling<sup>3</sup>, the GGDEF, EAL and HD-GYP domains are absent from the genomes of archaea and eukaryotes suggesting that c-di-GMP is a bacterial specific regulatory molecule<sup>19,30</sup>.

### Biological processes regulated by c-di-GMP

Analogous to cAMP and (p)ppGpp signaling, one of the hallmark themes of c-di-GMP signaling is the remarkable diversity of physiological processes in which this second messenger plays a regulatory role. c-di-GMP is often referred to as the 'lifestyle molecule' because one of its primary functions is to mediate the transition between the sessile, biofilm forming state and the motile, planktonic lifestyle by inversely regulating the expression of genes necessary for motility and cell surface adhesion factors<sup>3,10,13,31</sup>. It is now well established that high intracellular concentrations of c-di-GMP promote biofilm formation whereas low intracellular concentrations stimulate motility<sup>31</sup> (Figure 2). However, before it was directly demonstrated that GGDEF and EAL domain proteins synthesized and degraded c-di-GMP, the realization that this dinucleotide was a global second messenger emerged from genetic studies linking diverse biological phenotypes such as rugose colony formation, biofilm matrix development, virulence, and twitching motility to these protein domains<sup>20</sup>.

In addition to its prevalent role in regulating bacterial lifestyle processes, the importance of this second messenger in controlling bacterial pathogenesis is also now well established<sup>30,32-34</sup>. *Vibrio cholerae* was the first pathogenic organism for which a link between c-di-GMP and virulence response was demonstrated<sup>32</sup>. Low intracellular levels of c-di-GMP, promoted by the EAL domain phosphodiesterase VieA, were shown to enhance the production of cholera toxin, the virulence factor responsible for colonization of the mammalian intestine during cholera infection. Mutation of VieA to abolish its PDE activity decreased expression of cholera toxin and reduced the ability of the bacterium to colonize the intestine of a mouse model of cholera, suggesting that c-di-GMP repressed the transcription of virulence factors<sup>32</sup>. The critical function of c-di-GMP in regulating virulence response has now been demonstrated for other known human pathogens, including the causative agents of gastroenteritis<sup>33</sup> and Lyme disease<sup>33,34</sup>.

This second messenger also regulates other biological processes including cell differentiation, cell morphology, cell-cell communication, exopolysaccharide matrix production, expression of virulence factors, flagellum biosynthesis, and pilus assembly<sup>9</sup>. c-di-GMP has also been linked to quorum sensing, the process by which bacteria sense and communicate with each other<sup>35,36</sup>, and the cAMP signaling pathway<sup>37</sup>. Key components of both cAMP signaling and quorum sensing have been shown to alter the expression of genes encoding GGDEF and EAL domain proteins<sup>37,38</sup>, underscoring the complexity of c-di-GMP-mediated signaling.

## Regulation of c-di-GMP biosynthesis by GGDEF and EAL domain proteins

For c-di-GMP to trigger the appropriate behavioral responses, the intracellular concentration of this second messenger must be tightly controlled by the opposing enzymatic activities of DGCs and PDEs. However, one of the more puzzling aspects relating to the distribution of GGDEF and EAL domain proteins throughout bacterial genomes is the duplication of these enzymes within a single organism<sup>10,14</sup>. For example, *Escherichia coli* encodes 17 EAL domains and 19 GGDEF domains, while *Vibrio vulnificus* is predicted to have 33 EAL domains and 66 GGDEF domains<sup>9</sup>. This observation raises the important question of how the enzymatic activities of these proteins with identical function are regulated in parallel. In addition, many GGDEF-EAL dual domain proteins exist where one of the domains is usually catalytically inactive and instead possesses a regulatory function<sup>10,39,40</sup>. However, bi-functional GGDEF-EAL composite proteins that possess both cyclase and phosphodiesterase activity exist<sup>41,42</sup>. This observation raises the related question of how both enzymatic activities are regulated within a single protein. Bacteria have evolved two distinct mechanisms for controlling the activity of GGDEF and EAL domain proteins. The first is by linking these domains to additional regulatory domains that sense specific signals and control GGDEF and/or EAL function in response, and the second is by localizing these proteins to specific parts of the cell and inducing expression only at certain times<sup>10</sup>.

### Signal sensing and regulatory accessory domains

Based on genomic sequencing, it is estimated that 50 to 70% of all GGDEF and EAL domain proteins contain an additional accessory domain that functions to either sense a specific environmental signal or produce an output<sup>40</sup>. This facilitates the direct coupling of an external or intracellular signal input to either c-di-GMP synthesis or degradation<sup>10-13</sup>. A wide variety of either integral membrane or cytoplasmic sensory domains are linked to GGDEF and EAL domain proteins, which can sense signals such as light, gases (NO and O<sub>2</sub>), phosphorylation, redox conditions and binding of proteins, DNA and small molecule ligands<sup>9-11</sup>. In addition, a large majority of GGDEF-EAL dual domain proteins predicted to possess both enzymatic activities have an added regulatory domain that is thought to control the selection between these two activities<sup>40</sup>. However, the specific molecular details of how the opposing enzymatic activities of many of these bifunctional proteins are regulated within the cell have yet to be determined. Collectively, these observations suggest that DGCs and PDEs are functionally diversified such that numerous environmental and intracellular signals can be integrated into the c-di-GMP regulatory network<sup>9,10</sup>.

### Spatial and temporal control of c-di-GMP signaling modules

Another strategy employed by bacteria to manage the redundancy of GGDEF and EAL domain proteins is spatial sequestration of these c-di-GMP signaling components<sup>10,12</sup>. DGC and PDE enzymes control the local concentration of c-di-GMP, which is then only available to a specific, local target molecule<sup>12</sup>. Because specific PDEs and DGCs are associated with different phenotypic outputs, the co-localization of both c-di-GMP metabolic enzymes and receptors to specific parts of the cell is a useful strategy for delegating specific c-di-GMP signaling systems to carry out particular cellular tasks<sup>12,43</sup>.

Temporal regulation is also utilized by bacteria to coordinate the catalytic functions of the numerous GGDEF and EAL domains encoded in a single genome<sup>10</sup>. In this case, both the cellular expression levels and activities of these proteins are expected to change in response to specific environmental or intracellular stimuli at discreet times. Consistent with this idea, a recent study in *Escherichia coli* looking at the expression levels of all the GGDEF and EAL domain proteins under varying growth conditions demonstrated that these enzymes were differentially expressed in response to changes in temperature, growth phase, and

media<sup>44</sup>. The observation that the majority of GGDEF and EAL domains are linked to signal sensing domains is also consistent with this model, as this domain arrangement implies that certain proteins are not constitutively active, but are instead only activated in response to specific stimuli.

The spatial and temporal sequestration of c-di-GMP signaling components are not likely to be mutually exclusive. This is best demonstrated by the role of the specific DGC PleD from *Caulobacter crescentus* in mediating the transition from a motile swarmer cell, to a sessile stalked cell<sup>9,22,43</sup>. Its ability to synthesize c-di-GMP is dependent on an intracellular signal (temporal) whereas its physiological output requires it to be localized to a specific position in the cell to properly exert its function (spatial). It is clear that a great deal of complexity is involved in organizing c-di-GMP signaling within a particular organism.

## Macromolecular targets of c-di-GMP

As a second messenger, c-di-GMP directly interacts with macromolecular targets to communicate the extracellular signals received by cell surface receptors (Figure 2, Table 1). However, one of the greatest challenges in understanding the mechanisms of c-di-GMP signaling proves to be the identification of the receptor molecules that recognize and bind this second messenger. Unlike the second messengers cAMP and (p)ppGpp, which both act by primarily binding a single protein receptor<sup>3</sup>, the macromolecules that sense c-di-GMP are highly diverse and cannot be identified by a single binding site<sup>13,45</sup>. In addition to several classes of protein receptors that bind c-di-GMP, RNA receptors of the second messenger have also been identified as part of this signaling pathway (Table 1).

## Proteins that recognize and bind c-di-GMP

The first class of macromolecules found to directly bind c-di-GMP was the PilZ protein domain, which was originally thought to be the long sought after c-di-GMP binding partner<sup>46,47</sup> (Table 1). In accordance with this initial impression, the first protein identified to be regulated by c-di-GMP, the cellulose synthase protein BcsA from *Gluconoacetobacter xylinus*, contains a PilZ domain<sup>46,48</sup>. In addition, the phylogenetic distribution of these domains was similar to that of the GGDEF and EAL domain proteins<sup>46,47,49</sup>. The PilZ domain proteins are currently the most commonly occurring protein for c-di-GMP signaling with nearly 4000 sequences (Pfam Search: PF07238) identified in approximately 1100 bacterial species. In enterobacteria, PilZ domains are often found C-terminal to the protein domain YcgR (Pfam PF07317), which interacts with additional proteins that control flagellar rotation and motility<sup>47,50,51</sup>. In *Vibrio cholerae*, these c-di-GMP binding proteins have been implicated in biofilm formation, motility, and virulence<sup>49</sup>, and for *Pseudomonas aeruginosa*, in alginate biosynthesis<sup>52</sup>. However, not all species that utilized this second messenger for signaling encoded a PilZ domain in their genome, suggesting that other effectors must exist<sup>46</sup>. As predicted, additional classes of c-di-GMP binding proteins have now been identified that include transcription factors<sup>53–58</sup>, degenerate GGDEF and EAL domain proteins<sup>59–63</sup>, and I-site effectors<sup>64</sup>.

Thus far, five distinct classes of transcription factors have been identified that directly bind c-di-GMP and regulate the expression of target genes (Table 1). FleQ from *Pseudomonas aeruginosa* was the first transcription factor identified to directly interact with c-di-GMP<sup>53</sup>. FleQ inversely regulates the transcription of genes necessary for exopolysaccharide (EPS) matrix production and flagella biosynthesis<sup>53</sup>. Approximately 200 homologs of FleQ (Pfam PF06490) in 200 different bacterial species have been found, but their ability to bind the second messenger and regulate transcription in a c-di-GMP dependent manner remains to be verified.

In *Vibrio cholerae*, two classes of c-di-GMP binding transcription factors have been identified, VpsT and VpsR. VpsT upregulates the expression of EPS producing genes in a c-di-GMP dependent fashion<sup>54</sup>. VpsR was shown to be a transcriptional activator of biofilm formation<sup>57</sup>. c-di-GMP binding to VpsR activates expression of a protein known to both activate virulence gene expression and regulate the quorum sensing pathway at low cell density. Although the implication of this finding is not yet clear, it is thought that *Vibrio cholerae* integrates information from both quorum sensing and c-di-GMP signaling to appropriately adapt to its current environment<sup>57</sup>.

Another family of c-di-GMP binding transcription factors identified was Clp, which is a subset of the cyclic AMP receptor protein (CRP) super family. Clp is important for regulating genes involved in the pathogenesis of xanthomonads<sup>55,56</sup>. The cyclic nucleotide monophosphate (cNMP) binding domain of CRP contains six highly conserved amino acid residues that confer specificity for cAMP recognition. However, the nucleotide binding site of Clp differs from that of CRP by a single amino acid substitution, which defines its specificity for c-di-GMP<sup>56</sup>.

More recently, MrkH from *Klebsiella pneumoniae* has also been identified as a c-di-GMP dependent transcriptional regulator<sup>58</sup>. In the presence of c-di-GMP, MrkH activates the expression of genes encoding fimbriae production during biofilm formation. MrkH is distinct from other known c-di-GMP binding transcription factors because it contains a C-terminal PilZ domain which is hypothesized to be the second messenger binding site<sup>58</sup>.

The last general class of c-di-GMP binding proteins that have been identified to date are degenerate variants of the GGDEF and EAL domain proteins that synthesize and degrade the second messenger (Table 1). These proteins possess mutations to one or more of the conserved active site residues and are no longer catalytically active, but they retain the ability to bind c-di-GMP and instead function as receptors of the second messenger<sup>10,12,13,45</sup>. Degenerate EAL domain proteins bind c-di-GMP at the active site<sup>59,65</sup> whereas degenerate GGDEF proteins bind c-di-GMP at the regulatory I-site<sup>61</sup>. The I-site effectors are related to the degenerate GGDEF effectors because the c-di-GMP binding site of these proteins is homologous to the I-site, yet the I-site effectors lack a GGDEF site<sup>64</sup>. The only validated example of an I-site effector to date is the PelD protein from *Pseudomonas aeruginosa*, which controls PEL polysaccharide synthesis during pellicle formation<sup>64</sup>. Although sequence analysis can readily predict degenerate GGDEF and EAL domain proteins, the inability of these variants to metabolize c-di-GMP and their specific biological function must be verified experimentally.

Only a few examples of this class of c-di-GMP effectors have been validated. Degenerate GGDEF domain proteins include PopA from *Caulobacter crescentus*<sup>61</sup>, CdgG from *Vibrio cholerae*<sup>63</sup>, and SgmT from *Myxococcus xanthus*<sup>62</sup>. For both PopA and SgmT, c-di-GMP binding induces spatial sequestration of the protein<sup>61,62</sup>. PopA has been implicated in cell cycle control via c-di-GMP dependent degradation of an inhibitor of replication initiation<sup>61</sup> and SgmT is a histidine kinase that phosphorylates DigR, a DNA binding regulator of extracellular matrix production<sup>62</sup>. However, the functional importance of c-di-GMP dependent sequestration of SgmT is currently unknown. Although the ability of CdgG to bind c-di-GMP has not been confirmed, this protein has no DGC activity yet requires an intact I-site to regulate rugosity, suggesting that second messenger binding is important for its function<sup>63</sup>. The degenerate EAL domain proteins that have been identified are FimX<sup>59</sup> and LapD<sup>60</sup> from *Pseudomonas aeruginosa* and *Pseudomonas fluorescens*, respectively. FimX and LapD have been shown to control very different processes, with the former important for twitching motility<sup>66</sup> and the latter required for biofilm formation<sup>60</sup>. This

demonstrates the diversity of biological phenotypes regulated by these degenerate c-di-GMP metabolizing enzymes.

Collectively, the PilZ domain, degenerate GGDEF/EAL domain proteins, transcription factors, and I-site effectors have affinities for c-di-GMP that range from approximately 50 nM into the low  $\mu\text{M}$  range<sup>46,48,49,53–55,57,59–61,64</sup>. Despite the progress made towards elucidating the downstream targets of this second messenger, the specific mechanisms through which c-di-GMP interacts with many of these proteins to induce a physiological output remain largely obscure. Given the diversity of protein targets already identified, it is likely that additional classes of protein receptors for this second messenger remain to be discovered.

### RNA molecules that recognize and bind c-di-GMP

In addition to the proteins that bind c-di-GMP, two classes of riboswitch RNAs, termed class I and class II, have also been identified as part of this second messenger signaling pathway<sup>67,68</sup> (Table 1). Riboswitches are non-coding regulatory RNAs consisting of an aptamer domain and expression platform that recognize small molecule ligands with high affinity and specificity (for recent reviews see<sup>69–72</sup>). Over 20 different riboswitch classes have been identified to date that recognize diverse ligands including amino acids, nucleobases, modified sugars, vitamins and coenzymes, and even small anions. Ligand binding to the aptamer domain causes structural changes in the expression platform that lead to changes in the expression levels of the downstream genes in the absence of any protein cofactors, most often by affecting either transcription or translation. Although riboswitches are now recognized as a major form of bacterial genetic control, the c-di-GMP binding riboswitches represent the first known class of RNAs that directly bind a second messenger signaling molecule and function as a downstream effector in an intracellular signaling pathway<sup>67,68</sup>.

Approximately 500 examples of class I<sup>67</sup> and 45 examples of class II riboswitches<sup>68</sup> have been identified throughout the bacterial domain in many different species. In addition, these riboswitches are present in many prominent human pathogens, including the causative agents of anthrax and cholera. Although the use of the class I riboswitch for gene expression is more widespread, several organisms exclusively use the class II riboswitch for signaling, namely those of the class Clostridia<sup>68</sup>. Consistent with their predicted role in a second messenger signaling pathway, these aptamers are found upstream of a wide collection of genes that regulate cellular processes in which c-di-GMP is now known to play a role, including motility, pilus formation, virulence factor expression, flagellum biosynthesis, and exopolysaccharide matrix production<sup>67,68</sup>.

The affinities of these riboswitches for c-di-GMP are tighter than for any of the known protein receptors and range from as low as 10 pM for the class I riboswitch<sup>73</sup> to between 200 pM and 2 nM for the class II riboswitch<sup>68</sup>. For the Vc2 class I riboswitch from *Vibrio cholerae*, it has been demonstrated that the tight binding affinity (10 pM) results from very slow rates of ligand release ( $k_{\text{on}} = 1 \times 10^6 \text{ M}^{-1} \text{ min}^{-1}$  and  $k_{\text{off}} = 1 \times 10^{-5} \text{ min}^{-1}$ )<sup>73</sup>. Thus, the half life of the c-di-GMP bound complex is approximately 1 month, indicating that ligand binding is effectively irreversible on a biological timescale. Therefore, this riboswitch is thought to be kinetically controlled, where gene regulation is governed entirely by the on-rate and is highly responsive to changing intracellular levels of the second messenger<sup>73,74</sup>. However, the kinetic properties of other class I riboswitch sequences and that of the class II riboswitch has not been fully explored.

The identification of riboswitches as c-di-GMP receptors is of particular interest because unlike protein effectors, ligand binding is directly coupled to gene regulation, suggesting a

mechanism for how c-di-GMP induces a cellular response upon binding this class of effectors. Based upon sequence analysis, transcriptional terminator stems can easily be identified for a subset of class I and class II riboswitches, suggesting that c-di-GMP binding affects the expression levels of downstream genes through a transcriptionally controlled mechanism. However, this has only been experimentally validated for a few riboswitches. One example of a transcriptionally controlled switch is the Cd1 class I riboswitch from *Clostridium difficile*<sup>67</sup>. This riboswitch is located upstream of the operon encoding a large subset of the proteins and machinery necessary for flagellum biosynthesis in this organism. In the presence of increasing concentrations of c-di-GMP, transcription is terminated. This is consistent with the observation that high intracellular concentrations of c-di-GMP repress motility. The class II riboswitch encoded by *Bacillus halodurans* (Bha-1-1) was also determined to be transcriptionally regulated because full length RNA was produced only in the presence of c-di-GMP<sup>68</sup>.

Translationally controlled riboswitches have also been identified. More specifically, *in vivo* reporter assays indicated that the Vc2 riboswitch up-regulated expression of the VC1722 gene in the presence of c-di-GMP, while the full-length RNA was transcribed independent of c-di-GMP<sup>67</sup>. Although the exact function of this gene is unknown, it is hypothesized to be a transcription factor and rugose *Vibrio cholerae* mutants exhibited elevated levels of this mRNA<sup>75</sup>. This suggests that increased intracellular levels of c-di-GMP upregulate expression of this protein. These second messenger riboswitches can function as genetic 'on' switches, where c-di-GMP binding increases expression levels of the downstream genes, or genetic 'off' switches independent of whether they are transcriptionally or translationally controlled<sup>67,68</sup>.

## Structural basis of second messenger recognition

As anticipated from the diversity of its receptor molecules, the specific mechanisms by which this second messenger induces physiological responses are highly variable. In contrast to cAMP and (p)ppGpp, which both act primarily at the transcriptional level through a single protein effector, c-di-GMP acts on protein and RNA targets at the transcriptional, translational, and post-translational levels to alter cellular behavior and physiology<sup>3,11,13</sup>.

### Recognition of c-di-GMP by protein receptors

Given that c-di-GMP plays a regulatory role in many different physiological processes, it was hypothesized that protein receptors of c-di-GMP must also be adaptable to functioning in different biological pathways<sup>48</sup>. To further investigate how proteins recognize c-di-GMP and how c-di-GMP binding induces different physiological responses mediated through these receptors, the crystal structures of the second messenger in complex with several protein receptors have been solved (Figure 3). Overall, these structures reveal that proteins employ different modes of recognition and can bind the second messenger as a monomer<sup>48,59,65,76</sup> or as an intercalated dimer<sup>54,76–80</sup>. In addition, proteins that recognize c-di-GMP as a monomer can bind the second messenger in either an eclipsed conformation<sup>48,76</sup>, where the guanine bases are aligned overtop one another, or in an extended conformation where the guanine bases are splayed apart<sup>59,65</sup>. PilZ domain proteins have been shown to recognize both the monomer and dimer form of c-di-GMP<sup>48,76,77</sup>, demonstrating that different approaches to second messenger recognition are utilized by effectors of the same class.

Common c-di-GMP recognition strategies used by proteins include hydrogen bonding contacts with charged or polar side chains and base stacking interactions with aromatic residues. Cation- $\pi$  interactions are also expected to play a role in recognition as evidenced



by charged residues that sometimes stack directly above or below one of the c-di-GMP bases. For proteins that bind c-di-GMP as a dimer, several interactions between the individual c-di-GMP monomers are also observed. The structures of unbound and c-di-GMP bound protein complexes often adopt different conformations. This suggests that ligand-induced conformational changes mediate the specific molecular mechanisms used by these protein effectors to induce biological outputs<sup>65,76,77,81,82</sup>. However, these c-di-GMP receptors often interact with other proteins to induce a biological output and many of the molecular mechanisms of these downstream interactions remain unknown.

The PilZ domain protein PP4397 from *Pseudomonas putida* binds the second messenger as an intercalated dimer where both c-di-GMP molecules are in the eclipsed conformation<sup>76</sup> (Figure 3a). This allows for base stacking interactions to occur between one guanine base from each monomer, which are arbitrarily termed  $G_{\alpha}$  and  $G_{\beta}$  and labeled for each c-di-GMP unit. Recognition of the guanine bases in each individual monomer is asymmetric, however recognition of  $G_{\alpha 1}$  is similar to that of  $G_{\beta 2}$ , whereas that of  $G_{\alpha 2}$  matches the contacts made to  $G_{\beta 1}$ . Every atom with hydrogen bonding potential along the Watson-Crick and Hoogsteen faces of  $G_{\alpha 1}$  and  $G_{\beta 2}$  is contacted by an RNA atom. In contrast,  $G_{\alpha 1}$  and  $G_{\beta 2}$  are only contacted by an arginine residue on their Hoogsteen face. In addition, arginine residues stack with two of the four guanine bases and non-bridging phosphate oxygens hydrogen bond with guanine exocyclic amines or other charged and polar side chains. Overall, recognition of dimeric c-di-GMP by PilZ domain proteins is similar to second messenger recognition and binding by diguanylate cyclases at the inhibitory I-site<sup>29,79,80</sup>. In contrast, the PilZ domain protein VCA0042 (PDB ID 3KYG) from *Vibrio cholerae* recognizes c-di-GMP as a monomer in the eclipsed conformation<sup>48,76</sup> (Figure 3b). The top base ( $G_{\alpha}$ ) is more heavily recognized than the bottom base ( $G_{\beta}$ ) with contacts made to both its Watson-Crick and Hoogsteen faces. Similar to the base stacking interactions observed for PP4397, an arginine residue also stacks with one of the guanine bases,  $G_{\alpha}$ , in this protein.

While the PilZ domain proteins typically recognize c-di-GMP in the eclipsed conformation, the degenerate EAL domain protein FimX from *Pseudomonas aeruginosa* binds the second messenger in the extended conformation<sup>59</sup> (Figure 3c). Base recognition is achieved primarily through electrostatic interactions with charged side chains and stacking contacts with aromatic residues.  $G_{\alpha}$  is less heavily recognized than  $G_{\beta}$  and only one of the c-di-GMP phosphates is contacted. While these c-di-GMP binding proteins utilize similar contacts for recognition, the second messenger binding pockets can be quite diverse.

### Recognition of c-di-GMP by riboswitches

The identification of c-di-GMP binding riboswitches defines a new role for RNA as part of a second messenger signaling pathway. To understand the molecular basis of RNA recognition of c-di-GMP, the crystal structures of the second messenger bound to the aptamer domains of both the class I and class II riboswitch were determined by our laboratory and others<sup>73,83–85</sup> (Figure 4). As anticipated from secondary structure analysis, these RNAs adopt unique conformations and employ different strategies for ligand recognition. The class I riboswitch adopts a y-shape with c-di-GMP bound at the three helix junction incorporated as part of a duplex<sup>73,83</sup>, whereas class II folds into a more compact structure and incorporates the second messenger into a triplex at the edge of the pseudoknot region<sup>85</sup>. The ability of these riboswitches to incorporate the second messenger into the RNA structure likely contributes to the tighter binding affinities for these effectors as compared to protein receptors.

In the class I riboswitch, c-di-GMP binds at the three helix junction formed by P1, P2, and P3 with contacts to the ligand made by three critical nucleotides (G20, C92, and A47)<sup>83</sup>. The guanine bases, arbitrarily termed  $G_{\alpha}$  and  $G_{\beta}$  are recognized asymmetrically in canonical

pairings, with  $G_{\alpha}$  forming a Hoogsteen pair with G20 and  $G_{\beta}$  in a Watson-Crick pair with C92 (Figure 4b,c). Stacking between the two guanine bases is a highly conserved adenosine (A47) that serves to bridge the P1 and P2 helices with the c-di-GMP ligand resulting in extensive base stacking interactions (Figure 4a). The ribosyl-phosphate backbone is also extensively contacted, with recognition mediated by both hydrogen bonding and metal-phosphate contacts (Figure 4d).

In contrast to recognition by the class I riboswitch, the class II riboswitch also recognizes the guanine bases of c-di-GMP asymmetrically, however no canonical base pairings are observed and in general, fewer base contacts to the bases are made by binding pocket nucleotides<sup>85</sup> (Figure 4f, g). Instead,  $G_{\alpha}$  is recognized as part of a base triple with A69 and U37, and  $G_{\beta}$  is contacted by hydrogen bonds from RNA residues A70 and G73, as well as by a hydrated magnesium. Three conserved adenosine nucleotides that stack between (A70), above (A61) and below (A13) the bases of c-di-GMP indicate that base stacking is also important for ligand binding (Figure 4e).

Between the two riboswitches, there are four different solutions to guanine recognition by the RNA aptamer<sup>83,85</sup> (Figure 4b,c,f,g). Stacking interactions are the only parallels in ligand recognition between the two riboswitch classes<sup>85</sup> (Figure 4a,e). Minimal contacts are made to the ribosyl-phosphate backbone by the class II riboswitch with only a single hydrogen bond observed to one non-bridging phosphate oxygen (Figure 4e). It is clear that at least two independent molecular solutions for RNA recognition of the same second messenger ligand exist.

The mechanisms evolved for c-di-GMP recognition by RNA differ significantly from those utilized by protein receptors. Unlike proteins, RNA is able to incorporate the second messenger into structural elements, which likely contributes to the tight binding affinities observed for this class of receptors<sup>83,85</sup>. While both proteins and RNA employ stacking as a mechanism for second messenger binding, the stacking interactions made between the guanine bases of c-di-GMP and purine residues in the riboswitch binding pocket appear to be more extensive than those made with aromatic side chains of proteins. Although, the specific contacts made by the class I and class II riboswitches to c-di-GMP greatly differ, the ligand is bound in the same eclipsed conformation. This is in contrast to proteins, which can bind c-di-GMP in many different conformations, suggesting that the eclipsed conformation is functionally relevant for RNA recognition. Collectively, these observations indicate that the diverse receptors of c-di-GMP also employ diverse mechanisms for second messenger recognition and binding.

## **Chemical tools to study and target the c-di-GMP signaling pathway: Synthesis of c-di-GMP and second messenger analogs**

c-di-GMP is used by many prominent human pathogens to control biological processes and the lack of this second messenger in eukaryotes suggests that, c-di-GMP binding partners could be attractive new targets for the development of therapeutics against bacterial infection. Thus, the ability to synthesize c-di-GMP and second messenger analogs is useful for increasing our understanding of how macromolecular effectors recognize c-di-GMP and how to effectively design second messenger analogs to target the c-di-GMP receptors that function downstream in this signaling pathway. Towards this goal, several groups have developed chemical methods for synthesizing cyclic dinucleotide analogs of the second messenger. Many of these synthetic efforts have been undertaken with the ultimate goal of using these molecules to manipulate the biological processes regulated by c-di-GMP through the selective modulation of RNA and protein second messenger binding partners.

## Chemical synthesis of c-di-GMP and nucleotide analogs of the second messenger

The first chemical method for synthesizing c-di-GMP was reported in 1990 by the laboratories of Benziman and van Boom<sup>86</sup>. This method utilized phosphotriester chemistry to access c-di-GMP via the linear dinucleotide GpGp. Although producing pure c-di-GMP, this strategy suffered from low product yields due to difficulty in efficiently synthesizing the GpGp intermediate as well as product degradation upon multiple deprotection steps, some which required the use of harsh basic conditions<sup>86,87</sup>. Since then, and mostly in the past decade, several novel synthetic strategies have been developed that produce c-di-GMP in greater yields and can also be utilized to synthesize derivatives of the second messenger.

Hayakawa *et al.* utilized a phosphoramidite approach to make the linear dinucleotide GpGp intermediate, which overcomes the limitations of the first chemical method for c-di-GMP production<sup>87</sup> (Figure 5a). Using allyl-based protecting groups (allyl= All; alloxycarbonyl= AOC) on the phosphate and guanine base of the starting phosphoramidite monomer, oxidation in the presence of 2-cyanoethanol yielded the fully protected guanosine phosphate derivative. After 5'-deprotection, this building block was condensed with a second guanosine phosphoramidite monomer, producing the linear intermediate GpGp in much higher yield than previously observed<sup>86</sup>. This intermediate was then directly used in the cyclization reaction. Use of allyl protecting groups for both the guanine bases and the phosphates allowed for deprotection in a single step under much milder conditions, resulting in overall increased product yield. Hyodo and Hakayawa further improved this method by utilizing a guanosine phosphoramidite building block with different base and phosphate protecting groups<sup>88</sup> (Figure 5b). By employing a dimethylformamide (dmf) group for protection of the guanine exocyclic amine while leaving the O<sup>6</sup> unprotected the starting phosphoramidite monomer could be synthesized in fewer steps. In addition, the phosphate moiety was protected with a cyanoethyl group and then oxidized in the presence of allyl alcohol. After condensation to form the GpGp intermediate, the allyl group, was selectively removed prior to cyclization, leaving only the dimethylformamide and cyanoethyl protecting groups to be removed after cyclization, which could be accomplished in a single step<sup>88</sup>. Modification to the synthesis of the guanosine phosphoramidite monomer proved to increase overall yields beyond that of the first generation method. Furthermore, this method was readily adapted to afford base modified second messenger analogs containing a single inosine or adenine base in place of one guanine by using the appropriate base protecting groups compatible with the deprotection conditions for the dmf group<sup>89</sup>. A single phosphorothioate analog was also obtained using this method by replacing the oxidation reagent with a sulfuration reagent for one oxidation step.

A modified version of the phosphoramidite approaches described above using an H-phosphonate cyclization procedure was developed by the Jones group<sup>90</sup>, providing yet another synthetic route for accessing c-di-GMP (Figure 5c). In this method, the 5'-DMTr, 2'-TBDMS, N<sup>2</sup>-isobutyl protected guanosine was converted to both the methyl-phosphoramidite and H-phosphonate, which were then coupled together. The H-phosphonate serves as a protecting group in the coupling step, yielding only a single GpGp intermediate and eliminating additional protection and deprotection steps that were necessary for the previously developed methods. After 5'-detritylation, cyclization is catalyzed by the addition of adamantoyl chloride followed by the final oxidation step to give the fully protected cyclic dinucleotide. This method is advantageous over the previously used strategies due to the elimination of several protection/deprotection steps, resulting in fewer synthetic steps overall.

In addition to the solution phase syntheses discussed above, two additional synthetic methodologies have also been developed that have proven to be useful for accessing c-di-GMP and its analogs. The first method, developed by Amiot *et al.*, employs a novel strategy

of glycosylation of a glycan precursor<sup>91</sup>. In this method, the glycan precursor first synthesized is the cyclized ribosyl-phosphate backbone absent any base structure. Glycosylation of this intermediate introduces the desired base to produce the cyclic dinucleotide molecule. The second strategy, developed by Yan *et al.*, utilizes a modified H-phosphonate approach along with the 1-(4-chlorophenyl)-4-ethoxypiperidin-4-yl (Cpep) group for protection of the ribose 2'-hydroxyl functionality<sup>92</sup>. Both methods have been demonstrated to yield pure c-di-GMP in good yield.

All the aforementioned c-di-GMP synthetic strategies are solution phase methods and therefore require the purification of various intermediates before formation of the final product. To further simplify the synthesis of c-di-GMP, Kiburu *et al.* developed a simple solid-phase synthesis for this second messenger that required purification of only the final, fully deprotected product<sup>93</sup> (Figure 6a). The authors used two different approaches: one in which the linear dinucleotide was cyclized on the solid support (Figure 6a) and a second in which the linear intermediate was cleaved from the solid support and cyclization was performed in solution. For on-bead cyclization, the 5'-DMTr, 2'-TBDMS, O-methyl guanosine phosphoramidite was coupled to the solid support followed by oxidation and 5'-deprotection. For the second coupling, the more commonly available cyanoethyl-phosphate protected phosphoramidite was used. This allowed for the selective deprotection of the methyl group, providing only one free phosphate oxygen for the subsequent cyclization reaction. Following cyclization, the cyclic dimer was cleaved from the bead, deprotected and purified by high-performance liquid chromatography (HPLC) (Figure 6a). In contrast, the solution cyclization method used O-methyl phosphoramidites for both coupling steps. Because the O-methyl group is inert to  $\beta$ -elimination, triethylamine can be used to cleave the linear dinucleotide from the solid support without removing any protecting groups that are necessary for successful cyclization. The development of this solid-phase approach for c-di-GMP synthesis was a great advancement because naturally occurring bases are available as O-methyl phosphoramidites, allowing several dinucleotide analogs to easily be synthesized using this approach.

Our lab has further modified this solid-phase synthetic approach such that the most common, commercially available starting phosphoramidites employing a 5'-DMTr, 2'-OTBDMS, and cyanoethyl phosphate protecting groups could be used for both coupling steps<sup>94</sup> (Figure 6b). Phosphoramidites were coupled to a solid support that is inert to  $\beta$ -elimination under mild basic conditions. This allowed us to introduce a phosphate deprotection step after the first coupling step, providing the free phosphate oxygen for subsequent cyclization, without cleaving the nucleotide from the bead. After the second coupling reaction, cyclization was carried out on bead. Using this approach, we successfully synthesized novel base and ribose modified second messenger analogs, including 2'-fluoro, 2'-methoxy, 7-deaza guanine, and N1-methyl guanosine derivatives in addition to others previously synthesized by different methods (inosine, adenosine, and 2'-deoxy), in reasonable yield for biochemical experiments (Figure 7a,b).

The ability to synthesize the second messenger and its derivatives on a large scale at low cost is necessary for biological studies aimed at further elucidating its molecular mechanisms of action and the development of high yielding synthetic procedures is a continuing aim. Towards this goal, the Jones group recently reported a synthesis of c-di-GMP and its phosphorothioate analogs that yields gram-scale quantities of these compounds via eight steps in a single flask<sup>95,96</sup> (Figure 5d). The phosphoramidite and H-phosphonate approaches are combined as previously reported<sup>90</sup>, however the reagents for each synthetic step are compatible with those used for subsequent steps, including the removal of all protecting groups. In addition, pure c-di-GMP can be isolated by crystallization, eliminating the need for column purification of reaction intermediates or the final product.

Phosphorothioate derivatives of c-di-GMP were also successfully synthesized by this method using a sulfurization reagent in place of the oxidation reagent and column purification of these compounds was only necessary to separate the different diastereomers<sup>95</sup>. Adapting this method for the synthesis of other c-di-GMP analogs is expected to allow for the production of these derivatives in large enough yield to perform biological studies and probe second messenger signaling *in vivo*.

### Targeting c-di-GMP binding proteins with second messenger analogs

The ability to selectively bind protein receptors could be a useful strategy for altering bacterial behavior and developing compounds with anti-biofilm or anti-virulence activity<sup>97</sup>. Although the targeting of c-di-GMP binding proteins with second messenger analogs has not been as extensively explored as it has for riboswitch effectors (discussed below), one general paradigm that has emerged from efforts to design c-di-GMP derivatives that target protein receptors is the use of conformational biasing. c-di-GMP can exist in primarily two conformations, the open state where the guanine bases are splayed apart or the closed state where the guanine bases are aligned over top one another in an anti-parallel fashion, and different protein receptors of the second messenger recognize different conformations (discussed above). This observation has been exploited by the Sintim group to selectively target c-di-GMP binding proteins that recognize only one conformation of the second messenger using an analog with an altered conformer population relative to that of the native second messenger<sup>97</sup>.

The Sintim group replaced one of the bridging oxygens of a single c-di-GMP phosphate linkage with sulfur and demonstrated both computationally and experimentally that the resulting analog, termed endo-S-c-di-GMP, favored the open conformation in solution and was three times less likely than c-di-GMP to populate the closed conformation<sup>97</sup> (Figure 7b). This analog was not able to bind the PilZ domain protein Alg44 from *Pseudomonas aeruginosa*, which recognizes c-di-GMP as two intercalating dimers, both in the closed conformation, nor inhibit c-di-GMP synthesis of the DGC WspR, which binds c-di-GMP at the inhibitory I-site in the same fashion<sup>97</sup>. However, this compound was able to bind and inhibit c-di-GMP degradation by the EAL domain protein RocR, which binds c-di-GMP in the open conformation<sup>97</sup>. This suggests that conformational biasing may be a general strategy for designing second messenger analogs that target one class of binding proteins based on the oligomeric state in which they recognize c-di-GMP.

### Targeting c-di-GMP binding riboswitches with second messenger analogs

Our lab has also completed a structure-activity relationship for c-di-GMP binding to both the class I and class II riboswitches using base, ribose, and phosphate modified analogs (Figure 7a,b) to determine the structural features of the second messenger most important for recognition<sup>94</sup>. This has provided insights into how to design c-di-GMP analogs that can bind these regulatory RNAs and potentially be used to manipulate the biological processes under control of these riboswitches. These studies have revealed that the class I riboswitch is much more discriminatory in second messenger recognition than the class II riboswitch and utilizes nearly every structural feature of c-di-GMP for binding and recognition. In contrast, the class II riboswitch can accommodate many structural analogs of guanine into the binding pocket that the class I riboswitch either completely discriminated against or bound with a significantly weaker affinity compared to the native ligand. Furthermore, the class II riboswitch does not absolutely require the ribosyl-phosphate backbone for second messenger recognition. This lack of backbone recognition has been further confirmed by a binding study using a biotinylated version of c-di-GMP that demonstrated the class II riboswitch was highly tolerant to this bulky modification at the 2'-OH position of the ribose rings<sup>98</sup>.

The lack of c-di-GMP ribosyl-phosphate backbone recognition by the class II riboswitch lends this class of c-di-GMP receptors to be selectively targeted over the class I riboswitch. This could be a useful strategy for manipulating specific biological processes under control of the class II riboswitch without affecting processes regulated by the class I riboswitch. Taking advantage of the minimal contacts made to the ribosyl-phosphate backbone of c-di-GMP by the class II riboswitch, we demonstrated that a doubly modified 2'-O-methyl second messenger analog showed a 10,000-fold binding preference for this riboswitch over class I<sup>85</sup> (c-di-2'-OMe-GMP, Figure 7b). In addition, another study demonstrated that a 2'-OTBDMS version of c-di-GMP bound the class II RNA with an affinity only 50-fold weaker than that of the native second messenger<sup>99</sup> (Figure 7b). While this further demonstrates the tolerance of this RNA towards ribose modification, introduction of such large hydrophobic groups has the potential to increase the cell permeability of the second messenger, which is necessary for *in vivo* studies. Other analogs that also showed a greater than 100-fold binding preference for the class II riboswitch and still bound tightly to this regulatory RNA included c-di-IMP, c-di-dGMP, and c-(RpSp)-di-G<sub>ps</sub><sup>94</sup> (Figure 7a,b). This also underscores the observation that the class II riboswitch is tolerant to modification of the bases, ribose rings and phosphate linkages of c-di-GMP, whereas the class I riboswitch extensively utilizes all of these structural features of the second messenger for recognition and is much less tolerant of the same ligand perturbations<sup>94</sup>.

One of the only similarities in second messenger recognition noted from structural analysis among the class I and class II riboswitches is the extensive base stacking interactions made between the guanine bases of the ligand and conserved binding pocket nucleotides<sup>85</sup> (Figure 4a,e). Analog binding studies have further confirmed the importance of base stacking in ligand recognition, suggesting general principles for the design of second messenger analogs to target these c-di-GMP binding aptamers. Binding of 7-deaza guanine modified ligands to c-di-GMP riboswitches revealed that a much larger cost to the binding energy is observed for this modification than would be predicted from eliminating only the observed hydrogen bonding contacts to the N7 position in the crystal structure<sup>94</sup> (c-c<sup>7</sup>GMP-GMP and c-di-c<sup>7</sup>GMP, Figure 7a). These data were interpreted as a stacking effect based on the observation that nucleic acid duplexes containing 7-deaza guanine are less stable than the corresponding duplexes containing the native guanine base due to decreased base pairing and base stacking interactions<sup>100</sup>. This suggests that the extensive base stacking contacts made in the aptamer binding pocket are also being perturbed by this modification to c-di-GMP, contributing to the large loss in binding energy. To maintain these base stacking contacts, c-di-GMP must be bound in the closed conformation, suggesting that analogs populating the open conformation may be discriminated against by these RNA receptors. Consistent with these observations, it has also been demonstrated that the c-di-GMP analog favoring the open conformation, endo-S-c-di-GMP (introduced above) (Figure 7b), has a weaker affinity for both the class I and class II riboswitches compared to the native second messenger<sup>98</sup>. No contacts are made to the bridging phosphate oxygens by either riboswitch and replacing this atom with sulfur should not perturb any contacts, suggesting that the loss in affinity is due to a loss of efficient stacking. This is further supported by the observation that sulfur substitution of non-bridging phosphate oxygens has almost no effect on ligand binding by the class II riboswitch<sup>94</sup>, whereas the equally conservative modification of substituting a bridging phosphate oxygen with sulfur, which affects the conformation of the second messenger, has a much greater effect on ligand binding<sup>98</sup>.

All of the strategies for targeting c-di-GMP binding riboswitches discussed above utilize cyclic derivatives of the second messenger. Recently, Breaker and coworkers explored the utility of using linear dinucleotide analogs to target these second messenger riboswitches<sup>99</sup> (Figure 7c). The advantage of designing linear analogues over cyclic analogues to target these second messenger riboswitches is that they are synthetically much easier to access<sup>99</sup>.

Both RNA aptamers show a large preference for binding the cyclic dinucleotide over the linear dinucleotide (pGpG, Figure 7c), approximately 5 orders of magnitude for the class I riboswitch<sup>67,101</sup> and 3 orders for class II<sup>68</sup>. Despite this fact, several linear modified derivatives bound reasonably well to these RNAs. However, the tightest binding linear analog for both riboswitches proved to be the natural c-di-GMP breakdown product, pGpG<sup>99</sup>. Interestingly, removal of the terminal 5'-phosphate to yield GpG did not significantly affect the affinity for the class I riboswitch relative to pGpG<sup>99,101</sup> (Figure 7c). In contrast, slightly improved affinity for the class II riboswitch was observed if the terminal 5'-phosphate on the linear dinucleotide analogs was retained<sup>99</sup>. Structural studies from our group have revealed that pGpG can only bind the class I aptamer in a single orientation with the terminal 5'-phosphate placed 5' of G<sub>α</sub> in close proximity to G20, whereas GpG can bind in two different orientation with no preference for the placement of the linking phosphate<sup>101</sup>. This is likely due to potential steric and electrostatic clashes that may occur between the terminal 5'-phosphate and RNA atoms when placed at alternative positions. In addition, many of the metal-phosphate and metal-RNA contacts observed for c-di-GMP are lost in binding pGpG. Thus, if the molecule is not cyclic, there is little to no advantage in maintaining a 5'-phosphate<sup>101</sup>. Taken together, this suggests that cyclic versions of the second messenger are more effective at targeting c-di-GMP riboswitches over linear analogues, despite the ease of synthesis in obtaining the latter.

Several studies have reported that the exogenous addition of c-di-GMP and several of its second messenger analogs to bacteria induces biological phenotypes, specifically a decrease in biofilm formation<sup>102-105</sup>. While the mechanism of action of these compounds is unknown, it has been hypothesized that these second messenger derivatives enter cells and act intracellularly on known c-di-GMP target molecules. Thus, it will be interesting to see if these compounds modulate bacterial behavior through interactions with c-di-GMP RNA and protein effectors and if it is possible to design analogs that induce different phenotypic effects.

## Conclusions and Perspectives

The ubiquitous bacterial second messenger signaling molecule c-di-GMP controls a variety of biological processes including biofilm formation, virulence response, and motility. The absence of this small molecule in eukaryotes makes this pathway an attractive target for the development of new antibacterial drugs that would not kill the cells but instead modulate the virulence and biofilm forming behaviors of the microbe. In addition, we are far from a complete mechanistic understanding of c-di-GMP signaling pathways. Thus, the development of c-di-GMP analogs as chemical tools to study interactions of the second messenger with its targets needs to be further explored because these studies will guide the design of compounds with potential therapeutic efficacy.

Designing c-di-GMP analogs that tightly bind second messenger targets but simultaneously incorporate modifications that impart desirable properties for *in vivo* studies, such as increased cell permeability and increased resistance to enzymatic degradation, is a desirable goal. However, what remains a constant challenge is the development of more efficient and higher yielding synthetic routes to obtain modified versions of the second messenger. Given the complexity of c-di-GMP signaling and the steady emergence of new protein and RNA binding partners of the second messenger, chemical tools to isolate new effector molecules would also be valuable. Towards this aim, a biotinylated version of the second messenger has been employed to isolate c-di-GMP binding proteins<sup>106</sup>, however the placement of the biotin moiety inhibits binding to certain targets<sup>98</sup>. Thus, an array of analogs with chemical handles at different positions on c-di-GMP could be useful for identifying new targets with unique modes of second messenger recognition.

Furthermore, it will be necessary to explore the biological consequences of targeting c-di-GMP binding proteins and RNA *in vivo* with second messenger analogs. Multiple c-di-GMP targets often exist within a single organism and identifying analogs that selectively target proteins over RNA and vice versa will be necessary to disentangle signaling events that are mediated by these diverse macromolecular effectors. It is currently unknown if it is possible to achieve selectivity for one particular target within the complex environment of the cell. Identification of second messenger analogs that induce interesting phenotypic effects and have behavioral responses in bacteria, as well as elucidating the molecular mechanisms of action of such bioactive compounds, will continue to be an active area of research.

## Acknowledgments

We thank K. Smith for critical comments on this manuscript.

## References

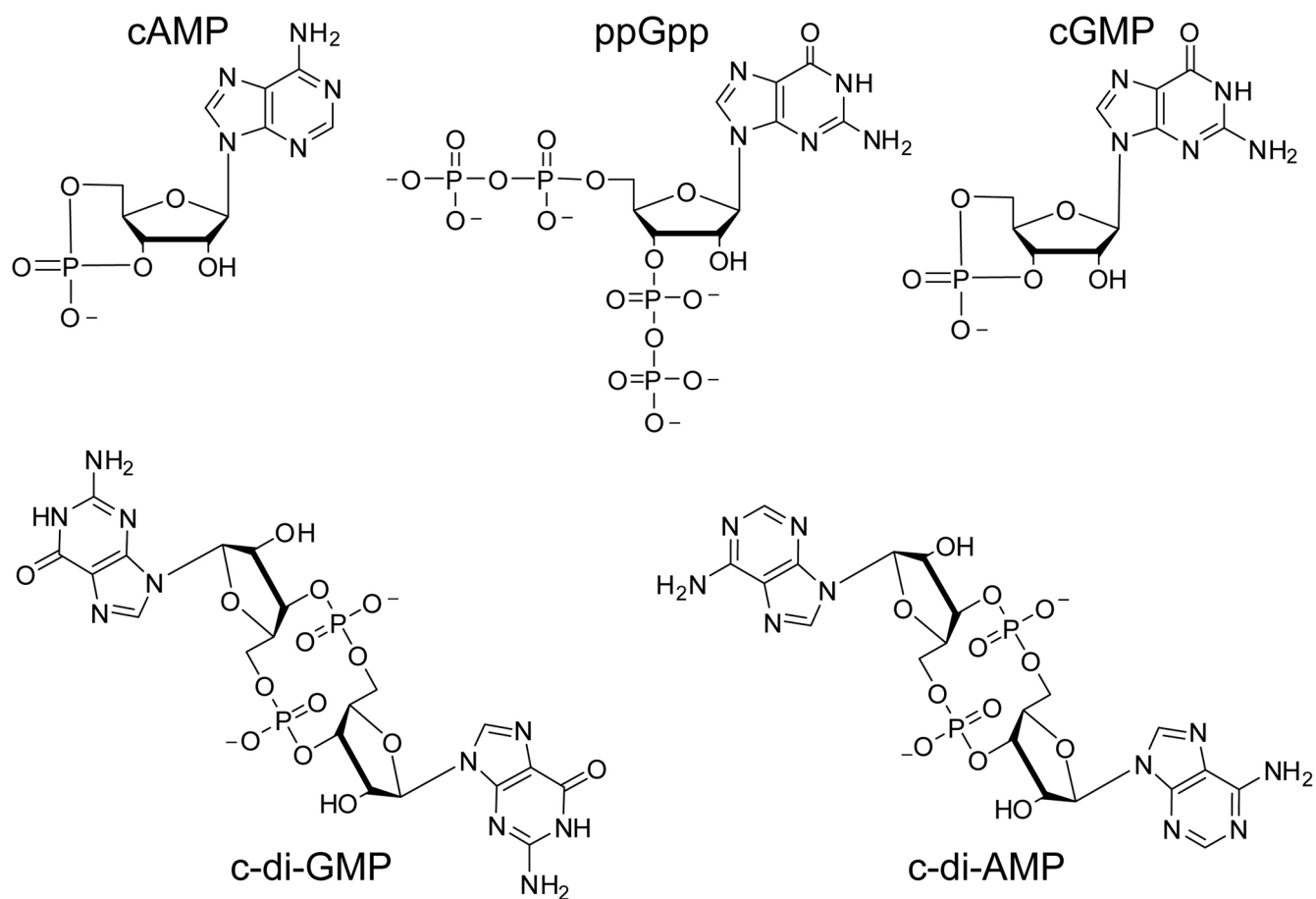
1. Galperin MY. *Environ Microbiol.* 2004; 6:552–567. [PubMed: 15142243]
2. Camilli A, Bassler BL. *Science.* 2006; 311:1113–1116. [PubMed: 16497924]
3. Pesavento C, Hengge R. *Curr Opin Microbiol.* 2009; 12:170–176. [PubMed: 19318291]
4. McDonough KA, Rodriguez A. *Nat Rev Microbiol.* 2012; 10:27–38. [PubMed: 22080930]
5. Magnusson LU, Farewell A, Nyström T. *Trends Microbiol.* 2005; 13:236–242. [PubMed: 15866041]
6. Potrykus K, Cashel M. *Annu Rev Microbiol.* 2008; 62:35–51. [PubMed: 18454629]
7. Gomelsky M. *Mol Microbiol.* 2011; 79:562–565. [PubMed: 21255104]
8. Jenal U. *Curr Opin Microbiol.* 2004; 7:185–191. [PubMed: 15063857]
9. Römling U, Gomelsky M, Galperin MY. *Mol Microbiol.* 2005; 57:629–639. [PubMed: 16045609]
10. Hengge R. *Nat Rev Microbiol.* 2009; 7:263–273. [PubMed: 19287449]
11. Schirmer T, Jenal U. *Nat Rev Microbiol.* 2009; 7:724–735. [PubMed: 19756011]
12. Mills E, Pultz IS, Kulasekara HD, Miller SI. *Cell Microbiol.* 2011; 13:1122–1129. [PubMed: 21707905]
13. Sondermann H, Shikuma NJ, Yildiz FH. *Curr Opin Microbiol.* 2012; 15:140–146. [PubMed: 22226607]
14. Jenal U, Malone J. *Annu Rev Genet.* 2006; 40:385–407. [PubMed: 16895465]
15. Oppenheimer-Shaanan Y, Wexselblatt E, Katzhendler J, Yavin E, Ben-Yehuda S. *EMBO Rep.* 2011; 12:594–601. [PubMed: 21566650]
16. Corrigan RM, Abbott JC, Burhenne H, Kaever V, Gründling A. *PLoS Pathog.* 2011; 7:e1002217. [PubMed: 21909268]
17. Ross P, et al. *Nature.* 1987; 325:279–281. [PubMed: 18990795]
18. Tal R, et al. *J Bacteriol.* 1998; 180:4416–4425. [PubMed: 9721278]
19. Galperin MY, Nikolskaya AN, Koonin EV. *FEMS Microbiol Lett.* 2001; 203:11–21. [PubMed: 11557134]
20. D'Argenio DA, Miller SI. *Microbiol.* 2004; 150:2497–2502.
21. Ryjenkov DA, Tarutina M, Moskvina OV, Gomelsky M. *J Bacteriol.* 2005; 187:1792–1798. [PubMed: 15716451]
22. Paul R, Weiser S, Amiot NC, Chan C, Schirmer T, Giese B, Jenal U. *Genes Dev.* 2004; 18:715–727. [PubMed: 15075296]
23. Schmidt AJ, Ryjenkov DA, Gomelsky M. *J Bacteriol.* 2005; 187:4774–4781. [PubMed: 15995192]
24. Tamayo R, Tischler AD, Camilli A. *J Biol Chem.* 2005; 280:33324–33330. [PubMed: 16081414]
25. Rao F, Yang Y, Qi Y, Liang Z-X. *J Bacteriol.* 2008; 190:3622–3631. [PubMed: 18344366]
26. Galperin MY, Natale DA, Aravind L, Koonin EV. *J Mol Microbiol Biotechnol.* 1999; 1:303–305. [PubMed: 10943560]



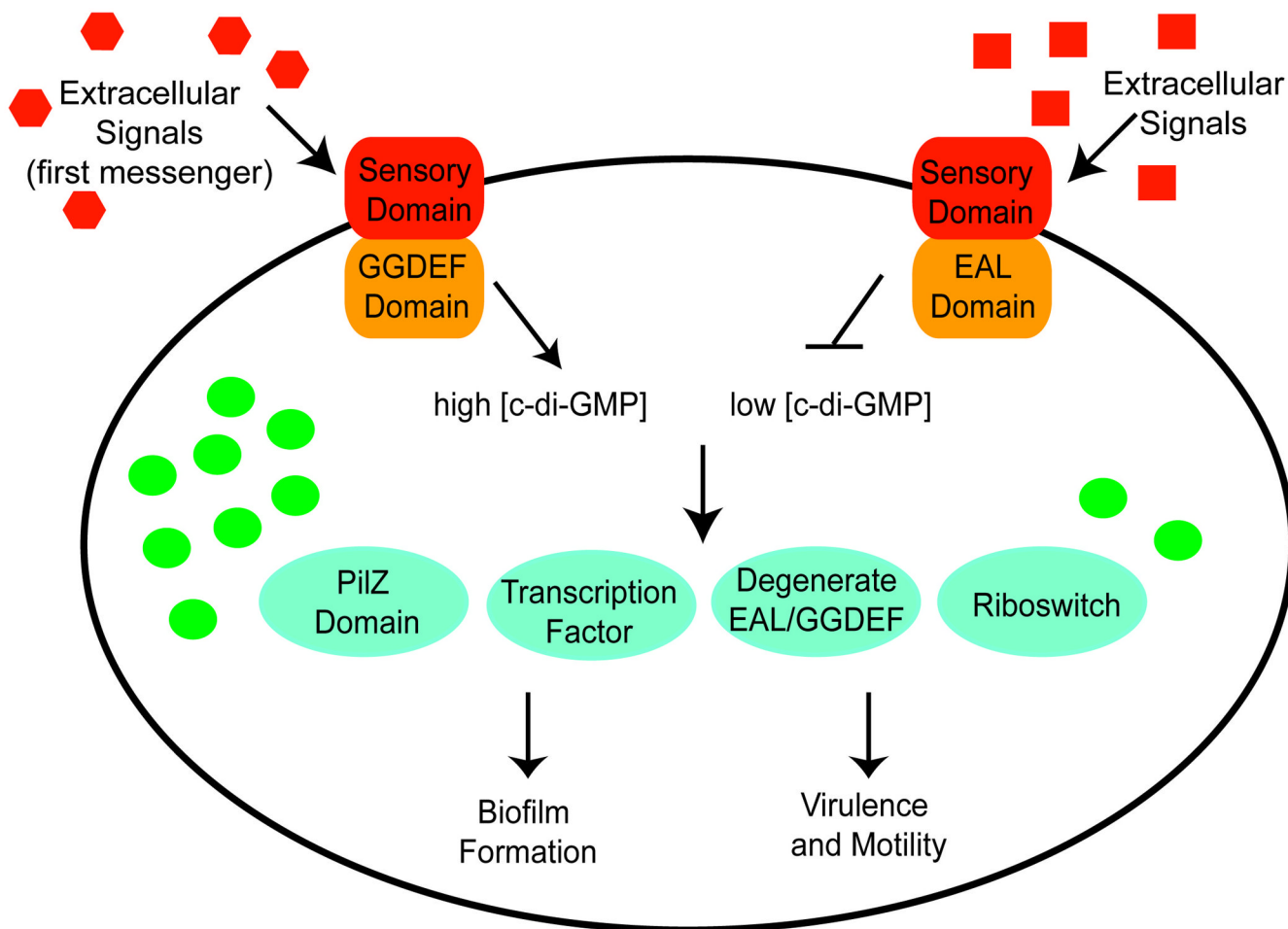
27. Ryan RP, Fouhy Y, Lucey JF, Crossman LC, Spiro S, He Y, Zhang L, Heeb S, Camara M, Williams P, Dow JM. *Proc Natl Acad Sci USA*. 2006; 103:6712–6717. [PubMed: 16611728]
28. Christen B, Christen M, Folcher M, Schauerte A, Jenal U. *J Biol Chem*. 2006; 281:32015–32024. [PubMed: 16923812]
29. Chan C, Paul R, Samoray D, Amiot NC, Giese B, Jenal U. *Proc Natl Acad Sci USA*. 2004; 101:17084–17089. [PubMed: 15569936]
30. Tamayo R, Pratt JT, Camilli A. *Annu Rev Microbiol*. 2007; 61:131–148. [PubMed: 17480182]
31. Simm R, Morr M, Kader A, Nimtz M, Römling U. *Mol Microbiol*. 2004; 53:1123–1134. [PubMed: 15306016]
32. Tischler AD, Camilli A. *Infect Immun*. 2005; 73:5873–5882. [PubMed: 16113306]
33. Ahmad I, Lamprokostopoulou A, Le Guyon S, Streck E, Barthel M, Peters V, Hardt W, Romling U. *PLoS ONE*. 2011; 6:e28351. [PubMed: 22164276]
34. Pitzer JE, Sultan SZ, Hayakawa Y, Hobbs G, Miller MR, Motaleb MA. *Infect Immun*. 2011; 79:1815–1825. [PubMed: 21357718]
35. Waters CM, Lu W, Rabinowitz JD, Bassler BL. *J Bacteriol*. 2008; 190:2527–2536. [PubMed: 18223081]
36. Srivastava D, Waters CM. *J Bacteriol*. 2012; 194:4485–4493. [PubMed: 22661686]
37. Fong JCN, Yildiz FH. *J Bacteriol*. 2008; 190:6646–6659. [PubMed: 18708497]
38. Beyhan S, Bilecen K, Salama SR, Casper-Lindley C, Yildiz FH. *J Bacteriol*. 2007; 189:388–402. [PubMed: 17071756]
39. Christen M, Christen B, Folcher M, Schauerte A, Jenal U. *J Biol Chem*. 2005; 280:30829–30837. [PubMed: 15994307]
40. Seshasayee ASN, Fraser GM, Luscombe NM. *Nucl Acids Res*. 2010; 38:5970–5981. [PubMed: 20483912]
41. Tarutina M, Ryjenkov DA, Gomelsky M. *J Biol Chem*. 2006; 281:34751–34758. [PubMed: 16968704]
42. Ferreira RBR, Antunes LCM, Greenberg EP, McCarter LL. *J Bacteriol*. 2008; 190:851–860. [PubMed: 17993539]
43. Ryan RP, Fouhy Y, Lucey JF, Dow JM. *J Bacteriol*. 2006; 188:8327–8334. [PubMed: 17028282]
44. Sommerfeldt N, Possling A, Becker G, Pesavento C, Tschowri N, Henge R. *Microbiol*. 2009; 155:1318–1331.
45. Krasteva PV, Giglio KM, Sondermann H. *Protein Sci*. 2012; 21:929–948. [PubMed: 22593024]
46. Amikam D, Galperin MY. *Bioinformatics*. 2006; 22:3–6. [PubMed: 16249258]
47. Ryjenkov DA, Simm R, Römling U, Gomelsky M. *J Biol Chem*. 2006; 281:30310–30314. [PubMed: 16920715]
48. Benach J, Swaminathan SS, Tamayo R, Handelman SK, Folta-Stogniew E, Ramos JE, Forouhar F, Neely H, Seetharaman J, Camilli A, Hunt JF. *EMBO J*. 2006; 26:5153–5166. [PubMed: 18034161]
49. Pratt JT, Tamayo R, Tischler AD, Camilli A. *J Biol Chem*. 2007; 282:12860–12870. [PubMed: 17307739]
50. Boehm A, Kaiser M, Li H, Spangler C, Kasper CA, Ackermann M, Kaefer V, Sourjik V, Roth V, Jenal U. *Cell*. 2010; 141:107–116. [PubMed: 20303158]
51. Paul K, Nieto V, Carlquist WC, Blair DF, Harshey RM. *Mol Cell*. 2010; 38:128–139. [PubMed: 20346719]
52. Merighi M, Lee VT, Hyodo M, Hayakawa Y, Lory S. *Mol Microbiol*. 2007; 65:876–895. [PubMed: 17645452]
53. Hickman JW, Harwood CS. *Mol Microbiol*. 2008; 69:376–389. [PubMed: 18485075]
54. Krasteva PV, Fong JC, Shikuma NJ, Beyhan S, Navarro MV, Yildiz FH, Sondermann H. *Science*. 2010; 327:866–868. [PubMed: 20150502]
55. Leduc JL, Roberts GP. *J Bacteriol*. 2009; 191:7121–7122. [PubMed: 19633082]
56. Tao F, He Y-W, Wu D-H, Swarup S, Zhang L-H. *J Bacteriol*. 2010; 192:1020–1029. [PubMed: 20008070]

57. Srivastava D, Harris RC, Waters CM. *J Bacteriol.* 2011; 193:6331–6341. [PubMed: 21926235]
58. Wilksch JJ, Yang J, Clements A, Gabbe JL, Short KR, Cao H, Cavaliere R, James CE, Whitchurch CB, Schembri MA, Chuah ML, Liang Z-X, Wijburg OL, Jenney AW, Lithgow T, Strugnell RA. *PLoS Pathog.* 2011; 7:e1002204. [PubMed: 21901098]
59. Navarro MVAS, De N, Bae N, Wang Q, Sondermann H. *Structure.* 2009; 17:1104–1116. [PubMed: 19679088]
60. Newell PD, Monds RD, O'Toole GA. *Proc Natl Acad Sci USA.* 2009; 106:3461–3466. [PubMed: 19218451]
61. Duerig A, Abel S, Folcher M, Nicollier M, Schwede T, Amiot N, Giese B, Jenal U. *Genes Dev.* 2009; 23:93–104. [PubMed: 19136627]
62. Petters T, Zhang X, Nesper J, Treuner-Lange A, Gomez-Santos N, Hoppert M, Jenal U, Soggard-Andersen L. *Mol Microbiol.* 2012; 84:147–165. [PubMed: 22394314]
63. Beyhan S, Odell LS, Yildiz FH. *J Bacteriol.* 2008; 190:7392–7405. [PubMed: 18790873]
64. Lee VT, Matewish JM, Kessler JL, Hyodo M, Hayakawa Y, Lory S. *Mol Microbiol.* 2007; 65:1474–1484. [PubMed: 17824927]
65. Navarro MVAS, Newell PD, Krasteva PV, Chatterjee D, Madden DR, O'Toole GA, Sondermann H. *PLoS Biol.* 2011; 9:e1000588. [PubMed: 21304926]
66. Kazmierczak BI, Lebron MB, Murray TS. *Mol Microbiol.* 2006; 60:1026–1043. [PubMed: 16677312]
67. Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, Breaker RR. *Science.* 2008; 321:411–413. [PubMed: 18635805]
68. Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR. *Science.* 2010; 329:845–848. [PubMed: 20705859]
69. Montange RK, Batey RT. *Annu Rev Biophys.* 2008; 37:117–133. [PubMed: 18573075]
70. Roth A, Breaker RR. *Annu Rev Biochem.* 2009; 78:305–334. [PubMed: 19298181]
71. Breaker RR. *Cold Spring Harb Perspect Biol.* 2012; 4:a003566. [PubMed: 21106649]
72. Garst AD, Edwards AL, Batey RT. *Cold Spring Harb Perspect Biol.* 2011; 3:a003533. [PubMed: 20943759]
73. Smith KD, Lipchock SV, Ames TD, Breaker RR, Strobel SA. *Nat Struct Mol Biol.* 2009; 16:1218–1223. [PubMed: 19898477]
74. Smith KD, Strobel SA. *Biochem Soc Trans.* 2011; 39:647–651. [PubMed: 21428955]
75. Beyhan S, Yildiz FH. *Mol Microbiol.* 2007; 63:995–1007. [PubMed: 17233827]
76. Ko J, Ryu KS, Kim H, Shin JS, Lee JO, Cheong C, Choi BS. *J Mol Biol.* 2010; 398:97–110. [PubMed: 20226196]
77. Shin J-S, Ryu K-S, Ko J, Lee A, Choi B-S. *Protein Sci.* 2011; 20:270–277. [PubMed: 21280119]
78. Whitney JC, Colvin KM, Marmount LS, Robinson H, Parsek MR, Howell PL. *J Biol Chem.* 2012; 287:23582–23593. [PubMed: 22605337]
79. Wassmann P, Chan C, Paul R, Beck A, Heerklotz H, Jenal U, Schirmer T. *Structure.* 2007; 15:915–927. [PubMed: 17697997]
80. De N, Navarro MVAS, Raghavan RV, Sondermann H. *J Mol Biol.* 2009; 393:619–633. [PubMed: 19695263]
81. Qi Y, Chuah ML, Dong X, Xie K, Luo Z, Tang K, Liang ZX. *J Biol Chem.* 2011; 286:2910–2917. [PubMed: 21098028]
82. Habazettl J, Allan MG, Jenal U, Grzesiek S. *J Biol Chem.* 2011; 286:14304–14314. [PubMed: 21310957]
83. Smith KD, Lipchock SV, Livingston AL, Shanahan CA, Strobel SA. *Biochemistry.* 2010; 49:7351–7359. [PubMed: 20690679]
84. Kulshina N, Baird NJ, Ferré-D'Amaré AR. *Nat Struct Mol Biol.* 2009; 16:1212–1217. [PubMed: 19898478]
85. Smith KD, Shanahan CA, Moore EL, Simon AC, Strobel SA. *Proc Natl Acad Sci USA.* 2011; 108:7757–7762. [PubMed: 21518891]

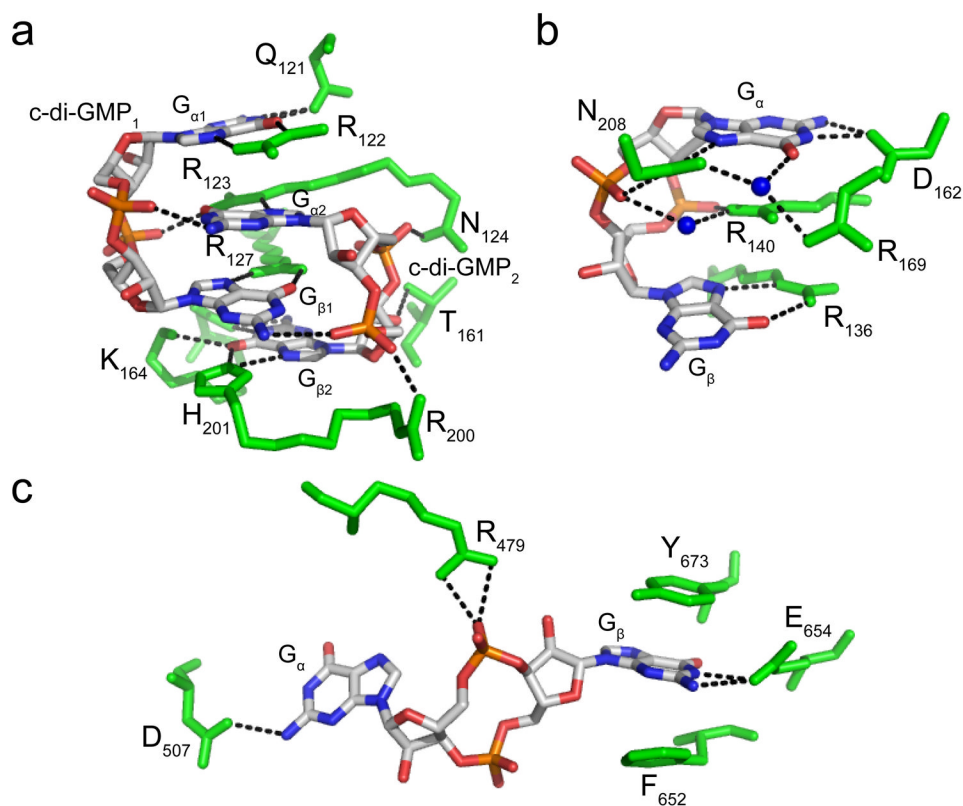
86. Ross P, Mayer R, Weinhouse H, Amikam D, Huggirat Y, Benziman M, de Vroom E, Fidder A, de Paus P, Sliedregt LA, van der Marel GA, van Boom JH. *J Biol Chem*. 1990; 265:18933–18943. [PubMed: 2172238]
87. Hayakawa Y, Nagata R, Hirata A, Hyodo M, Kawai R. *Tetrahedron*. 2003; 59:6465–6471.
88. Hyodo M, Hayakawa Y. *Bull Chem Soc Japan*. 2004; 77:2089–2093.
89. Hyodo M, Sato Y, Hayakawa Y. *Tetrahedron*. 2006; 62:3089–3094.
90. Zhang Z, Gaffney BL, Jones RA. *J Am Chem Soc*. 2004; 126:16700–16701. [PubMed: 15612689]
91. Amiot N, Heintz K, Giese B. *Synthesis*. 2006; 2006:4230–4236.
92. Yan H, Aguilar AL. *Nucleosides, Nucleotides and Nucleic Acids*. 2007; 26:189–204.
93. Kiburu I, Shurer A, Yan L, Sintim HO. *Mol Biosyst*. 2008; 4:518–520. [PubMed: 18493648]
94. Shanahan CA, Gaffney BL, Jones RA, Strobel SA. *J Am Chem Soc*. 2011; 133:15578–15592. [PubMed: 21838307]
95. Gaffney BL, Veliath E, Zhao J, Jones RA. *Org Lett*. 2010; 12:3269–3271. [PubMed: 20572672]
96. Gaffney BL, Jones RA. *Curr Protoc Nucleic Acid Chem*. 2012; Chapter 14(Unit 14.8):1–7.
97. Wang J, Zhou J, Donaldson GP, Nakayama S, Yan L, Lam YF, Lee VT, Sintim HO. *J Am Chem Soc*. 2011; 133:9320–9330. [PubMed: 21612220]
98. Luo Y, Zhou J, Watt SK, Lee VT, Dayie TK, Sintim HO. *Mol Biosyst*. 2012; 8:772–778. [PubMed: 22182995]
99. Furukawa K, Gu H, Sudarsan N, Hayakawa Y, Hyodo M, Breaker RR. *ACS Chem Biol*. 2012; 7:1436–1443. [PubMed: 22646696]
100. Grein T, Lampe S, Mersmann K, Rosemeyer H, Thomas H, Seela F. *Bioorg Med Chem Lett*. 1994; 4:971–976.
101. Smith KD, Lipchok SV, Strobel SA. *Biochemistry*. 2012; 51:425–432. [PubMed: 22148472]
102. Yan W, Qu T, Zhao H, Su L, Yu Q, Gao J, Wu B. *Microbiol Res*. 2010; 165:87–96. [PubMed: 19186044]
103. Karaolis DKR, Rashid MH, Chythanya R, Luo W, Hyodo M, Hayakawa Y. *Antimicrob Agents Chemother*. 2005; 49:1029–1038. [PubMed: 15728899]
104. Ishihara Y, Hyodo M, Hayakawa Y, Kamegaya T, Yamada K, Okamoto A, Hasegawa T, Ohta M. *FEMS Microbiol Lett*. 2009; 301:193–200. [PubMed: 20169626]
105. Mano E, Hyodo M, Sato Y, Ishihara Y, Ohta M, Hayakawa Y. *ChemMedChem*. 2007; 2:1410–1413. [PubMed: 17886852]
106. Nesper J, Reinders A, Glatzer T, Schmidt A, Jenal U. *J Proteom*. 2012; 75:4874–4878.



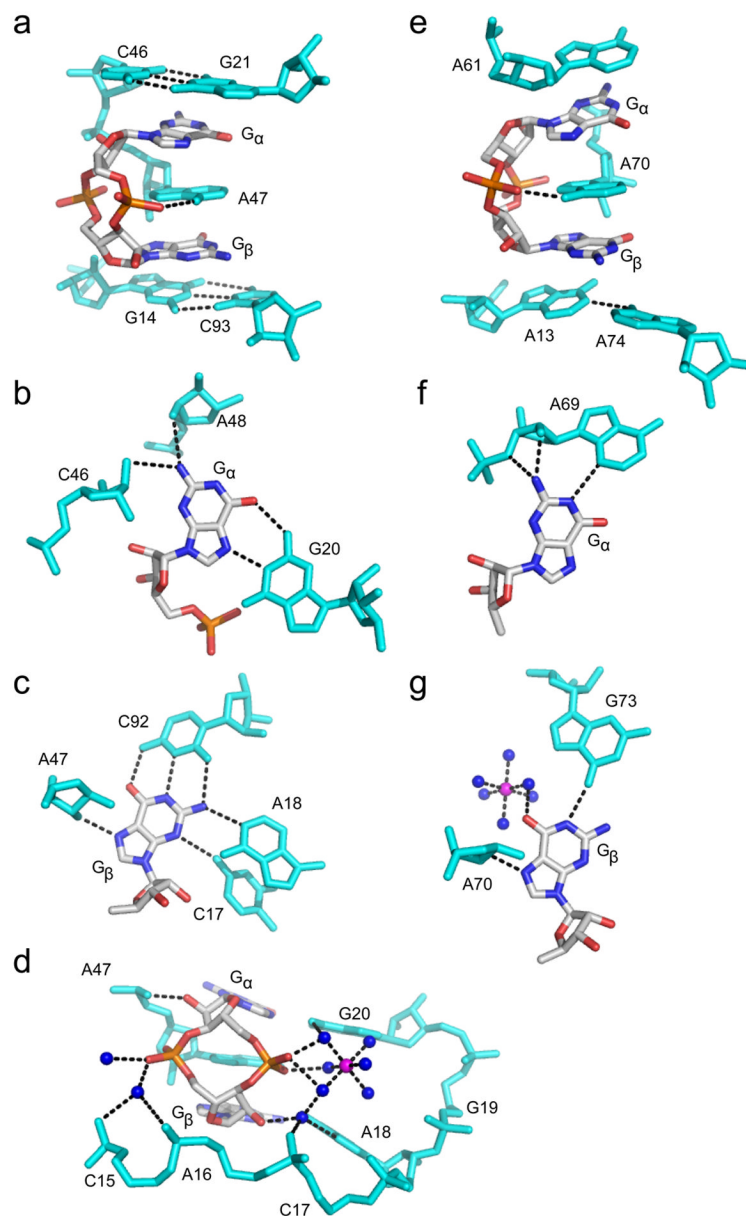
**Figure 1.** Chemical structures of known bacterial nucleotide based second messenger signaling molecules. Bacteria utilize both cyclic mononucleotides and cyclic dinucleotides for signaling.



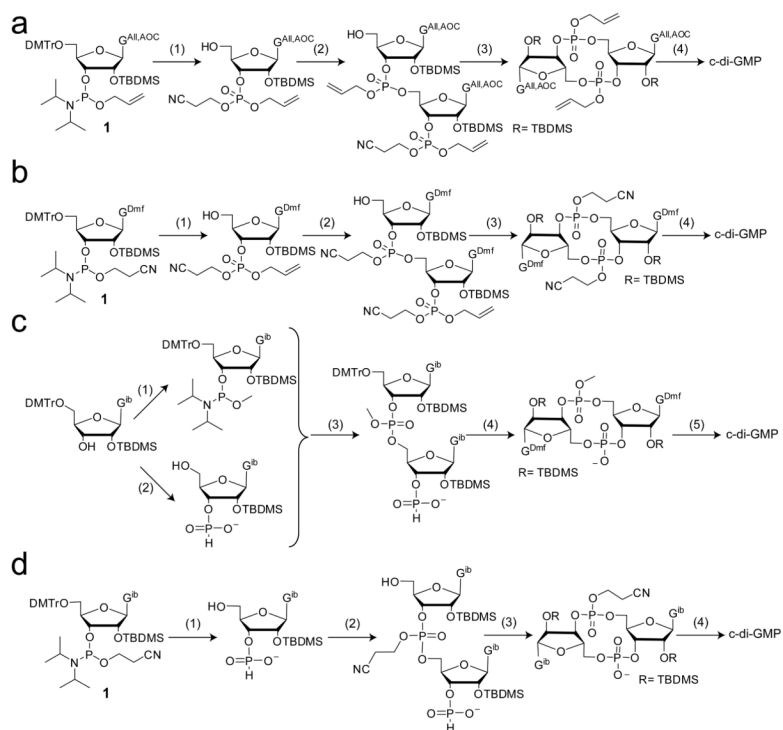
**Figure 2.** Schematic of c-di-GMP signaling in bacteria. c-di-GMP metabolizing enzymes are shown as orange rectangles, c-di-GMP molecules are represented by green circles, and receptors of the second messenger are indicated in teal circles. Similar to the EAL domain, some bacteria also utilize HD-GYP domain proteins to decrease intracellular levels of c-di-GMP. Phenotypic outputs induced by high concentrations of the second messenger are indicated on the bottom left and those induced by low concentrations are indicated on the bottom right.



**Figure 3.** Protein recognition of c-di-GMP. (a) Recognition by the PilZ domain protein PP4397 from *Pseudomonas putida* (PDB ID 3KYF). c-di-GMP is colored by atom with carbon shown in white, nitrogen in blue, oxygen in red and phosphorous in orange. Protein residues are colored green and hydrogen bonds are shown as black dashed lines. c-di-GMP is bound as a dimer and the monomeric units are labeled as c-di-GMP<sub>1</sub> and c-di-GMP<sub>2</sub>. The top guanine base of each c-di-GMP monomer is labeled G<sub>α1</sub> or G<sub>α2</sub>, depending on which c-di-GMP monomer it belongs to, and the bottom bases are labeled G<sub>β1</sub> or G<sub>β2</sub>. (b) Recognition by the PilZ domain protein VCA0042 from *Vibrio cholerae* (PDB ID 2RDE). Each base of the c-di-GMP monomer is arbitrarily labeled G<sub>α</sub> or G<sub>β</sub>. Coloring is the same as in part (a) with water molecules shown as blue spheres. (c) Recognition by the degenerate EAL domain protein FimX. Coloring is the same as in part (a) and the guanine bases are arbitrarily labeled as in part (b).

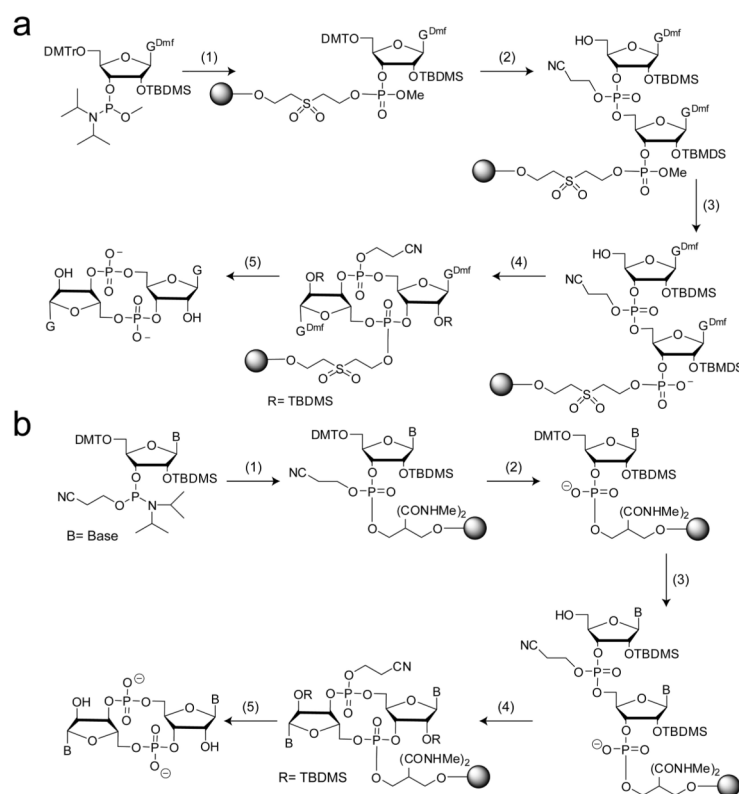


**Figure 4.** RNA recognition of c-di-GMP. c-di-GMP is colored by atom as in Figure 3. RNA residues are colored cyan, water molecules are shown as blue spheres and magnesium ions as magenta spheres. Hydrogen bonds are indicated by black dashed lines. (a) c-di-GMP bound to the Vc2 class I riboswitch from *Vibrio cholerae* (PDB ID 3MXH). (b) Recognition of G<sub>α</sub>, (c) G<sub>β</sub> and (d) the c-di-GMP ribosyl-phosphate backbone by the class I riboswitch. (e) c-di-GMP bound to the Cac-1-2 class II riboswitch from *Clostridium acetobutylicum* (PDB ID 3Q3Z). (f) Recognition of G<sub>α</sub> and (g) G<sub>β</sub> by the class II riboswitch.

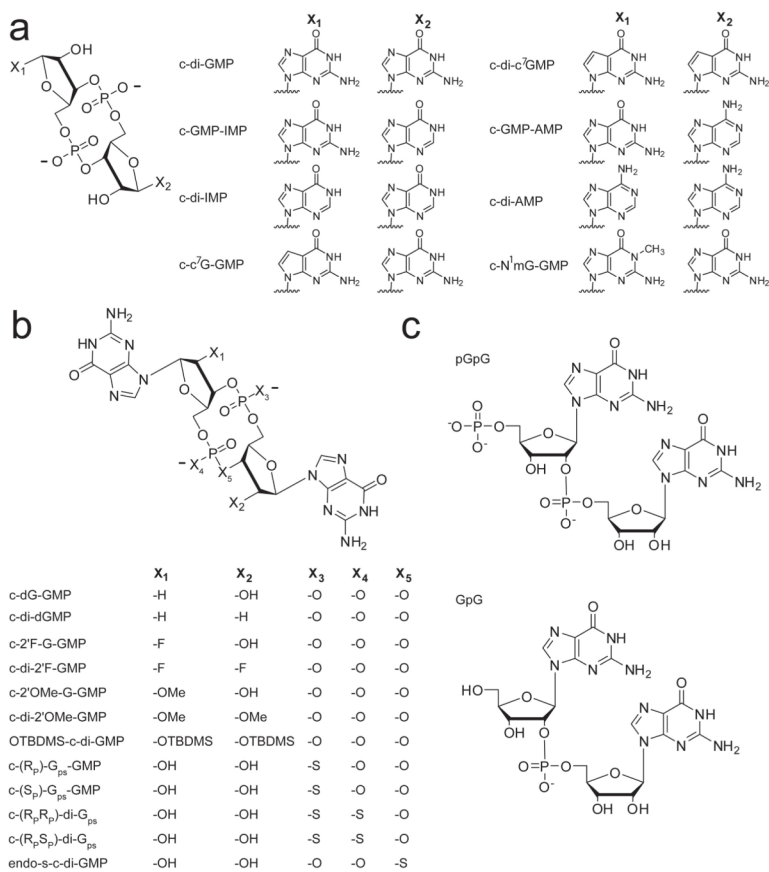
**Figure 5.**

Solution phase methods for c-di-GMP synthesis. (a) First phosphoramidite method developed by Hayakawa *et al.*<sup>87</sup>. 1) i) 2-cyanoethanol, IMP, ACN ii) 2-butanone peroxide (BPO)/toluene iii) DCA/DCM; 2) i) **1**, IMP, ACN ii) BPO/toluene iii) DCA/DCM; 3) i) NH<sub>3</sub>-MeOH ii) TPSCl, *N*-methylimidazole, THF; 4) i) palladium catalyst, triphenylphosphine, butylammonium formate, THF ii) HF-TEA. (b) Improved phosphoramidite method<sup>88</sup>. Phosphoramidite **1** is synthesized in higher yield and few steps than that used in Scheme (a). 1) i) allyl alcohol, IMP, ACN ii) BPO/toluene iii) DCA/DCM; 2) i) **1**, IMP, ACN ii) BPO/toluene iii) DCA/DCM; 3) i) NaI, acetone ii) TPSCl, *N*-methylimidazole; 4) i) NH<sub>3</sub>-MeOH ii) HF-TEA. (c) Method developed by the Jones group<sup>90</sup>. c-di-GMP is synthesized using phosphoramidite coupling and H-phosphonate cyclization. 1) Bis(diisopropylamino)methyl phosphoramidite and pyridinium trifluoroacetate; 2) 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one; 3) i) pyridinium trifluoroacetate ii) *tert*-butylhydroperoxide; 4) i) sulfonic acid resin ii) adamantoylcarbonyl chloride iii) methanol/NBS; 5) i) pyridine/aq. NH<sub>3</sub> ii) HF-TEA. (d) Synthesis of c-di-GMP in one pot<sup>95,96</sup>. 1) i) pyr-TFA/H<sub>2</sub>O ii) *t*BuNH<sub>2</sub> iii) DCA/H<sub>2</sub>O 2) i) pyridine/*t*BuOOH iii) DCA/H<sub>2</sub>O 3) i) DMOCP ii) I<sub>2</sub>/H<sub>2</sub>O 4) i) *t*BuNH<sub>2</sub> ii) TEA·HF. For synthesis of dithiophosphate analogs, *t*BuOOH was replaced with the sulfuration reagent 3-((dimethylamino-methylidene)amino)-3*H*-1,2-benzodithiol-3-one in the first oxidation step and I<sub>2</sub>/H<sub>2</sub>O was replaced with 3-*H*-1,2-benzodithiol-3-one in the second oxidation step<sup>95</sup>.



**Figure 6.**

Solid phase methods for c-di-GMP synthesis utilizing on-bead cyclization procedures. (a) Method developed by Kiburu *et al.*<sup>3</sup>. Two different phosphoramidites containing different phosphate protecting groups were used. 1) i) solid support, tetrazole/ACN ii) I<sub>2</sub>, pyridine, H<sub>2</sub>O; 2) i) DCA/DCM ii) cyanoethyl phosphoramidite, tetrazole/ACN iii) I<sub>2</sub> pyridine, H<sub>2</sub>O iv) DCA/DCM; 3) S<sub>2</sub>Na<sub>2</sub> 4) MSNT, pyridine 5) i) aqueous NH<sub>3</sub> ii) HF-TEA. For the solution phase cyclization method, O-methyl phosphoramidites were used in both coupling steps and the linear dinucleotide was cleaved from the bead prior to cyclization and global deprotection. (b) Modified solid-phase synthesis of c-di-GMP, base and ribose modified analogs. Cyanoethyl-protected phosphoramidites were used for both coupling reactions. 1) i) tetrazole/ACN ii) tBuOOH iii) acetic anhydride/methylamine; 2) i) 50% TEA/ACN, 2 hours; 3) i) 3% DCA/DCM ii) tetrazole/ACN + CNE phosphoramidite iii) tBuOOH iv) acetic anhydride/methylamine 4) 0.1M MSNT, 72–96 hours; 5) i) ammonium hydroxide ii) HF-TEA (for 2'-OH analogs only).



**Figure 7.** Structures of chemically synthesized dinucleotide second messenger analogs used to probe protein and RNA c-di-GMP effector molecules. (a) Cyclic analogs containing base modifications and (b) ribosyl-phosphate backbone modifications. (c) Linear dinucleotide analogs.

**Table 1**

Receptors that sense the second messenger c-di-GMP.

| Class of Receptor           | Example | Organism              | Biological Output                    | Level of action                 |
|-----------------------------|---------|-----------------------|--------------------------------------|---------------------------------|
| <b>PilZ Domain Protein</b>  | BcsA    | <i>G. xylinus</i>     | Cellulose synthesis                  | Post-Translational              |
|                             | YcgR    | <i>E. coli</i>        | Swimming velocity                    | Post-Translational              |
|                             | DcgR    | <i>C. crescentus</i>  | Motility                             | Post-Translational              |
|                             | Alg44   | <i>P. aeruginosa</i>  | Alginate synthesis                   | Post-Translational <sup>1</sup> |
| <b>Transcription Factor</b> | FleQ    | <i>P. aeruginosa</i>  | Flagella expression                  | Transcriptional                 |
|                             | VpsT    | <i>V. cholerae</i>    | EPS matrix synthesis                 | Transcriptional                 |
|                             | Clp     | <i>X. campestris</i>  | Virulence gene expression            | Transcriptional                 |
|                             | VpsR    | <i>V. cholerae</i>    | Virulence gene expression            | Transcriptional                 |
|                             | MrkH    | <i>K. pneumoniae</i>  | Fimbriae production                  | Transcriptional                 |
| <b>Degenerate EAL</b>       | FimX    | <i>P. aeruginosa</i>  | Twitching motility                   | Post-Translational              |
|                             | LapD    | <i>P. fluorescens</i> | Biofilm formation                    | Post-Translational              |
| <b>Degenerate GGDEF</b>     | PopA    | <i>C. crescentus</i>  | Cell cycle progression               | Post-Translational              |
|                             | CdgG    | <i>V. cholerae</i>    | Rugosity                             | Unknown                         |
|                             | SmgT    | <i>M. xanthus</i>     | Extracellular matrix production      | Post-Translational              |
| <b>I-Site</b>               | PelD    | <i>P. aeruginosa</i>  | Pellicle formation                   | Unknown                         |
| <b>Class I Riboswitch</b>   | Vc2     | <i>V. cholerae</i>    | Rugosity                             | Translational                   |
|                             | Cd1     | <i>C. difficile</i>   | Flagella expression                  | Transcriptional                 |
| <b>Class II Riboswitch</b>  | 84 Cd   | <i>C. difficile</i>   | Expression of a cell surface protein | Translational                   |
|                             | Bha-1-1 | <i>B. halodurans</i>  | Outer membrane protein synthesis     | Transcriptional                 |

<sup>1</sup>Hypothesized to regulate alginate biosynthesis via a protein-protein interaction.