Structural basis of differential ligand recognition by two classes of bis-(3'-5')-cyclic dimeric guanosine monophosphate-binding riboswitches

Kathryn D. Smith^a, Carly A. Shanahan^a, Emily L. Moore^b, Aline C. Simon^b, and Scott A. Strobel^{a,b,1}

^aDepartment of Chemistry, Yale University, New Haven, CT 06520; and ^bDepartment of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520

Edited by Juli Feigon, University of California, Los Angeles, CA, and approved March 18, 2011 (received for review December 20, 2010)

The bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) signaling pathway regulates biofilm formation, virulence, and other processes in many bacterial species and is critical for their survival. Two classes of c-di-GMP-binding riboswitches have been discovered that bind this second messenger with high affinity and regulate diverse downstream genes, underscoring the importance of RNA receptors in this pathway. We have solved the structure of a c-di-GMP-II riboswitch, which reveals that the ligand is bound as part of a triplex formed with a pseudoknot. The structure also shows that the guanine bases of c-di-GMP are recognized through noncanonical pairings and that the phosphodiester backbone is not contacted by the RNA. Recognition is quite different from that observed in the c-di-GMP-I riboswitch, demonstrating that at least two independent solutions for RNA second messenger binding have evolved. We exploited these differences to design a c-di-GMP analog that selectively binds the c-di-GMP-II aptamer over the c-di-GMP-I RNA. There are several bacterial species that contain both types of riboswitches, and this approach holds promise as an important tool for targeting one riboswitch, and thus one gene, over another in a selective fashion.

The ability of bacteria to adapt to their environment is essential for their survival. The bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) signaling system is one key pathway that allows bacteria to alter their behavior in response to changing environmental conditions (reviewed in refs. 1–4). Cyclic-di-GMP is a second messenger that is synthesized by diguanylate cyclases (DGCs) and degraded by specific phosphodiesterases in response to a variety of stimuli, including light, oxidative conditions, cell density, and antibiotics. Cyclic-di-GMP controls critical transitions such as the switch from a planktonic state to the formation of a biofilm (1, 2). It also controls virulence in many pathogenic bacterial species (5, 6).

To alter gene expression or protein activity in response to external or internal signals, c-di-GMP interacts with several proteins (1, 2, 7–17). In addition, two classes of RNAs have been reported that bind this second messenger, the c-di-GMP-I (class I) and c-di-GMP-II (class II) riboswitches (18, 19). Riboswitches are noncoding RNA elements usually located in the 5' UTR of a mRNA that regulate gene expression by selectively binding small molecules. These RNAs have a two-domain architecture, with the 5' domain (aptamer) responsible for ligand binding. Most often, genetic regulation is achieved when ligand binding in the first domain is coupled to an RNA structural rearrangement in the second domain leading to control at the transcriptional or translation level (reviewed in refs. 20–22).

Both class I and class II riboswitches have been shown to bind c-di-GMP and regulate gene expression (18, 19). The affinity of these RNAs for their ligand is very high; class I riboswitches have a K_d as tight as 10 pM (23) and the class II riboswitches have K_d s between 200 pM and 2 nM (ref. 19 and Table 1). These K_d s are substantially tighter than those of many of the known protein receptors (1, 2, 9–16). Whereas a few bacteria contain examples of both classes, several species lacking a class I riboswitch use

c-di-GMP for signaling and have exclusively a class II riboswitch (19). The majority of c-di-GMP-II riboswitches are found in organisms from the bacterial class Clostridia, including the prominent pathogen *Clostridium difficile*, which is responsible for severe gastrointestinal infection. One of the four class II riboswitches found in *C. difficile* strain 630 allosterically regulates a downstream group I intron, demonstrating that these RNAs can be involved in sophisticated forms of genetic control (19). Molecular understanding of conserved bacterial pathways, such as c-di-GMP signaling, is critical to manipulate the growth of these organisms, a particularly desirable goal in the case of pathogens like *C. difficile*.

We previously reported the structure and mutational analysis of the class I aptamer, showing that c-di-GMP is bound as part of a duplex (23, 24). The second messenger binds at the junction of three helices and is recognized asymmetrically through base pairing interactions. One guanine of the ligand, G_α , is contacted on the Hoogsteen face by a G of the riboswitch whereas the second guanine, G_β , forms a Watson–Crick base pair with a conserved C. This base pair formed with c-di-GMP is the first of a helical stem (23).

The predicted secondary structure of the class II riboswitch is distinct from that of class I. The class II aptamer was predicted to form three helices, a kink turn and a pseudoknot (19). The presence of these latter two elements, as well as the positions of the helices with respect to one another, suggested that the tertiary structure of the class II riboswitch is different from the class I aptamer. Phylogenetic sequence alignments showed that the majority of the conserved nucleotides were in the pseudoknot and in the regions immediately flanking this structural element. Additionally, in-line probing analysis revealed that these are the residues that become more structured upon c-di-GMP binding (19). Taken together these observations suggested that the pseudoknot region may have a role in ligand binding. If the class II aptamer recognizes c-di-GMP similarly to class I, we would expect that a conserved, unpaired C would be present within the secondary structure. However, there are no unpaired Cs in this RNA, suggesting that class II riboswitches recognize c-di-GMP through a different mechanism than used by class I (19).

Given these anticipated differences in second messenger recognition, we set out to determine the crystal structure of the class II riboswitch aptamer from *Clostridium acetobutylicum* bound to c-di-GMP. We used this structure to design a second

Author contributions: K.D.S., C.A.S., and S.A.S. designed research; K.D.S., C.A.S., E.L.M., and A.C.S. performed research; K.D.S. and C.A.S. analyzed data; and K.D.S. and S.A.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 3Q3Z).

¹To whom correspondence should be addressed. E-mail: scott.strobel@yale.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1018857108/-DCSupplemental.

messenger analog that is able to discriminate between class I and class II c-di-GMP-binding RNAs.

Results and Discussion

Structure Determination and Global Architecture of the c-di-GMP-II Aptamer. We solved the structure of the aptamer domain of a class II c-di-GMP riboswitch from *C. acetobutylicum* at 2.5 Å resolution (Fig. 1, Fig. S1, and Table S1). This riboswitch controls a gene that contains a carbohydrate binding domain along with cellulase and glycosyl hydrolase domains (19), implying a role in carbohydrate processing. To promote crystallization and phasing, we increased the GC content of the variable region of the P1 stem and added an optimal iridium hexamine binding site into a non-conserved region of the P3 helix (26). The RNA sequence that was amenable to structure determination contains a six base-pair P1 helix and one additional nucleotide on the 3' end (Fig. 1; see Fig. S2 for wild-type sequence). Iridium hexamine bound tightly to the engineered site in P3, as well as to several other G • U wobble pairs in the RNA. The structure was solved by single isomorphous replacement with anomalous scattering (SIRAS) phasing. In addition to binding within the RNA, c-di-GMP mediates intermolecular contacts in the crystal lattice (SI Text and Fig. S3). Two molecules are present in the asymmetric unit with an extensive RNA interface formed between them (Fig. S4).

The class II aptamer folds into a compact structure with c-di-GMP bound at the junction of P1, P2, and P4 (Fig. 1). The predicted helices, kink turn, and pseudoknot are all present in the structure. The pseudoknot is formed from nucleotides in the loops of the P3 and P1/P2 helices and is therefore an example of an H-H-type pseudoknot, also called an intramolecular kissing hairpin (27–30). Helices P1 and P4 form a stack, with P2 at a right angle. The kink turn bends the backbone, allowing P3 to span across the RNA and connect with P4. One of the nucleotides in J3/4, A34, forms a noncanonical base pair with U45, the nucleotide at the junction of P3 and P4. This base pair stacks directly on the P4 helix and forms a base quadruple with the U31/G46 pair in P3 (Fig. S4). This quadruple may help facilitate the sharp bend that occurs at the P3/P4 junction (Fig. 1B).

A triple helix is formed in the major groove of P4 by nucleotides from J2/4 as well as c-di-GMP and a residue from P4 itself (Fig. 2A and B). These triples are noncanonical interactions. Many additional noncanonical interactions are observed in this RNA. Twelve out of 37 pairings are not Watson–Crick or G • U wobble pairs, with six occurring in and around the binding pocket. Several of these are made through a single hydrogen bond.

The binding pocket is composed of nucleotides from P1, P2, J2/4, and P4, although the majority of the nucleotides that directly contact c-di-GMP are from P4 (Fig. 1C). Residues from P1 and J2/4 provide stacking contacts. P2 forms the edge of the pocket behind the phosphodiester backbone of c-di-GMP but does not directly contact the ligand.

Recognition of c-di-GMP in the Class II Aptamer. A unique feature of the pseudoknot allows the class II riboswitch to bind c-di-GMP as part of a triplex. Five noncanonical base triples are formed with P4 by nucleotides 61–63, G_α, and residue 70 (Fig. 2A and B and Fig. S5). These bases form a continuous stack that extends into the P1 helix. Even though G_α stacks with the other residues that form the triple helix, it is antiparallel with respect to these nucleotides. To accommodate this one antiparallel base, the c-di-GMP-II aptamer uses two RNA motifs, a U-turn and an S-turn, to insert a single base pair into P4 upside down and backward with respect to the rest of the helix (Fig. 2C, Figs. S6 and S7, and SI Text). The result is that whereas four out of the five base triple interactions are in the major groove of P4, the base triple formed between G_α, U37 and A69 is a minor groove contact (Fig. S5). This groove reversal results because U37 is inserted into the P4 helix from P3 in an antiparallel fashion. The turn in the RNA strand that follows U37 positions G38, the next nucleotide, to form the first base pair of P4 (Fig. 2C). The S-turn flips A69 so that it can form a Watson–Crick pair with U37 (Fig. 2C and Fig. S7). This A69/U37 pair therefore presents its minor groove edge to G_α (Fig. S5). All RNA structures in the Protein Data Bank (PDB) were examined for this feature using the program AMIGOS II (31) but no structures were found that contained a similar U-turn/S-turn architecture. Incorporation of c-di-GMP into this triplex is a key recognition feature of the class II aptamer and is facilitated by a unique groove-reversed pseudoknot.

Specific recognition of c-di-GMP is achieved through stacking interactions as well as noncanonical pairings to the guanine bases. Three conserved adenosines stack between and on either side of c-di-GMP. A70 intercalates between the two guanine bases of the ligand whereas A13 stacks below G_β and A61 stacks above G_α (Fig. 3A). Stacking is continued into the P1 helix as well as the third strand of the P4 triplex. In the few sequences where these nucleotides are not adenosines they are guanines, consistent with their role in stacking (Fig. S8). G_α forms a minor groove base triple by interacting with the sugar edge of A69 in the P4 helix. No contacts are made to the Hoogsteen face of G_α (Fig. 3B). G_β is contacted on its Watson–Crick and Hoogsteen faces by RNA

Fig. 1. Structure of the c-di-GMP-II aptamer domain from *C. acetobutylicum*. (A) Secondary structure of the c-di-GMP-II aptamer. P1 is shown in blue, the kink turn is shown in yellow, P3 is shown in purple, the pseudoknot helix (P4) is shown in green, and c-di-GMP is shown in red. Base pairs are shown using standard symbols (25). (B) Crystal structure of the c-di-GMP-II aptamer. Coloring is the same as in panel A. (C) Binding pocket. Coloring is the same as in panel A. $2F_o - F_c$ density is shown in gray, contoured at 1σ .

BIOCHEMISTRY

nitrogen. For iridium hexammine soaks, crystals were transferred to freezing solution supplemented with 1 mM iridium hexammine and allowed to equilibrate for 4–5 hours prior to flash freezing.

Structure determination. Native and single-wavelength anomalous dispersion (SAD) datasets were collected at Brookhaven National Laboratory beamline X29A. Data were processed using HKL2000 (33). The Matthews coefficient suggested that there were two molecules present in the asymmetric unit. Native Patterson maps revealed a translational twofold noncrystallographic symmetry (NCS) axis between the two molecules. SIRAS phasing was performed using a native dataset and a SAD dataset with a R_{iso} of 24%. The SHELX suite (through the hkl2map interface) was used to find four initial sites at a resolution cutoff of 3.2 Å (34). These sites were refined in MLPHARE (35) and initial phases were calculated. Difference Fourier methods were then employed to find additional sites. A total of 8 sites were included to produce the maps used for building. Density modification with twofold NCS averaging was performed with RESOLVE (36). The resolution was extended to 2.5 Å during density modification. Model building was done in COOT (37) and refinement was performed in Refmac (38). Figures were made using Pymol (39).

Synthesis of the 2'-O-methyl c-di-GMP analog. Cyanoethyl phosphate protected 2'-O-methoxy guanosine phosphoramidite and the 3-(4,4'-dimethoxytrityloxy)-2,2-(dicarboxymethylamido)propyl-1-O-succinoyl-long chain alkylamino-CPG (3'-CPR II CPG) solid support were purchased from Glen Research. Synthesis was performed on solid-phase (2 μmol scale) as previously described with slight adaptations (40–42) (see *SI Text*).

Measurement of dissociation constants. Radiolabeled c-di-GMP was enzymatically synthesized as previously described (23, 43) and incubated with RNA in folding buffer containing 10 mM NaCac, pH 6.8, 10 mM MgCl₂, and 10 mM KCl. Binding reactions were allowed to come to equilibrium at room temperature (approximately 21–23 °C) and bound and free c-di-GMP were separated by native PAGE and data were quantified as reported previously (23, 24). Competition experiments with the 2'-O-methyl c-di-GMP analog

were performed under similar conditions, only the radiolabeled c-di-GMP and unlabeled competitor ligand were premixed before adding wild-type class II RNA or GUAA mutant class I RNA (24) to a final concentration of 25 nM. Data were fit to the following equation as previously reported (44) to determine the dissociation constant of the 2'-O-methyl analog:

$$\text{FB} = \frac{1}{2[\text{*cdiG}]} \left\{ K_d^{\text{cdiG}} + \frac{K_d^{\text{cdiG}}[\text{Omethyl}]}{K_d^{\text{comp}}} + [\text{RNA}] + [\text{*cdiG}], \right. \\ \left. - \left[\left(K_d^{\text{cdiG}} + \frac{K_d^{\text{cdiG}}[\text{Omethyl}]}{K_d^{\text{comp}}} + [\text{RNA}] + [\text{*cdiG}] \right)^2 \right. \right. \\ \left. \left. - (4[\text{RNA}][\text{*cdiG}]) \right]^{\frac{1}{2}} \right\},$$

where FB = fraction bound; [*cdiG] = 5 pM; [Omethyl] = concentration of 2'-O-methyl analog; [RNA] = 25 nM.

In all cases, binding reactions were allowed to equilibrate for increasing amounts of time until no change in the final fraction bound was observed. This was approximately 1 hour for the wild-type class II sequence and mutants. For competition experiments, the class II RNA reached equilibrium after incubation overnight, whereas the class I RNA required a 48 hour incubation.

ACKNOWLEDGMENTS. We thank S. Myers, A. Soares, A. Héroux, and the beamline staff at X25 and X29A at the National Synchrotron Light Source at Brookhaven National Laboratory; M. Strickler and D. Keller (Yale); J. Wang for data processing and phasing help and advice; E. Butler for help with data collection; V. Singh for synthesis of iridium hexammine; Z. Weinberg, E. Lee, N. Sudarsan, R. Breaker, and other members of the Breaker Lab for helpful advice and discussions; K. Keating and A. Pyle for the AMIGOS II software; E. Butler, S. Lipchick, D. Hiller and other members of the Strobel Lab for helpful discussions; Z. Liang for the gift of the tDGC expression plasmid; and U. Jenal (University of Basel) for the gift of the PleD* expression plasmid. This work was supported by National Institutes of Health Grant GM022778 and US Department of Defense National Security Science and Engineering Faculty Fellowship Grant 1N00244-09-1-0070.

- Hengge R (2009) Principles of c-di-GMP signalling in bacteria. *Nat Rev Microbiol* 7:263–273.
- Schirmer T, Jenal U (2009) Structural and mechanistic determinants of c-di-GMP signalling. *Nat Rev Microbiol* 7:724–735.
- Römling U, Amikam D (2006) Cyclic di-GMP as a second messenger. *Curr Opin Microbiol* 9:218–228.
- Jenal U, Malone J (2006) Mechanisms of cyclic-di-GMP signaling in bacteria. *Annu Rev Genet* 40:385–407.
- Tamayo R, Pratt JT, Camilli A (2007) Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. *Annu Rev Microbiol* 61:131–148.
- Cotter PA, Stibitz S (2007) c-di-GMP-mediated regulation of virulence and biofilm formation. *Curr Opin Microbiol* 10:17–23.
- Paul K, Nieto V, Carlquist WC, Blair DF, Harshey RM (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.
- Boehm A, et al. (2010) Second messenger-mediated adjustment of bacterial swimming velocity. *Cell* 141:107–116.
- Krasteva PV, et al. (2010) *Vibrio cholerae* VpsT regulates matrix production and motility by directly sensing cyclic di-GMP. *Science* 327:866–868.
- Leduc JL, Roberts GP (2009) Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of *Xanthomonas axonopodis* pv. citri. *J Bacteriol* 191:7121–7122.
- Chin K-H, et al. (2009) The c-AMP receptor-like protein CLP is a novel c-di-GMP receptor linking cell-cell signaling to virulence gene expression in *Xanthomonas campestris*. *J Mol Biol* 396:646–662.
- Tao F, He Y-W, Wu D-H, Swarup S, Zhang L-H (2009) The cNMP domain of *Xanthomonas campestris* global regulator Clp defines a new class of c-di-GMP effectors. *J Bacteriol* 192:1020–1029.
- Hickman JW, Harwood CS (2008) Identification of FleQ from *Pseudomonas aeruginosa* as a c-di-GMP-responsive transcription factor. *Mol Microbiol* 69:376–389.
- Duerig A, et al. (2009) Second messenger-mediated spatiotemporal control of protein degradation regulates bacterial cell cycle progression. *Genes Dev* 23:93–104.
- Newell PD, Monds RD, O'Toole GA (2009) LapD is a bis-(3',5')-cyclic dimeric GMP-binding protein that regulates surface attachment by *Pseudomonas fluorescens* Pf0-1. *Proc Natl Acad Sci USA* 106:3461–3466.
- Lee VT, et al. (2007) A cyclic-di-GMP receptor required for bacterial exopolysaccharide production. *Mol Microbiol* 65:1474–1484.
- Amikam D, Galperin MY (2006) PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics* 22:3–6.
- Sudarsan N, et al. (2008) Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* 321:411–413.
- Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR (2010) An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science* 329:845–848.
- Garst AD, Edwards AL, Batey RT (2010) Riboswitches: Structures and mechanisms. *Cold Spring Harb Perspect Biol*, 10.1101/cshperspect.a003533.
- Serganov A (2009) The long and the short of riboswitches. *Curr Opin Struct Biol* 19:251–259.
- Roth A, Breaker RR (2009) The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem* 78:305–334.
- Smith K, et al. (2009) Structural basis of ligand binding by a c-di-GMP riboswitch. *Nat Struct Mol Biol* 16:1218–1223.
- Smith KD, Lipchick SV, Livingston AL, Shanahan CA, Strobel SA (2010) Structural and biochemical determinants of ligand binding by the c-di-GMP riboswitch. *Biochemistry* 49:7351–7359.
- Leontis NB, Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* 7:499–512.
- Keel AY, Rambo RP, Batey RT, Kieft JS (2007) A general strategy to solve the phase problem in RNA crystallography. *Structure* 15:761–772.
- Sperschneider J, Datta A, Wise M (2011) Heuristic RNA pseudoknot prediction including intramolecular kissing hairpins. *RNA* 17:27–38.
- Giedroc DP, Cornish PV (2009) Frameshifting RNA pseudoknots: Structure and mechanism. *Virus Res* 139:193–208.
- Brierley I, Pennell S, Gilbert RJC (2007) Viral RNA pseudoknots: Versatile motifs in gene expression and replication. *Nat Rev Microbiol* 5:598–610.
- Staple DW, Butcher SE (2005) Pseudoknots: RNA structures with diverse functions. *PLoS Biol* 3:956–959.
- Wadley L, Keating K, Duarte C, Pyle A (2007) Evaluating and learning from RNA pseudotorsional space: Quantitative validation of a reduced representation for RNA structure. *J Mol Biol* 372:942–957.
- Rao F, et al. (2009) Enzymatic synthesis of c-di-GMP using a thermophilic diguanylate cyclase. *Anal Biochem* 389:138–142.
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276:307–326.
- Pape T, Schneider T (2004) HKL2MAP: A graphical user interface for phasing with SHELX programs. *J Appl Crystallogr* 37:843–844.
- Collaborative Computational Project 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
- Tervilliger T (2000) Maximum likelihood density modification. *Acta Crystallogr D* 56:965–972.
- Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
- Winn MD, Isupov MN, Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D* 57:122–133.
- DeLano WL (2008) *Pymol* (DeLano Scientific, San Carlos, CA).
- Kiburu I, Shurer A, Yan L, Sintim HO (2008) A simple solid-phase synthesis of the ubiquitous bacterial signaling molecule, c-di-GMP and analogues. *Mol Biosyst* 4:518–520.

41. Alazzouzi E, Escaja N, Grandas A, Pedroso E (1997) A straightforward solid-phase synthesis of cyclic oligodeoxyribonucleotides. *Angew Chem Int Ed Engl* 36:1506–1508.
42. Micura R (1999) Cyclic oligoribonucleotides (RNA) by solid-phase synthesis. *Chemistry Eur J* 5:2077–2082.
43. Paul R, et al. (2004) Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev* 18:715–727.
44. Strobel SA, Cech TR (1993) Tertiary interactions with the internal guide sequence mediate docking of the P1 helix into the catalytic core of the Tetrahymena ribozyme. *Biochemistry* 32:13593–13604.