nature structural & molecular biology

Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch

Nadia Kulshina[1,2,4], Nathan J Baird[2–4] & Adrian R Ferré-D'Amaré[1–3]

The cyclic diguanylate (bis-(3′-5′)-cyclic dimeric guanosine monophosphate, c-di-GMP) riboswitch is the first known example of a gene-regulatory RNA that binds a second messenger. c-di-GMP is widely used by bacteria to regulate processes ranging from biofilm formation to the expression of virulence genes. The cocrystal structure of the c-di-GMP responsive GEMM riboswitch upstream of the *tfoX* gene of *Vibrio cholerae* reveals the second messenger binding the RNA at a three-helix junction. The two-fold symmetric second messenger is recognized asymmetrically by the monomeric riboswitch using canonical and noncanonical base-pairing as well as intercalation. These interactions explain how the RNA discriminates against cyclic diadenylate (c-di-AMP), a putative bacterial second messenger. Small-angle X-ray scattering and biochemical analyses indicate that the RNA undergoes compaction and large-scale structural rearrangement in response to ligand binding, consistent with organization of the core three-helix junction of the riboswitch concomitant with binding of c-di-GMP.

c-di-GMP is a second messenger widespread in bacteria, where it regulates complex physiological processes associated with transition between motile-planktonic and sedentary-adhesive lifestyles and influences phenomena such as the formation of biofilms, the expression of virulence genes and the persistence of infections (reviewed in refs. 1–4). Diguanylate cyclases (DGCs), characterized by a GGDEF domain, synthesize c-di-GMP from two GTP molecules, and c-di-GMP is degraded by phosphodiesterases (PDEs) with EAL and HD-GYP domains in response to various signals. Whole-genome sequencing demonstrates the ubiquity of proteins with these signatures in bacteria, where they are often found in large numbers in the genome of one organism. However, despite the ubiquity of DGCs and PDEs, only four types of effector proteins are known—the PilZ family of proteins, the FleQ transcription factor, PelD and the I-site effectors. Thus, the mechanisms by which global changes in gene expression are modulated by c-di-GMP are a subject of intense research[1–4].

Computational searches delineated a bacterial RNA motif widely associated with genes related to the environment, membranes and motility (GEMM)[5]. The diversity of genes associated with the GEMM motif is much broader than is typical of riboswitches, which generally only regulate a limited set of genes responsible for the biosynthesis, transport or degradation of their cognate metabolites. Therefore, it was hypothesized that the GEMM motif may respond to an intracellular signaling molecule[5]. Recently, it was discovered that the GEMM motif specifically binds c-di-GMP and serves a direct role in the regulation of gene expression[6], confirming predictions of the existence of such a riboswitch[1,2]. The c-di-GMP riboswitch is the first known example of a gene-regulatory RNA that recognizes a second messenger. It binds c-di-GMP with high specificity and K_d ~1 nM. This affinity is ~1,000 times tighter than that for the second messenger of an *Escherichia coli* PilZ c-di-GMP binding protein[6] and is comparable to the tightest RNA–small molecule interactions known[7]. The broad distribution among bacteria of this RNA domain[6], the physiological importance of c-di-GMP, and its involvement in controlling virulence of many pathogens make this riboswitch an important target for structure-function studies.

To provide a framework with which to understand how this RNA recognizes the bacterial second messenger, we have determined the cocrystal structure of a canonical c-di-GMP riboswitch from *V. cholerae*. The structure and previous biochemical characterization suggest that this riboswitch undergoes a conformational change upon ligand binding. We analyzed the solution behavior of the riboswitch through a combination of small-angle X-ray scattering (SAXS) and enzymatic probing. These studies indicate that allosteric binding of c-di-GMP modulates the global architecture of the RNA, which switches between markedly different free and c-di-GMP–bound conformations.

RESULTS

Crystal structure determination

Previous analyses of the *V. cholerae tfoX* c-di-GMP riboswitch[6] delineated the minimal 5′ and 3′ RNA boundaries (the 'aptamer domain') needed for tight binding ($K_d \leq 10$ nM) to the second messenger. A phylogenetically conserved helix (paired region P2) is closed by a loop (L2) that is highly variable[5]. Our crystallization construct encompasses these boundaries and incorporates[8] a binding site for the spliceosomal protein U1A in place of L2. Consistent with lack of conservation of L2, this construct binds both c-di-GMP and the U1A RNA-binding domain (RBD) (**Supplementary Fig. 1**).

[1]Molecular and Cellular Biology Program, University of Washington, Seattle, Washington, USA. [2]Division of Basic Sciences and [3]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to A.R.F.-D. (aferre@fhcrc.org).

Received 7 July; accepted 15 September; published online 8 November 2009; doi:10.1038/nsmb.1701

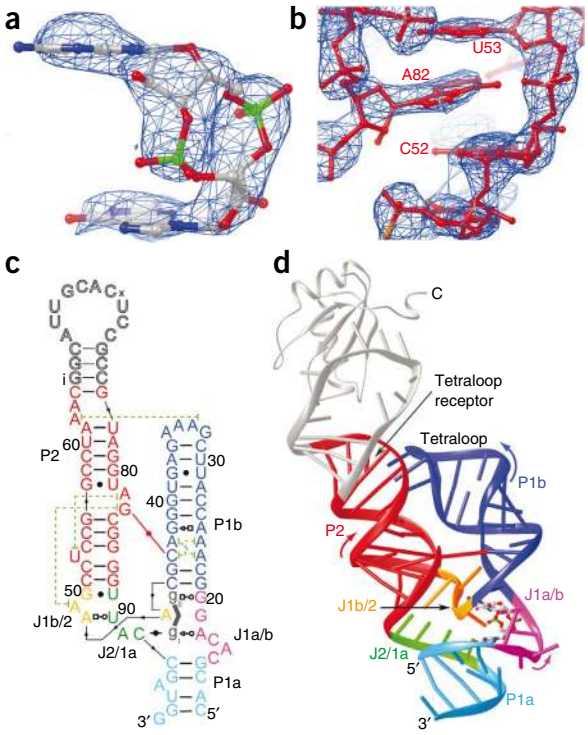

Figure 1 Overall structure of the c-di-GMP riboswitch. (**a**) Unbiased $|F_o| - |F_c|$ electron density corresponding to the bound c-di-GMP before it was included in the crystallographic model, superimposed on the refined model of the ligand (map contoured at 3.0 s.d.) (**b**) Portion of a composite simulated annealing-omit $2|F_o| - |F_c|$ Fourier synthesis[42] calculated with the final crystallographic model, contoured around A82 at 1.5 s.d. (**c**) Schematic representation of the structure of the riboswitch bound to c-di-GMP. Thin black lines with arrowheads depict connectivity, dashed green lines long-range stacking interactions and the red line the interhelical base pair. Leontis-Westhof symbols[43] depict noncanonical base pairs. Except for the U1A binding site, the numbering scheme is that of ref. 6, which is used throughout. (**d**) Cartoon representation of the three-dimensional structure. Color coding as in **c**. The U1A-RBD is shown as a gray ribbon.

We solved the structure *de novo* by molecular replacement using the U1A-RBD cognate complex[9] and short A-form duplexes of arbitrary sequence as search models, analogous to what has been reported for two ribozymes[10,11] and an unrelated riboswitch[12]. This procedure[13] unambiguously revealed the bound c-di-GMP in residual electron density maps phased with models that had never included the second messenger (**Fig. 1a**), and it allowed complete tracing of both RNA molecules in the asymmetric unit (**Fig. 1b**). The crystallographic model (R_{free} = 29.2% at 3.2 Å resolution) has a mean coordinate precision of 0.5 Å (Online Methods and **Table 1**).

Overall ligand-bound structure

The c-di-GMP riboswitch is h-shaped (**Fig. 1c**,**d**). In addition to the two predicted helices (hereafter, P1b and P2), a third helix, P1a, is formed by the 5′- and 3′-most nucleotides of the aptamer domain, which were thought[5,6] to be single stranded. The c-di-GMP binding site is formed by nucleotides that join the three helical segments (J1a/b, J1b/2 and J2/1a). One of the guanine bases of the bound c-di-GMP (hereafter, g_I) stacks on the uppermost base pair of P1a, while the other (g_{II}) stacks underneath the bottom base pair of P1b. The two nucleobases of the second messenger sandwich that of A47 from J1b/2, thus maintaining continuous base stacking between P1a and P1b. As predicted by phylogenetic analysis, P1b is capped by a GAAA tetraloop that makes canonical interactions (mediated by a ribose zipper and stacking of the base of A33 under that of A62) with a tetraloop receptor in P2 (refs. 5,6).

A second interhelical interaction mediating side-by-side packing of P1b and P2 is a *cis*-Watson-Crick pair between C44 (from P1b) and G83 (extruded from P2). Both nucleotides adopt the *anti* conformation, unlike in the hairpin ribozyme[14,15] where the *syn* conformation of the guanosine in the interhelical Watson-Crick pair enforces a closer approach of the two helices. The extrusion of G83 is accomplished by an S-shaped turn of the backbone, constructed by juxtaposition of the G57•U81 wobble pair and two base triples, that results in marked underwinding of P2 (**Supplementary Fig. 2a**). C44 projects into the minor groove of the overwound P1b. Overwinding results from cross-strand stacking of A24 on C44 and A43 on A23 and from formation of a sheared pair between A25 and G42 (**Supplementary Fig. 2b**). In addition to these nonstandard conformations of P1b and P2, the interhelical C44-G83 base pair appears to be stabilized by stacking of G83 on A49 (from J1b/2). Consistent with the importance of this interhelical base pair, nucleotides corresponding to C44 and G83 are conserved in at least 97% of the 503 c-di-GMP riboswitches discovered by Sudarsan *et al.*[6] in diverse bacterial genomes. Similarly, nucleotides whose interactions result in overwinding of P1b (corresponding to A23, A24 and A25) and underwinding of P2 (corresponding to U81 and A82) are 75–97% conserved[6].

Second messenger recognition

J1b/2 contributes to forming the c-di-GMP–binding site and to organizing the global architecture of the riboswitch. Nucleotides corresponding to A47, A48 and A49 are at least 97% conserved across all known c-di-GMP riboswitches and are typically (>75%) followed by a guanine[6]. J1b/2 alternates between the P1a/b and the P2 stacks, thereby joining them (**Figs. 1d** and **2a**). Whereas A47 together with the bound c-di-GMP forms part of the continuous stack between P1a

Table 1 Crystallographic data and refinement statistics

	Crystal I
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	31.8, 91.0, 280.1
Resolution (Å)	30.0–3.2 (3.25–3.2)*
R_{merge}	17.9 (43.8)
$I/\sigma I$	7.9 (2.2)
Completeness (%)	84.0 (75.4)
Redundancy	4.4 (2.7)
Refinement	
Resolution (Å)	30.0–3.2 (3.25–3.2)
No. reflections	11,257 (276)
R_{work} / R_{free} (%)	22.2/29.2
No. atoms	
RNA	3,984
Protein	1,429
c-di-GMP	92
B-factors	
RNA	46.9
Protein	11.6
c-di-GMP	31.8
r.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.4

One crystal was used for the entire project.
*Values in parentheses are for highest-resolution shell.

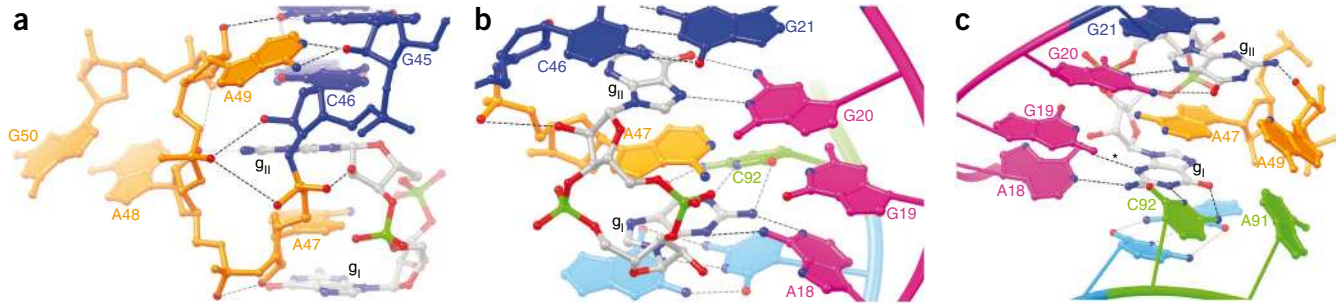

Figure 2 Specific binding of c-di-GMP. (**a**) The J1b/2 region stitches the three-helix junction and provides a nucleotide that intercalates between the c-di-GMP nucleobases. View from the direction of **Figure 1d**. Note alternation of J1b/2 residues between right (P1a/P1b) and left (P2). (**b**) View of the binding site from the major groove. (**c**) View from the minor groove. Additional interactions between the Watson-Crick faces of G19 and A47 and the sugar edge of A49 and g_{II} are suggested by the location of hydrogen-bond donors and acceptors (colored blue or red denoting nitrogen or oxygen, respectively) but remain uncertain at the current resolution limit. Dashed lines denote putative hydrogen bonds (distance cutoff = 3.8 Å; note that the mean precision of the atomic coordinates is 0.5 Å). In **c**, the hydrogen bond marked (*) is between the N6 of A18 and the N3 of g_I, not between G19 and the c-di-GMP.

and P1b, A48 forms part of the P2 stack by pairing with U90 through its Hoogsteen edge (**Supplementary Fig. 3**). The next residue of J1b/2, A49, flips back into P1b, stacking underneath the interhelical C44-G83 pair. Finally, G50 pairs with U89 and forms part of the P2 stack.

Binding to the riboswitch buries 69% of the solvent-accessible surface area of the c-di-GMP molecule. The guanine bases of the second messenger face the same direction (**Supplementary Fig. 4**) and are deeply buried, whereas the 12-member ribose phosphate ring of c-di-GMP is partially exposed in the broad major groove and is flanked on one side by nucleotides of J1a/b (**Figs. 1d** and **2b**). At the present resolution, cations and water molecules that may mediate additional contacts between the RNA and the ribose and phosphate moieties of c-di-GMP are unresolved. The g_I nucleobase of the c-di-GMP makes a Watson-Crick base pair with the highly conserved[6] C92 of J2/1a (**Fig. 2c**), and is also recognized through its sugar-edge face by A18. The latter nucleotide is not highly conserved, suggesting that this interaction varies among different c-di-GMP riboswitches. The Watson-Crick face of G20 recognizes the Hoogsteen edge of g_{II}, which is upside down relative to g_I. Across all known c-di-GMP riboswitches, either a guanosine or an adenosine residue is present at the position corresponding to G20 of this riboswitch[6], indicating that the specific interactions used to recognize g_{II} are somewhat variable.

Ligand binding-induced global conformational change

In-line probing analysis of the *V. cholerae tfoX* riboswitch[6] revealed reduction in self-scission of nucleotides 13–20, 47 and 93–96 upon c-di-GMP binding. Our structure shows that residues 13–20 and 93–96 assemble into P1a and J1a/b in the second messenger–bound form of the RNA, and that residue 47 intercalates between the two nucleobases of c-di-GMP. Together, these results suggest that the riboswitch undergoes folding induced by c-di-GMP binding. However, it is not clear whether the metabolite-induced folding is localized to the binding pocket or if the RNA undergoes a global rearrangement upon binding c-di-GMP. We tested the nature of the structural response elicited by c-di-GMP using SAXS.

Guinier analysis[16] revealed that the radius of gyration (R_g) of the c-di-GMP riboswitch aptamer compacts in response to both c-di-GMP and Mg^{2+} (**Supplementary Table 1**). At a Mg^{2+} concentration of 0.5 mM (below the physiologic level), the R_g (~32 Å) is insensitive to saturating c-di-GMP. However, in the presence of physiologic (2.5 mM) Mg^{2+}, the free RNA compacts (R_g = 28.5 Å) and undergoes a further c-di-GMP–dependent compaction (R_g = 23.9 Å). This latter R_g is the same, within experimental precision, as that calculated with a model of the riboswitch based on our crystal structure (with a wild type–length L2 modeled in; **Supplementary Table 2** and **Supplementary Fig. 5**). At highly stabilizing Mg^{2+} (10 mM), the RNA compacts even in the absence of c-di-GMP to an R_g that is comparable to that of its c-di-GMP–bound form at physiologic Mg^{2+} (**Supplementary Table 1**). However, Kratky analysis[16] (**Fig. 3a**) showed that even at 10 mM Mg^{2+}, the free RNA shows behavior at high q indicative of local disorder, suggesting that the c-di-GMP–binding pocket and/or P1a remain unfolded in the absence of ligand (q is the momentum transfer and is defined as $q = 4\pi \sin(\theta)/\lambda$, where λ is the X-ray wavelength and the scattering angle is 2θ). $P(r)$ analysis[16] (**Fig. 3b**) revealed a marked ligand-induced compaction of the RNA at physiologic Mg^{2+}, and also a modest c-di-GMP–induced conformational change at 10 mM Mg^{2+}, likely indicative of folding of the second messenger–binding pocket.

To gain further insight into the nature of the c-di-GMP–induced conformational change, we calculated low-resolution molecular envelopes[16] using SAXS data collected at physiologic Mg^{2+} concentration. Reconstructions based on SAXS data from the ligand-bound state using a range of starting maximum molecular dimension (D_{max}) values robustly converged into an envelope with high shape complementarity (correlation coefficient = 0.94) to our cocrystal structure (with the U1A binding site replaced by a model of length identical to that of

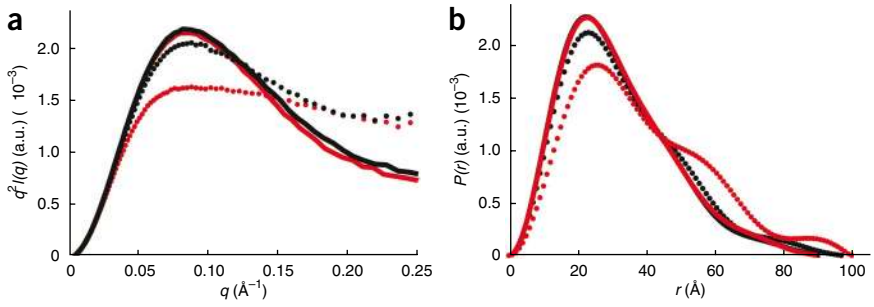

Figure 3 c-di-GMP- and Mg^{2+}-induced size and shape changes of the riboswitch, monitored by SAXS. Red indicates 2.5 mM Mg^{2+}; black, 10 mM Mg^{2+}. Dots denote c-di-GMP–free conditions; solid lines, c-di-GMP–bound conditions (a.u., arbitrary units). (**a**) Kratky plot suggests local disorder of the ligand-free riboswitch even under high $MgCl_2$ concentration (black dots). (**b**) Electron-pair probability analysis reveals compaction of the riboswitch upon binding the second messenger.

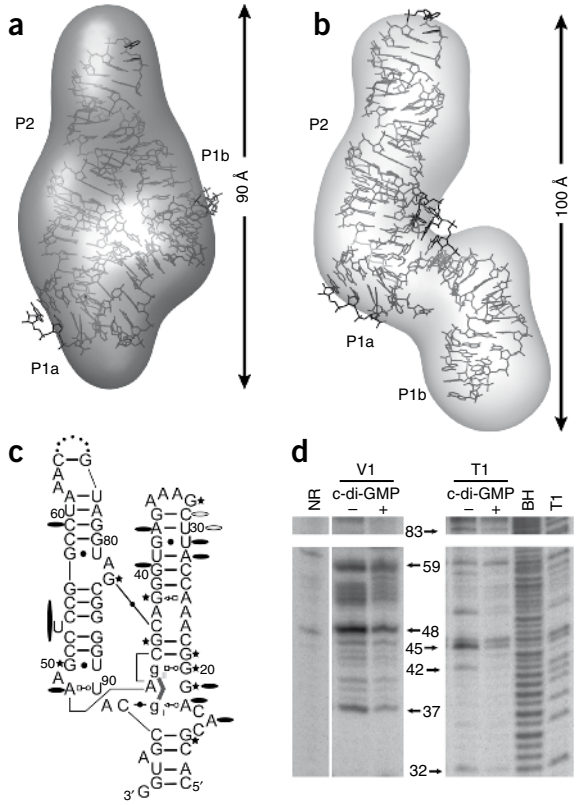

Figure 4 Global rearrangement of the riboswitch induced by c-di-GMP binding. (**a**) Low-resolution molecular envelope reconstruction based on SAXS data for the riboswitch in the presence of c-di-GMP (at 2.5 mM Mg^{2+}). The maximum linear dimension of the reconstruction is ~90 Å. (**b**) Molecular envelope based on SAXS data of the riboswitch in the absence of c-di-GMP (at 2.5 mM Mg^{2+}). The maximum linear dimension of the reconstruction is ~100 Å. The two arms of this reconstruction have dimensions approximately corresponding to those of P2 and P1b. (**c**) Results of nuclease probing summarized on the bound-state secondary structure. Stars denote RNase T1 protection upon c-di-GMP binding; filled ovals, RNase V1 protection upon c-di-GMP binding; open ovals, increased RNase V1 cleavage upon c-di-GMP binding. (**d**) Section of the autoradiogram highlighting protections observed due to their presence in the interhelical interface, or indirectly, from stabilization of the interface between P1b and P2. (See **Supplementary Fig. 10** for complete autoradiogram; this panel shows the lanes corresponding to 3.0 mM Mg^{2+}.) + and – denote conditions with 70 μM and 0 μM c-di-GMP, respectively. NR and BH denote no reaction and base hydrolysis. The numbers between the two gel images correspond to the sequence numbering scheme in **c** and indicate regions of protection from the nucleases due to the c-di-GMP induced stabilization of P1b and P2 in a side-by-side orientation.

wild-type L2), and D_{max} ~90 Å (**Fig. 4a** and **Supplementary Figs. 5** and **6a**,**b**). Reconstructions based on the SAXS data from the ligand-free state converged on a more elongated molecular envelope (D_{max} ~100 Å) with two prominent arms (**Fig. 4b** and **Supplementary Fig. 6c**,**d**). The shape of this envelope is distinctly different from that of the c-di-GMP–bound state. In addition to obvious limitations due to its low resolution, interpretation of this envelope has the caveat that the ligand-free form of the RNA might explore multiple conformations in solution. However, we found that reconstructions based on ligand-free SAXS data do converge to molecular envelopes that are self-consistent at each of the three Mg^{2+} concentrations analyzed. Moreover, the variability of the reconstructions for any one of the three ligand-free conditions is comparable to that of the reconstructions from the ligand-bound SAXS data (**Supplementary Fig. 7**). We conclude that building molecular models to fit the SAXS-based molecular envelope reconstruction of the ligand-free riboswitch is justified, keeping in mind that the level of detail implied by such models is limited to the approximate location and orientation of helices.

In-line probing of the ligand-free riboswitch suggested that P1a is destabilized in the absence of c-di-GMP[6]. A molecular model of the free state based on our cocrystal structure in which the two strands of P1a have simply been separated, however, fails to recapitulate the experimentally determined change in R_g between the free and c-di-GMP–bound states (ΔR_g of melted P1a model = 2.2 Å; experimental ΔR_g = 4.6 Å) and fits the SAXS envelope poorly (**Supplementary Table 2** and **Supplementary Fig. 8**). We noted that the dimensions of P2 fit the long arm of the free-state SAXS–based envelope, and the dimensions of P1b are a good fit for the short arm of this envelope (**Fig. 4b**, and **Supplementary Fig. 6c**,**d**). A model of the c-di-GMP–free state constructed by manually rotating P1b by ~180° with the center of rotation at A47 (**Fig. 4b**) has a calculated ΔR_g = 4.7 Å (**Supplementary Table 2**) and a correlation coefficient with the envelope of 0.91. Moreover, the scattering profile calculated from this model closely approximates the experimental scattering data, except at high q (**Supplementary Fig. 9**). Therefore, we assign the two arms of the ligand-free SAXS reconstruction to helices P1b and P2, while emphasizing that this model cannot capture molecular details (as underscored by the poor high-q fit).

Our tentative model of the ligand-free form of the RNA implies that in addition to unfolding of the c-di-GMP binding pocket (and possibly P1a), the interhelical interactions mediating side-by-side packing of P1b and P2 in the c-di-GMP–bound state are broken in the ligand-free state. To test this, we subjected the riboswitch to nuclease probing. Protection from nucleases T1 and V1 is indicative of structure formation[17,18]. Steric clash between nucleases and a structured substrate RNA can also protect RNA segments that would otherwise be readily cleaved[19–23]. For instance, in two separate studies of RNase P, it was found that relative protection from nuclease V1 was consistent with occlusion of nucleotides due to tertiary structure[17] or protein binding[24]. RNase V1 probing of the c-di-GMP riboswitch revealed protection patterns that suggest occlusion due to structure formation. Specifically, nucleotides in the c-di-GMP binding pocket and three nucleotides found at the interhelical interface (A37, A48, C59) became protected from cleavage by nuclease V1 upon addition of c-di-GMP (**Fig. 4c**,**d** and **Supplementary Fig. 10**). In addition, we found protection of guanosine residues of the riboswitch from RNase T1 in the presence of c-di-GMP, consistent with stabilization of P1a, J1a/b and interhelical regions of the RNA. Noteworthy are c-di-GMP–induced RNase T1 protections of G32 (of the tetraloop) and G83 (of the interhelical G83-C44 base pair). These nucleotides are likely protected by their participation in P1b-P2 interactions. Our nuclease probing results are therefore consistent with compaction of the riboswitch induced by c-di-GMP binding, and with a free-state structure that lacks the side-by-side packing of P1b and P2 of the bound state.

DISCUSSION

The ligand-bound structures of the purine[25,26] and c-di-GMP riboswitches show some general resemblance to each other: both RNAs are comprised of two helical stacks that pack side by side, their ligand-binding pockets are both at a three-helix junction, and in both cases the bound small molecule continues the helical stack between two of the three helices. Consistent with the dissimilarity between a free purine and c-di-GMP, the nature of the binding pockets

of the two corresponding riboswitches is completely different. The long-range interactions that hold their respective two helical stacks in side-by-side arrangements are distinctly different as well, suggesting independent evolutionary origins for these two riboswitch classes.

Although the c-di-GMP riboswitch uses separate binding pockets to recognize the two nucleobases of the second messenger, the presence of both, in cyclically linked form, is required for high-affinity binding (ref. 6 and **Supplementary Fig. 11a**). Our structure predicts that this riboswitch should discriminate against c-di-AMP, a recently discovered putative second messenger of archaea and bacteria[27]. If bound in an equivalent manner, the nucleobase of c-di-AMP corresponding to g_I would be forced to make unfavorable hydrogen bonds with the highly conserved C92, and the other nucleobase would not be able to make the interactions with G20 and A49 in which g_{II} engages (**Fig. 2c**). This is borne out by electrophoretic mobility shift analysis, in which the riboswitch does not bind c-di-AMP even at a ligand concentration of 70 μM (**Supplementary Fig. 11b**). Because bacteria use both c-di-GMP and c-di-AMP, it is possible that paralogs of the *V. cholerae tfoX* riboswitch might exist that recognize c-di-AMP or that do not discriminate between the two second messengers, allowing for cross-talk between pathways. The purine riboswitches are selective for either adenine or guanine, and their selectivity derives almost exclusively from a single pyrimidine residue that Watson-Crick base-pairs with their respective ligands[25,26]. Because the c-di-GMP riboswitch uses Watson-Crick pairing as only one of multiple strategies to recognize g_I of its cognate ligand, and does not use Watson-Crick pairing to recognize g_{II}, it is unlikely that two point mutations would result in high-affinity (K_d ~1 nM) binding of c-di-AMP. Sequence alignments[6] suggest that some GEMM riboswitches have a uridine at the position that corresponds to C92, raising the possibility that they bind c-di-AMP. Although the ligand specificity of such RNAs will have to be determined experimentally, riboswitches have been shown to have considerable plasticity in their binding sites[28,29], and it is possible that a few mutations may allow evolution of a high-affinity c-di-AMP riboswitch from a c-di-GMP riboswitch.

Except for the *glmS* ribozyme-riboswitch[30,31], most riboswitches function by a mechanism in which stabilization of the aptamer domain (through the free energy of binding to their cognate ligand) results in sequestration of an RNA segment that would otherwise form part of an alternative structure: for instance, a rho-independent transcriptional terminator or a ribosome binding site[32–34]. Typically, this sequestered segment is part of P1 in the ligand-bound state. Analyses of the genomic contexts of c-di-GMP riboswitches[5] and *in vivo* reporter assays[6] suggest that such a mechanism is operative for at least some of these riboswitches. Our combined crystallographic, SAXS and biochemical analyses of the c-di-GMP riboswitch show that it undergoes a pronounced global structural rearrangement in response to ligand binding (**Fig. 4**). This rearrangement probably involves folding of the J1a/b, J1b/2 and J2/1b regions (**Fig. 1c**) to form the second messenger-binding site, resulting in stabilization of helix P1a and, concomitantly, in stabilization of the side-by-side orientation of P1b and P2 through interhelical contacts and the ligand-induced structure of the three-helix junction itself. Comparison with other studies shows that such a conformational switch is not a general property of riboswitches. SAXS analysis of the glycine riboswitch[35] demonstrated that it undergoes pronounced Mg^{2+}- and ligand-dependent conformational changes. In contrast, the lysine riboswitch[36,37] adopts essentially indistinguishable three-dimensional structures in the presence or absence of the amino acid and even in the absence of free Mg^{2+}. Thus, it is possible that the large conformational switch we have uncovered is solely a consequence of the stabilization of the riboswitch aptamer domain structure by ligand binding. Alternatively, the conformational switch may have evolved to modulate the affinity[38,39] of the riboswitch or its rate of folding[40,41] so as to optimize its switching behavior in its physiological context.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. Protein Data Bank: Coordinates and structure factor amplitudes for the *V. cholerae tfoX* c-di-GMP riboswitch in complex with c-di-GMP have been deposited with accession code 3IWN.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank the staff of ALS beamline 8.2.2 and J. Bolduc for assistance with synchrotron and home laboratory single-crystal diffraction data collection, respectively, L. Guo from BioCAT at the Advanced Photon Source (APS) for assistance with SAXS data collection and T. Hamma, J. Pitt, J. Posakony, A. Roll-Mecak and H. Suga for discussions. Use of the APS was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract No. W-31-109-ENG-38. BioCAT is a US National Institutes of Health–supported Research Center (RR-08630). This work was supported by the Howard Hughes Medical Institute (HHMI) and the W.M. Keck Foundation. A.R.F.-D. is an Investigator of the HHMI.

AUTHOR CONTRIBUTIONS

N.K. designed and prepared RNA constructs, analyzed ligand binding, obtained crystals, carried out diffraction data collection and participated in structure determination and in SAXS data collection. N.J.B. participated in SAXS data collection, analyzed the SAXS data and designed, performed and analyzed the nuclease probing experiments. A.R.F.-D. participated in diffraction data collection, structure determination and refinement. All authors contributed to manuscript preparation.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Jenal, U. & Malone, J. Mechanisms of cyclic-di-GMP signaling in bacteria. *Annu. Rev. Genet.* **40**, 385–407 (2006).
2. Tamayo, R., Pratt, J.T. & Camilli, A. Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. *Annu. Rev. Microbiol.* **61**, 131–148 (2007).
3. Wolfe, A. & Visick, K. Get the message out: cyclic-di-GMP regulates multiple levels of flagellum-based motility. *J. Bacteriol.* **190**, 463–475 (2008).
4. Pesavento, C. & Hengge, R. Bacterial nucleotide-based second messengers. *Curr. Opin. Microbiol.* **12**, 170–176 (2009).
5. Weinberg, Z. *et al.* Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.* **35**, 4809–4819 (2007).
6. Sudarsan, N. *et al.* Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* **321**, 411–413 (2008).
7. Xiao, H., Edwards, T.E. & Ferré-D'Amaré, A.R. Structural basis for specific, high-affinity tetracycline binding by an *in vitro* evolved aptamer and artificial riboswitch. *Chem. Biol.* **15**, 1125–1137 (2008).
8. Ferré-D'Amaré, A.R. & Doudna, J.A. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.* **295**, 541–556 (2000).
9. Oubridge, C., Ito, N., Evans, P.R., Teo, C.-H. & Nagai, K. Crystal structure at 1.92Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438 (1994).
10. Robertson, M.P. & Scott, W.G. The structural basis of ribozyme-catalyzed RNA assembly. *Science* **315**, 1549–1553 (2007).
11. Xiao, H., Murakami, H., Suga, H. & Ferré-D'Amaré, A.R. Structural basis of specific tRNA aminoacylation by a small *in vitro* selected ribozyme. *Nature* **454**, 358–361 (2008).
12. Klein, D., Edwards, T. & Ferré-D'Amaré, A. Cocrystal structure of a class I $preQ_1$ riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat. Struct. Mol. Biol.* **16**, 343–344 (2009).
13. Robertson, M.P. & Scott, W.G. A general method for phasing novel complex RNA crystal structures without heavy-atom derivatives. *Acta Crystallogr D* **64**, 738–744 (2008).

14. Rupert, P.B. & Ferré-D'Amaré, A.R. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **410**, 780–786 (2001).
15. Ferré-D'Amaré, A.R. The hairpin ribozyme. *Biopolymers* **73**, 71–78 (2004).
16. Lipfert, J. & Doniach, S. Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 307–327 (2007).
17. Baird, N., Westhof, E., Qin, H., Pan, T. & Sosnick, T. Structure of a folding intermediate reveals the interplay between core and peripheral elements in RNA folding. *J. Mol. Biol.* **352**, 712–722 (2005).
18. Ehresmann, C. *et al.* Probing the structure of RNAs in solution. *Nucleic Acids Res.* **15**, 9109–9128 (1987).
19. Dock-Bregeon, A.C., Garcia, A., Giegé, R. & Moras, D. The contacts of yeast tRNA(Ser) with seryl-tRNA synthetase studied by footprinting experiments. *Eur. J. Biochem.* **188**, 283–290 (1990).
20. Mei, H.Y. *et al.* Inhibitors of protein-RNA complexation that target the RNA: specific recognition of human immunodeficiency virus type 1 TAR RNA by small organic molecules. *Biochemistry* **37**, 14204–14212 (1998).
21. Hallegger, M., Taschner, A. & Jantsch, M.F. RNA aptamers binding the double-stranded RNA-binding domain. *RNA* **12**, 1993–2004 (2006).
22. Pouch-Pélissier, M.N. *et al.* SINE RNA induces severe developmental defects in *Arabidopsis thaliana* and interacts with HYL1 (DRB1), a key member of the DCL1 complex. *PLoS Genet.* **4**, e1000096 (2008).
23. Savochkina, L., Alekseenkova, V., Belyanko, T., Dobrynina, N. & Beabealashvilli, R. RNA protections from RNase V1 due to RNA structure alone. *BMC Res. Notes* **1**, 15 (2008).
24. Esakova, O., Perederina, A., Quan, C., Schmitt, M.E. & Krasilnikov, A.S. Footprinting analysis demonstrates extensive similarity between eukaryotic RNase P and RNase MRP holoenzymes. *RNA* **14**, 1558–1567 (2008).
25. Batey, R.T., Gilbert, S.D. & Montange, R.K. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432**, 411–415 (2004).
26. Serganov, A. *et al.* Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.* **11**, 1729–1741 (2004).
27. Witte, G., Hartung, S., Büttner, K. & Hopfner, K.P. Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol. Cell* **30**, 167–178 (2008).
28. Edwards, T.E. & Ferré-D'Amaré, A.R. Crystal structures of the *thi*-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure* **14**, 1459–1468 (2006).
29. Gilbert, S.D., Reyes, F.E., Edwards, A.L. & Batey, R.T. Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs. *Structure* **17**, 857–868 (2009).
30. Collins, J.A., Irnov, I., Baker, S. & Winkler, W.C. Mechanism of mRNA destabilization by the *glmS* ribozyme. *Genes Dev.* **21**, 3356–3368 (2007).
31. Klein, D.J., Been, M.D. & Ferré-D'Amaré, A.R. Essential role of an active-site guanine in *glmS* ribozyme catalysis. *J. Am. Chem. Soc.* **129**, 14858–14859 (2007).
32. Edwards, T.E., Klein, D.J. & Ferré-D'Amaré, A.R. Riboswitches: small-molecule recognition by gene regulatory RNAs. *Curr. Opin. Struct. Biol.* **17**, 273–279 (2007).
33. Henkin, T. Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev.* **22**, 3383–3390 (2008).
34. Serganov, A. The long and the short of riboswitches. *Curr. Opin. Struct. Biol.* **19**, 251–259 (2009).
35. Lipfert, J. *et al.* Structural transitions and thermodynamics of a glycine-dependent riboswitch from *Vibrio cholerae*. *J. Mol. Biol.* **365**, 1393–1406 (2007).
36. Serganov, A., Huang, L. & Patel, D. Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature* **455**, 1263–1267 (2008).
37. Garst, A.D., Héroux, A., Rambo, R.P. & Batey, R.T. Crystal structure of the lysine riboswitch regulatory mRNA element. *J. Biol. Chem.* **283**, 22347–22351 (2008).
38. Mulhbacher, J. & Lafontaine, D.A. Ligand recognition determinants of guanine riboswitches. *Nucleic Acids Res.* **35**, 5568–5580 (2007).
39. Tomsic, J., Mcdaniel, B., Grundy, F. & Henkin, T. Natural variability in S-adenosylmethionine (SAM)-dependent riboswitches: S-box elements in *Bacillus subtilis* exhibit differential sensitivity to SAM *in vivo* and *in vitro*. *J. Bacteriol.* **190**, 823–833 (2008).
40. Wickiser, J.K., Cheah, M.T., Breaker, R.R. & Crothers, D.M. The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry* **44**, 13404–13414 (2005).
41. Wickiser, J.K., Winkler, W.C., Breaker, R.R. & Crothers, D.M. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol. Cell* **18**, 49–60 (2005).
42. Brünger, A.T. *et al.* Crystallography and NMR system: a new software system for macromolecular structure determination. *Acta Crystallogr D* **54**, 905–921 (1998).
43. Leontis, N.B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499–512 (2001).

ONLINE METHODS

Reagent preparation and biochemical assays. We carried out transcription and purification of the 93-nucleotide (nt) crystallization RNA construct, preceded by a hammerhead ribozyme and followed by a VS ribozyme substrate stem-loop, essentially as described[44,45]. We prepared the 88-nt wild-type *V. cholerae* c-di-GMP riboswitch aptamer domain[6] similarly. Ref. 8 describes U1A-RBD (Y31H, Q36R) expression and purification. We purchased c-di-GMP and c-di-AMP from Axxora, LLC and used both without further purification.

For electrophoretic mobility shift assays (EMSA), we diluted RNA to 35 μM in a buffer comprising 5 mM Tris-HCl (pH 8.0), 3 mM $MgCl_2$, 10 mM NaCl, and 100 mM KCl and small molecule (0 or 70 μM). After a 30-min incubation at 310 K, we resolved the samples on 10% polyacrylamide gels with 0.5× THE supplemented with 1 mM $MgCl_2$ as the running buffer. We stained the gels with toluidine blue. For nuclease probing, we prepared ^{32}P-labeled RNA in 50 mM Tris-HCl (pH 8.0), 30 mM NaCl, 30 mM KCl, appropriate final $MgCl_2$ concentration (1.8 mM, 2.4 mM, 3.0 mM) and either 0 or 70 μM c-di-GMP. We added 1 μl of nuclease (T1, 0.01 U μl^{-1}; V1, 0.0025 U μl^{-1}) to initiate structural probing. We performed reactions at 310 K for 5 min in a final volume of 5 μl, terminated them by addition of 5 μl gel loading buffer (Ambion) and loaded them immediately onto 10% polyacrylamide–8 M urea gels. We performed autoradiography by phosphorimaging. We performed nuclease mapping experiments in triplicate. We quantified cleavage bands that could be resolved as singular using ImageQuantTL software (GE Healthcare) and normalized them to a uniform band (nucleotide 17) across all lanes. We calculated protection ratios by dividing the intensity of the band in the absence of c-di-GMP by the intensity of the band in the presence of c-di-GMP. We report nucleotides with a protection ratio greater than 1.75 (T1 nuclease) and 1.25 (V1 nuclease) in **Figure 4c**,**d**. Similarly, we report increased cleavage by nuclease V1 for ratios less than 0.8.

Crystallization and structure determination. We mixed RNA incubated as for EMSA, but at a final concentration of 250 μM (with 1 molar equivalent each of c-di-GMP and U1A-RBD and 1 mM spermine) with an equal volume of a reservoir solution comprising 30% (w/v) PEG 3,350, 300 mM ammonium acetate (pH 7.0), 100 mM Tris-HCl (pH 8.5). We obtained thin, plate-shaped crystals by vapor diffusion at 303 K over the course of 1 d. We flash froze a crystal directly from the mother liquor by plunging it into liquid nitrogen. We thawed and then annealed[46] this crystal in a solution comprising 35% (w/v) PEG 3,350, 320 mM ammonium acetate (pH 7.0), 100 mM Tris-HCl (pH 8.5) and 3 mM $MgCl_2$, and flash froze it in the same manner as described above. We collected diffraction data by the oscillation method with 1-Å X-radiation at 100 K from one crystal at beamline 8.2.2 of the Advanced Light Source (ALS), Lawrence Berkeley National Laboratory, and reduced the data with the HKL package[47]. High crystal mosaicity (~0.9°) and the relatively long *c* unit cell dimension limited the completeness and precision of the data.

We initiated structure determination by molecular replacement with PHASER[48] with four independent search models (two U1A-RBDs and their binding sites, and two 5-bp duplexes). In the first round, PHASER successfully placed two U1A-RBDs and their binding sites and one of the duplexes (translation function *Z* scores of 10.0, 16.8 and 10.1, respectively, and an overall log-likelihood gain of 446). We carried out iterative rounds of simulated annealing, energy minimization and tightly restrained individual *B*-factor refinement[42] (with a random 10% of the data reserved for cross-validation), interspersed with further rounds of molecular replacement to achieve complete tracing[49] of both RNA-protein complexes in the asymmetric unit. This model had a free-*R* factor of 31.4%, and a residual Fourier synthesis calculated with phases derived from it revealed clear electron density for the two bound c-di-GMP molecules (**Fig. 1a**). We manually placed the second messengers into the residual electron density and carried out further refinement, which smoothly reduced the free-*R* factor to 29.2%. We used an anisotropic *B*-factor correction and an automatic solvent mask throughout. We did not use noncrystallographic symmetry restraints. We chose *B*-factor restraints to minimize the free-*R* factor (the r.m.s. deviation of *B*-factors of covalently bonded atoms is 0.49 $Å^2$). The current model has a cross-validated σA coordinate precision of 0.54 Å. 95.8% of residues of the U1A-RBDs lie in the most favored and additionally allowed regions of the Ramachandran plot[50]. There are no residues with disallowed backbone conformations. We prepared crystal structure figures with RIBBONS[51] using the coordinates of RNA chain A.

Small-angle X-ray scattering. We prepared SAXS samples similarly to the nuclease probing samples, with slight modifications (20 μM RNA, 100 μl final volume). We also prepared identical buffers (**Supplementary Methods**). We collected data under continuous flow at 310 K at the BioCAT beamline of the Advanced Photon Source (APS), Argonne National Laboratory. We analyzed buffer-subtracted data using IGOR (Wavemetrics, Inc.). We used Guinier analysis in the range $q^{\star}R_g \sim 1$ to calculate the radii of gyration. We averaged multiple DAMMIN[52] reconstructions using DAMAVER[53]. We converted the filtered, averaged reconstructions to electron density map using SITUS[54]. We display these molecular envelope reconstructions with CHIMERA[55]. We computed the scattering profile of molecular models using CRYSOL[56].

44. Ferré-D'Amaré, A.R. & Doudna, J.A. Use of *cis*- and *trans*-ribozymes to remove 5′ and 3′ heterogeneities from milligrams of in vitro transcribed RNA. *Nucleic Acids Res.* **24**, 977–978 (1996).
45. Rupert, P.B. & Ferré-D'Amaré, A.R. Crystallization of the hairpin ribozyme: illustrative protocols. *Methods Mol. Biol.* **252**, 303–311 (2004).
46. Heras, B. & Martin, J.L. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr D* **61**, 1173–1180 (2005).
47. Otwinowski, Z. & Minor, W. Processing of diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
48. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
49. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
50. Laskowski, R.J., Macarthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
51. Carson, M. Ribbons. *Methods Enzymol.* **277**, 493–505 (1997).
52. Svergun, D.I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
53. Volkov, V.V. & Svergun, D.I. Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Cryst.* **36**, 860–864 (2003).
54. Wriggers, W., Milligan, R.A. & McCammon, J.A. Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125**, 185–195 (1999).
55. Pettersen, E. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
56. Svergun, D.I., Bargerato, C. & Koch, M.H.J. CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* **28**, 768–773 (1995).

doi:10.1038/nsmb.1701