

Review

The World of Cyclic Dinucleotides in Bacterial Behavior

Aline Dias Purificação ¹, Nathalia Azevedo ¹, Gabriel Guarany de Araujo ¹, Robson Francisco Souza ¹ and Cristiane Rodrigues Guzzo ^{1,*}.

¹ Department of Microbiology, Institute of Biomedical Science, University of São Paulo, São Paulo, Brazil
All authors contributed equally to this article.

* Correspondence: crisguzzo@usp.br [and crisguzzo@gmail.com](mailto:crisguzzo@gmail.com); Tel.: +55 11 3091-7298

Running title: Bacterial second messengers c-di-GMP, c-di-AMP and cGAMP.

Abstract: The regulation of multiple bacterial phenotypes was found to depend on different cyclic di-nucleotides (CDNs) that constitute intracellular signaling second messenger systems. Most notably, c-di-GMP, along with proteins related to its synthesis, sensing and degradation, was identified as playing a central role in the switching from biofilm to planktonic modes of growth. Recently, this world has been under expansion, with the discoveries of other signaling CNDs in bacteria (c-di-AMP and cGAMP) and also in eukaryotes, novel protein and RNA receptors of CDNs, and the numerous functions related to these molecules. In this work, we comprehensively review and analyze the structural biology data about the systems that bacteria use to synthesize and recognize CDNs, detailing their interactions at molecular level with their products/ligands. Additional interesting observations were made, including that different receptor types can bind CDNs in similar conformations and that, based on genomic data, different CDN second messenger systems coexist in many organisms. The large amount of sequence and structural data available allows a broad view of the importance of CDNs in bacteria, but how cells coordinate these molecules to ensure adaptation to changing environmental conditions is still open for much further exploration.

Keywords: GGDEF; SMODS; DAC; c-di-GMP; c-GAMP; c-di-AMP.

1. Introduction

Cyclic dinucleotides (CDNs) are small molecules that play central roles as secondary messengers in bacteria¹. Different classes of cyclic oligonucleotides, most of which are essentially ubiquitous among [bacteria](#), such as c-di-GMP, c-di-AMP, cGAMP, c-UAMP, c-di-UMP, c-UGM, c-CUMP and c-AAGMP, have been found to be the products of different classes of enzymes²⁻⁵. These molecules include not only di-purines but also hybrids of purine and pyrimidines and cyclic trinucleotides are synthesized by different classes of enzymes, in which the cyclic dinucleotides are essentially ubiquitous among [bacteria](#)³.

The c-di-GMP molecules, cyclic-bis(3' → 5')-dimeric GMP, generally coordinate the transition of a bacterium's life style, from a mobile single cell undergoing planktonic growth to a multicellular community in biofilm structures, a form of sessile growth. Regulation of these transitions are mediated by controlling the bacterial motility through the regulation of the flagellar rotor⁶ and the twitching motility machinery⁷. Alternatively, in streptomyces, c-di-GMP regulates the transition from vegetative mycelial growth to the formation of reproductive aerial mycelium⁸. This di-nucleotide is also involved in the regulation of bacterial adhesion, cell cycle progression and division, biofilm formation, quorum sensing⁹, regulation of the type II (T2SS)¹⁰, type III (T3SS)¹¹ and type VI (T6SS)¹² secretion system machineries, as well as the synthesis and secretion of virulence factors and pathogenesis¹³⁻¹⁷. Similarities in the roles of eukaryotic cyclins and bacterial c-di-GMP molecules have

been suggested¹⁸. In eukaryotes, cyclins drive the cell cycle by regulating the activity of cyclin-dependent kinases and promoting the asymmetric replication of future cells¹⁸.

Some similar biological roles have been observed between c-di-GMP and c-di-AMP (cyclic-bis(3' → 5')-dimeric AMP) molecules¹⁹. Nevertheless, few c-di-AMP synthesizing enzymes have thus far been studied in bacteria and archaea, and the more well-known enzymes are more widely distributed and were better characterized in Gram positive bacteria, but homologs can be found in several Gram negative and a few archaeal lineages (**Supplementary Table 1**)^{20,21}. Given its abundance and widespread distribution, c-di-GMP stands out as the main second messenger in bacteria while other cyclic di-nucleotides seem to be specialized in particular roles. The c-di-AMP molecule regulates processes such as osmoprotection^{22,23}, cell-wall homeostasis²⁴, potassium ion channel expression and function²⁵, DNA repair to maintain genomic integrity⁴, diverse gene expression^{26,27}, biofilm formation^{28,29}, sporulation³⁰, antibiotic resistance³¹, and metabolism³². Another CDN, 3' - 5' cGAMP (cyclic guanosine (3' → 5') monophosphate - adenosine (3' → 5') monophosphate), modulates chemotaxis, virulence and exoelectrogenesis (the use of insoluble extracellular terminal electron acceptors)³³. 3' - 5' cGAMP and also c-UAMP activate the phospholipase activity of patatin-like lipase enzymes^{3,34}.

The first ubiquitous second messenger described was the c-di-GMP molecule, which is synthesized by GGDEF domain-containing proteins³⁵⁻³⁷ and degraded into pGpG or GMP by phosphodiesterase proteins containing HD-GYP or EAL domains. The released pGpG can then be degraded by an oligoribonuclease Orn^{38,39}. A divergent GGDEF family enzyme synthesizes not only c-di-GMP but also 3' - 5' cGAMP, cyclic guanosine (3' → 5') monophosphate - adenosine (3' → 5') monophosphate^{40,41}.

Bacteria synthesize 3' - 5' c-GAMP mainly by the activity of proteins containing SMODS domains (**S**econd **M**essenger **O**ligonucleotide or **D**inucleotide **S**ynthetase)^{5,42}, while eukaryotic cells synthesize cGAMP with a mixed 2' - 5' and 3' - 5' phosphodiester linkage (2' - 3' cGAMP) by GMP-AMP synthase (cGAS) enzymes⁴³⁻⁴⁵. cGAS are structurally similar to SMODS and they belong to **c**GAS/**D**ncV-like **n**ucleotidyl**t**ransfer**a**ses (CD-NTases) enzymes superfamily^{5,42}. The best characterized enzyme containing SMODS domain is the DncV protein (*Vibrio cholerae* dinucleotide cyclase), with its orthologues being able to synthesize different cyclic nucleotides, including 3' - 5' cGAMP, c-UAMP, c-di-UMP, and c-AAGMP, respectively produced by DncV, DncE, *LpCdnE02* and *EcCdnD02* proteins³. Bacterial CD-NTases enzymes also synthesize, as minor products, c-di-GMP, c-di-AMP, c-UGMP and c-CUMP³. This diverse array of products synthesized by this group of enzymes is thought to be related to a low energetic barrier for altering product specificity³.

The c-di-AMP molecules are synthesized by di-adenylyl cyclase (DAC) proteins and hydrolyzed into pApA or AMP by specific phosphodiesterase (PDE) that contain DHH-DHHA1 or HD domains^{46,47}. For examples, *L. monocytogenes* encodes PdeA and PgpH, which act cooperatively to hydrolyze c-di-AMP⁴⁷. Four classes of DAC proteins have been identified, DisA, CdaA, CdaS, and CdaS, and all of them contain DGA and RHR motifs, that are involved in the catalysis^{48,49}. Notably, GGDEF, SMODS and DAC domains do not share structural similarities and probably perform the nucleotide cyclization catalysis by different mechanisms^{19,50}.

Members of families within the CD-NTases superfamily, such as SMODS and cGAS, often do not share detectable primary sequence similarity but adopt a Pol-β-like nucleotidyl transferase fold, suggesting a common origin followed by divergent evolution^{3,5,51,52}. cGAS and enzymes containing-SMODS domains use a single active site to sequentially form two separate phosphodiester bonds and release one cyclic nucleotide product. On the other hand, proteins containing DAC or GGDEF domains require homodimerization to perform catalysis⁴, but are not related, as DACs adopt a unique, particular fold, while GGDEF domains are homologous to adenylyl/guanylyl cyclase catalytic domains and to the palm domain of DNA polymerases (**Figure 1**)^{21,53}.

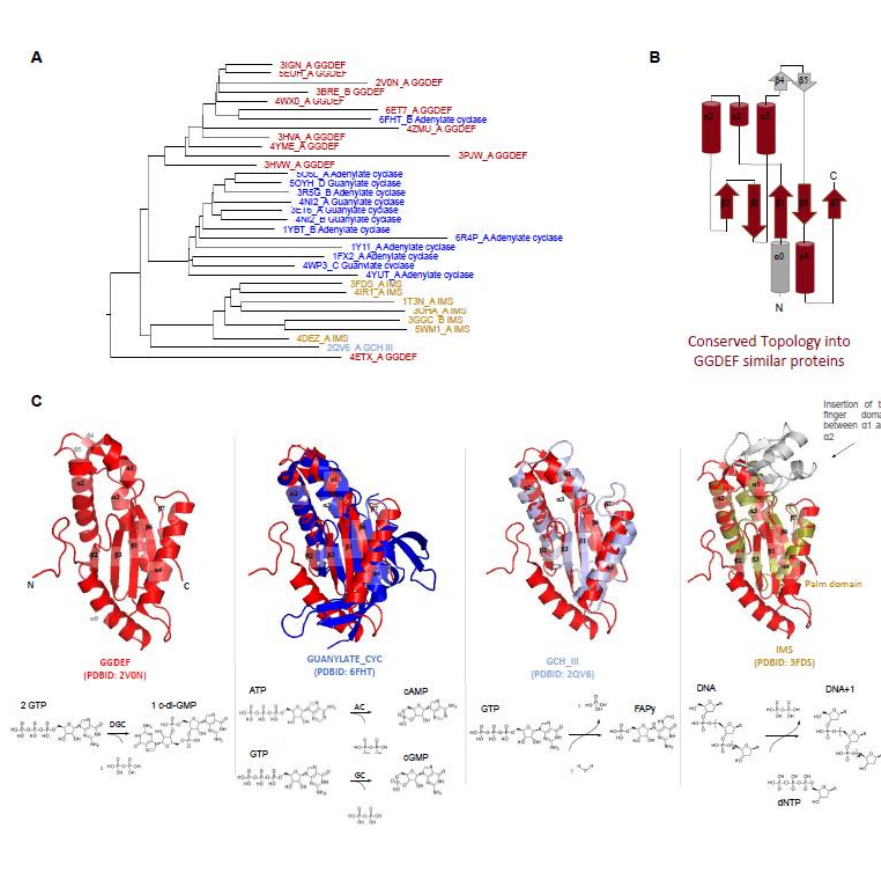


Figure 1. Structural similarities between GGDEF domains and adenylate/guanylate cyclase, GTP cyclohydrolase III and RRM-like palm domain of DNA polymerases. (A) Dendrogram showing structures similar to GGDEF domain made with the Dali server⁹¹ (query: PleD PDBID: 2V0N). Each domain is colored with different colors and the PDBID_chain and the Pfam name are shown for each branch. The conserved fold found in most of these structures is shown in brown, panel B, overlaid on the GGDEF domain of PleD topology. (C) Structural superposition of GGDEF domain of PleD (PDBID: 2V0N) with other class of domain shown in the panel C dendrogram, using the same colors to represent each domain. At the bottom of each structural alignment is shown the domain's name and the PDBID code, as well as the chemical reaction performed. DGC: diguanylate cyclase; AC: adenylate cyclase, GC: guanylate cyclase, and FAPy: 2-amino-5-uridylyl-5'-phosphate.

Proteins containing GGDEF domains requires an accessory domain that sense different signals to regulate the GGDEF homodimerization and consequently its enzymatic activity⁵⁴. Each GGDEF domain binds one molecule of GTP and its dimerization positions the two GTP molecules in an antiparallel manner to enable their condensation into c-di-GMP and the release of two molecules of pyrophosphate⁵⁵. Therefore, proteins containing GGDEF domains are Bi-Ter (two substrates, three products) enzymes and cannot be describe by a Michaelis-Menten model^{56,57}. A similar enzymatic mechanism seems to happen for proteins containing DAC domains.

Bacteria sense second messenger molecules in different ways and respond in different manners. The c-di-GMP, c-di-AMP and cGAMP molecules are sensed by both proteins and RNAs. Some examples of c-di-GMP effectors are mRNA riboswitches⁵⁸, transcription factors^{8,59,60}, and different classes of protein domains such as PilZ domains⁶¹⁻⁶³, degenerate GGDEF domains, degenerate EAL domains⁶⁴, and AAA+ ATPases domains^{10,11,65} (Table 1). When binding to its receptors, c-di-GMP can regulate bacterial physiology by altering gene transcription, protein expression and protein function, thus allowing the bacteria to respond to environmental stimuli at different speeds, be it a fast response, achieved by directly regulating protein activity, or a slower one, by regulating gene expression. Moreover, c-di-GMP can regulate the same physiological process in different ways, as

seen for bacterial mobility, when both flagellar gene expression and the flagellar motor function are regulated by c-di-GMP molecules^{61,66}.

Table 1: List of the bacterial c-di-GMP, c-di-AMP, c-GAMP and eukaryotic c-GAMP receptors that had their structure solved in complex with the ligand and deposited in the protein data bank. The Pfam/Rfam and in some cases the interpro domain are described. The residues involved in ligand binding are also described for one receptor representative.

Recept or class (Pfam/ Rfam)	Organism (PDBID)	Receptor function	Ligand binding site	Ref.
3' - 5' c-di-GMP				
STING (TMEM173, PF15009)	<i>Homo sapiens</i> (4EF4, 4EMT, 6RM0, 6S86, 4F9G, 4F5D, 4F5Y)	Members of Transmembrane Protein 173 (TMEM173) family, also known as Stimulator of Interferon Genes (STING), is an important component of the immune system. STING proteins are responsible for regulating the induction of type I interferon via activation of INF- β gene transcription. Human STING (carrying the more common R232 allele) binds eukaryotic 2' - 3' cGAMP with high affinity compared with bacterial CDNs such as c-di-GMP, c-di-AMP, and 3'- 5' cGAMP ¹⁴⁰ . Nevertheless, it is controversy whether STING binds 2' - 5' cGAMP preferentially since others STINGs binds CDNS with the same affinity ¹¹⁰ .	STING proteins interacts with c-di-GMP at the protein dimer interface in a perfectly symmetrical manner increasing the homodimer stability. This bind involves a hydrophilic core, that in the human STING (PDB 4F5D) correspond to, S162, G166, Y167, R238, Y240, S241, N242, E260, T267, and the presence of two Mg ²⁺ ions and two waters (Figure 7 A-C). STING proteins binds monomers of c-di-GMP that are stabilized in the protein pocket at the intermediate or in closed conformations, Figure 8 .	141-146
	<i>Sus scrofa</i> (6A04)			147
	<i>N. vectensis</i> (5CFL, 5CFP)			148
c-di-GMP I Riboswitch (RF01051)	<i>V. cholerae</i> (3MXH, 3MUT, 3MUR, 3MUM, 3IRW)	c-di-GMP Riboswitches, also known as GEMM (Genes for the Environment, Membranes and Motility), are structured RNAs located in the 5'-untranslated regions of mRNAs and sense c-di-GMP molecules to regulate	GEMM Riboswitches interacts with c-di-GMP by an uncharacterized motif with high affinity, picomolar, compared to c-di-GMP protein receptors, nanomolar to micromolar affinity. In the case of c-di-	149, 150
	<i>Geobacter</i> (4YB0)			151
	<i>E. coli</i> (3IWN)			152

c-di-GMP II Riboswitch (RF01786)	<i>C. acetobutylicum</i> (3Q3Z)	expression of downstream genes that could be involved with virulence, motility and biofilm formation. Despite having the same function, the c-di-GMP I Riboswitch and c-di-GMP II Riboswitch don't share any sequence motifs or structural features.	GMP I Riboswitch (PDB 3IRW) the nucleotides involved in ligand binding are: G14, C15, A16, C17, A18, G19, G21, C46, A47, A48, A49, G50. c-di-GMP II riboswitch (PDBID 3Q3Z) binds to c-di-GMP through the nucleotides: A13, A14, U37, G39, U60, A61, C68, A69, A70, C71, C72, G73, and A74. Riboswitches can recognize the guanine base of the ligand in different ways. The ligand was found as <i>closed monomers</i> , Figure 8 .	10
	<i>V. cholerae</i> (2RDE)	VCA0042 is an important protein for the efficient infection of mice by <i>V. cholerae</i> . This PilZ-containing protein sense the bacterial second messenger c-di-GMP and controls virulence factors.	This PilZ domain interact with monomeric c-di-GMP via two main sequence motifs: RxxxR and DxSxxG motifs (PDBID: 2RDE), Figure 7 D - E . The ligand was found as <i>intermediate monomers</i> , Figure 8 .	119
PilZ domain (PF07238)	<i>R. sphaeroides</i> (5EIY, 5EJ1, 5EJZ, 4P00, 4P02)	BcsA, Bacterial cellulose synthase A, is a component of a protein complex that synthesis and translocate the cellulose across the inner membrane. The bind of c-di-GMP to a complex BscA and BcsB releases an autoinhibited state of the enzyme, generating a constitutively active cellulose synthase.	The most PilZ domains interact with dimeric c-di-GMP, in which one molecule interacts with two main sequence motifs on the β -barrel surface, DxSxxG and RxxxR motifs (PDBI: 5EIY, 5EJ1, 5EJZ, 4P00, 4P02, 5Y6F, 5Y6G, 5VX6, 5KGO, 5EJL, 5XLY, 2L74, 5Y4R, 4RT0, 4RT1). In the PilZ domain of YcgR (PDBID: 5Y6F) the "DxSxxG" motif correspond to D145, S147 and G150, and the "RxxxR" motif correspond to R114 and R118, Figure 7 D - E . The ligand was found as <i>closed dimers</i> , Figures 8 . One PilZ	153, 154
	<i>E. coli</i> (5Y6F, 5Y6G)	YcgR like proteins such as the motility inhibitor (MotI) protein is a diguanylate receptor that binds c-di-GMP, acting as a molecular clutch on the flagellar	motif correspond to D145, S147 and G150, and the "RxxxR" motif correspond to R114 and R118, Figure 7 D - E . The ligand was found as <i>closed dimers</i> , Figures 8 . One PilZ	155
	<i>B. subtilis</i> (5VX6)			156
	<i>K. pneumoniae</i> . (5KGO, 5EJL)			157, 158

		<p>stator MotA to inhibit swarming motility.</p> <p>The PilZ domain of MrkH, also a YcgR like protein, is transcriptional regulator protein, and binds c-di-GMP as well as DNA sequence to regulate the type 3 fimbriae expression and biofilm formation.</p> <p>YcgR proteins regulate motility and biofilm formation by sensing c-di-GMP.</p>	interacts with a trimeric c-di-GMP (PDBID: 4XRN), Figures 8 B.	
	<i>P. aeruginosa</i> . (5XLY, 2L74, 5Y4R)	MapZ in complex with c-di-GMP interacts directly with a chemotaxis methyltransferase, CheR1, and inhibit your activity. By this way, regulates the chemotaxis in <i>Pseudomonas aeruginosa</i> .		62,1 59,1 60
	<i>P. aeruginosa</i> (4RT0, 4RT1)	Alginate biosynthesis protein Alg44 regulates the alginate secretion to promote the biofilm formation by sensing dimeric c-di-GMP molecules.		161
	<i>P. aeruginosa</i> (4XRN)	Unknown function	The ligand is in an unusual trimeric oligomerization state, in which the six guanine bases are oriented almost in a parallel way to each other, Fig. 8 B.	131
I-site of GGDEF domains	<i>P. fluorescens</i> (5EUH for GcbC)	Proteins containing GGDEF domains are DGCs and some of them are regulated by feedback regulation by interaction of c-di-GMP to your allosteric site (I-site).	Proteins with GGDEF domain act as receptor proteins when the c-di-GMP bind at the allosteric site of them via RxxD motif.	162
	<i>P. aeruginosa</i> (3BRE and 3I5C for WspR; 4EUV,			163- 165

	4ETZ, 4EU0 for PelD)		In the WspR GGDEF (PDB 3BRE) this motif corresponds to Arg242, Ser243, Ser244 and Asp245. The ligand was found as <i>closed dimers</i> , very similar at the PilZ proteins, Figure 8 .	
	<i>P. syringae</i> (3I5A for WspR)			163
	<i>M. hydrocarbonoclasticus</i> (3IGN for MqR89a)			166
	<i>T. maritima</i> (4URG, 4URS for TM1788)			167
	<i>C. vibrioides</i> (1W25, 2WB4, 2V0N for PelD)			72,8 1,16 8
	<i>E. coli</i> (3TVK, 4H54 for DgcZ)			169
	<i>P. aeruginosa</i> (4DN0)	PelD is a membrane protein, in which your cytoplasmatic GGDEF domain binds c-di-GMP to regulate the synthesis of a PEL exopolysaccharide.		170
Degenerate EAL domains (PF00563)	<i>X. citri</i> (4FOK, 4FOJ, 4FOU)	FimX protein regulates twitching motility by sensing c-di-GMP molecules through its EAL domain and regulates the type IV pilus machinery.	Proteins with EAL domain, as FimX (PDB 4FOK), interacts with the c-di-GMP by Q463, F479, L480, R481, S490, P491, M495, D508, R534, E653, F654, Q673, G674, D675 and T680. The A ₄₇₈ F ₄₇₉ L ₄₈₀ belongs to the degenerate EAL motif, Figure 7 H and I . The ligand was found always as <i>open or intermediate monomers</i> . Different EAL containing proteins bind with the most diverse c-di-GMP conformation states analyzed in this review, Figure 8 .	129
	<i>P. aeruginosa</i> (3HV8)			171
	<i>X. campestris</i> (4F3H, 4F48)			172
	<i>P. fluorescens</i> (3PJT, 3PJU)	Transmembrane receptor LapD is a multidomain protein, in which the C-terminal EAL domain binds c-di-GMP to prevent cleavage of the surface adhesin LapA and therefore, activate the biofilm formation.		121
C-termi	<i>S. venezuelae</i> (5TZD)	BldD is a master regulator of cell development. BldD represses the	The C-terminal domain of BldD (PDB 5TZD) interacts with a	132

	<i>S. coelicolor</i> (4OAZ)	transcription of close to 170 sporulation genes during vegetative growth controlling the morphological differentiation and also directly control expression of antibiotics. BldD has a N-terminus helix-turn-helix motif (HTH) and the C-terminal domain binds four c-di-GMP molecules to regulate the cell differentiation.	tetramer of c-di-GMP, BldD ₂ -(c-di-GMP) ₄ complex, by two motifs: R ₁₁₄ G ₁₁₅ D ₁₁₆ and R ₁₂₅ Q ₁₂₆ D ₁₂₇ D ₁₂₈ . The ligand was found as <i>closed tetramers</i> , Figure 8 .	8
REC domain (Response_reg, PF00072)	<i>V. cholerae</i> (3KLO)	VpsT is transcriptional regulator that binds c-di-GMP at your REC domain to control biofilm formation and motility. VpsT is described as a master regulator for biofilm formation and consists of a N-terminal REC domain and a C-terminal HTH domain.	A c-di-GMP ₂ binds into the VspD interface between two REC domains, the REC dimerization is required to ligand binding. Proteins with REC domain of VpsT (PDB 3KLO) interacts with two molecules of c-di-GMP by a K and W[F/L/M][T/S]R motif that correspond to: K120, W131, L132, T133 and R134. The ligand was found as <i>closed dimers</i> , Figure 8 .	59

T2SSE_N domain (PF05157)	<i>V. cholerae</i> (5HTL)	<p>MshE is an ATPases associated with bacterial type II secretion system and type IV pilus formation.</p> <p>The N-terminal domain binds c-di-GMP and c-GAMP with different affinities, while the C-terminal catalytic domain binds ATP.</p> <p>The MshE N-terminal domain (T2SSE_N) binds c-di-GMP (<i>K_d</i> of 0,5 M) with higher affinity than with cGAMP (<i>K_d</i> of 330 M).</p>	<p>The N-terminal domain of MshE (locus tag VC0405, PDB 5HTL) interacts with c-di-GMP by mainly two similar motifs spaced by five residues. These motifs have a similar sequence, RLGxx(L)(V/I)xxG(I/F)(L/V)xxx xLxxxLxxQ, and the residues involved to ligand binding are shown in bold and correspond to R₉L₁₀G₁₁ and L₂₅xxxL₂₉xxQ₃₂ for the motif I, and R₃₈L₃₀G₄₀ and L₅₄xxxL₅₈xxQ₆₁ for motif II. Other residues also important to ligand binding are: R7, D108 (from the C-terminal ATPase domain), and main chain of D41.</p> <p>The ligand was found as <i>open monomers</i> similar with those find into EAL domains, Figure 8.</p>	173
MerR domain (PF00376)	<i>P. aeruginosa</i> (5XQL)	<p>BrlR upregulate the expression of multidrug efflux pumps. C-di-GMP activates the BrlR expression and enhance your affinity to DNA binding. BrlR has a N-terminus DNA-binding motif (HTH_MerR domain described in the Pfam as MerR domain), and a C-terminus effector-binding domain (GyrI-like domain) linked by a coiled-coil region.</p>	<p>There are two different c-di-GMP binding sites located at the N-terminus of the protein mainly at the DNA binding domain of each BrlR protomer of the protein tetramer.</p> <p>Binding site 1 is composed by M1, R31, D35, Y40, and Y270. The binding site 2 is composed by: P61, A64, R67, R70, F83, R86).</p> <p>The ligand was found as <i>closed monomers</i>, Figure 8.</p>	174

Sigma54_activat (PF00158) or AAA+_ATPase (IPR003593)	<i>P. aeruginosa</i> (5EXX)	<p>FleQ is a transcription regulator and a contain three domains: a central AAA+ ATPase $\sigma(54)$-interaction domains, flanked by a divergent N-terminal receiver domain and a C-terminal helix-turn-helix DNA-binding motif. FleQ binds c-di-GMP through your AAA+ ATPase domain in a different binding site than the catalytic pocket site.</p> <p>FleQ regulates the expression of flagellar and exopolysaccharide biosynthesis genes in response to cellular levels of c-di-GMP.</p>	<p>FleQ binds c-di-GMP at the N-terminal part of the AAA+ ATPase through the L₁₄₂F₁₄₃R₁₄₄S₁₄₅ motif (R-switch), E₃₃₀XXXR₃₃₄ motif and residues R185 and N186 of the Pos-Walker A motif (KExxxRN).</p> <p>The ligand was found as <i>closed dimers</i>, Figure 8.</p>	65
HATPase_c (PF02518)	<i>C. vibrioides</i> (5IDM)	<p>Cell cycle kinase CckA is a bifunctional histidine kinase/phosphatase enzyme, mediating both phosphorylation and dephosphorylation of downstream targets. CckA binds c-di-GMP and drives the cell cycle progression by swapping the CckA kinase activity into phosphatase mode.</p>	<p>CckA is a membrane and multidomain protein, in which catalytically active (CA) domain binds c-di-GMP. The CA domain of cell cycle kinase CckA interacts with c-di-GMP by the residues: Y514, K518, W523, I524, E550, H551, H552, H553, H554 and H555.</p> <p>The ligand was found as <i>open monomer</i>, Figure 8.</p>	175
3'-5' cGAMP or 3'-3' cGAMP				
STING (TMEM173, PF15009)	<i>N. vectensis</i> (5CFM)	<p>STING regulates the induction of type I interferons via recruitment of protein kinase TBK1 and transcription factor IRF3 activating IFN-β gene transcription.</p> <p>cGAS-STING responses to cytosolic DNA via binding to 3'-5' cGAMP.</p>	<p>STING proteins interacts with cGAMP at the dimer interface. In the anemone STING (PDBID 5CFM), the residues involved with the ligand interaction are: Y206, R272, F276, R278, and T303 of each protomer of the dimer. Y280 binds the ligand by a water molecule.</p> <p>The ligand was found as <i>intermediate monomer</i>, Figure 8.</p>	148
c-di-GMP	<i>Geobacter</i> (4YAZ)	<p>Acts like a transcriptional factor, because when it binds cGAMP,</p>	<p>3'-5' cGAMP riboswitch binds cGAMP (PDBID 4YAZ) through</p>	151

	<i>Homo sapiens</i> (4YB1)	switching between RNA secondary structures and regulate the expression itself. A human c-di-GMP I Riboswitch mutant (G20A) can also bind cGAMP.	the nucleotides G8, A11, A12, U13, A14, C15, A41, A42, G74, C75, and C76. The ligand was found as <i>closed monomer</i> , Figure 8 .	151
2' - 3' c-GAMP				
STING (TMEM173, PF15009)	<i>Sus scrofa</i> (6A06)	STING regulates the induction of type I interferons via recruitment	STING proteins interacts with 2'-3' cGAMP produced by	147
	<i>Gallus gallus</i> (6NT7, 6NT8)	of protein kinase TBK1 and transcription factor IRF3	eukaryotic cGAS enzyme in the dimer interface. In the porcine	176
	<i>Rattus norvegicus</i> (5GRM)	activating IFN-β gene transcription.	STING (PDBID 6A06), the residues involved to ligand	177
	<i>N. vectensis</i> (5CFQ)	STING pathway plays an important role in the detection of	binding are: S162, Y167, I235, R232, R238, Y240, E260 and T263.	148
	<i>Homo sapiens</i> (4LOH, 4LOJ, 4KSY, 6DNK)	viral and bacterial pathogens in animals.	The ligand was found as <i>closed monomer</i> , Figure 8 .	140, 178, 179
3' - 5' c-di-AMP				
STING (TMEM173, PF15009)	<i>Sus scrofa</i> (6A03, 6IYF)	STING pathway plays an important role in the detection of viral and bacterial pathogens in animals. STING binds eukaryotic	STING proteins interacts with the c-di-AMP differently of the c-di-GMP but in the same dimer interface. In the porcine STING	147
	<i>N. vectensis</i> (5CFN)	2' - 3' cGAMP with high affinity compared with bacterial CDNs	(PDBID 6A03), the amino acids involved with the interaction are:	148
	<i>Mus moluscus</i> (4YP1)	such as c-di-GMP, c-di-AMP, and 3'- 5' cGAMP.	S162, Y167, I235, R232, R238, Y240, T263. The ligand was found as <i>closed monomers</i> , Figure 8 .	180

Aldo-keto reductase (PF00248)	<i>Mus musculus</i> (5UXF)	RECON (reductase controlling NF-κB) is an aldo-keto reductase, a STING antagonist. They negatively regulate NF-κB activation, that induce the expression of IFN-induced genes. RECON recognize c-di-AMP by the same site that binds co-substrate nicotinamide. One AMP molecule (AMP1) of c-di-AMP has essentially the same position as the AMP portion of the NAD ⁺ co-substrate, while another AMP (AMP2) has a shift of position.	RECON binds c-di-AMP by the residues: E276, E279, N280, L219, and A253 in contact with AMP1, and Y24, Y216, Y55, and L306 are in contact with AMP2. L219, T221, and G217 are also involved to ligand binding. The ligand was found as open monomers, Figure 8 .	70
	c-di-GMP I Riboswitch (RF01051)	<i>E. coli</i> (G20A/C92U mutant Riboswitch, 3MUV)	Bacterial c-di-AMP is involved in cell wall stress and signaling DNA damage through interactions with several protein receptors and a widespread <i>ydaO</i> -type riboswitch, one of the most common riboswitches in various bacterial species. This riboswitch is found in the vicinity of genes involved in cell wall metabolism, synthesis and transport of osmoprotectants, sporulation and other important biological processes ^{122,123} .	149
	ydaO-yuaA Riboswitch (RF00379)	<i>T. pseudethanolicus</i> (4QK8 and 4QKA)	The ligand was found as closed monomers, Figure 8 .	181
		<i>T. lienii</i> (4QK9)		181
		<i>B. subtilis</i> (4W92 and 4W90)		182
		<i>C. subterraneus</i> (4QLM and 4QLN)		183
Cyclic-di-AMP receptor	<i>S. aureus</i> (4WK1 and 4D3H)	Pil-like signal transduction protein (PtsA) is a c-di-AMP receptor. Pil-like proteins are associated with nitrogen metabolism using different pathways. PtsA binds c-di-AMP	PstA (PDBID 4D3H) forms trimers and binds to c-di-AMP in the interface between two molecules, through the interaction with the residues	184, 185
	<i>L. monocytogenes</i> (4RWW)			69

	<i>B. subtilis</i> (4RLE)	with a K_d of 0.37 μ M (intracellular c-di-AMP is in M range). Others c-di-AMP receptors bind the ligand with a K_d range of 0.1 to 8 M.	N24, R26, T28, A27, F36, L37, N41, G47, F99, and Q108. The ligand was found as <i>intermediate monomer</i> , Figure 8 .	186
Pyruvate carboxylase (HMGL-like, PF00682)	<i>L. lactis</i> (5VYZ and 5VZ0)	<i>L. monocytogenes</i> pyruvate carboxylase (LmPC) or <i>L. lactis</i> pyruvate carboxylase (LIPC) are inhibited by c-di-AMP. LmPC is biotin-dependent enzyme with biotin carboxylase (BC) and carboxyltransferase (CT) activities. C-di-AMP causes conformational changes in the CT dimer that may explain the molecular mechanism for its inhibitory activity.	LIPC forms a tetramer and each c-di-AMP molecule binds in a protein dimer interface at the carboxyltransferase (CT) domain (HMGL-like domain in the Pfam) (PDBID 5VYZ) in a binding site not well conserved among pyruvate carboxylases. The residues involved in the interaction are: Q712, Y715, I742, S745, G746 and Q749 from both monomers. The ligand was found as <i>intermediate monomers</i> , Figure 8 .	128
TrkA_C (PF02080)	<i>S. aureus</i> (4YS2, 4XTT, and 5F29)	Potassium transporter A (KtrA) and Bacterial cation-proton antiporter (CpaA) are members of the RCK domain family of proteins (Regulator of conductance of K ⁺) and regulates the cellular conductance of K ⁺ . The C-terminal domain (RCK_C or TrkA_C) binds specifically c-di-AMP molecules, K_d of 43.1 nM, causing inactivation of the KtrA.	c-di-AMP binds at the RCK_C domain of KtrA in the interface of a dimer (PDBID 4XTT). The amino acids involved in the interaction are I163, I164, D167, I168, R169, A170, N175, I176, and P191 from both monomers. R169 and the isoleucines residues (hydrophobic pocket) are well conserved in other species. The ligand was found as <i>closed monomers</i> , Figure 8 .	67,1 80,1 87

CBS domain (PF00571)	<i>L. monocytogenes</i> (5KS7)	Intracellular pathogen <i>L. monocytogenes</i> synthesizes and secretes c-di-AMP during growth in culture and also in host cells. Overexpression of c-di-AMP is toxic to the cell. C-di-AMP binds to OpuC carnitine transporter at the CBS domain (K_d of 4.8 M), probably inhibiting carnitine uptake. OpuC is the ATPase subunit of the transporter complex OpuCA.	The c-di-AMP binds to the cystathionine γ -synthase domain (CBS) of OpuC at the dimer interface. The residues involved to ligand binding are well conserved among OpuCA orthologues and are composed by the following residues: V260, V280, T282, Y342, I355, I357, R358 and A359. The ligand was found as <i>open monomers</i> , Figure 8 .	23
----------------------	--------------------------------	---	--	----

The c-di-AMP molecule is a ligand of different classes of proteins and also some RNA riboswitches. Some examples of c-di-GMP effectors are proteins containing RCK domain⁶⁷, the universal stress protein (USP) domain⁶⁸, and PstA proteins⁶⁹. In the case of the products of CD-NTases enzymes, genes located adjacent to CD-NTases genes in the genome encode nucleotide receptors and act as effectors in biological conflicts, such as phospholipases, nucleases and pore-forming agents^{34,42}. The CD-NTases genes and the cyclic nucleotide receptors are generally found on mobile genetic elements, while genes encoding GGDEF domains are widespread in the chromosomes of different bacteria and c-di-GMP receptors are not show a tendency to be located close to genes encoding proteins containing GGDEF domains.

CDNs are probably present in almost all kinds of bacteria and their impact on an organism's physiology is probably determined by their concentration and the type of the second messenger. Interestingly, eukaryotic host cells evolved ways to sense some bacterial CDNs as a strategy to detect the presence of a pathogen and thus trigger a counter-attack to avoid or fight invasion. The stimulator of interferon genes (STING) protein binds bacterial CDNs 3' - 5' cGAMP, 3' - 5' c-di-AMP and 3' - 5' c-di-GMP molecules, as well as eukaryotic 2' - 3' cGAMP molecules. Binding of STING to these cyclic di-nucleotides activates expression of type I interferon in infected cells and initiates the innate immunity response for successful pathogen elimination⁷⁰. On the other hand, c-UAMP and c-AAGMP are not recognized by STING proteins but are a ligand of mammalian CDN sensor reductase controlling NF- κ B (RECON)⁷⁰. Binding of c-di-AMP, c-UAMP or c-AAGMP to RECON inhibits its enzymatic activity, leading to increased activation of the proinflammatory transcription factor NF- κ B, redirecting the cellular response toward an antibacterial reaction⁷⁰. These new discoveries suggest that different classes of cyclic oligonucleotides molecules may have a role in bacterial signaling and pathogen recognition than previously understood. Consequently, STING could be a target for new drugs for the treatment of bacterial infections⁵³. Interestingly, STING is currently being explored as a candidate stimulant for anticancer immune activity⁷¹.

In this review, we focus on recent advances in relation to the enzymes involved in the production of bacterial CDNs. We describe the conserved residues important to perform the catalysis of proteins containing GGDEF, SMODS or DAC domains. Some bacterial CDNs receptors that had their structure solved in complex with the ligand are also presented, with a focus on the residues involved in ligand recognition. We also highlight the conformation of the CDNs inside of the protein binding pocket. Surprisingly, different kinds of receptors bind CDNs with similar conformations. Additional interesting observations, based on genomic data, suggest different CDN second messenger systems tend to coexist in many organisms. The large amount of sequence and structural data available allows a broad view of the importance of CDNs in bacteria, but how cells coordinate these molecules to ensure adaptation to changing environmental conditions is still open for much further exploration.

2. Synthesis of Cyclic Di-nucleotides Molecules by Proteins Containing Ggdef, Dac and Smods Domains

At the moment, three different classes of prokaryotic proteins are known to synthesize CDN molecules: (i) proteins containing GGDEF domains (Pfam family: PF00990); (ii) CD-NTases enzymes that have the catalytic domain known as SMODS, (PF18144)⁴²; and (iii) DAC proteins that have a catalytic domain called DAC domain (DisA_N domain, PF2457).

Proteins containing GGDEF domains synthesizes mainly 3' - 5' c-di-GMP (c-di-GMP) molecules, while proteins containing SMODS domain synthesize preferentially 3' - 5' cGAMP (cGAMP) molecules and proteins containing DAC domain synthesize mainly 3' - 5' c-di-AMP (c-di-AMP) molecules. Even though CDNs molecules are mainly synthesized by prokaryotic cells, eukaryotic cells also synthesize CDNs such as 2' - 3' cGAMP by cGAS enzymes. These three classes of CDN synthetases do not share structural similarities, have different residues involved to substrate binding and possess different catalytic mechanisms. Therefore, they are not homologues and probably evolved independently to catalyze analogous chemical reactions.

GGDEF domain structure and catalysis.

GGDEF structure and structural similarities with other protein domains. The GGDEF domain has an overall structure composed of a central five-stranded β sheet surrounded by five α helices⁷² and one hairpin (**Figure 1 B - C**). The GGDEF domain has structural similarities to three others catalytic domains: a) the class III adenylate and guanylate cyclase catalytic domains (Guanylate_cyc, PF002111), b) the GTP cyclohydrolase III (GCH_III, PF05165), and c) the palm domain of family Y DNA polymerases, such as IMS domain - impB/mucB/samB family domain, PF00817) (**Figure 1**). All of these families have a similar structural core composed by a β - α - α - β - β - α - β - α - β topology (**Figure 1 B**), which contains the Alpha-beta Plait topology (β - α - β - β - α - β), as defined by the CATH database⁷³. The specific version of the Alpha-beta Plait topology embedded in this group is better known as the RNA Recognition Motif-like fold (RRM-like fold) and corresponds to the so-called "palm domain" shared by archaeo-eukaryotic primases, reverse transcriptases, viral RNA-dependent RNA polymerases and families A, B, and Y of DNA polymerases⁷⁴.

The catalytic domains of class III adenyl cyclases (AC) and guanylyl cyclases (GC), are involved in the conversion of adenosine triphosphate (ATP) to 3' - 5' cyclic AMP (cAMP) and in the conversion of guanosine triphosphate (GTP) to 3' - 5' cyclic GMP (cGMP), respectively (**Figure 1 C**)^{75,76}. Class III AC and GC are well characterized: they are widely present in eukaryotic and prokaryotic cells and perform important function in many human tissues, being involved in signal transduction⁵⁴.

The GCH_III domain (GTP cyclohydrolase III) catalyzes the conversion of GTP to 2-amino-5-formylamino-6-ribosylamino-4(3H)-pyrimidinone 5'-phosphate (FAPy)⁷⁷. GCH III catalyzes two modifications on the GTP molecule that involves two hydrolysis reactions, one at the base (a cyclohydrolase activity) and another in the phosphodiester bond (phosphotransferase reaction) that causes the release of a pyrophosphate molecule that is converted to inorganic phosphates⁷⁷ (**Figure 1 C**). The palm domains recognized by the IMS model of the Pfam database are the catalytic domains of DNA polymerases such as prokaryotic DNA polymerase IV and eukaryotic DNA polymerases eta and kappa⁷⁸. All of them are Family Y DNA polymerases involved in DNA repair and exhibit error-prone behaviour⁷⁹. In these enzymes, the palm domain has deoxynucleotidyltransferase activity (**Figure 1 C**).

Given their conserved structural similarity to GGDEF domains, the class III adenyl/guanylyl cyclases (AC/GC), GTP cyclohydrolase III and the palm domain of DNA polymerases have been shown to be ancient homologous domains⁸⁰ that evolved from a common ancestor to perform different biological functions while preserving some core similarities such as: binding of nucleotides or deoxynucleotides and release pyrophosphate or phosphate during the enzymatic reaction course (**Figure 1 C**).

Residues important to GGDEF catalysis. The GGDEF domains are diguanylate cyclases (DGCs) that convert two molecules of GTP into one molecule of c-di-GMP. The GGDEF active site is thought to

be assembled only when two GGDEF domains come together in such a manner that permits the nucleophilic attack of the 3' OH groups on the α -phosphate groups of each GTPs, leading to the synthesis of one molecule of c-di-GMP and two molecules of pyrophosphate^{72,81,82}. Therefore, DGCs are Bi Ter enzymes (two substrates, three products)^{56,57}.

The catalytic activity of GGDEF domains is often regulated by input domains that precede the GGDEF domain, of which most are known or predicted to form dimers or heterodimers and to be sensor domains (**Figure 2 A**). Isolated GGDEF domains have little or no detectable enzymatic activity^{36,83} and require the dimerization of the input domain to assemble a catalytically competent GGDEF domain. Two hypothesis of the GGDEF activity regulation were reported, one of them suggests that the input domain binds its ligand and promote its dimerization and the GGDEF dimerization, the other hypothesis suggests that the input domain binds its ligand it causes a reorientation of the GGDEF domains from a non-catalytically competent GGDEF orientation to an active GGDEF dimer, or vice versa⁸⁴. The signal transduction from the input domain to the GGDEF domain is predicted to be relayed by a S-helix (signaling-helix) that connects the two domains and forms a two-helical parallel coiled coil (stalk) in the dimer form of the protein^{84,85}. Some proteins containing GGDEF domains possess a more complex activation mechanism and may involve formation of higher oligomers^{86,90,91}.

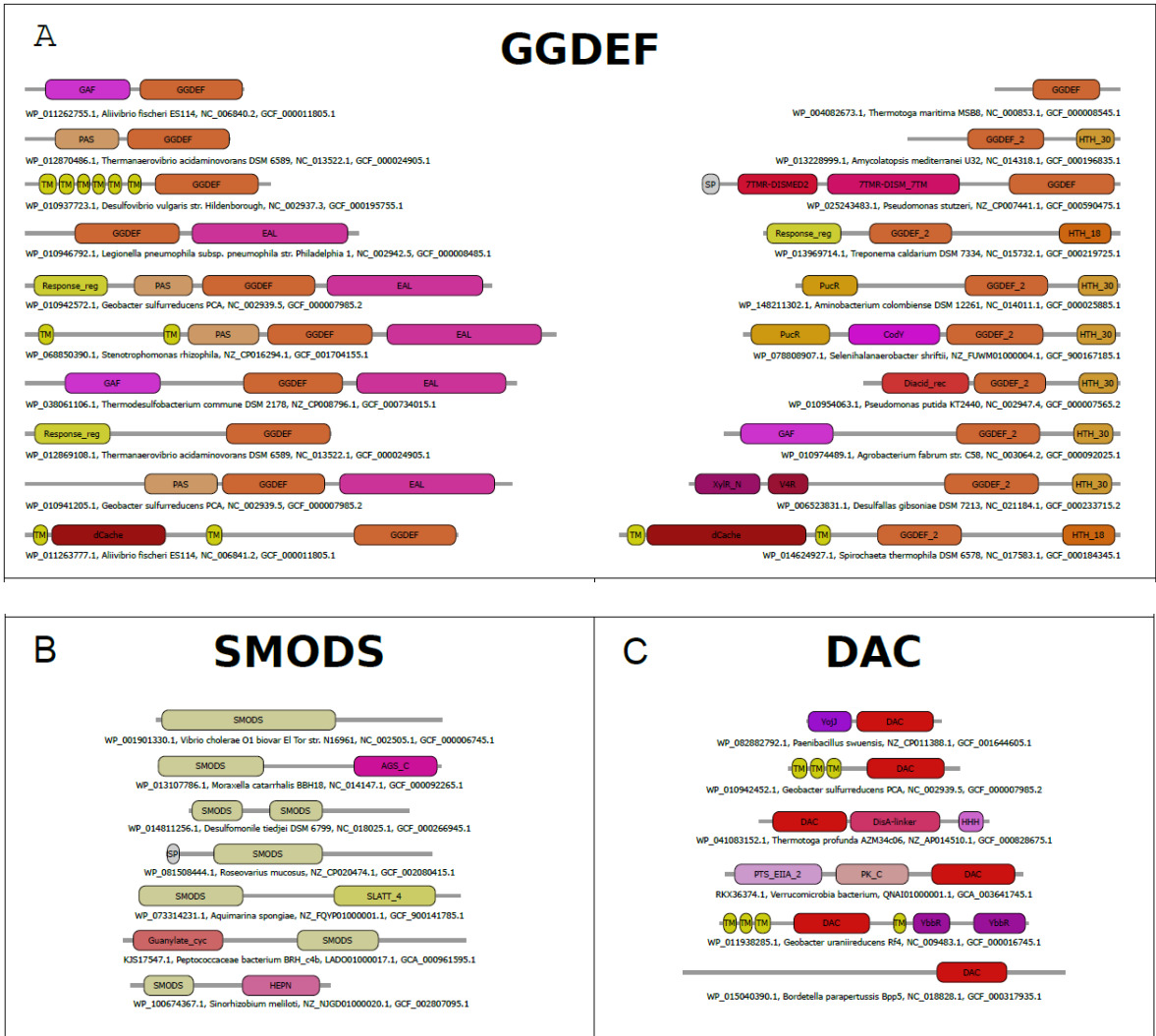


Figure 2: Domain architectures of proteins containing GGDEF, SMODS, and DAC domains. The most frequency domain architectures of proteins containing GGDEF (A), SMODS (B) and DAC (C) domains. The analysis was done using a non-redundant dataset (<80% identity) of proteins sequences built from sequences retrieved from the NCBI protein database¹³⁴. The name of the domains is based on the Pfam database.

The GG(D/E)EF motif (glycine, glycine, aspartic acid, glutamic acid and phenylalanine residues) is located in the loop between $\beta 2$ and $\beta 3$ (**Figure 3 A**), in which the glutamic acid residue binds to the α -phosphate group of GTP molecule as well as coordinates one of the cations located in the binding site (**Figure 3 B**). In the case of the PleD GGDEF domain, two magnesium cations are located in the binding site and are coordinated by E370 (from the GG(D/E)EF motif), D327, and the main chain of I328. The PleD residues D344 and N335 bind the guanosine base of the substrate, while the side chains of E370, K442, R446 and the main chains of F330, F331, and K332 bind the phosphate moieties of the GTP molecule (**Figure 3 B**)⁸¹. The GG(D/E)EF consensus sequence and most of the residues important to catalysis are very well conserved within GGDEF family members (**Figure 3 A**). This includes the D327, N335 and D344 residues, which have been reported to be essential to GGDEF domain activity^{72,86}

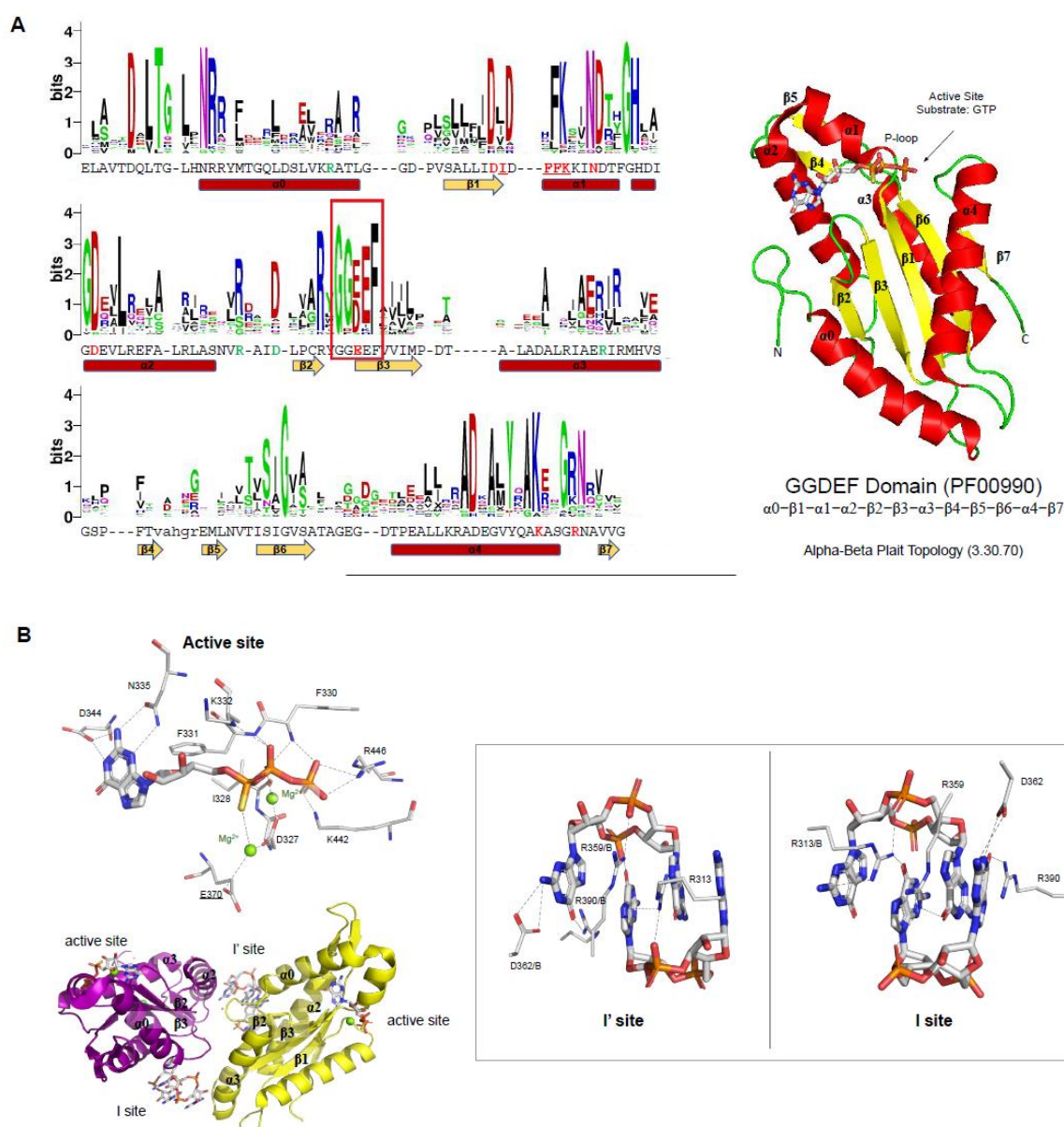


Figure 3. Conserved sequence within GGDEF members and active and I-sites pockets. (A) Residue frequency present in GGDEF domain. Using the Dali server⁹¹, 23 sequences of GGDEF domain structures were used to create a multiple sequence alignment, and the sequence logo was made with the WebLogo server¹³⁵. The sequence shown below the logo and the secondary structure elements belong to PleD of *Caulobacter vibrioides* (PDBID: 2V0N). Residues colored in red are involved in ligand or magnesium binding (for underlined residues, only the main chain is involved) and those colored in green are located in the I-sites. The GGDEF motif is placed in a red box. In the right the structure

of the GGDEF domain of PleD is shown as a cartoon. The topology of the GGDEF is shown below the structure, and the CATH topology name and code are also shown¹³⁶. (B) Interaction network between the GGDEF domain of PleD binding pocket with the substrate, GTP. In the bottom is shown the PleD structure in the inactive conformation, in which two inhibitory sites are shown (I-site and I'-site). In the right, it is shown in more detail the residues involved in the (c-di-GMP)₂ interactions at the inhibitory sites. Gray dotted lines represent hydrogen bonds. The magnesium ions are colored in green. The GTP and the protein residues involved in its binding are shown as sticks. Carbons are colored white, oxygens are red, nitrogen atoms are blue, and phosphorous atoms are orange.

A subclass of GGDEF domain, called Hybrid promiscuous (Hypr) GGDEF enzymes, synthesizes predominantly cGAMP molecules, but also synthesizes c-di-AMP and c-di-GMP molecules⁴⁰. The change in the substrate specificity seems to be related with the substitution of an aspartate (D344 of PleD located at the $\alpha 2$) by a serine, as this residue binds the guanine base of the GTP in PleD, but may also involve other yet unknown features of this subclass.

Some GGDEF domains have diverged from the canonical GG(D/E)EF amino acid sequence and are described as degenerate GGDEF domains, due to the loss of their catalytic activity. These degenerate GGDEF domains can evolve to possess different biological functions, and two examples have been described in the literature: 1- a degenerated GGDEF domain that is a sensor domain and binds GTP to activate the PDE activity in the neighboring EAL domain of the *Caulobacter crescentus* CC3396 protein³⁷ and; 2- a degenerated GGDEF domain of the *Bacillus subtilis* YybT protein that has unexpected ATPase activity⁸⁷.

GGDEF domain allosteric inhibition. The DGC activity of proteins containing GGDEF domain are inhibited by an allosteric noncompetitive product inhibition. A GGDEF dimer contains two symmetrical allosteric sites (I and I' sites), in which each allosteric site binds a c-di-GMP dimer (c-di-GMP)₂ (Figure 3 B). Both sites are formed by 4 residues, three of them from one GGDEF molecule, the RxxD motif (R359 and D362 of PleD) and an arginine (R390 of PleD), and the fourth residue is an arginine from the adjacent GGDEF molecule (R313 of PleD). The two (c-di-GMP)₂ dimers are expected to crosslink allosteric sites on opposite GGDEF domains, resulting in their immobilization in an inactive orientation^{72,81,86,88} (Figure 3 B). The RxxD motif and the positively charged residue (R390 of PleD) are conserved in GGDEF members (Figure 3 A).

SMODS Domain Structure and Catalysis.

SMODS structure and structural similarities with other protein domains. The *Vibrio cholerae* dinucleotide cyclase (DncV, the gene product of VC0179) has two domains, a SMODS domain located at its N-terminus and an Adenylyl/Guanylyl and SMODS C-terminal sensor domain (AGS-C)⁴² at its C-terminus⁸⁹. The first 23 residues of the protein are part of the AGS-C domain which is mainly an α -helical domain. The SMODS domain has two β -sheets connected by one β -strand ($\beta 3$) and it also has 6 α -helices that do not make part of the interface between the two domains. The two β -sheets are composed by the strands: $\beta 2$ - $\beta 3$ - $\beta 7$ - $\beta 8$ - $\beta 9$ and $\beta 3$ - $\beta 6$ - $\beta 5$ - $\beta 4$ (Figure 4 A). The substrate binding site is located in the interface between the two domains, in which the SMODS β -sheets make close contacts with the substrate (Figure 5 A). Proteins containing SMODS domains are also found associated with other domains (Figure 2 B) and, in rare cases, can be found in proteins containing two enzymatic domains: a SMODS and a class III AC/GC catalytic domain, both domains related with synthesis of cyclic nucleotide second messengers.

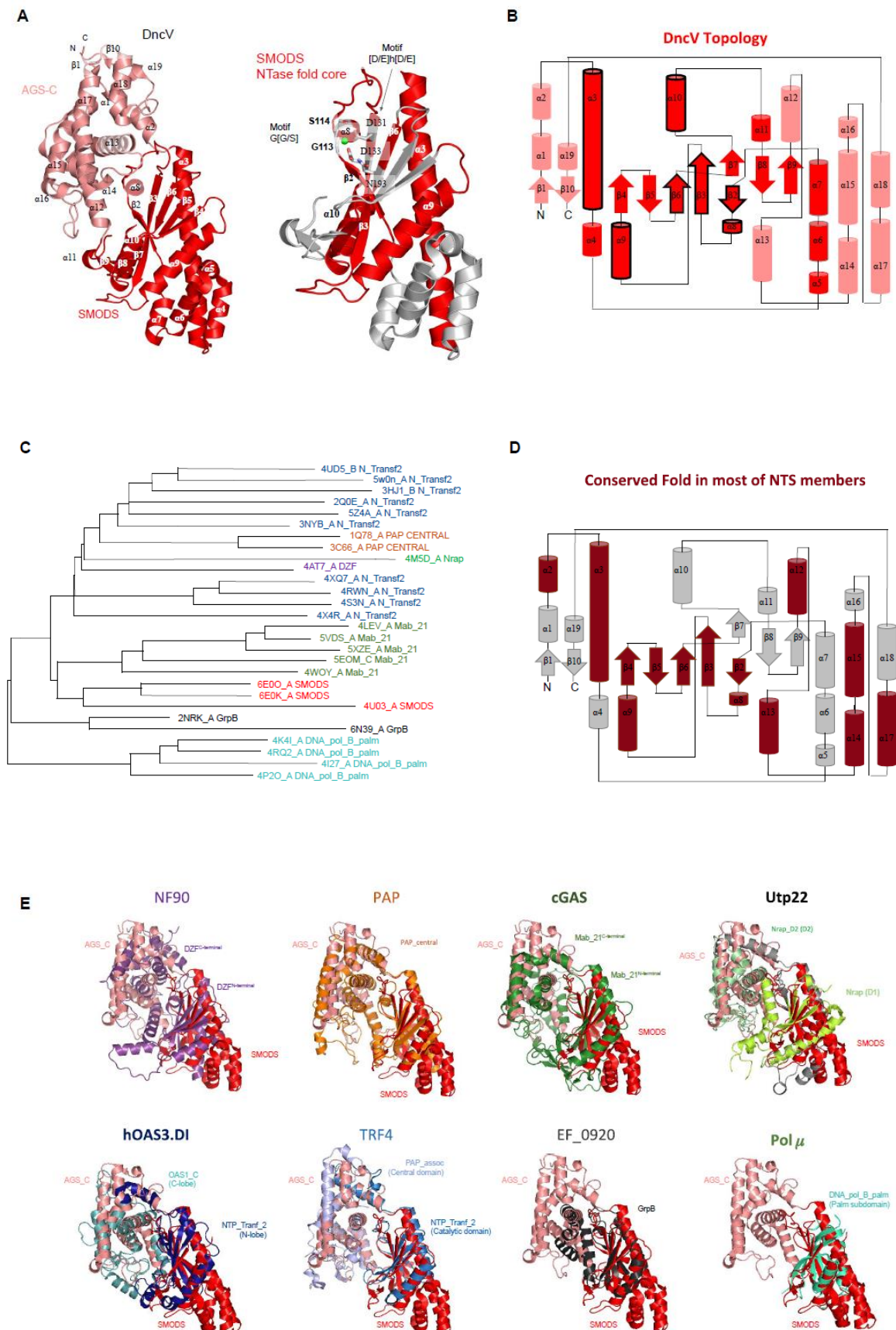
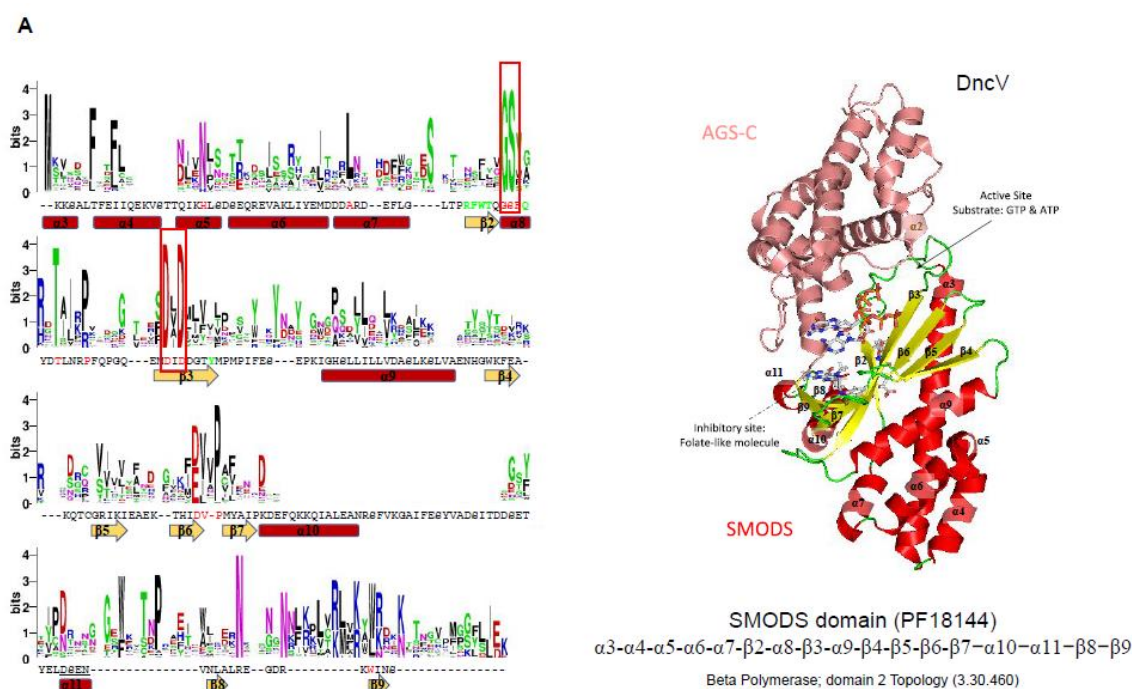


Figure 4. Structural similarities between SMODS domain and other nucleotidyltransferases superfamily members. (A) The N-terminal domain SMODS (colored in red) and the C-terminal ACS_C (light pink) of the DncV protein structure are shown as cartoons (PDBID: 4U0M). The SMODS

domain is involved in cGAMP synthesis and belongs to the nucleotidyltransferase superfamily (NTS). NTS fold structure is characterized by the presence of a minimal conserved core of a mixed β -sheet flanked by α -helices with $\alpha 1-\beta 1-\alpha 2-\beta 2-\alpha 3-\beta 3-\alpha 4$ topology that correspond to $\alpha 3-\beta 2-\alpha 8-\beta 3-\alpha 9-\beta 6$ (colored in red), missing the $\alpha 4$ element. Various insertions are observed and are colored in grey (right panel). Members of NTS contain three conserved motifs located at the active site: i) G[G/S] located at $\alpha 8$, ii) [D/E]h[D/E] (h indicates a hydrophobic amino acid) located at $\beta 3$, and (iii) [D/E] located at $\beta 6$. Two of these motifs are conserved in proteins containing SMODS domains (**Figure 5 A**). (**B**) Topology of DncV, showing the location of the NTS fold core (bold outlines). The secondary structure elements from SMODS are colored in red and the AGS-C domain in light pink. (**C**) Dendrogram showing structures similar to DncV made with the Dali server⁹¹ (query: DncV, PDBID: 4U03). Each domain is colored with different colors and the PDBID_chain and the Pfam name are shown for each branch. The conserved fold found in most of these structures, all NTS members, is shown in brown **panel D** overlaid on the DncV topology. (**E**) Structural superposition of DncV (PDBID: 4U03) with other NTS members found in the **panel C** dendrogram, using the same colors to represent each domain. At the top of each structural alignment is shown the protein's name. NF90 is colored in purple (PDBID: 4AT7), PAP is colored in orange (PDBID: 1Q78), cGAS is colored in dark green (PDBID: 5XZE), Utp22 is colored in light green (PDBID: 4M5D), hOAS3.DI is colored in blue (PDBID: 4S3N), TRF4 is colored in light blue (PDBID: 3NYB), EF_0920 (PDBID: 2NRK), Pol m is colored in green (PDBID: 2IHM).



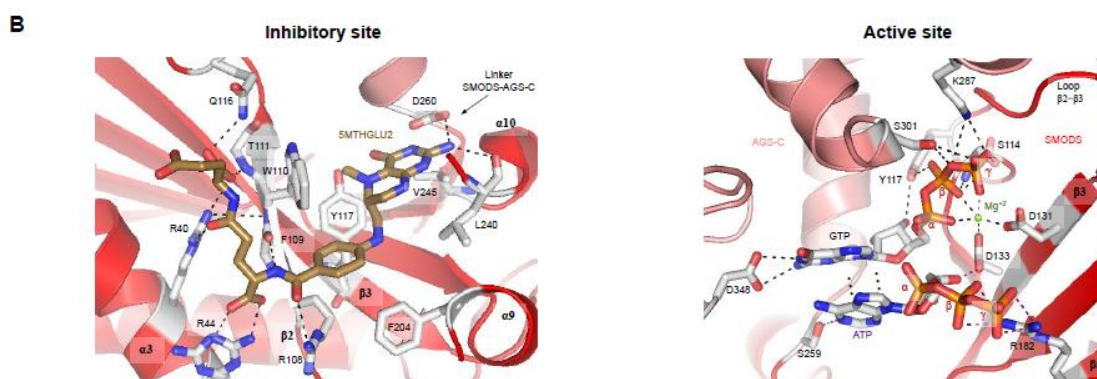


FIGURE 5: Conserved sequence within SMODS members and substrate and inhibitor binding pockets.

(A) Residue frequency present in SMODS proteins. 45 sequences were used from the Pfam database to create a multiple sequence alignment of the SMODS domain of different proteins using ClustalW¹³⁷, and the sequence logo was done using the WebLogo server¹³⁵. The sequence shown below in the logo and the secondary structure elements belongs to *Vibrio cholerae* di-nucleotide cyclase (DncV, PDBID: 4U03). Residues colored in green are involved in the interaction between DncV SMODS domain with the folate-like inhibitor, 5-methyltetrahydrofolate diglutamate (5MTHGLU2) molecule (residues Arg36, Arg40, Arg44 and Asp260 are not shown to be located at the AGS-C domain). Residues located at the SMODS domain involved in the catalytic activity are colored in red (residues Ser259, Lys287, Ser301, and Asp348 are not shown to be located at the AGS-C domain). The red boxes consist of G(G/S) and Dx(D/E) motifs found in members of NTS. The structure shown at the right belongs to the DncV protein (PDBID: 4U03), in which the AGS-C is colored in salmon and the SMODS domain is colored by secondary structure (β -strands in yellow and α -helices in red). The SMODS domain topology is shown below the structure, and the CATH topology name and code are also shown¹³⁶ (B) Interaction network between the DncV binding pocket (Active site) and its inhibitory site with substrate and inhibitor molecules, respectively. The substrates GTP and ATP are found bound at the active site, while the folate-like molecule (5MTHGLU2) binds at the allosteric site, inactivating the protein. The residues interacting with substrate and inhibitor molecules are shown as sticks and colored by element. The inhibitor carbons are colored brown. The magnesium ion is shown as a green sphere and gray dotted lines represent hydrogen bonds.

DncV have structural similarities with proteins belonging to the nucleotidyltransferase (NTase) fold, a highly diverse superfamily of proteins (Figure 4C and E)⁹⁰. NTase fold structure is characterized by the presence of a minimal conserved core of a mixed β -sheet flanked by α -helices with $\alpha 1-\beta 1-\alpha 2-\beta 2-\alpha 3-\beta 3-\alpha 4$ topology that correspond to $\alpha 3-\beta 2-\alpha 8-\beta 3-\alpha 9-\beta 6$, missing the $\alpha 4$ element (Figure 4 A and B). This common core is usually decorated by various additional structural elements depending on the family. The NTase fold core is present in the DncV SMODS domain and various insertions are observed (Figure 4 A and B). Members of NTases superfamily contain three conserved motifs located at the active site: (i) hG[G/S] located at $\alpha 2$; (ii) [D/E]h[D/E]h (h indicates a hydrophobic amino acid) located at $\beta 2$, and (iii) h[D/E]h located at $\beta 3$ ⁹⁰. These three conserved motifs in DncV protein correspond to: (i) G₁₁₃-S₁₁₄ (located at $\alpha 8$); (ii) D₁₃₁-I₁₃₂-D₁₃₃ (located at $\beta 3$); and (iii) N₁₉₃ (located at $\beta 6$). Two of these motifs are conserved in proteins containing SMODS domains, the hG[G/S] and [D/E]h[D/E]h motifs (Figure 4 A).

More specifically, as revealed by Dali searches⁹¹, the DncV structure shares strong structural similarities to: (i) proteins containing DZF domain (Domain associated with Zinc Fingers, PF07528) such as nuclear factor NF90 and NF45; (ii) the catalytic domain of Poly(A) polymerase, PAP, (PAP_central domain, PF04928)⁹²; (iii) the catalytic domain of eukaryotic cGAS enzymes (Mab-21 domain, PF03281); (iv) the D1 and D2 domains of U3 small nucleolar RNA-associated protein 22 (Utp22)⁹³, in which the D1 domain matches the PFAM model named nucleolar RNA-associated proteins domain (Nrap, PF03813) and the D2 domain matches the Nrap protein PAP/OAS-like model

(Nrap_D2, PF17403)⁹⁴; (v) the OAS1-like domain of the dsRNA-activated oligoadenylate synthase (OAS) protein, which is described in the PFAM database as two domains: a N-terminal domain matching the NTP_transf_2 model (PF01909), also known as N-lobe of human OAS3 pseudoenzymatic domain DI (hOAS3.DI), and a C-terminal OAS1_C domain (PF10421), also known as C-lobe of hOAS3.DI⁹⁵; (vi) the catalytic and central domains of Poly(A) RNA polymerase protein 2 (TRF4 gene), where the catalytic domain matches the NTP_transf_2 PFAM model (PF01909) and the central domain matches the PAP_assoc model (Cid1 family poly A polymerase, PF03828)^{96,97}; (vii) proteins matching the GrpB domain (PF04229)⁹⁸, this domain is found in uncharacterized proteins such as EF_0920 from *Enterococcus faecalis*; and (viii) the palm subdomain (DNA_pol_B_palm, PF14792)⁹⁹ of the DNA polymerase μ (Pol μ) from the family X¹⁰⁰ that includes DNA polymerase β , γ , and μ ¹⁰¹ (**Figure 4 E**).

All of these proteins belong to the Nucleotidyltransferase superfamily (NTS) and most of them share not only the NTase fold core but also some secondary structures from the AGS-C domain (**Figure 4 D**) suggesting that the domain interface is conserved across these families. It is worth mentioning that the AGS-C domain shares structural similarities with domains that are commonly associated with the catalytic domain of members of Ntase, such as DZF C-terminal domain, OAS1_C, PAP_assoc and Nrap_D2 (**Figure 4 E**).

DZF domain forms dimers and heterodimers and they are found in proteins involved in gene expression and RNA metabolism such as NF90 that forms a complex with NF45 and regulates genes expression¹⁰². Poly(A) polymerase (PAP) is involved in eukaryotic mRNA processing by its polyadenylation at the end of transcription process, so PAP incorporates ATP at the 3' end of mRNA¹⁰³. In metazoans, the cGAS enzyme, which has a Mab_21 domain, binds cytoplasmatic double-stranded DNA (dsDNA) to activate synthesis of 2' – 3' cGAMP molecules and initiate host innate immune responses. Endogenous or exogenous dsDNA in the cytoplasm, which could be from damaged mitochondria or from an invasion of pathogenic bacteria or viruses, indicates major danger to eukaryotic cells. The cytosolic accumulation of 2' – 3' cGAMP activates type-1 mediated stress-responses via STING and regulates autoimmunity in human cells¹⁰⁴. Human dsRNA-activated oligoadenylate synthase (OAS), which matches both NTP_transf_2 and OAS1_C PFAM models, is a mammalian dsRNA sensor, which is increased during pathogen infections, and activates the synthesis of a second messenger 2' – 5'-linked RNA molecules to cause RNA decay^{105,106}.

TRF4 protein, which also contains a region most similar to the NTP_Transf_2 model, makes part of a polyadenylation TRAMP complex that recognizes aberrant eukaryotic RNAs and target them for degradation¹⁰⁷. Members of DNA polymerase family X, which have a “palm” subdomain, play essential roles in the base-excision repair mechanism, a process that repair cell DNA base damage, being responsible for DNA synthesis and 5'-deoxyribose-phosphate (dRP) removal (dRP lyase activity). These enzymes can be also involved in other DNA repair processes such as non-homologous end-joining and lesion bypass^{101,108,109}. Utp22, which has an Nrap_D2 domain, forms a complex with Rrp7 and they are present in early precursors of the small ribosomal subunit. Utp22 is a structural building block and apparently lacks any enzymatic activity⁹³.

Among all described enzymes, cGAS enzymes are the only ones that share functional similarities with the prokaryotic DncV enzymes. While DncV synthesizes 3' - 5' cGAMP molecules, cGAS enzymes synthesize 2' - 3' cGAMP. Most of the other enzymes with structural similarity with the DncV SMODS domain have functions related to DNA or RNA processing.

Active site of SMODS Domains. DncV synthesizes preferentially cGAMP, but also produces c-di-AMP and c-di-GMP molecules. DncV regulates the expression of more than 80 genes in *Vibrio cholerae* and its DGC activity is inhibited by folate-like molecules *in vitro*^{5,89,110}. Orthologs of DncV are found in Gram-negative and Gram-positive bacterial species and the residues involved in the ligand binding and to folate-like molecule binding are conserved among them⁸⁹ (**Figure 5 A**). The active site is located in the interface between the SMODS and AGS-C domains, while the folate-like molecules bind at the opposite of the substrate-binding pocket at the flat side of the molecule (**Figure 5 A**). The folate-like molecule binds mainly at the SMODS domain and makes few interactions with the linker between the two domains (**Figure 5**). The inhibitory site of the DncV, which binds folate-like

molecules, such as 5-methyltetrahydrofolate diglutamate (5MTHGLU2), is formed by side chains of Arg36, Arg40, Arg44, Arg108, Trp110, Gln116, Tyr137, Phe204 and Asp260, by the main chain of Phe109, Thr111 and Leu240 and by a hydrophobic pocket formed mostly by the side chains of Leu240 and Val245 (**Figure 5 B**). These residues are not conserved in members of SMODS (**Figure 5 A**).

The active site of DncV is built by 9 residues: Ser114, Tyr117, Asp131, Asp133, Arg182, Ser259, Lys287, Ser301, and Asp348. Asp348 and Ser259 interact with the guanine and adenine bases, respectively. Tyr117 and Asp133 are involved in the interactions with the ribose of the guanine and the adenine nucleotides, respectively. Arg182 binds the β and γ phosphate groups of the ATP. Tyr117, Ser114, Lys287 and Ser301 interact with the β and γ phosphate groups of GTP. The magnesium ion is coordinate by Asp131, Asp133, and the α , β and γ -groups of the GTP ligand (**Figure 5 B**). These two aspartic residues belong to the G(G/S)X₉₋₁₃DX(D/E) motif conserved in members of NTS and are key residues in (2' - 5') oligoadenylate synthetase (OAS1) and poly(A)polymerase activities¹¹¹.

DAC domain structure and catalysis.

DAC structure and structural similarities with other protein domains. c-di-AMP is synthesized by DAC enzymes that convert two molecules of ATP into one c-di-AMP and two pyrophosphate molecules. In the case of DisA, the Rv3586 protein from *Mycobacterium tuberculosis*, the synthesis of c-di-AMP is made using ATP or ADP¹¹². Production of c-di-AMP has been described as essential for the growth of some Gram-positive bacteria due to it being involved in crucial cellular activities, such as cell wall metabolism, maintenance of DNA integrity, ion transport, cell division, and cell size control^{21,113-115}. *Bacillus subtilis* encode three DAC enzymes, DisA, CdaA and CdaS. Two of them, DisA and CdaA, are constitutively expressed during vegetative growth while CdaS is required for efficient germination of spores. Other Gram-positive bacteria encode only one DAC protein that is essential for their growth, as observed in *Listeria monocytogenes*, *Streptococcus pneumoniae* and *Staphylococcus aureus*, thus making this enzyme a likely target for constructing new inhibitors that may serve as antibiotics for pathogenic Gram-positive bacteria.

At the moment three structures of proteins containing DAC domains have been solved: DisA from *T. maritima* (PDBID 3C1Y)⁴, DacB from *B. cereus* (PDBID 2FB5)¹¹⁶; and DacA (CdaA-APO Y187A Mutant) from *L. monocytogenes* (PDBID: 6HVN)¹¹⁷. As described before, no structural similarities with other domain were detected so far. The overall DAC domain structure exhibits a globular α/β fold with a slightly twisted central β -sheet, made up of seven mixed-parallel and antiparallel β -strands (β 1 – β 7) surrounded by five α -helices (α 1 - α 5), in which the N-terminal helix (α 1) can be split in two parts (α 1' and α 1). Like GGDEF domains, two DAC domains must be correctly oriented to allow the conversion of two ATP molecules into one c-di-AMP molecule and two pyrophosphates. Therefore, DAC domains are also Bi Ter enzymes (two substrates, three products). The regulation of the catalytic activity of DAC domains may be regulated by input domains (**Figure 2 C**), and in the case of DisA from *T. maritima* (PDB code 3C23, 3C1Z, and 3C1Y)⁴, the protein forms a homo-octamer and the DAC domains are oriented in a such way that two active sites are oriented face to face to allow the catalysis. Linear DNA or DNA ends do not affect the protein activity but branched nucleic acids (such as in Holliday junctions) strongly suppress the DAC activity of DisA by binding to its C-terminal domain⁴.

Active site of DAC Domains.

In the case of the DAC domain of CdaA from *Listeria monocytogenes* (PDBID 4RV7) the ATP ligand is located in a well-defined cavity made up by N-terminal of α 4, loop β 5 – β 6, loop β 4- α 4, and loop α 3- β 3 (**Figure 6 C**), in which many conserved residues of the DAC domains are also located: the GALI motif, GxRHRxA motif, an absolutely conserved serine, DGAhh motif (h is a hydrophobic residue), and (V/I)SEE motif (**Figure 6 A**).

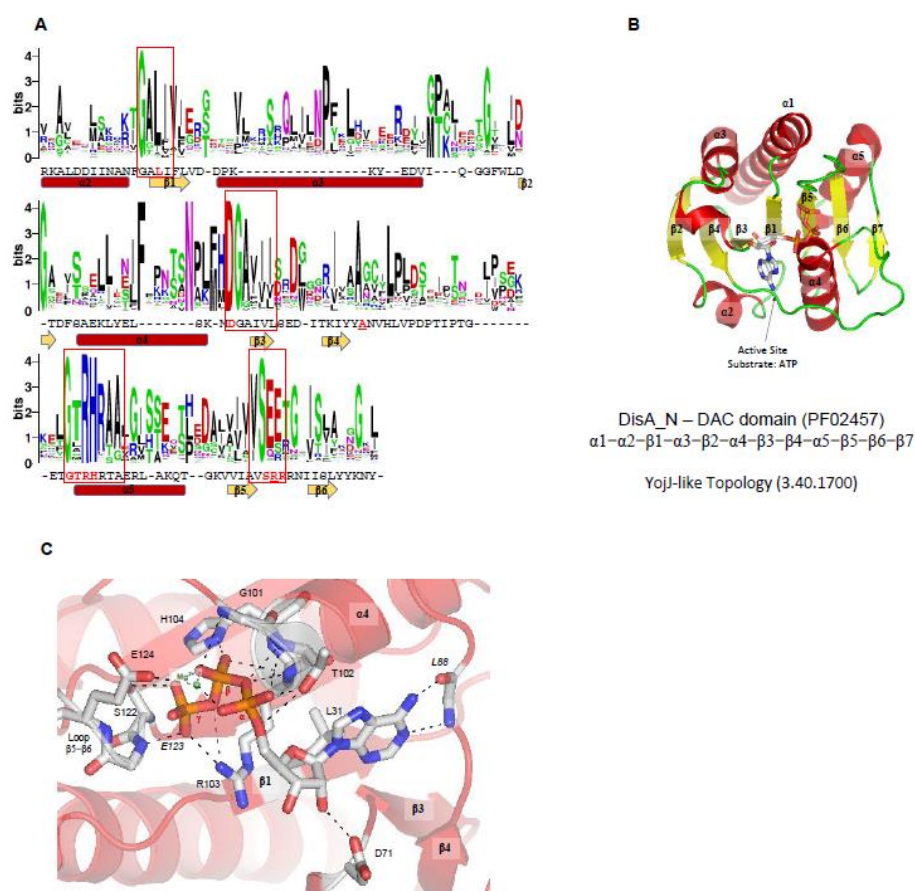


Figure 6. Conserved sequence within DAC domain members and substrate binding pocket. (A)

Residue frequency present in DAC proteins. 609 sequences were used from the Pfam database to create a multiple sequence alignment of the DAC domain of different proteins using ClustalW¹³⁷, and the sequence logo was done using the WebLogo server¹³⁵. The sequence shown below in the logo and the secondary structure elements belongs to DisA protein from *T. maritima* (PDBID: 3C1Y). Residues that bind ATP or the magnesium cation are colored in red, underlined residues bind mainly by the main chains. Conserved motifs within DAC members are placed in red boxes: GALI, DGAhh, GxRHRxA, and (V/I)SEE motifs. **(B)** Structure of the DAC domain of CdaA from *Listeria monocytogenes* (PDBID: 4RV7). The substrate ATP is found bound at the active site. The DAC domain topology is shown below the structure, and the CATH topology name and code are also shown¹³⁶ **(C)** Interaction network between the CdaA binding pocket with the substrate, ATP (PDBID: 4RV7). Gray dotted lines represent hydrogen bonds. The magnesium ion is colored in green. The ATP and the protein residues involved in its binding are shown as sticks. Carbons are colored white, oxygens are red, nitrogen atoms are blue, and phosphorous atoms are orange.

The active site of CdaA DAC domain is thought to be built by 10 residues: the side chains of Leu31, Asp71, Thr102, Arg103, His104, Ser122, and Glu124 and the main chain of Leu88, Gly101, and Glu123 (**Figure 6 C**). Leu31 belongs to the GALI motif, the Asp71 belongs to the DGAhh motif, and Thr102, Arg103, and His104 belongs to GxRHRxA motif. The main chain of Leu88 and the side chain of L31 interacts with the adenine base, Asp71 binds the ribose, and Thr102, Arg103, His104, Ser122, Gly101, and Glu123 bind the phosphate groups of the ATP molecule. The Glu124 coordinates one magnesium cation that binds the α and β phosphate groups of the ATP molecule (**Figure 6 C**)⁴⁹. CdaA from *L. monocytogenes* is active in the presence of Mn^{+2} or Co^{+2} but inactive in the presence of Mg^{2+} ions¹¹⁷. However, in the case of DisA from *M. tuberculosis*¹¹² and *T. maritima* the enzymes are active in the presence of Mg^{2+} ions.

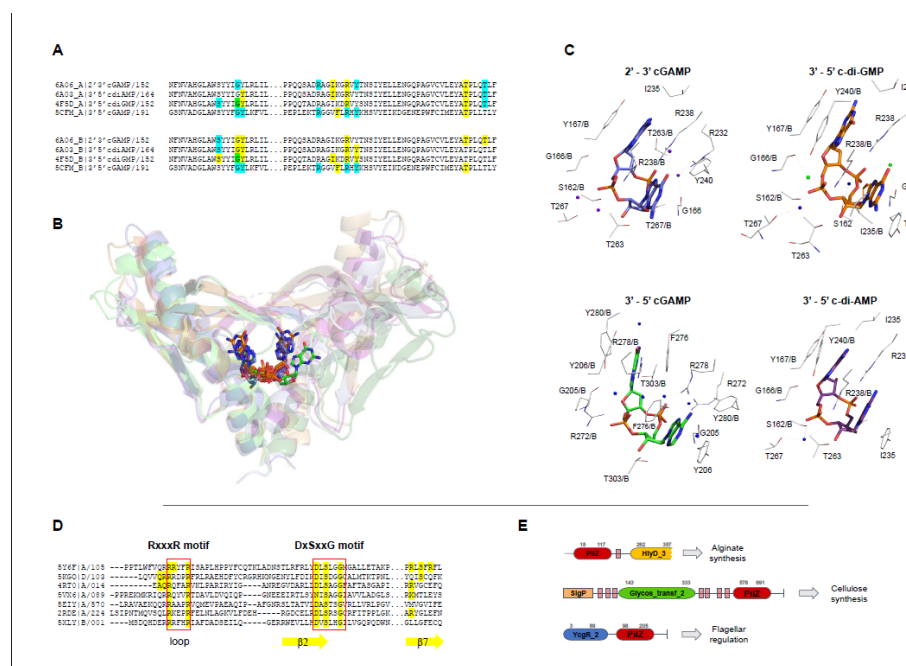
CYCLIC DI-NUCLEOTIDE RECEPTORS

Different classes of CDN receptors have been described and are involved in the regulation of a broad range of bacterial behaviors, while in eukaryotic cells they are associated with the activation of innate immune response through interactions with STING proteins. In order to analyze the residues involved in the CDN ligand and to compare the ligand structure inside of the protein pocket, we focused on the CDN receptors with three-dimensional structures solved and deposited in the Protein Data Bank (PDB) (**Table 1**).

Receptors of c-di-GMP are very well studied due to it being the first CDN identified as a bacterial second messenger. Therefore, most of the receptors with the structure solved in complex with the ligand analyzed in this review are c-di-GMP receptors. Examples of these are STING proteins and proteins containing PilZ domains, degenerate EAL domains, the allosteric site of the DGC enzymes containing GGDEF domains, and RNA structures such as c-di-GMP I and c-di-GMP II riboswitches. The C-terminal domain of the master regulator of *Streptomyces* cell development BldD, the REC domain of the transcriptional regulator VspT protein, the N-terminal domain of the ATPase of the Type II secretion system MshE protein (T2SSE_N domain), the MerR domain of BrlR protein, the AAA+ ATPase domain (Sigma54_activat, PF00158) of the transcriptional regulator FleQ protein, and the CA domain (HATPase_c, PF02518) of the cell cycle kinase CckA protein are also c-di-GMP receptors and their function are regulated by the CND ligand (**Table 1**).

Different classes of c-di-AMP receptors are also identified, such as proteins containing STING domains, the Aldo-keto reductase domain of the oxidoreductase RECON proteins, the cyclic-di-AMP receptor domain of PII-like signal transduction protein (PtsA), the pyruvate carboxylase domain (HMGL-like domain, PF00682) of *L. monocytogenes* pyruvate carboxylase (LmPC) or *L. lactis* pyruvate carboxylase (LIPC), the TrkA_C domain of the potassium transporter A (KtrA), the CBS domain of OpuC carnitine transporter, and RNA structures such as ydaO-yuaA riboswitches (**Table 1**). The cGAMP, the receptors analyzed in this review are STING proteins and c-di-GMP I riboswitches.

The function of each CDN receptor, as well as the residues involved in ligand binding, is described in more detail at **Table 1**. It is notable that most of the CDN receptors are specific to their ligands, with exception of receptors involved in mammalian cell innate immunity, such as STING, that interact with c-di-GMP, c-di-AMP and 3' - 5' and 2' - 3' cGAMP molecules using the same binding pocket (**Figure 7 B**). Interestingly, even though the CDNs are chemically different, the STING binding pockets for each kind of CDN are very similar and the residues involved in ligand binding for each CDN are almost the same (**Figure 7 A and C**). This suggests that STING adjusts the ligand binding site for each CDN by placing or removing water or magnesium molecules.



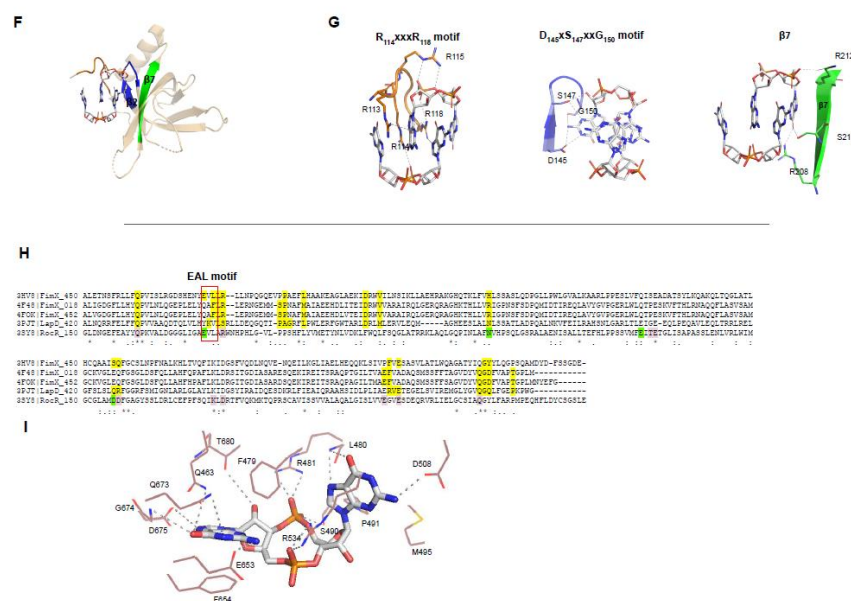


Figure 7. Some CDN receptors: STING, PilZ and degenerate EAL domains. **(A)** Multiple sequence alignment of STING proteins in complex with different CDNs: monomeric 3' - 5' c-di-GMP (PDBID: 4F5D), 2' - 3' cGAMP (PDBID: 6A06), 3' - 5' c-di-AMP (PDBID: 6A03), and 3' - 5' cGAMP (PDBID: 5CFM). The residues highlighted in yellow are involved with direct interaction with the ligand, in green are involved with interaction with the ligand via magnesium ions and in cyan via water molecules. **(B)** Structural superposition of STING proteins (domain TMEM173) in complex with different CDNs bound at the same protein interface. The ligands are colored by element, nitrogen atoms are in dark blue and oxygens are in red. Carbons are colored according to the ligand: 3' - 5' c-di-GMP in orange; 2' - 3' cGAMP in blue, 3' - 5' c-di-AMP in purple and; 3' - 5' cGAMP in green. **(C)** Residues involved with interaction with the different ligands are the same highlighted in A. The residues are colored by element, with carbon in white, nitrogen in dark blue and oxygen in red, while the ligands are colored as described in B. Water molecules are shown as blue spheres while magnesium cations are shown as green spheres. **(D)** Multiple sequence alignment of the PilZ domains that had their structures solved in complex with c-di-GMP molecules. Residues highlighted in yellow are involved in interactions with the ligand. The secondary structure elements shown belong to the PilZ domain of YcgR from *E. coli* (PDBID: 5Y6F). The motifs conserved within PilZ members are placed in red boxes. **(E)** Some domain organizations found in proteins containing PilZ domains (Alg44, BcsA and YcgR) and their related functions. **(F)** Structural representation of a PilZ domain as a cartoon (PilZ domain of YcgR, PDBID: 5Y6F). Residues belonging to the "RxxxR" motif are colored in orange, the "DxSxxG" motif are colored in blue and the β strand 7 is colored in green. The dimeric c-di-GMP is shown as sticks. The interaction network presented in this figure is shown in more details in G. **H** Multiple sequence alignment of proteins containing degenerate EAL domains (PDBID: 4F48, 4FOK, and 3PJT), and a catalytic EAL domain (PDBID: 3SY8) that had their structure solved in complex with monomeric c-di-GMP. The residues highlighted in yellow are involved with direct interaction with the ligand, in green are involved with interaction with the ligand via magnesium ions. In the case of RocR, which has a catalytic EAL domain, the residues highlighted in salmon were experimentally demonstrated to be important to catalysis¹³⁸. The consensus "EAL" motif is placed in a red box. **(I)** Interaction network of the binding site of a FimX degenerate EAL domain (PDBID: 4FOK). Gray dotted lines represent hydrogen bonds. The residues and the c-di-GMP are colored by element. The multiple sequence alignment were performed using the server ¹³⁷.

STING proteins are localized on the endoplasmic reticulum membrane of eukaryotic cells and are CDNs sensors that, when bound, regulate the induction of type I interferons (IFN- α and IFN- β), thus eliciting the intracellular signals of the invasion by bacteria and/or viruses, and activating the

innate immune response to attack the pathogen. The STING protein can directly sense the pathogen invasion by interaction with bacterial CDNs (3' - 5' c-di-GMP, 3'-5' c-di-AMP or 3'-5' cGAMP) or indirectly by binding to eukaryotic 2' - 3' cGAMP through its C-terminal domain (TMEM173, PF15009). It is controversial whether STING binds 2' - 5' cGAMP preferentially¹¹⁰. 2' - 3' cGAMP is synthesized by a eukaryotic cGAS enzyme, as described above, by sensing cytoplasmic endogenous and exogenous dsDNA that could be from damaged mitochondria or from invasion of pathogenic bacteria or viruses, and thus indicates major problems in the eukaryotic cells.

Other important c-di-GMP receptors are proteins containing the PilZ domain (PF07238). Proteins containing PilZ domains are involved in various metabolic pathways. Bioinformatics studies were the first to suggest that the PilZ domain could be related with sensing c-di-GMP and controlling the transition between sessile growth and motility in bacteria¹¹⁸. The first experimental evidences of how this occurs came from studies in *Vibrio cholerae*, where a conformational shift at the N-terminal portion of VCA0042 in the presence of the ligand indicated that this protein binds c-di-GMP and VCA0042 and another PilZ-containing protein (VC2344/PilZC) were shown to be important to *Vibrio cholerae* virulence¹¹⁹.

PilZ domains regulate twitching and swarming motility via the flagellar regulator YcgR protein⁶, but proteins containing different domain architectures are related with other functions, such as the regulation of the synthesis of cellulose by BcsA in *Rhodobacter sphaeroides*, or chemotaxis by MapZ protein and alginate secretion by Alg44 to promote the biofilm formation in *Pseudomonas aeruginosa* (**Figure 7 E** and **Table 1**). PilZ domain is found associated with different domains that could be sensor domains, such as GAF, Cache, and PAS domains, and catalytic domains such as GGDEF, EAL, and Peptidase_S8. The Pfam database describes 221 different domain architectures containing PilZ domains¹²⁰, showing the diversity of signaling networks in which c-di-GMP can be involved and have not yet been explored.

It is interesting that two proteins containing PilZ domains that are c-di-GMP receptors have been found to be involved in the production of different exopolysaccharides to produce bacterial biofilms, such as cellulose and alginate. Moreover, another c-di-GMP receptor is also involved in exopolysaccharide production, the PelD of *Pseudomonas aeruginosa* that regulates the synthesis of the Pel exopolysaccharide (**Table 1**).

PilZ proteins interact with dimeric c-di-GMP by two conserved sequence motifs: RxxxR and DxSxxG motifs (**Figure 7 D**). In the RxxxR motif located in a loop at the N-terminal part of the PilZ domain, each arginine is interacting with the phosphate group and the base of the ligand. In the case of the DxSxxG motif, the aspartic acid, serine and glycine residues bind the base and/or the pentose ring of the c-di-GMP molecules (**Figure 7 G**). Other residues not conserved within members of the PilZ family are also involved in ligand binding and some of them are located at the β -strand 7 of the PilZ protein (**Figure 7 D, F** and **G**). Some PilZ proteins lost their canonical residues to bind c-di-GMP and are not c-di-GMP receptors anymore but may work as protein-protein adaptors, as happens with the complex FimX-PilZ-PilB that regulates the twitching motility in *Xanthomonas citri*⁷. This ternary complex is an example of a full set of "degenerate" GGDEF, EAL and PilZ domains, in which GGDEF does not synthesize c-di-GMP, PilZ does not bind c-di-GMP and the EAL domain does not cleave c-di-GMP but kept the ability to bind it⁷.

Degenerate EAL proteins lost their ability to cleave c-di-GMP to pGpG, but still bind c-di-GMP molecules and the residues involved in this interaction are described in the **table 1** and **Figure 7 H** and **I**. The loss of the EAL domain catalytic function seems to be related with a change in the residues important for the coordination of a magnesium cation (**Figure 7 H - I**).

In *Xanthomonas citri*, *Xanthomonas campestris* and *Pseudomonas aeruginosa*, FimX protein regulates twitching motility by sensing c-di-GMP levels via interaction with degenerate EAL domain and regulates the type IV pilus machinery⁷. LapD from *Pseudomonas fluorescens*, is a transmembrane receptor of c-di-GMP, in which the C-terminal degenerate EAL domain binds c-di-GMP to prevent cleavage of the surface adhesin LapA and therefore, activates biofilm formation¹²¹.

Different classes of RNA riboswitches sense different kinds of CDNs (**Table 1**). Riboswitches are structured RNAs located in the 5'-untranslated regions of mRNAs and some can sense CDNs

molecules to change its structure and therefore regulate expression of downstream genes that could be involved with virulence, motility, biofilm formation, cell wall metabolism, synthesis and transport of osmoprotectants, sporulation and other important biological processes^{122,123}.

There are three classes of riboswitches that bind CDNs and have had their structures solved in complex with their ligand and deposited in the Protein Data Bank: c-di-GMP I riboswitch (RF01051), c-di-GMP II riboswitch (RF01786), and c-di-AMP riboswitch (ydaO-yuaA riboswitch, RF00379). The c-di-GMP I riboswitch was originally annotated as a conserved RNA-like structure of Genes Related to the Environment, Membranes and Motility (GEMM motif) and later another c-di-GMP riboswitch class was identified, the c-di-GMP II riboswitch. They have the same function but do not share any sequence motif or structural similarities. The c-di-AMP riboswitch is one of the most common riboswitches in various bacterial species and is found in the vicinity of genes related to cell wall metabolism, sporulation in Gram-positive bacteria, and other important biological processes^{122,123}.

The TetR-like transcriptional factor, DarR, from *Mycobacterium smegmatis* was the first c-di-AMP receptor discovered²⁶, where c-di-AMP stimulate the DNA binding activity of this protein. DarR is a repressor that negatively regulates the expression of its target genes²⁶. Another protein that interacts with c-di-AMP by a poorly understood mechanism is KdpD/KdpE that controls the potassium uptake in situations where the potassium concentrations are extremely low and other uptake systems wouldn't be enough to give the cell all potassium it requires. In *Escherichia coli*, there are three systems responsible for potassium uptake, namely, Trk, Kdp, and Kup. In the case of Trk system, four genes are constitutively expressed and TrkA is the predominant potassium transporter at neutral pH. The Kdp-ATPase system is induced at low potassium concentrations and under conditions of osmotic stress. The Kup, formerly TrkD, is activated when TrkA and Kdp activities are not sufficient¹²⁴⁻¹²⁶. In *Bacillus subtilis*, a novel high-affinity transporter KimA (formerly YdaO) has recently been characterized and the expression of KimA and KtrAB is negatively regulated by c-di-AMP riboswitches²⁷. When the concentration of potassium is high in the cell, the concentration of c-di-AMP increases inhibiting potassium uptake by two ways, by binding to c-di-AMP riboswitches that will avoid the expression of proteins involved in transport, and by direct interactions with regulatory subunits of KtrAB and KtrCD causing the inhibition of potassium transport¹²⁷. A similar process seems to happen in *Staphylococcus aureus*, where c-di-AMP binds to KtrA proteins and to the universal stress protein (USP) domain of the KdpD sensor kinase inhibiting the expression of Kdp potassium transporter components. In this manner, c-di-AMP appears to be a negative regulator of potassium uptake in different Gram-positive bacteria, such as *S. aureus*^{68,127}.

One of the most well understood receptors for c-di-AMP is KtrA, which binds c-di-GMP through its C-terminal domain (RCK_C or TrkA_C) causing inactivation of the KtrA function (**Table 1**). C-di-AMP binds to the interface of the KtrA homodimer and the residues involved in the ligand interaction are described in the **Table 1**. Another c-di-AMP receptor is the c-di-AMP receptor domain (PF06153) of the PII-like signal transduction protein, PstA. PstA is a homotrimer and in each protein interface one c-di-AMP molecule is bound. The residues involved in ligand binding in PstA are also described in **Table 1**.

c-di-AMP is also related with negative control of aspartate and pyruvate pools in *Listeria lactis* by a pyruvate carboxylase, LIPC protein, and *Listeria monocytogenes* pyruvate carboxylase, LmPC protein, respectively. In both cases, c-di-AMP binds to the pyruvate carboxylase domain (HMGL-like domain in the Pfam) (**Table 1**). LIPC forms a tetramer and each c-di-AMP molecule binds the protein dimer interface at the carboxyltransferase (CT) domain in a binding site pocket containing residues that are poorly conserved among pyruvate carboxylases¹²⁸.

CONFORMATION OF CYCLIC DI-NUCLEOTIDES INSIDE THE BINDING SITE OF RECEPTORS

The cyclisation between the two nucleotides of the most common CDNs involves the formation of a phosphodiester bond that links the C3' of one pentose ring with the C5' of another, resulting in a 3' - 5' cyclic di-nucleotide. This kind of cyclisation creates a two-fold symmetry between the two pentose rings of the di-nucleotides. Only c-GAMP has been reported to present not only a 3' - 5' linkage, but also being found with a 2' - 3' one that contains two distinct phosphodiester linkages, one between C3' of AMP and C5'-phosphate of GMP, and the other between C5'-phosphate of AMP and C2' of GMP (Figure 8 A).

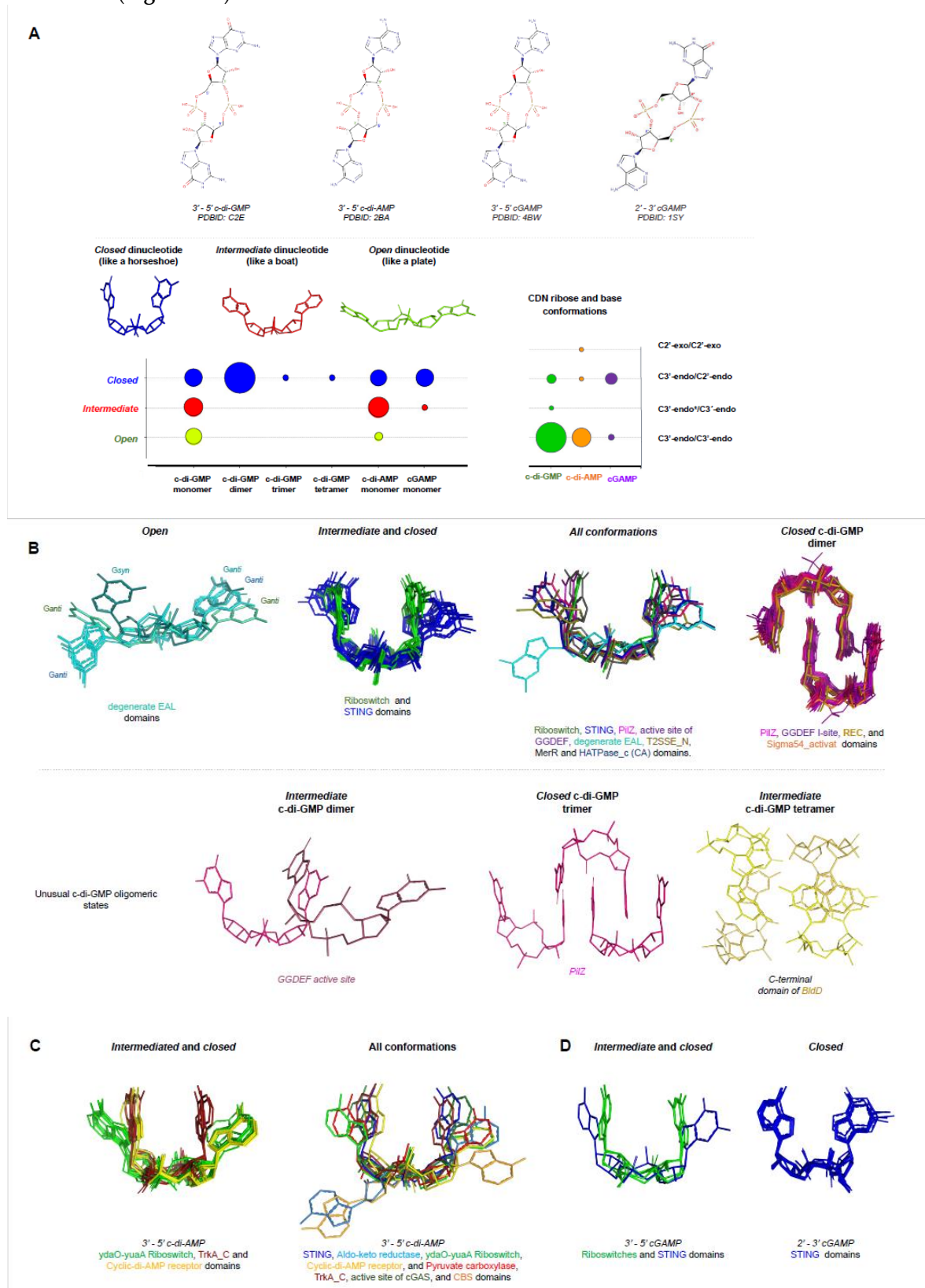


FIGURE 8: Diversity of cyclic di nucleotides produced by different kind of cells. All structures were observed in the three-dimensional protein structure solved and deposited in the protein data bank. A) Up panel: two-dimensional view (2D) representation of different cyclic di-nucleotides produced mainly by bacteria, with exception of 2'-3' cGAMP that is produced by eukaryotic cells by cGAS enzymes. The linkages between the pentoses and the phosphates are shown in green, blue or red and those carbons colored in grey are not involved in the phosphate linkage cyclisation. The structures of each CDNs were initially downloaded as SDF format at the CHEBI website¹³⁹ and edited using the MarvinSketch version number 19.18. The PDBID of each CDNs are also described for each ligand. Down panel: The three main conformations found of CDNs in the protein binding pocket in respect to the base proximity: 1- *closed* conformation (shaped as a horseshoe) is when the two base rings are face-to-face, colored as blue; 2- *intermediate* conformation (shaped as a boat) is when the bases are not in the *closed* conformation neither in the *open* conformation, colored by red and; 3- *open* conformation is when the two base rings are far from each other in an elongated conformation, colored by lemon. At the bottom of these structures are shown a Bubble Chart showing the frequency of each CDNs in the *Closed*, *Intermediate* and *Open* conformations. The c-di-GMP is the only one that have been found in the protein structure in different oligomeric states: as monomers, as dimers, as trimers, and as tetramers. The right Bubble chart shows the ribose conformation that can be in 3 different configurations, C3'-endo, C2'-endo or C2'-exo and the frequency of each of these conformation in the CDNs found in the protein binding pocket. * *Gsyn* conformation of the base ring in relation to the pentose, found only in one c-di-GMP structure (PDBID: 4FOK). Panels B to D show different CDNs superpositions showing the heterogeneity of conformations found in the protein and riboswitches binding pockets. B) Up panel: Different conformations for c-di-GMP found in different protein binding pocket and Riboswitches shown as a superposition between them. 3'-5' c-di-GMP found in degenerate EAL domains are colored in cyan (PDBID: 3HV8, 4F3H, 4F48, 3PJT, 3PJU), in Riboswitches are colored in green (PDBID: 3Q3Z, 3MXH, 3MUT, 3MUR, 3MUM, 3IRW, 4YB0, 3IWN), in STING protein are colored in blue (PDBID: 4EF4, 4EMT, 6RM0, 6S86, 4F9G, 4F5D, 4F5Y, 6A04, 5CFL, 5CFP), in PilZ domains are colored in pink (c-di-GMP as dimers: PDBID: 4ZMN, 5EUH, 3BRE, 3I5C, 1W25, 2WB4, 2V0N, 3TVK, 3I5A, 3IGM, 4URG, 4URS. C-di-GMP as trimer, PDBID: 4XRN), in the active site of GGDEF domain (monomer of c-di-GMP PDBID: 4RT1) and in the GGDEF I-site (c-di-GMP dimer PDBID: 5E1Y, 5EJ1, 5EJZ, 4P00, 4P02, 5KGO, 5EJL, 5VX6, 5Y4R, 5XLY, 2L74, 4RT0, 5Y6F, 5Y6G) are colored in purple, in the T2SSE_N domain is colored in brown (PDBID: 5HTL), in the HATPase_c (CA) domain is colored in blue (PDBID: 5IDM), in the REC domain is colored in yellow (PDBID: 3KLO), in the Sigma54_activat domain is colored in orange (PDBID: 5EXX). The unusual c-di-GMP oligomeric states found in one GGDEF active site is colored in pink and brown (PDBID: 3QYY), and in the C-terminal domain of BldD is colored in yellow (PDBID: 4OAZ). C) Different conformations for 3'-5' c-di-AMP found in different protein binding pocket and Riboswitches shown as a superposition between them. 3'-5' c-di-AMP found in Riboswitches are colored in green (PDBID: 4QK8, 4QK9, 4W92, 4W90, 4QLM, 4QLN, 4QKA), in TrkA_C domain are colored in brown (PDBID: 4YS2, 4YP1, 5F29), in Cyclic-di-AMP receptor domain are colored in yellow (PDBID: 4WK1, 4D3H, 4RWW, 4RLE), in STING proteins is colored in dark blue (PDBID: 6IYF), in Aldo-keto reductase domain is colored in light blue (PDBID: 5UXF), in Pyruvate carboxylase domain is colored in red (PDBID: 5VZ0), in the active site of cGAS is colored in dark green (PDBID: 3C1Y), and CBS domains is colored in orange (PDBID: 5KS7). D) 3'-5' cGAMP are found in *intermediate* and *closed* conformations in the ligand binding pocket of Riboswitches, colored in green (PDBID: 4YAZ and 4YB1) and in STING protein colored in blue (PDBID: 5CFM). 2'-3' cGAMP are found in *closed* conformations in the ligand binding pocket of STING proteins colored in blue (PDBID: 6NT7, 6NT8, 5CFQ, 4LOH, 4LOJ, 5GRM, 4KSY, and 6A06).

The dinucleotides can assume different conformations in the binding site of different receptors that can be described in relation to the base and the ribose conformations. The ribose ring can assume 3 different configurations, C3'-endo, C2'-endo or C2'-exo. When taking into account receptor structures in complex with cyclic dinucleotides (**Table 1**), more than 80% of the ligands have the two pentose rings in C3'-endo, almost 15% have one of the pentoses in C3'-endo and the other in C2'-endo and only one structure has the two pentoses in C2'-exo configuration (**Figure 8 A**). Furthermore,

the base can assume a *syn* or *anti* conformation in relation to the pentose by the N-glycosidic bond, and only one of the structures, the FimX EAL domain from *Xanthomonas citri* (PDBID: 4FOK)¹²⁹ has one of the pentoses at the *syn* conformation, which is the less stable state of the molecule. The conformation C3'-endo/C3'-endo is the more representative for the c-di-GMP and c-di-AMP molecules, while cGAMP is preferentially in C3'-endo/C2'-endo conformation (**Figure 8 A**).

The overall conformation of the ligand can be classified in three conformations in respect to the base proximity: 1- *closed* conformation (shaped as a horseshoe) is when the two base rings are face-to-face; 2- *open* conformation is when the two base rings are far from each other in an elongated conformation and; 3- *intermediate* conformation (shaped as a boat) is when the bases are not in the *closed* conformation neither in the *open* conformation (**Figure 8 A**).

The c-di-GMP molecule in solution is found in a fast equilibrium between a monomeric state and as a dimer with intercalated bases with a *K_d* of about 1mM under physiological salt conditions¹³⁰. Nevertheless, the intracellular concentration of c-di-GMP is about the μ M range, suggesting that free c-di-GMP molecules are monomeric inside of the cells. Looking at the conformation of c-di-GMP when it is bound to proteins, which includes its receptors and the active site of DGCs enzymes that contain GGDEF domains, most of them are found as monomers or dimers, though trimeric and tetrameric structures were also observed in PilZ¹³¹ and the C-terminal domain of BldD proteins¹³², respectively (**Figure 8 B**). Interestingly, PilZ is the only one that binds c-di-GMP in monomer, dimer and trimer forms, while EAL domain binds c-di-GMP monomers with the largest conformational divergences. Proteins containing STING domain and RNA riboswitches are bound to CDN monomers that share similar conformations (**Figure 8 B - D**). Looking at the conformation of c-di-AMP and cGAMP when bound to proteins or riboswitches, all of them are found as monomers (**Figure 8 C - D**).

DISTRIBUTION OF PROTEINS CONTAINING-GGDEF AND DAC DOMAINS IN BACTERIA

Initial reviews of the distribution of DisA homologs across bacterial clades suggested that c-di-AMP would play a more important role in Gram-positive bacteria than in Gram-negative and that, in general, bacteria would avoid allowing these two signaling networks to co-exist, so as to avoid unintended cross-talking and to easy regulation of the balance of these metabolites within the cell^{20,21}. Subsequent surveys on the distribution of DAC and GGDEF homologs don't support the idea that DAC homologs are rare among Gram-negative bacteria, as members of lineages such as Cyanobacteria, Spirochaetes and Deltaproteobacteria often carry both DAC and GGDEF genes, a profile compatible with the complex lifestyles and genomes of these lineages. Also, among Gram-positives, most members of Firmicutes and Actinobacteria, including model organisms such as *Bacillus*, *Clostridium*, *Streptomyces*, *Listeria*, and *Mycobacterium*, produce both signaling molecules and possess a wide array of GGDEF genes, following the general trend of having close to as many genomes with both DAC and GGDEF as possible (see **Supplementary Table 1** and **Figure 9**). The only lineages were several of the genomes sampled seem to have at least one DAC homolog but no or very few and rare recognizable GGDEF homologs are Bacteroidetes and the Archaea. In both lineages, the number of genomes with both DAC and GGDEF falls below 50% of the maximum allowed, *i.e.* the smallest between the number of genomes carrying DAC or GGDEF. Genomic data strongly suggests that there is a tendency for bacterial cells to use both c-di-AMP and c-di-GMP signaling networks simultaneously, which would imply that both the control of their synthesis and turnover and the specificity of their sensors is carefully tuned.

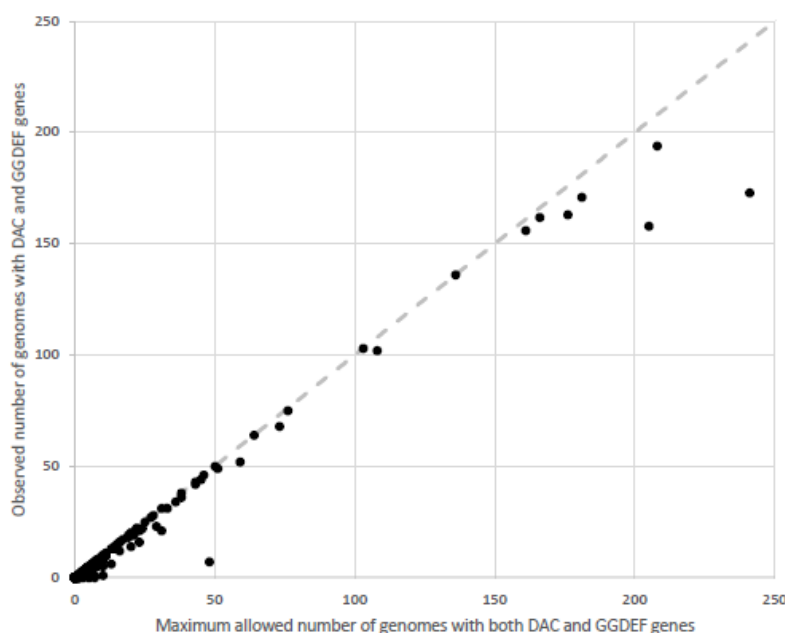


FIGURE 9: Lack of anti-correlation in the distribution of DAC and GGDEF genes per prokaryotic clades. Each dot represents a prokaryotic class, such as Gammaproteobacteria or Bacilli, as defined in the NCBI's Taxonomy Database. For each class, the number of genomes harboring at least one DAC and one GGDEF gene and the number of genomes harboring both was calculated. If, for a given class, we consider the number of genomes with DACs and the number of genomes with GGDEF, the smallest of these numbers is the maximum number of genomes that could, in principle, carry both genes. That number is seen on the horizontal axis while the actual number of genomes carrying both genes is on the y-axis. These numbers are very close to the diagonal line, indicating that, in most cases, if members of a given lineage are carrying both DAC and GGDEF, they tend to keep both genes, instead of having to choose between them.

3. Conclusions

Not long time ago, in the mid-2000s, emerged the idea that c-di-GMP molecules could be a second messenger ubiquitous in bacteria, in which proteins containing GGDEF and EAL or HD-GYP domains were at the center of this regulation, being involved in the synthesis and degradation of c-di-GMP, respectively. In the following years, different papers were published showing the central role of c-di-GMP orchestrating different signaling networks such as the regulation of the flagellar rotor, bacterial motility such as twitching, exopolysaccharide synthesis and regulation of bacterial biofilm formation. Nevertheless, c-di-GMP was first identified in 1987 as an allosteric activator of cellulose synthase in the fruit-degrading bacterium *Gluconacetobacter xylinus*². It was the first c-di-GMP receptor described, and nowadays a huge range of different receptors have been identified including RNA structures known as riboswitches. Therefore, a cyclic di nucleotide neglected in the microbiology area for 20 years emerged as regulator of the bacterial cell lifestyle, which can decide to live as a free single cell or within a community with many other cells.

Recently, this world has been under expansion, with the discoveries of other signaling CNDs in bacteria. In 2008, it was demonstrated that bacteria can also produce c-di-AMP by an enzyme known as DisA that possess a DAC domain⁴, this CDN can also be a second bacterial messenger that regulates a huge range of bacterial functions. Recently, a novel cyclic di nucleotide has been found to be a second bacterial messenger, cGAMP, synthesized by proteins containing SMODS domain such as the DncV protein.

In this review, additional interesting observations were made, including that different receptor types can bind CDNs in similar conformations and that, based on genomic data, different CDN second messenger systems may coexist in many organisms. The large amount of sequence and structural data available allows a broad view of the importance of CDNs in bacteria, but how cells

coordinate these molecules to ensure adaptation to changing environmental conditions is still open for much further exploration. These new discoveries open questions about how bacteria coordinate these three CDNs: Are they interconnected to regulate the same bacterial phenotype, or do they act independently? Do bacteria use the three CDNs as second messengers or is one chosen? Are the CDNs signaling pathways conserved in different bacteria? Others CDNs will be discovery to be also second messengers?

Interestingly, mammalian cells evolved to sense bacterial infection by detecting CDNs and stimulate the immune system to counterattack the infection. Thus, the use of CDNs as adjuvants in vaccines has been considered but can be also used as a stimulator of the innate immune system ^{71,133}.

Acknowledgments: This review was supported by grants of FAPESP (project numbers: 2019/00195-2; 2018/21076-9; 2018/08996-1, 2017/10611-8, 2016/08414-7), CNPQ (project number: 403965/2016-0), and CAPES (project number: 23038.015071/2018-12).

Conflicts of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions: Guzzo, C.R. and Purificação A.D. contributed to the analysis of the structures and ligand conformations. Azevedo, N. contributed to the CDN receptors analysis. Francisco, R.S. performed the bioinformatic analysis and Guarany, G.A. contributed to the ligand graph preparation and analysis. All authors contributed to the writing and to the revising of the text.

Supplementary table 1: Distribution of DAC, GGDEF and SMODS homolog across prokaryotic lineages. This data is the basis of **Figure 9**.

REFERENCES

1. Ryan, R.P., Fouhy, Y., Lucey, J.F. & Dow, J.M. Cyclic di-GMP signaling in bacteria: recent advances and new puzzles. *J Bacteriol* **188**, 8327-34 (2006).
2. Ross, P. et al. Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* **325**, 279-81 (1987).
3. Whiteley, A.T. et al. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* (2019).
4. Witte, G., Hartung, S., Buttner, K. & Hopfner, K.P. Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol Cell* **30**, 167-78 (2008).
5. Davies, B.W., Bogard, R.W., Young, T.S. & Mekalanos, J.J. Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for *V. cholerae* virulence. *Cell* **149**, 358-70 (2012).
6. Paul, K., Nieto, V., Carlquist, W.C., Blair, D.F. & Harshey, R.M. The c-di-GMP binding protein YcgR controls flagellar motor direction and speed to affect chemotaxis by a "backstop brake" mechanism. *Mol Cell* **38**, 128-39 (2010).
7. Guzzo, C.R., Salinas, R.K., Andrade, M.O. & Farah, C.S. PILZ protein structure and interactions with PILB and the FIMX EAL domain: implications for control of type IV pilus biogenesis. *J Mol Biol* **393**, 848-66 (2009).
8. Tschowri, N. et al. Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces* development. *Cell* **158**, 1136-1147 (2014).
9. Srivastava, D. & Waters, C.M. A tangled web: regulatory connections between quorum sensing and cyclic Di-GMP. *J Bacteriol* **194**, 4485-93 (2012).
10. Roelofs, K.G. et al. Systematic Identification of Cyclic-di-GMP Binding Proteins in *Vibrio cholerae* Reveals a Novel Class of Cyclic-di-GMP-Binding ATPases Associated with Type II Secretion Systems. *PLoS Pathog* **11**, e1005232 (2015).
11. Trampari, E. et al. Bacterial rotary export ATPases are allosterically regulated by the nucleotide second messenger cyclic-di-GMP. *J Biol Chem* **290**, 24470-83 (2015).
12. Moscoso, J.A., Mikkelsen, H., Heeb, S., Williams, P. & Filloux, A. The *Pseudomonas aeruginosa* sensor RetS switches type III and type VI secretion via c-di-GMP signalling. *Environ Microbiol* **13**, 3128-38 (2011).
13. D'Argenio, D.A. & Miller, S.I. Cyclic di-GMP as a bacterial second messenger. *Microbiology* **150**, 2497-502 (2004).
14. Dow, J.M., Fouhy, Y., Lucey, J.F. & Ryan, R.P. The HD-GYP domain, cyclic di-GMP signaling, and bacterial virulence to plants. *Mol Plant Microbe Interact* **19**, 1378-84 (2006).

15. Jenal, U. Cyclic di-guanosine-monophosphate comes of age: a novel secondary messenger involved in modulating cell surface structures in bacteria? *Curr Opin Microbiol* **7**, 185-91 (2004).
16. Jenal, U. & Malone, J. Mechanisms of cyclic-di-GMP signaling in bacteria. *Annu Rev Genet* **40**, 385-407 (2006).
17. Romling, U., Gomelsky, M. & Galperin, M.Y. C-di-GMP: the dawning of a novel bacterial signalling system. *Mol Microbiol* **57**, 629-39 (2005).
18. Morgan, D.O. et al. Control of eukaryotic cell cycle progression by phosphorylation of cyclin-dependent kinases. *Cancer J Sci Am* **4 Suppl 1**, S77-83 (1998).
19. Fahmi, T., Port, G.C. & Cho, K.H. c-di-AMP: An Essential Molecule in the Signaling Pathways that Regulate the Viability and Virulence of Gram-Positive Bacteria. *Genes (Basel)* **8**(2017).
20. Pesavento, C. & Hengge, R. Bacterial nucleotide-based second messengers. *Curr Opin Microbiol* **12**, 170-6 (2009).
21. Corrigan, R.M. & Grundling, A. Cyclic di-AMP: another second messenger enters the fray. *Nat Rev Microbiol* **11**, 513-24 (2013).
22. Schuster, C.F. et al. The second messenger c-di-AMP inhibits the osmolyte uptake system OpuC in *Staphylococcus aureus*. *Sci Signal* **9**, ra81 (2016).
23. Huynh, T.N. et al. Cyclic di-AMP targets the cystathionine beta-synthase domain of the osmolyte transporter OpuC. *Mol Microbiol* **102**, 233-243 (2016).
24. Corrigan, R.M., Abbott, J.C., Burhenne, H., Kaeffer, V. & Grundling, A. c-di-AMP is a new second messenger in *Staphylococcus aureus* with a role in controlling cell size and envelope stress. *PLoS Pathog* **7**, e1002217 (2011).
25. Bai, Y. et al. Cyclic di-AMP impairs potassium uptake mediated by a cyclic di-AMP binding protein in *Streptococcus pneumoniae*. *J Bacteriol* **196**, 614-23 (2014).
26. Zhang, L., Li, W. & He, Z.G. DarR, a TetR-like transcriptional factor, is a cyclic di-AMP-responsive repressor in *Mycobacterium smegmatis*. *J Biol Chem* **288**, 3085-96 (2013).
27. Nelson, J.W. et al. Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nat Chem Biol* **9**, 834-9 (2013).
28. Gundlach, J., Rath, H., Herzberg, C., Mader, U. & Stulke, J. Second Messenger Signaling in *Bacillus subtilis*: Accumulation of Cyclic di-AMP Inhibits Biofilm Formation. *Front Microbiol* **7**, 804 (2016).
29. Peng, X., Zhang, Y., Bai, G., Zhou, X. & Wu, H. Cyclic di-AMP mediates biofilm formation. *Mol Microbiol* **99**, 945-59 (2016).
30. Mehne, F.M. et al. Control of the diadenylate cyclase CdaS in *Bacillus subtilis*: an autoinhibitory domain limits cyclic di-AMP production. *J Biol Chem* **289**, 21098-107 (2014).
31. Dengler, V. et al. Mutation in the C-di-AMP cyclase *dacA* affects fitness and resistance of methicillin resistant *Staphylococcus aureus*. *PLoS One* **8**, e73512 (2013).
32. Sureka, K. et al. The cyclic dinucleotide c-di-AMP is an allosteric regulator of metabolic enzyme function. *Cell* **158**, 1389-1401 (2014).
33. Krasteva, P.V. & Sondermann, H. Versatile modes of cellular regulation via cyclic dinucleotides. *Nat Chem Biol* **13**, 350-359 (2017).
34. Severin, G.B. et al. Direct activation of a phospholipase by cyclic GMP-AMP in El Tor *Vibrio cholerae*. *Proc Natl Acad Sci U S A* **115**, E6048-E6055 (2018).
35. Paul, R. et al. Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel diguanylate cyclase output domain. *Genes Dev* **18**, 715-27 (2004).
36. Schmidt, A.J., Ryjenkov, D.A. & Gomelsky, M. The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. *J Bacteriol* **187**, 4774-81 (2005).
37. Christen, M., Christen, B., Folcher, M., Schuete, A. & Jenal, U. Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *J Biol Chem* **280**, 30829-37 (2005).
38. Orr, M.W. et al. Oligoribonuclease is the primary degradative enzyme for pGpG in *Pseudomonas aeruginosa* that is required for cyclic-di-GMP turnover. *Proc Natl Acad Sci U S A* **112**, E5048-57 (2015).
39. Cohen, D. et al. Oligoribonuclease is a central feature of cyclic diguanylate signaling in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* **112**, 11359-64 (2015).
40. Hallberg, Z.F. et al. Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3', 3'-cGAMP). *Proc Natl Acad Sci U S A* **113**, 1790-5 (2016).
41. Nelson, J.W. et al. Control of bacterial exoelectrogenesis by c-AMP-GMP. *Proc Natl Acad Sci U S A* **112**, 5389-94 (2015).

42. Burroughs, A.M., Zhang, D., Schaffer, D.E., Iyer, L.M. & Aravind, L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res* **43**, 10633-54 (2015).
43. Xiao, T.S. & Fitzgerald, K.A. The cGAS-STING pathway for DNA sensing. *Mol Cell* **51**, 135-9 (2013).
44. Barber, G.N. STING-dependent cytosolic DNA sensing pathways. *Trends Immunol* **35**, 88-93 (2014).
45. Cai, X., Chiu, Y.H. & Chen, Z.J. The cGAS-cGAMP-STING pathway of cytosolic DNA sensing and signaling. *Mol Cell* **54**, 289-96 (2014).
46. He, Q. et al. Structural and Biochemical Insight into the Mechanism of Rv2837c from *Mycobacterium tuberculosis* as a c-di-NMP Phosphodiesterase. *J Biol Chem* **291**, 3668-81 (2016).
47. Huynh, T.N. et al. An HD-domain phosphodiesterase mediates cooperative hydrolysis of c-di-AMP to affect bacterial growth and virulence. *Proc Natl Acad Sci U S A* **112**, E747-56 (2015).
48. Gundlach, J. et al. An Essential Poison: Synthesis and Degradation of Cyclic Di-AMP in *Bacillus subtilis*. *J Bacteriol* **197**, 3265-74 (2015).
49. Rosenberg, J. et al. Structural and biochemical analysis of the essential diadenylate cyclase CdaA from *Listeria monocytogenes*. *J Biol Chem* **290**, 6596-606 (2015).
50. Punta, M. et al. The Pfam protein families database. *Nucleic Acids Res* **40**, D290-301 (2012).
51. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z.J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786-91 (2013).
52. Kranzusch, P.J. et al. Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. *Cell* **158**, 1011-1021 (2014).
53. Jenal, U., Reinders, A. & Lori, C. Cyclic di-GMP: second messenger extraordinaire. *Nat Rev Microbiol* **15**, 271-284 (2017).
54. Zhang, G., Liu, Y., Ruoho, A.E. & Hurley, J.H. Structure of the adenylyl cyclase catalytic core. *Nature* **386**, 247-53 (1997).
55. Schirmer, T. C-di-GMP Synthesis: Structural Aspects of Evolution, Catalysis and Regulation. *J Mol Biol* (2016).
56. da Costa Vasconcelos, F.N. et al. Structural and Enzymatic Characterization of a cAMP-Dependent Diguanylate Cyclase from Pathogenic *Leptospira* Species. *J Mol Biol* **429**, 2337-2352 (2017).
57. Oliveira, M.C. et al. Cooperative substrate binding by a diguanylate cyclase. *J Mol Biol* **427**, 415-32 (2015).
58. Hengge, R. Cyclic-di-GMP reaches out into the bacterial RNA world. *Sci Signal* **3**, pe44 (2010).
59. Krasteva, P.V. et al. *Vibrio cholerae* VpsT regulates matrix production and motility by directly sensing cyclic di-GMP. *Science* **327**, 866-8 (2010).
60. Baraquet, C. & Harwood, C.S. Cyclic diguanosine monophosphate represses bacterial flagella synthesis by interacting with the Walker A motif of the enhancer-binding protein FleQ. *Proc Natl Acad Sci U S A* **110**, 18478-83 (2013).
61. Boehm, A. et al. Second messenger-mediated adjustment of bacterial swimming velocity. *Cell* **141**, 107-16 (2010).
62. Habazettl, J., Allan, M.G., Jenal, U. & Grzesiek, S. Solution structure of the PilZ domain protein PA4608 complex with cyclic di-GMP identifies charge clustering as molecular readout. *J Biol Chem* **286**, 14304-14 (2011).
63. Schumacher, M.A. & Zeng, W. Structures of the activator of *K. pneumoniae* biofilm formation, MrkH, indicates PilZ domains involved in c-di-GMP and DNA binding. *Proc Natl Acad Sci U S A* **113**, 10067-72 (2016).
64. Duerig, A. et al. Second messenger-mediated spatiotemporal control of protein degradation regulates bacterial cell cycle progression. *Genes Dev* **23**, 93-104 (2009).
65. Matsuyama, B.Y. et al. Mechanistic insights into c-di-GMP-dependent control of the biofilm regulator FleQ from *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* **113**, E209-18 (2016).
66. Russell, M.H. et al. Integration of the second messenger c-di-GMP into the chemotactic signaling pathway. *MBio* **4**, e00001-13 (2013).
67. Kim, H. et al. Structural Studies of Potassium Transport Protein KtrA Regulator of Conductance of K⁺ (RCK) C Domain in Complex with Cyclic Diadenosine Monophosphate (c-di-AMP). *J Biol Chem* **290**, 16393-402 (2015).

68. Moscoso, J.A. et al. Binding of Cyclic Di-AMP to the *Staphylococcus aureus* Sensor Kinase KdpD Occurs via the Universal Stress Protein Domain and Downregulates the Expression of the Kdp Potassium Transporter. *J Bacteriol* **198**, 98-110 (2016).
69. Choi, P.H., Sureka, K., Woodward, J.J. & Tong, L. Molecular basis for the recognition of cyclic-di-AMP by PstA, a PII-like signal transduction protein. *Microbiologyopen* **4**, 361-74 (2015).
70. McFarland, A.P. et al. Sensing of Bacterial Cyclic Dinucleotides by the Oxidoreductase RECON Promotes NF-kappaB Activation and Shapes a Proinflammatory Antibacterial State. *Immunity* **46**, 433-445 (2017).
71. Li, L. et al. Hydrolysis of 2'3'-cGAMP by ENPP1 and design of nonhydrolyzable analogs. *Nat Chem Biol* **10**, 1043-8 (2014).
72. Chan, C. et al. Structural basis of activity and allosteric control of diguanylate cyclase. *Proc Natl Acad Sci U S A* **101**, 17084-9 (2004).
73. Sillitoe, I. et al. CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res* **43**, D376-81 (2015).
74. Iyer, L.M., Abhiman, S. & Aravind, L. A new family of polymerases related to superfamily A DNA polymerases and T7-like DNA-dependent RNA polymerases. *Biol Direct* **3**, 39 (2008).
75. Garbers, D.L. The guanylyl cyclase receptor family. *New Biol* **2**, 499-504 (1990).
76. Steer, M.L. Adenyl cyclase. *Ann Surg* **182**, 603-9 (1975).
77. Morrison, S.D., Roberts, S.A., Zegeer, A.M., Montfort, W.R. & Bandarian, V. A new use for a familiar fold: the X-ray crystal structure of GTP-bound GTP cyclohydrolase III from *Methanocaldococcus jannaschii* reveals a two metal ion catalytic mechanism. *Biochemistry* **47**, 230-42 (2008).
78. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res* **47**, D427-D432 (2019).
79. Goodman, M.F. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* **71**, 17-50 (2002).
80. Aravind, L., Mazumder, R., Vasudevan, S. & Koonin, E.V. Trends in protein evolution inferred from sequence and structure analysis. *Curr Opin Struct Biol* **12**, 392-9 (2002).
81. Wassmann, P. et al. Structure of Bef3- modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. *Structure* **15**, 915-27 (2007).
82. Paul, R. et al. Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization. *J Biol Chem* **282**, 29170-7 (2007).
83. Tischler, A.D. & Camilli, A. Cyclic diguanylate (c-di-GMP) regulates *Vibrio cholerae* biofilm formation. *Mol Microbiol* **53**, 857-69 (2004).
84. Schirmer, T. C-di-GMP Synthesis: Structural Aspects of Evolution, Catalysis and Regulation. *J Mol Biol* **428**, 3683-701 (2016).
85. Anantharaman, V., Balaji, S. & Aravind, L. The signaling helix: a common functional theme in diverse signaling proteins. *Biol Direct* **1**, 25 (2006).
86. Christen, B. et al. Allosteric control of cyclic di-GMP signaling. *J Biol Chem* **281**, 32015-24 (2006).
87. Rao, F. et al. YybT is a signaling protein that contains a cyclic dinucleotide phosphodiesterase domain and a GGDEF domain with ATPase activity. *J Biol Chem* **285**, 473-82 (2010).
88. Seshasayee, A.S., Fraser, G.M. & Luscombe, N.M. Comparative genomics of cyclic-di-GMP signalling in bacteria: post-translational regulation and catalytic activity. *Nucleic Acids Res* **38**, 5970-81 (2010).
89. Zhu, D. et al. Structural biochemistry of a *Vibrio cholerae* dinucleotide cyclase reveals cyclase activity regulation by folates. *Mol Cell* **55**, 931-937 (2014).
90. Aravind, L. & Koonin, E.V. DNA polymerase beta-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. *Nucleic Acids Res* **27**, 1609-18 (1999).
91. Holm, L. Benchmarking Fold Detection by DaliLite v.5. *Bioinformatics* (2019).
92. Martin, G., Keller, W. & Doublié, S. Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. *EMBO J* **19**, 4193-203 (2000).
93. Lin, J., Lu, J., Feng, Y., Sun, M. & Ye, K. An RNA-binding complex involved in ribosome biogenesis contains a protein with homology to tRNA CCA-adding enzyme. *PLoS Biol* **11**, e1001669 (2013).
94. Utama, B., Kennedy, D., Ru, K. & Mattick, J.S. Isolation and characterization of a new nucleolar protein, Nrap, that is conserved from yeast to humans. *Genes Cells* **7**, 115-32 (2002).
95. Donovan, J., Whitney, G., Rath, S. & Korennykh, A. Structural mechanism of sensing long dsRNA via a noncatalytic domain in human oligoadenylate synthetase 3. *Proc Natl Acad Sci U S A* **112**, 3949-54 (2015).

96. Motamedi, M.R. et al. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789-802 (2004).
97. Hamill, S., Wolin, S.L. & Reinisch, K.M. Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. *Proc Natl Acad Sci U S A* **107**, 15045-50 (2010).
98. Kuchta, K., Knizewski, L., Wyrwicz, L.S., Rychlewski, L. & Ginalski, K. Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic Acids Res* **37**, 7701-14 (2009).
99. Sawaya, M.R., Pelletier, H., Kumar, A., Wilson, S.H. & Kraut, J. Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism. *Science* **264**, 1930-5 (1994).
100. Moon, A.F. et al. Structural insight into the substrate specificity of DNA Polymerase mu. *Nat Struct Mol Biol* **14**, 45-53 (2007).
101. Yamtich, J. & Sweasy, J.B. DNA polymerase family X: function, structure, and cellular roles. *Biochim Biophys Acta* **1804**, 1136-50 (2010).
102. Wolkowicz, U.M. & Cook, A.G. NF45 dimerizes with NF90, Zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold. *Nucleic Acids Res* **40**, 9356-68 (2012).
103. Shatkin, A.J. & Manley, J.L. The ends of the affair: capping and polyadenylation. *Nat Struct Biol* **7**, 838-42 (2000).
104. Chen, Q., Sun, L. & Chen, Z.J. Regulation and function of the cGAS-STING pathway of cytosolic DNA sensing. *Nat Immunol* **17**, 1142-9 (2016).
105. Sadler, A.J. & Williams, B.R. Interferon-inducible antiviral effectors. *Nat Rev Immunol* **8**, 559-68 (2008).
106. Han, Y. et al. Structure of human RNase L reveals the basis for regulated RNA decay in the IFN response. *Science* **343**, 1244-8 (2014).
107. Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763-76 (2009).
108. Hubscher, U. & Maga, G. DNA replication and repair bypass machines. *Curr Opin Chem Biol* **15**, 627-35 (2011).
109. Gallego, C., Estevez, A.M., Farez, E., Ruiz-Perez, L.M. & Gonzalez-Pacanowska, D. Overexpression of AP endonuclease protects Leishmania major cells against methotrexate induced DNA fragmentation and hydrogen peroxide. *Mol Biochem Parasitol* **141**, 191-7 (2005).
110. Danilchanka, O. & Mekalanos, J.J. Cyclic dinucleotides and the innate immune response. *Cell* **154**, 962-970 (2013).
111. Holm, L. & Sander, C. DNA polymerase beta belongs to an ancient nucleotidyltransferase superfamily. *Trends Biochem Sci* **20**, 345-7 (1995).
112. Bai, Y. et al. Mycobacterium tuberculosis Rv3586 (DacA) is a diadenylate cyclase that converts ATP or ADP into c-di-AMP. *PLoS One* **7**, e35206 (2012).
113. Witte, C.E. et al. Cyclic di-AMP is critical for Listeria monocytogenes growth, cell wall homeostasis, and establishment of infection. *MBio* **4**, e00282-13 (2013).
114. Luo, Y. & Helmann, J.D. Analysis of the role of Bacillus subtilis sigma(M) in beta-lactam resistance reveals an essential role for c-di-AMP in peptidoglycan homeostasis. *Mol Microbiol* **83**, 623-39 (2012).
115. Mehne, F.M. et al. Cyclic di-AMP homeostasis in bacillus subtilis: both lack and high level accumulation of the nucleotide are detrimental for cell growth. *J Biol Chem* **288**, 2004-17 (2013).
116. Zhang, R., Zhou, M., Ginell, S., Abdullah, J., Collart, F., Joachimiak, A. The crystal structure of the hypothetical membrane spanning protein from Bacillus cereus.
117. Heidemann, J.L., Neumann, P., Dickmanns, A. & Ficner, R. Crystal structures of the c-di-AMP-synthesizing enzyme CdaA. *J Biol Chem* **294**, 10463-10470 (2019).
118. Amikam, D. & Galperin, M.Y. PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics* **22**, 3-6 (2006).
119. Benach, J. et al. The structural basis of cyclic diguanylate signal transduction by PilZ domains. *EMBO J* **26**, 5153-66 (2007).
120. Finn, R.D. et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279-85 (2016).
121. Navarro, M.V. et al. Structural basis for c-di-GMP-mediated inside-out signaling controlling periplasmic proteolysis. *PLoS Biol* **9**, e1000588 (2011).
122. Barrick, J.E. et al. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci U S A* **101**, 6421-6 (2004).

123. Block, K.F., Hammond, M.C. & Breaker, R.R. Evidence for widespread gene control function by the ydaO riboswitch candidate. *J Bacteriol* **192**, 3983-9 (2010).
124. Trchounian, A. & Kobayashi, H. K⁺ uptake by fermenting Escherichia coli cells: pH dependent mode of the TrkA system operating. *Biosci Rep* **20**, 277-88 (2000).
125. Ballal, A., Basu, B. & Apte, S.K. The Kdp-ATPase system and its regulation. *J Biosci* **32**, 559-68 (2007).
126. Trchounian, A. & Kobayashi, H. Kup is the major K⁺ uptake system in Escherichia coli upon hyper-osmotic stress at a low pH. *FEBS Lett* **447**, 144-8 (1999).
127. Corrigan, R.M. et al. Systematic identification of conserved bacterial c-di-AMP receptor proteins. *Proc Natl Acad Sci U S A* **110**, 9084-9 (2013).
128. Choi, P.H. et al. Structural and functional studies of pyruvate carboxylase regulation by cyclic di-AMP in lactic acid bacteria. *Proc Natl Acad Sci U S A* **114**, E7226-E7235 (2017).
129. Guzzo, C.R., Dunger, G., Salinas, R.K. & Farah, C.S. Structure of the PilZ-FimXEAL-c-di-GMP Complex Responsible for the Regulation of Bacterial Type IV Pilus Biogenesis. *J Mol Biol* **425**, 2174-97 (2013).
130. Rao, F. et al. Enzymatic synthesis of c-di-GMP using a thermophilic diguanylate cyclase. *Anal Biochem* **389**, 138-42 (2009).
131. Chi, K.K., Yuan, Z.L. Pilz domain with c-di-gmp of a protein from Pseudomonas aeruginosa.
132. Schumacher, M.A. et al. The Streptomyces master regulator BldD binds c-di-GMP sequentially to create a functional BldD2-(c-di-GMP)₄ complex. *Nucleic Acids Res* **45**, 6923-6933 (2017).
133. Ogunniyi, A.D. et al. c-di-GMP is an effective immunomodulator and vaccine adjuvant against pneumococcal infection. *Vaccine* **26**, 4676-85 (2008).
134. Pruitt, K.D., Tatusova, T. & Maglott, D.R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**, D61-5 (2007).
135. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-90 (2004).
136. Dawson, N.L. et al. CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res* **45**, D289-D295 (2017).
137. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80 (1994).
138. Rao, F., Yang, Y., Qi, Y. & Liang, Z.X. Catalytic mechanism of cyclic di-GMP-specific phosphodiesterase: a study of the EAL domain-containing RocR from Pseudomonas aeruginosa. *J Bacteriol* **190**, 3622-31 (2008).
139. Hastings, J. et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res* **44**, D1214-9 (2016).
140. Zhang, X. et al. Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol Cell* **51**, 226-35 (2013).
141. Ouyang, S. et al. Structural analysis of the STING adaptor protein reveals a hydrophobic dimer interface and mode of cyclic di-GMP binding. *Immunity* **36**, 1073-86 (2012).
142. Shu, C., Yi, G., Watts, T., Kao, C.C. & Li, P. Structure of STING bound to cyclic di-GMP reveals the mechanism of cyclic dinucleotide recognition by the immune system. *Nat Struct Mol Biol* **19**, 722-4 (2012).
143. Smola, M., Birkus, G. & Boura, E. No magnesium is needed for binding of the stimulator of interferon genes to cyclic dinucleotides. *Acta Crystallogr F Struct Biol Commun* **75**, 593-598 (2019).
144. Yin, Q. et al. Cyclic di-GMP sensing via the innate immune signaling protein STING. *Mol Cell* **46**, 735-45 (2012).
145. Huang, Y.H., Liu, X.Y., Du, X.X., Jiang, Z.F. & Su, X.D. The structural basis for the sensing and binding of cyclic di-GMP by STING. *Nat Struct Mol Biol* **19**, 728-30 (2012).
146. Shang, G. et al. Crystal structures of STING protein reveal basis for recognition of cyclic di-GMP. *Nat Struct Mol Biol* **19**, 725-7 (2012).
147. Cong, X. et al. Crystal structures of porcine STING(CBD)-CDN complexes reveal the mechanism of ligand recognition and discrimination of STING proteins. *J Biol Chem* **294**, 11420-11432 (2019).
148. Kranzusch, P.J. et al. Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2',3' cGAMP Signaling. *Mol Cell* **59**, 891-903 (2015).
149. Smith, K.D., Lipchock, S.V., Livingston, A.L., Shanahan, C.A. & Strobel, S.A. Structural and biochemical determinants of ligand binding by the c-di-GMP riboswitch. *Biochemistry* **49**, 7351-9 (2010).

150. Smith, K.D., Shanahan, C.A., Moore, E.L., Simon, A.C. & Strobel, S.A. Structural basis of differential ligand recognition by two classes of bis-(3'-5')-cyclic dimeric guanosine monophosphate-binding riboswitches. *Proc Natl Acad Sci U S A* **108**, 7757-62 (2011).
151. Ren, A. et al. Structural basis for molecular discrimination by a 3',3'-cGAMP sensing riboswitch. *Cell Rep* **11**, 1-12 (2015).
152. Kulshina, N., Baird, N.J. & Ferre-D'Amare, A.R. Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch. *Nat Struct Mol Biol* **16**, 1212-7 (2009).
153. Morgan, J.L. et al. Observing cellulose biosynthesis and membrane translocation in crystallo. *Nature* **531**, 329-34 (2016).
154. Morgan, J.L., McNamara, J.T. & Zimmer, J. Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nat Struct Mol Biol* **21**, 489-96 (2014).
155. Hou, Y.J., Yang, W.S., Hong, Y., Zhang, Y., Wang, D.C., Li, D.F. C-di-GMP plays a dual role in the YcgR-mediated motility regulation.
156. Subramanian, S., Gao, X., Dann, C.E., 3rd & Kearns, D.B. MotI (DgrA) acts as a molecular clutch on the flagellar stator protein MotA in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **114**, 13537-13542 (2017).
157. Wang, F. et al. The PilZ domain of MrkH represents a novel DNA binding motif. *Protein Cell* **7**, 766-772 (2016).
158. Schumacher, M., Zeng, W. To be published: Structures of *K. pneumoniae* MrkH: dual utilization of the PilZ fold for c-di-GMP and DNA binding by a novel activator of biofilm genes.
159. Yan, X.F. et al. Structural analyses unravel the molecular mechanism of cyclic di-GMP regulation of bacterial chemotaxis via a PilZ adaptor protein. *J Biol Chem* **293**, 100-111 (2018).
160. Zhu, Y., Yuan, Z. & Gu, L. Structural basis for the regulation of chemotaxis by MapZ in the presence of c-di-GMP. *Acta Crystallogr D Struct Biol* **73**, 683-691 (2017).
161. Whitney, J.C. et al. Dimeric c-di-GMP is required for post-translational regulation of alginate production in *Pseudomonas aeruginosa*. *J Biol Chem* **290**, 12451-62 (2015).
162. Dahlstrom, K.M., Giglio, K.M., Collins, A.J., Sondermann, H. & O'Toole, G.A. Contribution of Physical Interactions to Signaling Specificity between a Diguanylate Cyclase and Its Effector. *MBio* **6**, e01978-15 (2015).
163. De, N., Navarro, M.V., Raghavan, R.V. & Sondermann, H. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. *J Mol Biol* **393**, 619-33 (2009).
164. De, N. et al. Phosphorylation-independent regulation of the diguanylate cyclase WspR. *PLoS Biol* **6**, e67 (2008).
165. Li, Z., Chen, J.H., Hao, Y. & Nair, S.K. Structures of the PelD cyclic diguanylate effector involved in pellicle formation in *Pseudomonas aeruginosa* PAO1. *J Biol Chem* **287**, 30191-204 (2012).
166. Vorobiev, S.M. et al. Crystal structure of a catalytically active GG(D/E)EF diguanylate cyclase domain from *Marinobacter aquaeolei* with bound c-di-GMP product. *J Struct Funct Genomics* **13**, 177-83 (2012).
167. Deepthi, A., Liew, C.W., Liang, Z.X., Swaminathan, K. & Lescar, J. Structure of a diguanylate cyclase from *Thermotoga maritima*: insights into activation, feedback inhibition and thermostability. *PLoS One* **9**, e110912 (2014).
168. Yeo, H.J., Yuan, Q., Beck, M.R., Baron, C. & Waksman, G. Structural and functional characterization of the VirB5 protein from the type IV secretion system encoded by the conjugative plasmid pKM101. *Proc Natl Acad Sci U S A* **100**, 15947-52 (2003).
169. Zahringer, F., Lacanna, E., Jenal, U., Schirmer, T. & Boehm, A. Structure and signaling mechanism of a zinc-sensory diguanylate cyclase. *Structure* **21**, 1149-57 (2013).
170. Whitney, J.C. et al. Structure of the cytoplasmic region of PelD, a degenerate diguanylate cyclase receptor that regulates exopolysaccharide production in *Pseudomonas aeruginosa*. *J Biol Chem* **287**, 23582-93 (2012).
171. Navarro, M.V., De, N., Bae, N., Wang, Q. & Sondermann, H. Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimX. *Structure* **17**, 1104-16 (2009).
172. Chin, K.H. et al. Structural polymorphism of c-di-GMP bound to an EAL domain and in complex with a type II PilZ-domain protein. *Acta Crystallogr D Biol Crystallogr* **68**, 1380-92 (2012).
173. Wang, Y.C. et al. Nucleotide binding by the widespread high-affinity cyclic di-GMP receptor MshEN domain. *Nat Commun* **7**, 12481 (2016).
174. Raju, H. & Sharma, R. Crystal structure of BrIR with c-di-GMP. *Biochem Biophys Res Commun* **490**, 260-264 (2017).

175. Dubey, B.N. et al. Cyclic di-GMP mediates a histidine kinase/phosphatase switch by noncovalent domain cross-linking. *Sci Adv* **2**, e1600823 (2016).
176. Shang, G., Zhang, C., Chen, Z.J., Bai, X.C. & Zhang, X. Cryo-EM structures of STING reveal its mechanism of activation by cyclic GMP-AMP. *Nature* **567**, 389-393 (2019).
177. Zhang, H., Han, M.J., Tao, J.L., Ye, Z.Y., Du, X.X., Deng, M.J., Zhang, X.Y., Li, L.F., Jiang, Z.F., Su, X.D. Crystal structure of rat STING in complex with cyclic GMP-AMP with 2'5'and 3'5'phosphodiester linkage(2'3'-cGAMP).
178. Gao, P. et al. Structure-function analysis of STING activation by c[G(2',5')pA(3',5')p] and targeting by antiviral DMXAA. *Cell* **154**, 748-62 (2013).
179. Ergun, S.L., Fernandez, D., Weiss, T.M. & Li, L. STING Polymer Structure Reveals Mechanisms for Activation, Hyperactivation, and Inhibition. *Cell* **178**, 290-301 e10 (2019).
180. Chin, K.H. et al. Structural Insights into the Distinct Binding Mode of Cyclic Di-AMP with SaCpaA_RCK. *Biochemistry* **54**, 4936-51 (2015).
181. Gao, A. & Serganov, A. Structural insights into recognition of c-di-AMP by the ydaO riboswitch. *Nat Chem Biol* **10**, 787-92 (2014).
182. Jones, C.P. & Ferre-D'Amare, A.R. Crystal structure of a c-di-AMP riboswitch reveals an internally pseudo-dimeric RNA. *EMBO J* **33**, 2692-703 (2014).
183. Ren, A. & Patel, D.J. c-di-AMP binds the ydaO riboswitch in two pseudo-symmetry-related pockets. *Nat Chem Biol* **10**, 780-6 (2014).
184. Campeotto, I., Zhang, Y., Mladenov, M.G., Freemont, P.S. & Grundling, A. Complex structure and biochemical characterization of the Staphylococcus aureus cyclic diadenylate monophosphate (c-di-AMP)-binding protein PstA, the founding member of a new signal transduction protein family. *J Biol Chem* **290**, 2888-901 (2015).
185. Muller, M., Hopfner, K.P. & Witte, G. c-di-AMP recognition by Staphylococcus aureus PstA. *FEBS Lett* **589**, 45-51 (2015).
186. Gundlach, J. et al. Identification, characterization, and structure analysis of the cyclic di-AMP-binding PII-like signal transduction protein DarA. *J Biol Chem* **290**, 3069-80 (2015).
187. Chin, K.H., Chou, S.H. RCK domain with CDA.