Critical Review

Cyclic Di-nucleotide Signaling Enters the Eukaryote Domain

Pauline Schaap*

College of Life Sciences, University of Dundee, Dundee, UK

Abstract

Cyclic (c-di-GMP) is the prevalent intracellular signaling intermediate in bacteria. It triggers a spectrum of responses that cause bacteria to shift from a swarming motile phase to sessile biofilm formation. However, additional functions for c-di-GMP and roles for related molecules, such as c-di-AMP and c-AMP-GMP continue to be uncovered. The first usage of cyclic-di-nucleotide (c-di-NMP) signaling in the eukaryote domain emerged only recently. In dictyostelid social amoebas, c-di-GMP is a secreted signal that induces motile amoebas to differentiate into sessile stalk cells. In humans, c-di-NMPs, which are either produced endogenously in response to foreign DNA or by invading bacterial pathogens, trigger the innate immune system by activating the expression of interferon genes. STING, the human c-di-NMP receptor, is conserved throughout metazoas and their closest unicellular relatives, suggesting protist origins for human c-di-NMP signaling. Compared to the limited number of conserved protein domains that detect the second messengers cAMP and cGMP, the domains that detect the c-di-NMPs are surprisingly varied.

© 2013 The Authors. IUBMB Life published by Wiley Periodicals, Inc. on behalf of International Union of Biochemistry and Molecular Biology. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited. Additional Supporting Information may be found in the online version of this article.

Keywords: cyclic di-GMP; diguanylate cyclase; biofilm; cyclic di-AMP; diadenylate cyclase; stalk cell differentiation; Dictyostelium; STING; cGAS; 2′,3′-cGAMP

Introduction

Nucleotide polymers are probably the oldest molecules of life, as proposed in the RNA-world hypothesis (1). The RNA building block ATP is the equally ancient carrier of cellular energy, while its cyclized form, 3′,5′-cyclic adenosine monophosphate (cAMP) is a carrier of sensory information in all domains of life. While the second messenger roles of cAMP and its sister molecule 3′,5′-cyclic guanosine monophosphate (cGMP) have been studied for over 50 years, the cyclic di-nucleotides 3′,5′-cyclic diguanylic acid (c-di-GMP), 3′,5′-cyclic diadenylic acid (c-di-AMP) and 3′,5′-cyclic adenylic-guanylic acid (c-AMP-GMP) were more recently identified as information carriers in prokaryotes (2–4). As outlined in an excellent recent review (5), c-di-GMP has now far surpassed the prevalence and importance of cAMP as a second messenger in prokaryotes (2–4). As outlined in an excellent recent review (5), c-di-GMP has now far surpassed the prevalence and importance of cAMP as a second messenger in prokaryotes, while novel roles still emerge at a rapid rate. In this work, I will compare the recently discovered roles of c-di-nucleotides in eukaryotes with their established roles in prokaryotes and provide an overview of proteins that mediate synthesis, detection, and degradation of cyclic di-nucleotides in the two domains.

Cyclic-Di-Nucleotide Signalling in Prokaryotes

C-di-GMP

C-di-GMP was identified in the late 80s as a cofactor that was required for activating the cellulose synthase of Gluconoacetobacter xylinus (2). However, it took an almost 15 year effort to determine how c-di-GMP was synthesized, detected, and degraded. Cellulose synthase detects c-di-GMP with its associated PilZ domain, which is also present in other proteins that are directly regulated by c-di-GMP (6,7). The synthetic enzyme, diguanylate cyclase, is related to the type III nucleotidyl...
transfersases, but contains a unique signature GGDEF sequence (8–11). The GGDEF domain proteins often harbor an EAL domain, which proved to be a 3’5’-phosphodiesterase (PDE) that hydrolyzes one of the phosphodiester bonds of c-di-GMP to form pGpG (8,12). In addition, c-di-GMP can also be converted to GMP by a phosphohydrolase of the HD family that also counts most of the cAMP and cGMP PDEs among its members. The c-di-GMP PDE is characterized by a signature HD-GYP sequence motif that differs from the HD motif of the cAMP and cGMP PDEs (13).

While mechanisms and genes responsible for c-di-GMP metabolism were being identified, the abundance of GGDEF and EAL encoding sequences in emerging bacterial genome sequences suggested that c-di-GMP mediated signaling must be very widely used. While this was experimentally confirmed for c-di-GMP activation of cellulose synthase, cellulose production appeared to be only one aspect of a suite of c-di-GMP mediated responses that cause bacteria to shift from a swarming planktonic state to a sessile biofilm-associated lifestyle (11,14,15). The c-di-GMP induced synthesis of cellulose and other exopolysaccharides contributes to the formation of the biofilm matrix, while c-di-GMP induced loss of flagella or other motility apparatus, formation of adhesive curli fimbriae or a stalk-like holdfast mark the developmental transition to a sessile life-style (16–18). For pathogenic bacteria, the planktonic state represents the virulent acute form of the infection, while the biofilm state represents the less virulent chronic stage, which is however more resistant to immune clearance and antibiotics.

Other developmental roles for c-di-GMP continue to emerge. In the predatory bacterium Bdellovibrio bacteriovorus, c-di-GMP induces the transition from the axenic mode of feeding to invasion of prey bacteria (19). In the cyanobacterium Anabaena sp. c-di-GMP triggers differentiation of nitrogen-fixing heterocysts (20), while in Streptomyces coelicolor c-di-GMP regulates the formation of aerial hyphae (21).

The targets for c-di-GMP appeared to be numerous and are listed in Table 1, which also summarizes the entire repertoire of cyclic-di-nucleotide (c-di-NMP) synthetic and hydrolyzing enzymes and targets that are discussed below. As mentioned earlier, c-di-GMP binds to the PilZ domain of cellulose synthase, but the PilZ domain can also be part of other proteins, such as the Klebsiella transcription factor MrkH, which activates the expression of fimbriae or the Pseudomonas aeruginosa protein Alg44, which is involved in the biosynthesis of the extracellular polysaccharide alginate (22). Additionally, the Pilz domain can be expressed on its own, such as PilZa from the Lyme disease spirochaete Borrelia burgdorferi, which mediates c-di-GMP induced motility and infectivity of this pathogen (23). In addition to the PilZ domain, c-di-GMP has a steadily growing number of other targets, such as the transcriptional regulators FleQ (24), Vpst (25), CLP (26), Bcam1349 (27), which each interact with c-di-GMP in a distinctive, but unresolved manner. For CLP and Bcam1349, this may involve the intrinsic cNMP domain, the deeply conserved cAMP binding domain of the bacterial catabolite repressor, CRP, and eukaryote PKA (28). Another c-di-GMP target is the I-site of some GGDEF domains, which allosterically inhibits c-di-GMP synthesis by active DGCs. However, both the GGDEF and EAL domains are often degenerate, lacking DGC and PDE activity respectively. In three proteins, PelD, PopA, and CdgG, the I-site of a defunct GGDEF domain acts as a c-di-GMP receptor, regulating protein function (29–31). In others, such as FinX and LapD, the substrate binding site of a defunct EAL domain functions as a c-di-GMP receptor (32,33).

This enumeration does not exhaust the variety of c-di-GMP targets (Table 1). The biofilm dispersal protein BecA acts as a c-di-GMP sink by binding c-di-GMP (41). c-di-GMP also mediates signal dependent RNA processing by binding directly to RNA polynucleotide phosphorylase (PNPase) (35) and c-di-GMP regulates gene expression by binding with high affinity to riboswitches. Riboswitches are sequences in the 5’ untranslated regions of bacterial mRNAs that fold into structures capable of binding small molecules, which then act to regulate gene transcription or translation (36,42).

This vast, and probably only partially explored, versatility in c-di-GMP receptors forms a stark contrast with only two intracellular receptors for the more ubiquitous signals, cAMP and cGMP, that is the cNMP domain (28) and the GAF domain (43).

Cyclic-di-AMP and Cyclic AMP-GMP

Bacteria use two other c-di-NMPs as second messengers, c-di-AMP and c-AMP-GMP, with again unique sets of metabolic enzymes and targets. The diadenylate cyclase, DAC or DisA_N, is part of the DNA scanning protein DisA that monitors Bacillus subtilis DNA for double strand breaks. The presence of Holliday junctions, the hallmark of partially repaired breaks, inhibits c-di-AMP synthesis by DAC. Further hydrolysis of c-di-AMP to pApA by YybT/GdpP prevents B. subtilis differentiation into spores (3,37). DAC is a novel catalyst with little structural similarity to other mono- or di-nucleotidyl cyclases, and YybT/GdpP is a member the DHH domain phosphatases. The DAC domain was found to be present in most bacterial phyla and even in Archaea (3). Involvement of c-di-AMP in responses ranging from cell size regulation (44), bacterial cell growth (45), peptidoglycan cell wall homeostasis (46) to pathogenicity (47) in a range of bacterial species has since been uncovered. The first known target for c-di-AMP is the transcriptional repressor DarR (38), again binding c-di-AMP with sequences unrelated to previously known cNMP or c-di-NMP binding domains.

The latest addition to the prokaryote c-di-NMPs is c-AMP-GMP. This hybrid molecule is synthesized from ATP and GTP by DncV, a cyclase of Vibrio cholerae, that is unrelated to any other nucleotidyl cyclases (4). In V. cholerae, c-AMP-GMP promotes intestinal colonization by down-regulating chemotaxis, but the DncV gene is also present in other proteobacteria and, like c-di-AMP and c-di-GMP, may regulate a range of cellular functions.

Cyclic-Di-Nucleotide Signaling in Eukaryotes

Dictyostelia

Until 2012, c-di-NMP signaling was considered to be confined to the prokaryote domain. DNA sequences with GGDEF
### TABLE 1

*Functional domains involved in the synthesis, detection and degradation of c-di-NMPs*

| Organism/c-di-NMP/process | Protein name | Functional domain | Interpro/cd identifier | Ref. |
|---------------------------|--------------|-------------------|------------------------|------|
| **Bacteria**               |              |                   |                        |      |
|                           | c-di-GMP     |                   |                        |      |
| Synthesis                 | diguanylate cyclase (DGC) | GGDEF/Class III nucleotidyl cyclase | IPR000160|IPR001054 | (8) |
| Degradation               | EAL          | Diguanylate PDE, EAL domain | IPR001633 | | (8) |
|                           | HD-GYP       | HD PDEase         | IPR003607 | | (13) |
| Detection                 | PilZ         | PilZ              | IPR009875 | | (7) |
|                           | FleQ         | unknown           | | | (24) |
|                           | Clp          | unknown           | | | (26) |
|                           | Bcam1349     | unknown           | | | (27) |
|                           | FimX         | EAL substrate binding site | IPR001633 | | (34) |
|                           | PelD         | RxxD motif in GGDEF domain | IPR000160 | | (29) |
|                           | PNPase       | unknown           | | | (35) |
|                           | Vpst         | unknown           | | | (25) |
|                           | Vc2 Class I Riboswitch | GEMM | | | (36) |
|                           |              |                   |                        |      |
|                           | c-di-AMP     |                   |                        |      |
| Synthesis                 | diadenylate cyclase | DisA_N | IPR003390 | | (3) |
| Degradation               | YybT/GdpP    | RecJ-like phosphoesterase | IPR001667 | | (37) |
| Detection                 | DarR         | GATase1_AraC_1     | cd03137 | | (38) |
|                           |              |                   |                        |      |
|                           | c-AMP-GMP    |                   |                        |      |
| Synthesis                 | DncV (VC0179) | no previously identified domain | | | (4) |
| Degradation               | unknown      |                   |                        |      |
| Detection                 | unknown      |                   |                        |      |
|                           |              |                   |                        |      |
|                           | Dictyostelia |                   |                        |      |
|                           | c-di-GMP     |                   |                        |      |
| Synthesis                 | diguanylate cyclase (DgcA) | GGDEF/Class III nucleotidyl cyclase | IPR000160|IPR001054 | (39) |
| Degradation               | unknown      |                   |                        |      |
| Detection                 | unknown      |                   |                        |      |
|                           |              |                   |                        |      |
|                           | Metazoa      |                   |                        |      |
|                           | 2'3'cGAMP    |                   |                        |      |
| Synthesis                 | Cyclic GMP-AMP synthase (cGAS) | MAB21 (male abnormal 21) | IPR024805 | | (40) |
| Degradation               | unknown      |                   |                        |      |
| Detection                 | STING        | STING-C           | cd12146 | | (40) |
domains were occasionally detected in the emerging genomes of disparate eukaryotes. However, in almost all cases, the sequences were very similar to prokaryote homologs and absent from close relatives of the eukaryote in question, indicating that they probably originated from contaminating bacterial DNA. The exception is a conserved GGDEF domain that is present in six genomes, sampled across the full genetic diversity of dictyostelid social amoebas (39). Social amoebas chemotactically aggregate in response to nutrient stress to form multicellular aggregates which transform into fruiting structures. These structures consist of an aerial mass of spores that is supported by a column of stalk cells, encased in cellulose. Several species, such as D. discoideum have an intermediate light-sensitive migrating “slug” stage, which in nature serves to bring the organism to the top layer of the soil for optimal spore dispersal. Dictyostelids extensively use cAMP both in the traditional second messenger role and as a secreted first messenger for induction of chemotaxis and spore differentiation (48). Disruption of the DgcA gene, which encodes the D. discoideum GGDEF domain yielded dgcA-amoeobas that form normal migrating slugs, but never initiate fruiting body formation. DgcA is expressed at the slug tip, where the stalk starts to form and was shown to synthesize c-di-GMP. However, unlike prokaryote c-di-GMP, D. discoideum c-di-GMP does not act as a second messenger, but is secreted to locally induce stalk cell differentiation (39). It is as yet unclear how cells detect and process the c-di-GMP signal, and whether c-di-GMP is hydrolyzed. No EAL or HD-GYP domains are present in dictyostelid genomes and only weak homologies are found to bacterial c-di-GMP binding proteins (Chen, Z. and Schaap, P., unpublished results). As a secreted signal, it may have entirely different targets, such as one of the 15 sensor histidine kinases or the 48 G-protein coupled receptors of D. discoideum. The most intriguing question about c-di-GMP function in dictyostelids is whether it came in through the back door of lateral gene transfer (LGT), or whether it has deep origins in the first eukaryotes. Because a Dictyostelium amoeba consumes thousands of bacteria every day, LGT is not unlikely. One piece of evidence is suggestive of deep origins. The D. discoideum cellulose synthase, which is essential for stalk formation is more related to bacterial cellulose synthases than to other eukaryote (plant) enzymes. Although it does not have a PilZ domain, this domain, and its activation by c-di-GMP, may have been part of a signaling cassette that early eukaryotes “inherited” from bacteria, with the function of c-di-GMP in stalk formation changing over time.

Humans (and Other Animals)
The detection of viral or bacterial nucleic acids by cellular receptors is a critical step in the activation of the human innate immune system upon infection. The transmembrane protein STING (stimulator of interferon genes) is an important link in the pathway that leads from nucleic acid detection to expression of interferon genes. While STING itself does not interact with DNA, it was found to bind c-di-GMP derived from bacterial infections, and this interaction also induces interferon gene expression. c-di-GMP binds at a deep cleft between the monomers of the constitutively dimeric STING receptor (49).

The question how STING is activated by DNA was recently resolved by the finding that human cells synthesize the hybrid cyclic dinucleotide c-GMP-AMP using the enzyme cGAS, in response to the presence of cytosolic DNA (50,51). This enzyme belongs to the nucleotidyl transferase family, which includes class III nucleotidyl cyclases, polyadenylate polymerase, DNA polymerase, and oligoadenylate synthase (OAS1), with the latter being most similar to cGAS. cGAS harbors two N-terminal DNA binding domains, of which the second one is essential for activation of c-GMP-AMP synthesis by cytosolic DNA (50). c-GMP-AMP subsequently binds to STING to activate interferon gene expression.

Further studies showed that unlike c-di-AMP and c-di-GMP, in which both nucleotides are linked from the 3’OH of one to the 5’-phosphate of the other molecule, cGAS synthesizes a molecule in which the 2’OH of GMP is linked to the 5’-phosphate of AMP, and the 3’OH of AMP to the 5’-phosphate of GMP. This molecule is alternatively named 2’3’cGAMP (40) or cG[2’,5’]pA(3’,5’p) (52). 2’3’cGAMP binds to STING with 300-fold higher affinity than c-di-GMP or 3’3’cGAMP, however, its efficacy for interferon gene induction is only 30 times that of c-di-GMP. Structural studies revealed that 2’3’cGAMP sits deeper in the cleft between the STING dimers than c-di-GMP, and has three additional polar contacts with STING, explaining its higher affinity (40). Structural analysis of cGAS showed that it binds double stranded DNA in a non-specific manner by electrostatic interactions with the phosphodiester backbone. These interactions open the catalytic pocket of cGAS and repopulation several catalytic residues. The formation of a 2’5’ linkage by cGAS is unusual for nucleotidyl cyclases, but similar to formation of 2’5’ oligoadenylates generated by OAS1 (53).

Detection of foreign DNA by cGAS and subsequent activation of STING represents a much broader mechanism for the detection of bacterial, viral, and fungal infections than the direct interaction of STING with bacterial c-di-NMPs. It is therefore likely that the cGAS/STING system evolved independently to detect foreign DNA, and that STING activation by bacterial c-di-NMPs represents fortuitous use of an existing pathway.

Evolutionary Origins of cGAS and STING To detect co-evolution of cGAS and STING, we screened representative genomes of the major phyla or metazoa and their closest unicellular relative, the choanoflagellate Monosiga brevicollis, for homologs of the cGAS and STING genes. STING genes were found in all animal phyla, except Porifera, and also in M. brevicollis (Fig. 1A). Alignment of the sequences with the structurally characterized human STING, showed that at least 9 out of 10 residues that are required for binding c-di-NMPs are conserved in all phyla, except in the lophotrochozoan Capitella.
telata (Supporting Information Fig. S1), which also has an aberrant functional domain architecture (Fig. 1A).

The cGAS catalytic and DNA binding sequences are contained within the PFAM Mab21 domain, which was first detected in proteins involved in embryonic development. Blastp search with human cGAS detects Mab21 proteins throughout metazoa, but arthropod/lophotrochozoan Mab21 proteins are much more related to mammalian sensu stricto Mab21 proteins than to cGAS (Fig. 1B). Sequence alignment (Supporting Information Fig. S2) shows that the vertebrate cGAS homologs share all residues required for DNA and substrate binding with human cGAS, while the invertebrate proteins lack most of the DNA binding residues and several of the substrate binding residues. Interestingly, conservation of essential residues in the chordates Branchiostoma floridae and Saccoglossus kowalevski is intermediate between cGAS and Mab21, with the proteins of the hemichordate S. kowalevski and the cephalochordate B. floridae being somewhat more similar to Mab21 and cGAS, respectively (Supporting Information Figs. S2 and 1B). This probably marks the emergence of cGAS in the chordate/vertebrate lineage.

Contrary to expectation, the comparative analysis shows that cGAS and STING did not evolve together and that metazoa and their protist ancestors detected c-di-NMP signaling long before they could synthesize 2’3’cGAMP. The function of this early metazoan c-di-NMP detection system presents an intriguing avenue for further study.

Concluding Remarks

Similar to cAMP signaling, the rapidly expanding field of c-di-NMP signaling owes its existence to the original painstaking identification of a small biologically active molecule (2). Whereas molecular genetics and genomics are currently the most powerful tools to unravel gene function, they have limited ability to predict the repertoire of small molecules that many of the cognate enzymes might synthesize. Our current understanding of particularly developmental signaling is therefore biased toward peptide signals that were discovered through genetics, leaving potentially important non-peptide mediated signaling undiscovered. The diverse functions of monomeric and polymeric nucleotides in information storage, energy transfer, and signaling depend on the making and breaking of phosphodiester bonds. All life forms have a very large repertoire of enzymes to perform these functions and their potential for producing biologically active molecules may as yet be vastly underestimated. The expertise of dedicated biochemists, rather than geneticists, will however be needed to identify these molecules.

References

[1] Gilbert, W. (1986) Origin of life—the RNA world. Nature 319, 618.
[2] Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., et al. (1987) Regulation of cellulose synthesis in Acetobacter xylinaum by cyclic diguanylic acid. Nature 325, 279–281.
[32] Wu, T., Lee, S., Chang, E., and Lee, J. (2007) Characterization of the Cbr-1 protein as a regulator of cyclic GMP metabolism in V. cholerae. PLoS Pathog. 3, e175.

[33] Hoshino, Y., Yoshimura, T., and Arakawa, H. (2007) Identification of the CrtR protein as a cyclic di-AMP-responsive regulator in Mycobacterium smegmatis. J. Bacteriol. 189, 5434–5443.

[34] Takeuchi, T., Yoshida, K., and Takeuchi, K. (2007) Identification of the CdiA protein as a cyclic di-AMP-responsive regulator in Mycobacterium smegmatis. J. Bacteriol. 189, 5434–5443.

[35] Shi, H., Tanaka, M., and Ohtsubo, E. (2007) Identification of the CrtR protein as a cyclic di-AMP-responsive regulator in Mycobacterium smegmatis. J. Bacteriol. 189, 5434–5443.

[36] Takeuchi, T., Yoshida, K., and Takeuchi, K. (2007) Identification of the CdiA protein as a cyclic di-AMP-responsive regulator in Mycobacterium smegmatis. J. Bacteriol. 189, 5434–5443.

[37] Shi, H., Tanaka, M., and Ohtsubo, E. (2007) Identification of the CrtR protein as a cyclic di-AMP-responsive regulator in Mycobacterium smegmatis. J. Bacteriol. 189, 5434–5443.
Corrigan, R. M., Abbott, J. C., Burhenne, H., Kaeever, V., and Grundling, A. (2011) 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–e1002217.

Mehne, F. M. P., Gunka, K., Eilers, H., Herzberg, C., Kaeever, V., et al. (2013) Cyclic Di-AMP Homeostasis in Bacillus subtilis both lack and high level accumulation of the nucleotide are detrimental for cell growth. J. Biol. Chem. 288, 2004–2017.

Luo, Y. and Helmann, J. D. (2012) Analysis of the role of Bacillus subtilis sM in ss-lactam resistance reveals an essential role for c-di-AMP in peptidoglycan homeostasis. Mol. Microbiol. 83, 623–639.

Witte, C. E., Whiteley, A. T., Burke, T. P., Sauer, J.-D., Portnoy, D. A., et al. (2013) Cyclic di-AMP is critical for Listeria monocytogenes growth, cell wall homeostasis, and establishment of infection. mBio 4, e00282–e00313.

Schaap, P. (2011) Evolutionary crossroads in developmental biology: Dictyostelium discoideum. Development 138, 387–396.

Burdette, D. L. and Vance, R. E. (2013) STING and the innate immune response to nucleic acids in the cytosol. Nat. Immunol. 14, 19–26.

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

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

Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B.L., et al. (2013) Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. Cell 153, 1094–1107.

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

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., et al. (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. 39, W13–W17.

Ronquist, F. and Huelsenbeck, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.