Second messenger molecules allow cells to amplify signals and rapidly control downstream responses. This concept is exemplified in human cells, where mislocalized double-stranded (ds)DNA stimulates the cytosolic enzyme cyclic GMP–AMP synthase (cGAS) to synthesize the CDN 2′-5′/3′-5′-cyclic GMP–AMP (2′,3′-cGAMP)1,2,12. 2′,3′-cGAMP diffuses throughout the cell, activates the receptor stimulator of interferon genes (STING) and induces type-I interferon and NF-κB responses to elicit protective anti-viral immunity1. Recently, synthetic CDN analogues have emerged as promising lead compounds for immune modulation and cancer immunotherapy2. CDNs were first identified in bacteria1 and this discovery established the foundation for later recognition of the importance of CDN signalling in mammalian cells3. Nearly all bacterial phyla encode CDN signalling pathways; however, all known natural CDN signals are constructed only from purine nucleotides6. CDNs control diverse responses in bacterial cells. For example, cyclic di-GMP coordinates the transition between planktonic and sessile growth, cyclic di-AMP controls osmoregulation, cell-wall homeostasis and DNA-damage responses, and 3′-5′/3′-5′-cGAMP (3′,5′-cGAMP) modulates chemotaxis, virulence and exoelectrogenesis7. The human receptor STING also senses these bacterial CDNs as pathogen (or microbe)-associated molecular patterns (PAMPs), revealing a direct, functional connection between these bacterial CDNs and their animal and plant hosts.

Cyclic dinucleotides (CDNs) have central roles in bacterial homeostasis and virulence by acting as nucleotide second messengers. Bacterial CDNs also elicit immune responses during infection when they are detected by pattern-recognition receptors in animal cells. Here we perform a systematic biochemical screen for bacterial signalling nucleotides and discover a large family of cGAS/DncV-like nucleotidyltransferases (CD-NTases) that use both purine and pyrimidine nucleotides to synthesize a diverse range of CDNs. A series of crystal structures establish CD-NTases as a structurally conserved family and reveal key contacts in the enzyme active-site lid that direct purine or pyrimidine selection. CD-NTase products are not restricted to CDNs and also include an unexpected class of cyclic tri nucleotide compounds. Biological and cellular analyses of CD-NTase signalling nucleotides demonstrate that these cyclic di- and tri nucleotides activate distinct host receptors and thus may modulate the interaction of both pathogens and commensal microbiota with their animal and plant hosts.

Discovery of a pyrimidine-containing CDN
The enzyme DncV synthesizes 3′,5′-cGAMP and controls a signalling network on the Vibrio seventh pandemic island-1 (VSP-I), a horizontally acquired genetic element present in all current V. cholerae pandemic isolates11,12. While investigating homologues of dncV outside the Vibrionales, we identified an unexpected partial operon in Escherichia coli where dncV is replaced with a gene of unknown function (WP_001593458, here named cdnE). The operon architecture suggests that cdnE may be an alternative CDN synthase (Fig. 1a). We tested this hypothesis by incubating purified CdnE protein with 32P-radiolabelled ATP, CTP, GTP and UTP and visualized the reaction products using thin-layer chromatography (TLC). CdnE synthesized a product distinct from currently known CDNs (Fig. 1b and Extended Data Fig. 1a, b). Notably, biochemical deconvolution using pairwise assessment of necessary NTPs revealed that ATP and UTP were necessary and sufficient for product formation (Fig. 1c). We analysed the purified product using nuclease digestion, mass spectrometry and NMR (Fig. 1d and Extended Data Fig. 1d–f), and confirmed that the product of CdnE is cyclic UMP–AMP (cUMP–AMP), a hybrid purine–pyrimidine CDN.

DncV is a structural homologue of cGAS, and each enzyme uses a single active site to sequentially form two separate phosphodiester bonds and release a CDN product10. In spite of no overall sequence homology, careful inspection of the CdnE sequence revealed potential cGAS/DncV-like active-site residues (GSYXGH3, here named cGAS/DncV-like nucleotidyltransferase cdnE). The operon architecture of dncV is also preceded by a gene that may be an alternative CDN synthase (Fig. 1a). The operon architecture of dncV is also preceded by a gene that may be an alternative CDN synthase (Fig. 1a).
CdnE was incubated with WP_001593458). The ECOR31 strain’s genomic island encodes a second phospholipase. Enzyme activity is reported in phospholipase A_1 units (U) ml$^{-1}$. Expected [M + H]$^+$ refers to mass spectrometry and NMR (see Extended Data). The CdnE product was confirmed by mass spectrometry and NMR (see Extended Data). The CdnE product represents three independent experiments.

Cyclic UMP–AMP, Cyclic dA–dU–AMP and Cyclic dA–dI–AMP (cGAMP) are formed through nucleotide addition (Ori., origin). Biochemical deconvolution of the CdnE reactions in b after incubation with $^{32}$P-labelled and unlabelled NTPs. Data are representative of three independent experiments. d, The CdnE product was confirmed by mass spectrometry and NMR (see Extended Data). e, Activation of CapE and CapE by Cdns, tested with no nucleotide added (−) or at 0.1-, 1- and 10-fold molar ratios of nucleotide to phospholipase. Enzyme activity is reported in phospholipase A_1 units (U) ml$^{-1}$. Data are mean ± s.e.m. for n = 3 technical replicates and are representative of three independent experiments.

Mechanism of pyrimidine discrimination
We determined a series of X-ray crystal structures of a CdnE homologue from the thermophilic bacterium Rhodothermus marinus (RmCdnE; Fig. 2, Extended Data Fig. 3a and Supplementary Table 1). CdnE adopts a Pol-β-like nucleotidyltransferase fold that is highly similar to cGAS and the core of DncV, confirming a shared structural and evolutionary relationship (Fig. 2d). CdnE is more distantly related to other nucleotidyltransferases, including non-templated CCA-adding enzymes, poly(A) polymerases and templated polymerases, such as DNA polymerase β and μ. The human innate immune enzymes cGAS and oligoadenylate synthase 1 (OAS1) are activated through a conformational change induced by binding to a double-stranded nucleic acid. CdnE, like DncV, is structurally more similar to the ‘activated’ conformation of these two enzymes, consistent with biochemical analyses that demonstrate that CdnE is constitutively active and does not require a cognate stimulus in vitro.

The structure of RmCdnE in complex with non-hydrolysable ATP and UTP reveals an asparagine side chain (Asn166) that forms hydrogen bonds with the uracil base and positions the ATP $^\gamma$P for attack by the $^\gamma$-hydroxyl of UTP (Fig. 2a). Asn166 is located in the same position as a serine residue in the acceptor nucleotide pocket of both DncV and cGAS (Extended Data Fig. 3b, d), and we hypothesized that this asparagine substitution might be sufficient to dictate CdnE product specificity. Whereas wild-type CdnE robustly synthesized cUMP–AMP, CdnE$^{166S}$ incorporated almost no UTP and instead predominantly synthesized c-di-AMP (Fig. 2c and Extended Data Fig. 3c). We surveyed CdnE homologues and determined that Asn166 is nearly universally conserved (Fig. 2b and Extended Data Fig. 4a). An exception is CdnE from the emerging nosocomial pathogen Elizabethkingia meningoseptica (EmCdnE, Fig. 2b), which encodes a serine at the analogous position to Asn166. In contrast to the other CdnE homologue, EmCdnE robustly synthesized cyclic dipurine products (Fig. 2c and Extended Data Fig. 4b–h). Crystal structures of EmCdnE bound to its nucleotide substrates demonstrated natural N to S reprogramming in the active-site lid, and re-introduction of the ancestral asparagine at this position reverted EmCdnE back to preferential production of pyrimidine-containing products (Fig. 2c, Extended Data Fig. 4f–i and Supplementary Table 1). These data reveal a remarkably low barrier for altering specificity of CdnE and demonstrate that organisms, such as E. meningoseptica, harbour mutations at Asn166 that reprogram purine and pyrimidine product specificity.

CD-NTases and cross-kingdom signalling
Many bacteria that encode CD-NTases thrive in close proximity to eukaryotic hosts, including humans, plants and fungi (Fig. 2b). CdnE homologues are found in the intracellular pathogen Shigella sonnei and commensal genera such as Bacteroides (Fig. 2b). Mammals have evolved a sophisticated surveillance system for detecting and initiating immune responses to bacterial products, including CDNs that are secreted or released during bacteriolysis. STING detects bacterial c-di-AMP, c-di-GMP and 3′,3′-c-di-GMP in addition to endogenously produced 2′,3′-c-cGAMP. We determined whether cUMP–AMP was also recognized by STING or other receptors of the innate immune system. Mouse STING bound to all four cyclic dipurine molecules with high affinity and activated type-I interferon signalling in cells. However, STING was unable to recognize cUMP–AMP in vitro at concentrations known to be sufficient for cyclic dipirurine agonists and cUMP–AMP failed to activate STING-dependent type-I interferon signalling in cells (Fig. 3a and Extended Data Fig. 5a–d). These data are consistent with previous experiments using chemically synthesized nucleotides and were not due to differences in CD-NTase expression levels (Extended Data Fig. 5c, d). By contrast, the recently described mammalian CDN...
Fig. 2 | Conserved active-site residues dictate CD-NTase specificity.

**a.** Crystal structure of RmCdnE in complex with non-hydrolysable ATP and UTP analogues and magnified inset of key Asn166-uridine contacts that control pyrimidine specificity. Green dotted lines indicate hydrogen bonds and the 2Fα – Fc electron density map is contoured at 1σ.

**b.** Phylogram of CdnE sequence homologues and their Asn166 analogous residue determined by sequence alignment (Extended Data Fig. 4a). Residues indicated in red (S) highlight cGAS/DncV-like serine residues. CdnE homologue and mutant reactions analysed by TLC as in Fig. 1b. N and red S indicate asparagine or cGAS/DncV-like serine residues at the Asn166 analogous position in the tested allele. Side chains are numbered according to the RmCdnE sequence. Data are representative of three independent experiments. For detailed deconvolution and purified versus pyrimidine migration pattern analysis see Extended Data Figs. 3, 4d. Structure-based comparisons define that RmCdnE and EmCdnE are CD-NTases with a similar architecture to DncV (Protein Data Bank (PDB) accession 4TY0), cGAS (PDB 6CTA) and OAS1 (PDB 4RWO). CD-NTases are more distantly related to Pol-β-like NTases: Pol-β (PDB 4YD1), Pol-β (PDB 4KLQ), CCA-adding enzyme (PDB 4X4T) and poly(A) polymerase γ (PAP, PDB 4LT6). The NTase core domain for each enzyme is clustered according to z-score and coloured in magenta and blue.

**CD-NTases synthesize diverse nucleotide products**

DncV and CdnE are likely to have evolved from a common ancestor, but exhibit marked divergence in primary amino acid sequence. We hypothesized that these enzymes comprise only a small fraction of existing bacterial CD-NTase diversity, and that kingdom-wide analysis of the protein family would allow systematic identification of bacterial signalling nucleotides as well as agonists/antagonists of the innate immune system. We therefore coupled bioinformatics analysis with a large-scale, forward biochemical screen to directly uncover additional nucleotide products. Previously, a hidden Markov model was derived from cGAS and DncV and conserved operon structures to identify potentially related bacterial proteins. Building upon this previous analysis, we identified more than 5,600 unique bacterial enzymes predicted to share common CD-NTase structural features (Fig. 4a, Extended Data Fig. 6a and Supplementary Table 2). CD-NTases were identified in over 10% of bacterial genomes available in the NCBI database, within taxa that span nearly every bacterial phylum (Extended Data Fig. 6b). Bacteria harbouring CD-NTase genes include human commensal organisms (for example, Clostridiales and Fusobacteria), human pathogens (for example, *Listeria, Shigella* and *Salmonella* species), extemophiles and agriculturally important bacteria (for example, rhizobia and plant pathogens, such as *Xanthomonas*). Although CD-NTases are found in many different organisms, they are typically not encoded in the core genome and are encoded by specific strains from each species. Sequence alignments revealed that CD-NTases cluster into roughly eight clades that we designated A–H starting with A for the DncV-harbouiring clade, E for the CdnE-containing clade and continued to the letter H. We further divided highly related sequences into clusters, which often grouped bacterial species that occupy a similar niche, such as the plant rhizobia in cluster G10 (Fig. 4a and Supplementary Table 2).

We purified 66 CD-NTase proteins and tested each for nucleotide product synthesis (Fig. 4a and Extended Data Fig. 6c–g). These proteins were selected as type enzymes from each cluster based on the relevance of the organism from which they were isolated (pathogens, commensals, and bacteria predicted to interact with eukaryotes) and

Fig. 3 | Immune detection of a pyrimidine-containing CDN.

**a.** In-cell STING reporter assay. Induction of an interferon-β (IFNβ) reporter in HEK293T cells transfected with a concentration gradient of plasmid-overexpressing enzymes as indicated. cGAS synthesizes 2′,3′-cGAMP, DncV synthesizes 3′,3′-cGAMP, DisA synthesizes c-di-AMP, WspR synthesizes c-di-GMP and CdnE synthesizes CUMP-AMP. Data are fold induction over vector only (−). Data are mean ± s.e.m. for n = 3 technical replicates and are representative of three independent experiments.

**b.** Nucleotide inhibition of RECON enzymatic activity, as measured by the oxidation of the NADPH co-substrate. The x axis is a log scale and data are representative of three independent experiments.
CD-NTases synthesize diverse nucleotide products and form a family of enzymes abundant in many bacterial phyla. a. Bioinformatic identification and clustering of ~5,600 predicted CD-NTases found in nearly every bacterial phylum, shown as an unrooted tree. Sequence-related enzymes that are approximately 10% identical are grouped by lettered clade and are similarly coloured, enzymes that are about 25% identical are grouped by cluster in a similar colour shade. Circles represent CD-NTases 001–066 that were selected as type CD-NTases for a biochemical screen (see Extended Data Fig. 6 for additional details). Blue circles denote CD-NTases selected for in-depth characterization and are labelled with CD-NTase numbers or letters from the biochemical screen (see b, CdnE is ‘56’ and DncV is ‘D’). For additional information, see Supplementary Discussion and Supplementary Table 2. Source Data are provided in the online version of the paper. b. c, PEI–cellulose and silica TLC analysis of 16 CD-NTases selected for in-depth characterization. Activity was analysed with 32P-radiolabelled NTPs as in Fig. 1b. Wild-type (WT) and catalytically inactive mutant (MUT) DncV reactions are included as controls. Screened CD-NTases were numbered CD-NTases 001–066. CD-NTase056 is CdnE, CD-NTase057 is LpCdnE02 and CD-NTase038 is EsCdnD02. These appear as ‘56’, ‘57’ and ‘38’ in respectively. Data are representative of three independent experiments. d, Identification of CD-NTase products by combining TLC and mass spectrometry data. CD-NTases that synthesize a major product that could not be matched to a predicted CDN are denoted as ‘unknown’. CD–NTases are encoded in conserved operons A unifying characteristic of almost all CD-NTase-encoding genes is their location within similar operons in predicted mobile genetic elements (Extended Data Fig. 8a). Often genes that encode identical CD-NTase proteins are found in specific strains of unrelated bacterial species, reflecting that these genes are members of the ‘mobilome’ (for detail on identifying CD-NTases within an organism of interest, see Supplementary Discussion). The horizontal acquisition of CD-NTases suggests that they are likely to provide a selective advantage by altering bacterial physiology through adjacently encoded receptors, similar to capV–dncV and capE–cdnE, rather than integrating into species-specific nucleotide signalling networks. It was previously demonstrated that genes adjacent to CD-NTases are effector-like and are generally involved in biological conflict, including phospholipases, nucleases and pore-forming agents. Coexpression of dncV and capV is toxic to E. coli and we tested whether coexpression of each CD-NTase with its adjacently encoded, putative receptor was also toxic to E. coli. Expression of dncV–capV was uniquely able to inhibit colony formation, whereas other CD-NTase-predicted receptor pairs, including the cdnE–capE pair, did not impair bacterial growth (Extended Data Fig. 8b, c). It is unclear whether CD-NTases are constitutively active in vivo or exhibit regulated enzymatic activity similar to the metazoan second messenger synthase cGAS. These findings demonstrate that phenotypes observed with Vibrio dncV–capV may not be indicative of general CD-NTase function, and that CD-NTase-containing islands...
may vary in their functions such as mediating bacteriophage resistance, modulating bacteria–host interactions, functioning as addiction modules or regulating bacteriolysis for dissemination of mobile genetic elements.

CD–NTase products include cyclic trinucleotides

Notably, we were unable to identify expected CDNs by mass spectrometry in some reactions despite visualizing robust product formation by PEI–cellulose TLC. Using orthogonal silica TLC conditions, these unknown products exhibited distinct migration patterns that suggested existence of unique non-CDN species (Fig. 4c, d). We focused on an orphan product of CD–NTase038 from the Enterobacter cloacae complex (strain UCI 50) (WP_032676400, here named EcCdnD02) for identification. The EcCdnD02 product initially appeared to be a cyclic diphosphate by PEI–cellulose TLC, but the major EcCdnD02 product displayed a unique migration pattern when analysed by silica TLC (Fig. 5a and Extended Data Fig. 9a). ATP and GTP were necessary and sufficient for product formation; however, roughly two thirds of the total $^32$P was incorporated from ATP and the remaining third from GTP. Consistent with this pattern, evaluation of mass spectrometry data and subsequent biochemical and NMR validation revealed that cyclic AMP–AMP–GMP (cAAG), a cyclic trinucleotide, is the major product of EcCdnD02 (Fig. 5b and Extended Data Fig. 9b–j).

Similar to cUMP–AMP, the bacterial cyclic trinucleotide cAAG escaped STING recognition but was detected by RECON, confirming our new definition of STING and RECON ligand specificity (Fig. 5c and Extended Data Fig. 10a, f). We next determined a co-crystal structure of mouse RECON in complex with the EcCdnD02 cyclic trinucleotide product (Fig. 5d, Extended Data Fig. 10 and Supplementary Table 1). The structure confirms that the bacterial cyclic trinucleotide is cAAG and exclusively contains $3'$–$5'$-phosphodiester linkages. The two adenine bases are coordinated in the same adenine and nicotinamide pockets observed in the previous structure of RECON bound to bacterial c-di-AMP$^{21}$; however, RECON Glu28 makes additional contacts with the third guanine base of the cAAG species as part of an extended base platform that is not required for enzymatic activity, as measured by the oxidation of the NADPH co-substrate. The x axis is a log scale and data are representative of three independent experiments. d, Co-crystal structure of the host receptor RECON in complex with cAAG. The inset highlights the cAAG $2F_o - F_c$ electron density contoured at 1.3σ. Green dotted lines indicate hydrogen bonds, some RECON–cAAG contacts are omitted for clarity, see Extended Data Fig. 10.

CD–NTases in health and disease

Our data demonstrate that bacterial CD–NTases are widespread and synthesize diverse CDNs that include pyrimidine nucleotides and additional cyclic trinucleotides like cAAG. Recent evidence demonstrates that divergent GGDEF family synthases join the GGDEF and DAC/DisA domains, responsible for c-di-GMP and c-di-AMP synthesis,$^{17,18}$ as a third major family of enzymes that control downstream signalling using CDN signals. Distinguishing features of CD–NTases are their location on mobile genetic elements, extreme sequence diversity and reaction mechanism that is reliant on a monomeric enzyme active site. Recent evidence demonstrates that divergent GGDEF family enzymes produce $3'$,$3'$-cGAMP in addition to c-di-GMP$^{25,26}$, suggesting that the selective pressures driving CD–NTase diversity may also be in effect for GGDEF and DAC/DisA-like synthases.

Understanding the functional role of CD–NTase genes in the biology of bacteria and host–microorganism interactions is a major challenge for future studies. Mammalian receptors recognize diverse CD–NTase products and CD–NTase genes may provide a selective advantage for some bacterium–eukaryote interactions. Our data show that a single mutation in a CD–NTase enables incorporation of pyrimidines and indicate that bacteria may evade or enhance STING signalling by modulating enzyme specificity. The possibility that diverse CDNs and related nucleotide signals produced by prokaryotic CD–NTases act as agonists and inhibitors of innate immunity and other host metabolic...
pathways provides an important new reservoir of compounds with biotechnological and therapeutic applications.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-0953-5.

Received: 6 July 2018; Accepted: 15 January 2019; Published online 20 February 2019.

1. Wu, J. & Chen, Z. J. Innate immune sensing and signalling of cytosolic nucleic acids, Annu. Rev. Immunol. 32, 461–488 (2014).
2. Corrales, L. et al. Direct activation of STING in the tumor microenvironment leads to potent and systemic tumor regression and immunity. Cell Rep. 11, 1018–1030 (2015).
3. Fu, J. et al. STING agonist formulated cancer vaccines can cure established tumors resistant to PD-1 blockade. Sci. Transl. Med. 7, 283ra52 (2015).
4. Ross, P. et al. Regulation of cell cycle progression in Acetobacter xylinum by cyclic diguanylic acid. Nature 325, 279–281 (1987).
5. Danilchanka, O. & Mekalanos, J. J. Cyclic dinucleotides and the innate immune response. Cell 154, 962–970 (2013).
6. Nelson, J. W. & Breaker, R. R. The last language of the RNA world. Sci. Signal. 10, eaax882 (2017).
7. Krasteva, P. V. & Sondermann, H. Versatile modes of cellular regulation via cyclic dinucleotides. Nat. Chem. Biol. 13, 350–359 (2017).
8. Burdette, D. L. et al. STING is a direct innate immune sensor of cyclic di-GMP. Nature 478, 515–518 (2011).
9. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP–AMP synthase is a cyclic DNA sensor that activates the type I interferon pathway. Science 339, 786–791 (2013).
10. Kranzusch, P. J. et al. Structure-guided reprogramming of human cGAS-dinucleotide linkage specificity. Cell 158, 1011–1021 (2014).
11. 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, 358–370 (2012).
12. Hu, D. et al. Origins of the current seventh cholera pandemic. Proc. Natl Acad. Sci. USA 113, E7730–E7739 (2016).
13. Dziejmin, M. et al. Comparative genomic analysis of Vibrio cholerae: genes that correlate with cholera endemic and pandemic disease. Proc. Natl Acad. Sci. USA 99, 1556–1561 (2002).
14. Severin, G. B. et al. Direct activation of a phospholipase by cyclic GMP–AMP in El Tor Vibrio cholerae. Proc. Natl Acad. Sci. USA 115, E6048–E6055 (2018).
15. Hornung, V., Hartmann, R., Ablasser, A. & Hopfer, K.-P. cGAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids. Nat. Rev. Immunol. 14, 521–528 (2014).
16. Jean, S. S., Lee, W. S., Chen, F. L., Ou, T. Y. & Hsueh, P. R. Elizabethkingia meningoseptica: an important emerging pathogen causing healthcare-associated infections. J. Hosp. Infect. 86, 244–249 (2014).
17. Jenal, U., Reinders, A. & Lori, C. Cyclic di-GMP: second messenger extraordinaire. Nat. Rev. Microbiol. 15, 271–284 (2017).
18. Corrigan, R. M. & Gründling, A. Cyclic di-AMP: another second messenger enters the fray. Nat. Rev. Microbiol. 11, 513–524 (2013).
19. Woodward, J. J., lavarone, A. T. & Porthoy, D. A. cyclic di-AMP secreted by intracellular Listeria monocytogenes activates a host type I interferon response. Science 328, 1703–1705 (2010).
20. Wang, C. et al. Synthesis of all possible canonical (3′-5′-linked) cyclic dinucleotides and evaluation of riboswitch interactions and immune-stimulatory effects. J. Am. Chem. Soc. 139, 16154–16160 (2017).
21. McFarland, A. P. et al. Sensing of bacterial cyclic dinucleotides by the oxoreductase RECON promotes NF-κB activation and shapes a proinflammatory antibacterial state. Immunity 46, 433–445 (2017).
22. Burrows, A. M., Zhang, D., Schaffer, D. E., Iyer, L. M. & Aravin, L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic Acids Res. 43, 10633–10654 (2015).
23. Kuzalskiene, M., Kostiku, G., Venclovos, C., Tamularis, G. & Siksnyš, V. A cyclic oligonucleotide signaling pathway in type III CRISPR–Cas systems. Science 357, 605–609 (2017).
24. A. H. Woehner, O. et al. Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. Nature 548, 543–548 (2017).
25. Hallberg, Z. F. et al. Hybrid promiscuous (Hypr) GDDE enzymes produce cyclic AMP–GMP (3′–3′cGAMP). Proc. Natl Acad. Sci. USA 113, 1790–1795 (2016).
26. Nelson, J. W. et al. Control of bacterial eukaryotic cell death by c-AMP–GMP. Proc. Natl Acad. Sci. USA 112, 5389–5394 (2015).

Acknowledgements
We thank K. McCarty and V. Cabrera for technical assistance; S. Wilson for advice; M. Reniere for reading the manuscript; T. Levin for input on CD-NTase tree construction; T. Wyche and M. Henke from the HMS ICBB Longwood Screening Facility and the staff at the HMS East Quad HMR Facility for screening support and technical support; A. Govande and W. Zhou for providing purified proteins; C. Miller for assistance with X-ray crystallography data collection; and the members of the Mekalanos and Kranzusch laboratories for advice and discussions. This work was supported by the Claudia Adams Barr Program for Innovative Cancer Research (P.J.K.), Richard and Susan Smith Family Foundation (P.J.K.), Charles H. Hood Foundation (P.J.K.), a Cancer Research Institute CLIP Grant (P.J.K.), NIH/NIADD R01AI108045 and R01AI026289 (J.J.M.), the Saratleon Scholars Program (A.S.Y.L) and a Sloan Research Fellowship (A.S.Y.L.). A.T.W. is supported as a fellow of the Jane Coffin Childs Memorial Fund for Medical Research, C.C.O.D.M. is supported as a Cancer Research Institute/Eugene V. Weissman Fellow and B.R.M. is supported by the NIH T32 Cancer Immunology training grant (ST32CA07021-02). X-ray data were collected at the Lawrence Berkeley National Laboratory Advanced Light Source beamlines 8.2.1 and 8.2.2, this work used Northeastern Collaborative Access Team beamlines 24-ID-C and 24-ID-E (P30 GM124165), a Pilatus detector (S1ORRR29205), an Eiger detector (S100DD21527) and Argonne National Laboratory Advanced Photon Source (DE-AC02-06CH11357).

Reviewer information
Nature thanks Regine Henge, Joshua Woodward and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions
Experiments were designed and conceived by A.T.W., J.J.M. and P.J.K. Gene identification, activity screening and bacterial assays were performed by A.T.W. Biochemical experiments and structural analysis were performed by J.B.E. and P.J.K. Structural experiments and analysis were performed by J.B.E. and P.J.K. Nucleotide synthesis and purification were performed by A.T.W., B.L. and P.J.K. Mass spectrometry was performed by C.C.O.D.M. and D.S.K. NMR was performed by B.R.M. Cloning assistance was provided by E.A.N. Observations and strains were contributed by O.D. Bioinformatics and cell assays were performed by A.S.Y.L. Figures were prepared by A.T.W., J.B.E. and P.J.K. The manuscript was written by A.T.W., J.J.M. and P.J.K. All authors contributed to editing the manuscript and support the conclusions.

Competing interests
Harvard Medical School and the Dana-Farber Cancer Institute have patents pending for CD-NTase technologies on which A.T.W., J.B.E., J.J.M. and P.J.K. are listed as inventors.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-0953-5.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-0953-5.
Reprints and permissions information is available at http://www.nature.com/reprints/index.html.
Correspondence and requests for materials should be addressed to J.J.M. or P.J.K.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019.
Bacterial strains and growth conditions. *E. coli* was cultivated at 37 °C, shaking, in LB medium (1% tryptone, 0.5% yeast extract and 0.5% NaCl w/v) and stored in LB plus 30% glycerol at −80 °C unless otherwise indicated. When appropriate, carbenicillin (100 μg ml−1), ampicillin (100 μg ml−1) and chloramphenicol (20–34 μg ml−1) were used. BL21 E. coli (strain CodonPlus (DE3)-RIL transformed with pRARE2, Agilent) was used for all protein expression and DH10B E. coli (strain Top10, Invitrogen) was used for cloning and plasmid propagation. For repression of protein expression from PET vectors, BL21 E. coli was cultivated in MDG medium (0.5% glucose, 25 mM NaHPO4, 25 mM KH2PO4, 50 mM NH4Cl, 5 mM Na2SO4, 2 mM MgSO4, 0.25% aspartic acid and trace metals) with ampicillin and chloramphenicol. For optimum protein expression from PET vectors, BL21 E. coli was cultivated in M9ZB medium (0.5% glycerol, 1% Cas-amino acids, 47.8 mM NaHPO4, 22 mM KH2PO4, 18.7 mM NH4Cl, 85.6 mM NaCl, 2 mM MgSO4 and trace metals) with ampicillin and chloramphenicol. Cloning and plasmid construction. Cloning and plasmid construction were performed as previously described[8]. In brief, for vectors constructed in this study, genes were either amplified from genomic DNA or synthesized as gBLOCKs (Integrated DNA Technologies) with 18 base pairs of homology flanking the insert sequence and ligated into a restriction endonuclease linearized vector by Gibson assembly. Reactions were transferred into electrocompetent DH10B cells and selected with appropriate antibiotic plates. Sanger sequencing confirmed each vector was free of mutations within the multiple cloning site. N-terminal 6×His-MBP tag and 6×His-SUMO2 tag fusions were constructed using custom PET61MBP[19] or PETSUMO2[20] vectors, respectively. cGAS standards used Homo sapiens Bacillus subtilis and Bacillus thuringiensis with approximately 1 cm from the top to 10-cm standards used Pseudomonas aeruginosa with a D70E constitutively activating mutation. CD-NTases and their effector coding sequences were codon-optimized for bacterial expression (Integrated DNA Technologies) with the exception of genes derived from *E. coli* strain ECO3315 (ATCC 35330) and *V. cholerae* strain C6706. Synthases were overexpressed in mammalian cells from pcDNA4 plasmids as previously described[37]. For expression of MBP N-terminally tagged dncV and cdnE in mammalian cells, MBP and the fused CD-NTase were codon-optimized for expression in human cells. For coexpression with their putative effector genes, N-terminal MBP-tagged CD-NTases were cloned into pBAD33 modified with a ribosomal binding site and oriT for conjugation. For cloned CD-NTase details see Supplementary Table 2 and for cloned CD-NTase effector details see Supplementary Table 3.

Recombinant protein purification. Proteins were purified as previously described[38]. In brief, chemically competent BL21 E. coli was transformed with a protein-expression plasmid, recovered on MDG plates overnight, cultivated as a 30 ml starter culture in MDG liquid medium overnight at 37 °C with 230 r.p.m. shaking and used to seed an M9ZB culture at approximately 1:1000. Then, 25 ml or two 1-l M9ZB cultures were cultivated for approximately 5 h at 37 °C with 230 r.p.m. shaking until the optical density (OD) at 600nm was 2–3.5 at which time cells were chilled on ice for 20 min, IPTG was added at 0.5 mM and cultures were shifted to 16 °C with shaking 230 r.p.m. overnight. Harvested *E. coli* cultures were washed in 1× PBS and stored as a flash-frozen pellet at −80 °C or cultures were immediately disrupted by sonication in lyss buffer (20 mM HEPES-KOH pH 7.5, 400 mM NaCl 30 mM imidazole, 10% glycerol and 1 mM DTT). Lysates were clarified by centrifugation, filtered through glass wool and proteins were purified by affinity chromatography using Ni-NTA (Qiagen) resin and a gravity elution buffer (250 mM CAPSO pH 9.4, 175 mM K/NaCl, 5 mM Mg2+, 32P]NTPs in Fig. 1b are 250 μM individual NTPs and 1 μM enzyme. A 20-μl aliquot of the larger reaction was removed and (65–70 °C) were added to monitor reaction progress. Reactions were incubated for 24 h at which time 5 μl of 1:1000 of alkaline phosphatase (New England Biolabs) was added and the reaction was further incubated for 2–24 h. Reactions were heat-inactivated at 65 °C for 30 min, diluted to a final salt concentration of 12.5 mM and purified by anion exchange chromatography (either 1 ml Q-sepharose column or Mono Q 4.6/100 PE, GE Healthcare). The column was washed with water and 1 ml fractions were collected during a gradient elution with 50 mM NaCl. The 20-cm × 20-cm F-coated PEI–cellulose TLCs (Millipore) were developed in 1.5 M KHPO4 (pH 3.8) until the buffer front reached approximately 1 cm from the top; 20-cm × 20-cm F-coated silica HP–TLC plates (Millipore) were developed in 11:7:2 1-propanol:NH4OH:H2O in a chemical fume hood for 1 h. Plates were dried and exposed to a phosphor screen before detection by a Typhoon Trio Variable Mode Imaging system (GE Healthcare).

Nucleotide synthesis and purification. Cyclic dinucleotides and oligonucleotides were produced in large scale using previously described methods[22] with the following changes. Small-scale nucleotide synthesis assays were scaled up to 10–40 ml reactions with final conditions of 50 mM CAPSO pH 9.4, 12.5–50 mM NaCl, 5–20 mM Mg(OAc)2, 1 mM DTT, 5% glycerol, 250 μM individual NTPs and 1 μM enzyme. A 20-μl aliquot of the larger reaction was removed and (6530 QTOF mass spectrometer coupled to a 1290 infinity binary LC system operated using the electrospray source in positive ionization mode. All samples were chromatographed and identified by A260 nm (Fractions containing the appropriate product were identified by absorbance at 260 nm (A260 nm) or U260, pooled and evaporated. Concentrations of purified nucleotides were estimated from A260 nm and the extinction coefficients based on RNA oligonucleotides: cUMP-AMP, ε = 22,800 l mol−1 cm−1; cAAG, ε = 37,000 l mol−1 cm−1.

Mass spectrometry. Liquid chromatography coupled with electrospray ionization-mass spectrometry (LC/ESI-MS) analysis was performed using an Agilent 6530 QTOF mass spectrometer coupled to a 1290 infinity binary LC system operating the electrospray source in positive ionization mode. All samples were chromatographed on an Agilent ZORBAX Bond-RP C18 column (4.6 mm × 150 mm; 3.5 μm particle size) at 50 °C column temperature. The solvent system consisted of 10 mM ammonium acetate (A) and methanol (B). The HPLC gradient with a flow rate of 1 ml min−1 starts at 5% B, holds for 2 min and then increases over 12 min to 100% B. Identification of CDNs and cAAG was performed by targeted mass analysis for exact masses and formulae for all possible CDNs and cAAG using Profinder software (version B.06.00 build 6.0.600.6, Agilent). NMR. All NMR experiments were conducted on a Varian 400 MR spectrometer (9.4 T, 400 MHz). Samples were prepared by resuspending evaporated nucleotide samples in 500 μl D2O supplemented with 5 mM TMSp (3-(trimethylsilyl)propionic-2,2,3,3-d4) at 27 °C. Data were processed and figures were generated using VnmrJ software (version 2.2C). 1H and 31P chemical shifts are reported in parts per million (p.p.m.).
Phospholipase assay. Patatin-like lipases were assayed as previously described38. In brief, CapV and CapE were produced recombinantly and catalytic activity was measured using the EnzChek Phospholipase A1 Assay Kit (Invitrogen) according to the manufacturer’s instructions. Phospholipases (250 nM) were incubated with 2.5, 0.25 or 0.025 μM CDN, c-di-AMP (Invitrogen), 3′,5′-cGAMP (Invitrogen) and c-di-GMP (Biolabs) were purchased as chemical standards, cUMP—AMP was purified as described above. Assays were monitored fluorometrically (excitation, 460 nm emission, 515 nm) for 60 min at approximately 90-s intervals at room temperature using a Biotek Synergy plate reader. Slope of each reaction in the linear range was used to calculate activity (linear regression/straight line analysis, GraphPad Prism 7.0c). A PLAA standard curve from 20 to 0.02 U was used to interpolate phospholipase activity. Emission was monitored at a gain of 100 and/or 50 in order to extend the linear range of the assay.

Crystallization and structure determination. CdnE homologues were crystallized in in situ form or in complex with nucleotide substrates at 18°C using hanging drop vapour diffusion. Purified RmCdnE and CdnE were diluted on ice to 7–10 mg/ml and used immediately to set trays. Alternatively, co-complex crystals were grown by first incubating RmCdnE and CdnE in the presence of a total concentration of combined nucleotides of approximately 10 mM and 10.5 mM MgCl2, on ice for 30 min. RECON–cAAG co-complex crystals were grown by pre-incubating approximately 10 mg ml−1 RECON (K68A, K70A) with 1 mM of purified EcCdnD02 product. Following incubation, 2 μl hanging drops were set at a ratio of 1:1 or 1:20:0.8 (protein/reservoir) over 350 μl of reservoir in Easy-Xtal 15–Well Trays (Qagen). Optimized crystallization conditions were as follows: apo RmCdnE, 100 mM Tris–HCl pH 7.5, 10–20% ethanol; RmCdnE–Ap CCP– Upnpp, 0.24 M sodium malonate, 24% PEG-3350; apo EcCdnE, 150 mM pH 7.0, 100 mM HEPES-KOH pH 7.5, 16% PEG-5000; MME; CdnE–CdnE G–GTP– AcPcp, 100 mM tri-sodium citrate pH 6.4, 10% PEG-3550; EcCdnE–pppApA, 100 mM tri-sodium citrate pH 7.0, 8% PEG-3350; RECON–cAAG, 0.1 M NaOAc, 1.0 M LiCl, 30% PEG-6000. Crystals grew in 3–30 days, and all crystals were harvested using reservoir solution supplemented with 10–25% ethylene glycol using a nylon loop, except apo RmCdnE crystals, which were harvested using NVH oil. X-ray diffraction data were collected at the Advanced Light Source (beamlines 8.2.1 and 5.0.1) and the Advanced Photon Source (beamlines 24-ID-C and 24-ID-E).

Data were processed with XDS and AIMLESS39 using the SRRL autosol script (A. Gonzalez, Stanford SRRL). Experimental phase information for RmCdnE was determined using data collected from crystals grown with selenomethionine-substituted protein as previously described40. Four sites were identified with HySS in PHENIX40, and an initial map was calculated using SOLVE/RESOLVE41. Model building was completed in Coot42 before refinement in PHENIX. Following model completion, the apo RmCdnE structure was used for molecular replacement to determine the nucleotide-bound structures. RmCdnE models were not sufficient to phase EcCdnE data, but a minimal core RmCdnE active-site model was able to successfully substructure and assist with data collected from a native crystal using sulfur single-wavelength anomalous dispersion at a minimal accessible wavelength (approximately 7,235 eV). The 16 heavy atoms were identified in HySS that correspond to 12 sulfur and 4 phosphate sites.

Western blot analysis. CD-Ntase in-cell expression levels were verified by western blot of lysed cells. Confluent HEK293T cells were seeded 24 h before transfection at a dilution of 1:4 in a six-well dish. Cells were transfected with 2 μg of plasmid using Lipofectamine 2000 in 96-well format with: a control plasmid constitutively expressing Renilla luciferase (2 ng pRL-TK), a reporter plasmid expressing interferon-β inducible firefly luciferase (20 ng), a plasmid expressing M. musculus STING (5 ng) and a fivefold dilution series of pcDNA4-based plasmids expressing a nucleotidyltransferase (1.2, 6, 30, and 150 ng). 2′,3′-cGAMP was produced with human cGAS, 3′,5′-cGAMP was produced with V. cholerae DncV, cyclic di-AMP (CDA) was produced with Bacillus subtilis DisA, cyclic di-GMP (cGg) was produced with P. aeruginosa WspR. Luciferase production was quantified after 24 h and firefly luciferase was normalized to Renilla, which was then used to normalize to nucleotidyltransferase expression.

Cell lines. HEK293T cells were used to measure a reporter plasmid expressing interferon-β inducible firefly luciferase. Cell lines were originally provided by the ATCC, no methods were used for authentication, and cell lines were not tested for mycoplasma.

Bioinformatics and tree construction. To bioinformatically map CD-Ntase-like enzymes in bacteria, we extended a previous analysis43 that combined iterative BLAST analysis, secondary structure predictions and hidden Markov models to collect DncV-like proteins and their genomic context. We identified homologues of each of these 1,300 identified proteins by BLAST analysis of the NCBI non-redundant protein database, then combined these datasets to identify more than 5,600 CD-Ntase-like genes. The dataset was then manually curated (Genious Software) according to shared CD-Ntase structural and active-site features. Bacterial genomes and sequences were aligned using MAFFT FFT-NS-2 algorithm version 7.3884, a BLOSUM62 scoring matrix, an open offset penalty of 2 and an offset value of 0.123. Proteins with large truncations or lacking the essential DNA polynucleotide (β-like nucleotidyltransferase residues) were removed. The tree was generated from the MAFFT alignment using a Jukes–Cantor genetic distance model, neighbour-joining method, no outgroup was used, and the tree was visualized using the Geneious Software.
than 10 CD-NTases that share over 24.5% identity to the sequence preceding each in the alignment. For clarity, 14 poorly aligned CD-NTases were excluded from the tree and are indicated in Supplementary Table 2. The full dataset organized by order from the alignment and containing pairwise comparison of protein identity to each preceding gene is available as Supplementary Table 2 and as Source Data for Fig. 4a. For details on identifying CD-NTases within an organism of interest or organisms encoding a CD-NTase of interest, see Supplementary Discussion.

Each sequence was identified from the nonredundant database of protein sequences and, at times, represents identical proteins translated from genes found in multiple bacteria. For this reason, additional metadata were extracted for each sequence from the NCBI identical protein groups (IPG) database. The number of ‘protein accessions’ in IPG was used as a surrogate quantification of the number of isolated bacterial genomes that contain each CD-NTase. At the time of access (3 February 2018), 5,686 nonredundant CD-NTases sequences were identified representing a total of 16,717 genomes. At that time, 130,155 bacterial genomes had been deposited in the NCBI Genome database, leading to the crude approxima-
tion of 12.8% of genomes containing CD-NTase genes. As some of these genomes may have more than one CD-NTase and the IPG database can overestimate the number of genomes encoding a given protein, we have estimated that more than 10% of bacterial genomes sequenced encode CD-NTases. Taxonomic analysis was performed using metadata associated with each nonredundant CD-NTase record in NCBI and when multiple bacteria were represented by one identical sequence the highest common taxonomical group was used. IPG and taxonomic data can also be found in Supplementary Table 2.

Type CD-NTase enzymes were manually selected from clusters based on the relevance of the organism from which they were isolated (that is, bacterial or plant pathogen/commensal organism), their predicted aptness for in vitro expression (thermophilic organisms or isolates of E. coli), the similarity of their operon to the DncV/Cdne operons, and the number of identical protein sequences represented by each unique sequence. CD-NTase001 was selected as and additional control and is encoded by dncV1_ccl in ECOR318.

CD-NTase screen. Each type CD-NTase gene was codon-optimized for E. coli, synthesized (IDT), cloned as an N-terminal 6 × His-MBP fusion and the protein was purified from a 25-mL culture. E. coli growth, protein induction and bacterial disruption were performed as described above. Lysates were clarified by centrifugation and Ni-NTA affinity purification was performed as described above with gravity columns replaced by spin columns, centrifuged at 100 g. Buffer exchange of eluted proteins was performed by concentrating the eluate using a 0.5-ml 10-kDa cut-off spin column (Amicon) followed by dilution with storage buffer and reconstitu-
tion 3 × (final iodamide concentration, approximately 0.3 mM). Proteins were analysed for nucleotide synthesis fresh and flash-frozen for storage at −80 °C. For biochemical screens, ATP/CTP/GTP/UTP were used at 25 μM each and incubated overnight with the reaction conditions indicated using methods described above. Type IV proteinase A (approximately 1 μg) was added to the reaction and the same volume assessed by SDS–PAGE followed by Coomasie staining, shown in Extended Data Fig. 6g.

Figure 4d was manually constructed based on known TLC migration patterns that guided CDN identification in each sample. The quantity of ions detected by mass spectrometry relative to other CD-NTases was used to determine whether products were a major or minor constituent. On PEI–cellulose, cyclic dipyrinides migrate similarly, cyclic purine–pyrimidine hybrids migrate similarly and cyclic dipyrinides migrate similarly; on silica c-di-AMP migrates uniquely, cyclic UMP–AMP and cGMP–UMP migrate similarly, c-di-GMP and c-di-UMP migrate similarly, and cGMP–UMP and cCMP–UMP migrate similarly; these cannot be distinguished.

Coexpression of CD-NTases and effectors in E. coli. CD-NTases chosen for in-depth analysis were cloned into an arabinose-inducible, chloramphenicol-resistant pBAD33 plasmid. Putative CD-NTase effector genes were selected based on proximity, if they were classified as involved in biological conflict22 and based on analogous operon architecture to known effector phospholipases. Effector genes were codon-optimized for E. coli and cloned into pETSUMO22, a carboxicillin-resistant vector that is IPTG inducible in BL21-DE3 E. coli (ThermoFisher Scientific). Three pairs of vectors were assessed for each CD-NTase–effector pair: (1) cognate CD-NTase and effector, (2) CD-NTase and GFP, (3) mCherry and effector. Fluorescent proteins were used as negative controls. Vectors were co-transformed into electrocompetent BL21-DE3 E. coli, selected with both rele-

fectors and using the arabinose-inducible, chloramphenicol-resistant pBAD33 plasmid. Putative CD-NTase effector genes were selected based on proximity, if they were classified as involved in biological conflict22 and based on analogous operon architecture to known effector phospholipases. Effector genes were codon-optimized for E. coli and cloned into pETSUMO22, a carboxicillin-resistant vector that is IPTG inducible in BL21-DE3 E. coli (ThermoFisher Scientific). Three pairs of vectors were assessed for each CD-NTase–effector pair: (1) cognate CD-NTase and effector, (2) CD-NTase and GFP, (3) mCherry and effector. Fluorescent proteins were used as negative controls. Vectors were co-transformed into electrocompetent BL21-DE3 E. coli, selected with both rele-

methods to regulate growth, antibiotic resistance and osmoregulation. Mol. Microbiol. 104, 212–233 (2017).

Kranzusch, P. J. & Whelan, S. P. J. Arenavirus Z protein controls viral RNA synthesis by locking a polymerase–promoter complex. Proc. Natl Acad. Sci. USA 108, 19743–19748 (2011).

Zhou, W. et al. Structure of the human cGAS–DNA complex reveals enhanced control of immune surveillance. Cell 174, 300–311 (2018).

Kulasakara, H. et al. Analysis of Pseudomonas aeruginosa diguanylate cyclases and phosphodiesterases reveals a role for bis(3’-5’)-cyclic-GMP in virulence. Proc. Natl Acad. Sci. USA 103, 2839–2844 (2006).

Schubert, S., Durfe, S., Sorsa, J. & Heesemann, J. A novel integrative and conjugative element (ICE) of Escherichia coli: the putative progenitor of the Yersinia high-pathogenicity island. Mol. Microbiol. 51, 837–848 (2004).

Kranzusch, P. J. et al. Ancient origin of cGAS–STING reveals mechanism of universal 2’3’ cGAMP signaling. Mol. Cell 59, 891–903 (2015).

Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. 177, 4121–4130 (1995).

Reverter, D. & Lima, C. D. Structural basis for SENP2 protease interactions with SUMO precurors and conjugated substrates. Nat. Struct. Mol. Biol. 13, 1060–1068 (2006).

Stetson, D. B. & Medzhitov, R. Recognition of cytosolic DNA activates an IRF3-dependent innate immune response. Immunity 24, 93–100 (2006).

Surekai, K. et al. The cyclic dinucleotide c-di-AMP is an allosteric regulator of metabolic enzyme function. Cell 158, 1389–1401 (2014).

Gaspar, A. H. & Machner, M. P. Vipd is a Rab5–activated phospholipase A1 that protects Legionella pneumophila from endothelial fusion. Proc. Natl Acad. Sci. USA 111, 4560–4565 (2014).

Kabosh, W. D. ACS. Acta Crystallogr. D 66, 125–132 (2010).

Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D 66, 213–221 (2010).

Terwilliger, T. C. Reciprocal-space solvent flattening. Acta Crystallogr. D 55, 1875–1879 (1999).

Emssly, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D 60, 2126–2132 (2004).

Holm, L. & Lasko, L. M. Dalii server update. Nuclic Acids Res. 43, W351–W355 (2015).

Lohofener, J. et al. The activation mechanism of 2’-5’ oligoadenylate synthetase gives new insights into OAS/cGAS triggers of innate immunity. Structure 23, 851–862 (2015).

Yang, Q., Naumc, L. W. M., Martin, G., Keller, W. & Dubikl, S. Crystal structure of human poly(A) polymerase gamma reveals a conserved catalytic core for canonical poly(A) polymers. J. Mol. Biol. 426, 43–50 (2014).

Kuhn, C. D., Wiusz, J. E., Zheng, Y., Beal, P. A. & Joshua-Tor, L. On-enzyme refolding permits small RNA and tRNA surveillance by the CCA-adding enzyme. Cell 160, 644–658 (2015).

Freudenthal, B. D., Beard, W. A., Shock, D. D. & Wilson, S. H. Observing a DNA polymerase choose right from wrong. Cell 154, 157–168 (2013).

Moon, A. F., Gosavi, R. A., Kunkel, T. A., Pedersen, L. G. & Bebenek, K. Creative template-dependent synthesis by human polymerase mu. Proc. Natl Acad. Sci. USA 112, E4530–E4536 (2015).

Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780 (2013).

Kato, K., Ishi, R., Hirano, S., Ishitani, R. & Nureki, O. Structural basis for the catalytic mechanism of DncV, bacterial homolog of cyclic GMP–AMP synthase. Structure 23, 843–850 (2015).

Gao, P. et al. 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 (2013).
Extended Data Fig. 1 | Detailed characterization of CdnE, a cUMP–AMP synthase.

**a, b**, Titration of reaction buffer pH in steps of 0.2 pH units. Recombinant DncV or CdnE was incubated with α32P-radiolabelled ATP, CTP, GTP and UTP at varying pH and the reactions were treated with alkaline phosphatase and visualized by PEI–cellulose TLC. CdnE activity is optimal at a pH of approximately 9.4, and this reaction condition was used in further experiments. Data are representative of two independent experiments.

**c**, PEI–cellulose TLC of products after incubation of the indicated enzyme, wild-type CdnE or active site mutant CdnE with α32P-radiolabelled ATP, CTP, GTP and UTP as in Fig. 1b. Mutations that ablate the CdnE Mg2+–coordinating, active-site residues eliminate all detectable activity. Data are representative of three independent experiments.

**d**, Nuclease P1 sensitivity of CDN products. Nuclease P1 specifically cleaves 3′,3′-cGAMP, whereas only one bond of 2′,3′-cGAMP is susceptible to digestion, producing the linear G[2′–5′]pA product. Data are representative of three independent experiments.

**e**, Workflow of nucleotide production for mass spectrometry and NMR analysis.

**f**, Anion-exchange chromatography of a CdnE reaction with ATP and UTP, eluted with a gradient of buffer B (2 M ammonium acetate) by FPLC. Individual fractions were concentrated before pooling for further analysis.

**g**, Anion-exchange chromatography (IEX) fractions from f were separated by silica TLC, visualized by ultraviolet-light shadowing and compared to a radiolabelled reaction to confirm the appropriate absorbance at 254 nm (A254) peak. Fractions were pooled and concentrated before mass spectrometry and NMR analysis.

**h, i, j**, 3′,3′-Cyclic uridine monophosphate–adenosine monophosphate proton-NMR spectrum (h) and associated magnified spectrum (i). 31P{1H} NMR (162 MHz): δP −1.59 (s, 1P), −1.65 (s, 1P).

**k, l**, 3′,3′-Cyclic uridine monophosphate–adenosine monophosphate proton–NMR spectrum (k) and associated magnified spectrum (l).
Extended Data Fig. 2 | Order of the DncV, cGAS and CdnE reactions. 

a–c, Incubation of CD-NTase enzymes with non-hydrolysable nucleotides traps reaction intermediates and identifies the order of the reaction. Left, PEI–cellulose TLC analysis of reactions as in Fig. 1b, in which individual NTPs have been replaced by their non-hydrolysable nucleotide analogues. Exposed γ-phosphates are labelled in parentheses to indicate that these are removed by phosphatase treatment before analysis. Data are representative of three independent experiments. Right, previously determined reaction mechanisms (DncV and cGAS)\textsuperscript{10,51} and proposed reaction mechanism for CdnE.
Extended Data Fig. 3 | Detailed analysis of RmCdnE. a, A thermophilic homologue of CdnE from R. marinus (RmCdnE) also synthesizes CUMP–AMP. Recombinant proteins were incubated with α-[32P]-radiolabelled NTPs as indicated at either 37 °C (CdnE) or 70 °C (RmCdnE) and the reactions were visualized by PEI–cellulose TLC as in Fig. 1b. Data are representative of two independent experiments. b, Active site of RmCdnE superimposed with structures of cGAS (6CTA25) and DncV (4TYU25). The analogous position to Asn166 was mutated in CdnE to a serine and that protein, CdnEN166S, was characterized further. Reactions, which were analysed as in Fig. 1b, demonstrate that CdnEN166S loses pyrimidine specificity. Data are representative of two independent experiments. c, Structure-based sequence alignment of CD-NTases, annotated with secondary structure features of RmCdnE and human cGAS (6CTA25). Mg2+-coordinating active-site residues are highlighted in red; blue arrows, cyan/pink annotations and purple highlights represent secondary structure; and the analogous residues to RmCdnE Asn166 are highlighted in orange.
Extended Data Fig. 4  See next page for caption.
Extended Data Fig. 4 | Detailed structural analysis of EmCdnE.

a, Sequence alignment of CdnE homologues in Fig. 2c, annotated with RmCdnE secondary structure features. Mg$^{2+}$-coordinating active-site residues are highlighted in red and the analogous residues to RmCdnE Asn166 are highlighted in orange. Versinia enterocolitica (WP_050915017); P. aeruginosa (WP_096075289); Xanthomonas arboricola (WP_104644370); Xenorhabdus nematophila (WP_010848498); Bordetella parapertussis (WP_015040391); Burkholderia cepacia complex (WP_006482377); R. marinus (RmCdnE, WP_014072508); L. pneumophila (WP_042646516); Mycobacterium avium (WP_062886322); E. meningoseptica (EmCdnE, WP_016200549); Staphylococcus aureus (WP_031901603); Enterococcus faecalis (WP_050492554); Bacteroides thetaiotaomicron (WP_062695386). b, Biochemical deconvolution of EmCdnE, which has a natural serine substitution at the Asn166 analogous site. Recombinant protein was incubated with NTPs as indicated and reactions were visualized as in Fig. 1b. Data are representative of three independent experiments. c, Reactions of EmCdnE incubated with α$^{32}$P-radiolabelled NTPs and non-hydrolysable nucleotide analogues as indicated and visualized as in Fig. 1b. Data are representative of three independent experiments. d, Anion exchange chromatography of an EmCdnE reaction with ATP and GTP, eluted with a gradient of buffer B (2 M ammonium acetate) by FPLC. Individual fractions were concentrated before pooling for further analysis. e, Anion exchange chromatography fractions from d were separated by silica TLC, visualized by ultraviolet-light shadowing and compared to a radiolabelled reaction to confirm the appropriate peak. Fractions were pooled and concentrated before mass spectrometry analysis. Mass spectrometry confirmed synthesis of products with masses corresponding to c-di-AMP, cGAMP and c-di-GMP. f, Crystal structure of EmCdnE in complex with GTP and non-hydrolysable ATP capturing the ‘first state’ structure before NTP hydrolysis. Mg$^{2+}$ ions are omitted for clarity. g, Magnified cut-away of the active site of the complex shown in f, confirming the position of a serine at the analogous site to RmCdnE Asn166. Nucleotide and metal $2F_o - F_c$ electron densities are contoured at 1σ. h, Magnified cut-away of the active site of the EmCdnE–pppApA structure, capturing the ‘second state’ after the first reaction has occurred to form a linear intermediate, but before CDN formation. Nucleotide and metal $2F_o - F_c$ electron densities are contoured at 1σ. g, h, Mg$^{2+}$ ions are shown in green. i, Biochemical deconvolution of mutant EmCdnE reverted to the ancestral asparagine at the Asn166 analogous site. This mutant loses preference for producing cyclic dipurine molecules and instead produces more pyrimidine-containing CDN products. Reactions were visualized as in Fig. 1b. Data are representative of two independent experiments.
Extended Data Fig. 5 | cUMP–AMP recognition helps to define innate immune receptor specificity. a, e, Quantification of nucleotide interactions with the host receptors STING (a) or RECON (e), measured using a radiolabelled nucleotide bound to a concentration gradient of host protein, separated in a native PAGE gel shift (0, 4, 20, 100 μM protein). Quantification of the gels shown in b and f. Data are representative of n = 2 independent experiments. b, f, Native PAGE gel shift analysis of STING (b) or RECON (f) complex formation with indicated radiolabelled CDNs. Proteins are titrated at 0 (−), 4, 20 and 100 μM. STING readily binds all cyclic dipurine species, but does not form a high-affinity complex with cUMP–AMP. RECON readily binds all 3′,3′-CDN species that contain at least one adenine base, including cUMP–AMP. Data are representative of two independent experiments. c, In-cell STING reporter assay. Induction of an IFNβ reporter in HEK293T cells transfected with a concentration gradient of plasmid-overexpressing enzymes as indicated. DncV and CdnE were expressed with N-terminal MBP tags and IFNβ reporter induction was compared as fold over empty vector, shown as (−). Data are mean ± s.e.m. for n = 3 technical replicates and are representative of two independent experiments. d, Western blot of MBP-tagged DncV and CdnE expressed from plasmids analysed in c to validate in vivo expression. Data are representative of two independent experiments. Gel source data are available in Supplementary Fig. 1. g, Gel shift analysis as in f, with protein titration to measure the relative affinity of the RECON–cUMP–AMP interaction. Protein concentrations listed below. Data are representative of 2 independent experiments.
Extended Data Fig. 6 | A biochemical screen of CD-NTases from diverse bacterial genera. a, Chart of the number of bacterial genomes \((n = 16,717)\) that have CD-NTases from clusters in Fig. 4a. See also Supplementary Table 2. b, Taxa of genome-sequenced bacteria isolated with unique CD-NTase genes, phyla are indicated in bold; Proteobacteria and Firmicutes are further divided by order and visualized by shades of colour. c–f, Type CD-NTases interrogated for product synthesis. Purified proteins were incubated with \(\alpha^{32}\)P-radiolabelled NTPs under different reaction conditions (indicated pH and divalent cation) and reaction products were visualized by either PEI–cellulose or silica TLC as in Fig. 1b and Fig. 4c, respectively. g, CD-NTase expression level and purity. Coomassie-stained SDS–PAGE analysis of purified CD-NTase enzymes used in each reaction. Data are representative of two independent experiments.
Extended Data Fig. 7 | Detailed biochemical analysis of \( \text{LpCdnE02} \).

**a**, Biochemical deconvolution of \( \text{LpCdnE02} \) (CD-NTase057), analysed as in Fig. 1c, demonstrates specific synthesis of cyclic dipyrimidine products. Data are representative of three independent experiments.

**b**, Nuclease sensitivity of the \( \text{LpCdnE02} \) product, as described in Extended Data Fig. 1d. Data are representative of three independent experiments.

**c**, Incubation of \( \text{LpCdnE02} \) with non-hydrolysable nucleotides, as described in Extended Data Fig. 2. Non-hydrolysable UTP completely blocks the reaction, indicating the first step requires attack of the \( \alpha P \) from UTP. However, the product formed when non-hydrolysable CTP is present cannot be distinguished from c-di-UMP in this assay, and it is unclear whether the reaction proceeds through a pppCpU reaction intermediate. Data are representative of three independent experiments.

**d**, Anion exchange chromatography of a \( \text{LpCdnE02} \) reaction with UTP and CTP, eluted with a gradient of buffer B (2 M ammonium acetate) by FPLC. Individual fractions were concentrated before pooling for further analysis.

**e**, Anion exchange chromatography fractions from **d** were separated by silica TLC, visualized by ultraviolet-light shadowing and compared to a radiolabelled reaction to confirm the appropriate peak. Fractions were pooled and concentrated before mass spectrometry analysis.

**f**, Mass spectrometry confirmed synthesis of c-di-UMP as the major product of \( \text{LpCdnE02} \).

**g**, Mass spectrometry confirmed synthesis of cCMP–UMP as a minor product of \( \text{LpCdnE02} \).
Extended Data Fig. 8 | CD-NTases are encoded in conserved, poorly understood operons on mobile genetic elements. a, Operon structure for CD-NTases selected for in-depth characterization showing the conserved protein domains in CD-NTase adjacent genes (see Fig. 4a–c). Conserved operons were identified previously and operons are vertically organized by similarity to one another22. Where found, linked genes demonstrating that CD-NTases are encoded on mobile genetic elements are indicated.
b, CD-NTases and their adjacently encoded 'effector' proteins were coexpressed in *E. coli* and bacterial colony formation was quantified by spot-dilution analysis. CD-NTases were inducibly expressed from a chloramphenicol-resistance (Cm<sup>R</sup>) vector and effectors were inducibly expressed from a carbenicillin-resistance (Carb<sup>R</sup>) vector. Bacteria that express cognate CD-NTase–effector plasmids or control plasmids were plated on medium containing both inducers and incubated for 24 h at 37 °C. Data were not determined (N.D.) for CD-NTase036, because the effector was toxic to *E. coli* under non-inducing conditions. Data are the mean ± s.e.m. of three independent experiments.
c, Spot-dilution analysis of bacteria expressing the indicated cognate CD-NTase–effector pairs, quantified as in b. Colony morphology indicates a potential interaction for some combinations; however, it is unclear how specific or meaningful this may be. Data are representative of three independent experiments.
Extended Data Fig. 9 | See next page for caption.
Extended Data Fig. 9 | Detailed biochemical analysis of EcCdnD02.

a, Titration of reaction buffer pH in steps of 0.2 pH units. Recombinant EcCdnD02 was incubated with α32P-radiolabelled NTPs at varying pH and the reactions were analysed and visualized by PEI–cellulose or silica TLC as in Fig. 1b and Fig. 4c, respectively. Silica TLC identified two products, denoted the major (blue triangle) and minor (red triangle) product. Quantification of TLC spots is shown at the bottom. Data are representative of two independent experiments. b, Biochemical deconvolution of EcCdnD02, recombinant protein was incubated with NTPs as indicated and analysed by TLC as in a. Data are representative of three independent experiments. c, Nuclease digestion of the EcCdnD02 product. Conventional nuclease digestion includes addition of a phosphatase, in this experiment reactions were first treated with Antarctic phosphatase to remove remaining NTPs and were then heat-inactivated. Next, reactions were untreated, treated with nuclease P1 (specific for 3′–5′-phosphodiester bonds) only or treated with nuclease P1 and phosphatase to remove exposed phosphate groups. 3′,3′-cGAMP (DncV) and EcCdnD02 product are digested into AMP and GMP constituents, which are phosphatase-sensitive. cAMP (CyaA) is insensitive to P1 digestion and cyclic monophosphates are phosphatase-resistant. These data demonstrate that the EcCdnD02 product does not contain a cyclic monophosphate. Data are representative of three independent experiments. d, Incubation of EcCdnD02 with non-hydrolysable nucleotides, as described in Extended Data Fig. 2. Non-hydrolysable ATP completely blocks the reaction, indicating the first step requires attack of the αP from ATP. However, when non-hydrolysable GTP is present the possible intermediates (pp(c)pGpA, pp(c)pGpApA, or pppApA) cannot be distinguished in this assay and it is unclear how the reaction proceeds. Silica TLC is not suited for analysing non-hydrolysable nucleotides, because they do not migrate beyond the origin. Data are representative of three independent experiments. e, Anion exchange chromatography of a EcCdnD02 reaction with ATP and GTP, eluted with a gradient of buffer B (2 M ammonium acetate) by FPLC. Individual fractions were concentrated before pooling for further analysis. f, g, 3′,3′,3′-Tricyclic adenosine monophosphate–adenosine monophosphate–guanosine monophosphate phosphate-NMR spectrum (f) and associated magnified spectrum (g). 31P{1H} NMR (162 MHz): δP −0.65 (s, 1P), −0.70 (s, 1P), −0.75 (s, 1P). h–j, 3′,3′,3′-Tricyclic adenosine monophosphate–adenosine monophosphate–guanosine monophosphate proton-NMR spectrum (h) and associated magnified spectra (i, j). 1H NMR (400 MHz): δH 8.43 (s, 1H), 8.39 (s, 1H), 8.19 (s, 1H), 8.12 (s, 1H), 8.01 (s, 1H), 6.15 (d, J = 7.0 Hz, 1H), 6.12 (d, J = 7.0 Hz, 1H), 5.92 (d, J = 7.5 Hz, 1H), 5.00–4.78 (m, 6H), 4.69–4.58 (m, 3H), 4.3–4.2 (m, 6H).
Extended Data Fig. 10 | Structural analysis of cAAG inhibition of RECON. 
a, cAAG interactions with STING or RECON, radiolabelled nucleotides incubated with a concentration gradient of each protein, separated in a native PAGE gel shift (0, 4, 20, 100 μM protein). Data are representative of two independent experiments. 
b, Co-crystal structure of the RECON–cAAG complex shown as cartoon (left) and surface (right). 
c, Overlay and orientation of RECON ligands cAAG, c-di-AMP (PDB 5UXF), co-substrate NAD (PDB 3LN3) demonstrate three individual binding pockets. 
d, Schematic representation of residues from RECON that interact with cAAG. Green dotted lines indicate hydrogen bonds, grey dotted lines indicate hydrophobic interactions. 
e, Magnified cutaways of individual RECON binding pockets as in d. 
f, Mammalian innate-immune sensors recognize CD-NTase products with overlapping specificities. 2′,3′-cGAMP and c-di-GMP are detected by STING; 3′,3′-cGAMP and c-di-AMP are detected by both STING and RECON; and cUMP–AMP and cAAG are detected by RECON.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐ Only common tests should be described solely by name; describe more complex techniques in the Methods section.
☐ A description of all covariates tested
☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
☐ Give P values as exact values whenever suitable.
☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐ Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
☐ Clearly defined error bars
☐ State explicitly what error bars represent (e.g. SD, SE, CI)

Software and code

Policy information about availability of computer code

Data collection

- Profinder version B.06.00 build 6.0.606.0 (Agilent), VnmrJ version 2.2C NMR software

Data analysis

- Phenix 1.13-2998, Coot 0.8.9, PyMOL 1.7.4.4, Geneious v11, Prism 7, VnmrJ, Profinder version B.06.00 build 6.0.606.0 (Agilent)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Atomic coordinates and structure factors for the reported crystal structures are deposited at the Protein Data Bank (PDB) under accession codes 6E0K, 6E0L, 6E0M, 6E0N, 6E0O, and 6M7K.
Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | N/A |
|-------------|-----|
| Data exclusions | N/A |
| Replication | All experiments were performed with independent replicates as described in the figure legends. |
| Randomization | N/A |
| Blinding | N/A |

Reporting for specific materials, systems and methods

| Materials & experimental systems | Methods |
|---------------------------------|---------|
| n/a | n/a |
| □ × Unique biological materials | □ × ChIP-seq |
| □ × Antibodies | □ × Flow cytometry |
| □ × Eukaryotic cell lines | □ × MRI-based neuroimaging |
| □ × Palaeontology | |
| □ × Animals and other organisms | |
| □ × Human research participants | |

Unique biological materials

Policy information about availability of materials

Obtaining unique materials: All unique CD-NTase enzyme genes and described plasmids are available from the authors.

Antibodies

Antibodies used:

- 1:5,000 Rabbit anti-MBP (Millipore Cat# AB3596, RRID:AB_91531);
- 1:10,000 Mouse anti-Tubulin (Millipore Cat# MABT205, RRID:AB_11204167);
- 1:10,000 IRDye 680RD Goat anti-Rabbit IgG (LI-COR Biosciences Cat# 925-68071, RRID:AB_2721181);
- 1:10,000 IRDye 800CW Goat anti-Mouse IgG (LI-COR Biosciences Cat# 925-32210, RRID:AB_2687825)

Validation:

Appropriate negative controls and target validation as stated by the manufacturers.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s):

- HEK293T were originally obtained from the ATCC.

Authentication:

No methods were used for authentication.

Mycoplasma contamination:

The cell lines were not tested for mycoplasma.
No commonly misidentified cell lines were used in this study.