Published online 9 April 2021

Nucleic Acids Research, 2021, Vol. 49, No. 8 ***4725–4737***
doi: 10.1093/nar/gkab165

Crystal structure and functional implication of a bacterial cyclic AMP–AMP–GMP synthetase

Tzu-Ping Ko [1,†], Yu-Chuan Wang [2,†], Chia-Ling Tsai [2,†], Chia-Shin Yang [2], Mei-Hui Hou [2] and Yeh Chen [2,3,4,*]

[1]Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan, [2]Institute of New Drug Development, China Medical University, Taichung 406, Taiwan, [3]Research Center for Cancer Biology, China Medical University, Taichung 406, Taiwan and [4]New Drug Development Center, China Medical University, Taichung 406, Taiwan

Received November 09, 2020; Revised February 24, 2021; Editorial Decision February 25, 2021; Accepted April 07, 2021

ABSTRACT

Mammalian cyclic GMP-AMP synthase (cGAS) and its homologue dinucleotide cyclase in *Vibrio cholerae* (*Vc*DncV) produce cyclic dinucleotides (CDNs) that participate in the defense against viral infection. Recently, scores of new cGAS/DncV-like nucleotidyltransferases (CD-NTases) were discovered, which produce various CDNs and cyclic trinucleotides (CTNs) as second messengers. Here, we present the crystal structures of *Ec*CdnD, a CD-NTase from *Enterobacter cloacae* that produces cyclic AMP-AMP-GMP, in its apo-form and in complex with ATP, ADP and AMPcPP, an ATP analogue. Despite the similar overall architecture, the protein shows significant structural variations from other CD-NTases. Adjacent to the donor substrate, another nucleotide is bound to the acceptor binding site by a non-productive mode. Isothermal titration calorimetry results also suggest the presence of two ATP binding sites. GTP alone does not bind to *Ec*CdnD, which however binds to pppApG, a possible intermediate. The enzyme is active on ATP or a mixture of ATP and GTP, and the best metal cofactor is Mg^{2+}. The conserved residues Asp69 and Asp71 are essential for catalysis, as indicated by the loss of activity in the mutants. Based on structural analysis and comparison with *Vc*DncV and RNA polymerase, a tentative catalytic pathway for the CTN-producing *Ec*CdnD is proposed.

INTRODUCTION

Cyclic nucleotides have long been known to serve as second messengers that regulate cellular activities in response to environmental variations. For example, 3′,5′-cyclic adenosine monophosphate (cAMP; Figure 1A) promotes glycogen breakdown in human (1). In bacteria, cAMP activates genes for catabolic pathways (2). Many bacterial cell processes are regulated also by 3′,5′-linked cyclic dinucleotides (CDNs) such as c-di-GMP and c-di-AMP (3,4). In mammals, 2′,3′-cyclic GMP-AMP (2′,3′-cGAMP; Figure 1B) with the guanosine C2′ forming a non-canonical phosphodiester bond to the adenosine C5′, and 2′,5′-oligoadenylate with similar linkages, are synthesized by cGAS and OAS upon activation by double-strand DNA (dsDNA) and dsRNA (5,6). Innate immune responses are triggered by 2′,3′-cGAMP binding to STING (stimulator of interferon genes) and by 2′,5′-oligoadenylate activation of RNase L (7,8). STING, serving as a pattern recognition receptor (PRR), also recognizes the bacterial CDNs as pathogen associated molecular patterns (PAMPs), which include 3′,3′-cGAMP in addition to c-di-GMP and c-di-AMP (9). 3′,3′-cGAMP is synthesized by dinucleotide cyclase in *Vibrio cholerae* (*Vc*DncV), a cGAS homologue (10). It binds to a CDN-activated phospholipase (*Vc*CapV), which degrades the bacterial cell membrane (11). These enzymes all belong to the SMODS (second messenger oligonucleotide or dinucleotide synthase) family of proteins (12).

Recently, scores of newly discovered cGAS/DncV-like nucleotidyltransferases (CD-NTases) were shown to produce a variety of CDNs and cyclic trinucleotides (CTNs) (13). For example, CdnC from *Escherichia coli* (*Ec*CdnC) and CdnD from *Pseudomonas aeruginosa* (*Pa*CdnD) synthesize cyclic tri-AMP (cAAA) as a component of cyclic oligonucleotide-based anti-phage signaling system (CBASS) (14). cAAA binds to and activates NucC, a DNA endonuclease, which leads to cell death as a mechanism of abortive infection (15). The above DncV-CapV pair is also an example of CBASS (16). Another CTN-synthesizing CD-NTase from *Enterobacter cloacae*, *Ec*CdnD02, produces a mixed species of cyclic AMP-AMP-GMP (cAAG; Figure 1C) (13). In bacterial cell cAAG functions as a CBASS effector that binds to and activates Cap4 (CD-NTase-associated protein 4), a DNA endonuclease with SAVED domain (SMODS-associated to various effector domain) (17). In mammalian cell it also serves as a PAMP

*To whom correspondence should be addressed. Tel: +88 64 2205 3366 (Ext 6513); Fax: +88 64 2233 3641; Email: bluecrystalprotein@gmail.com
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

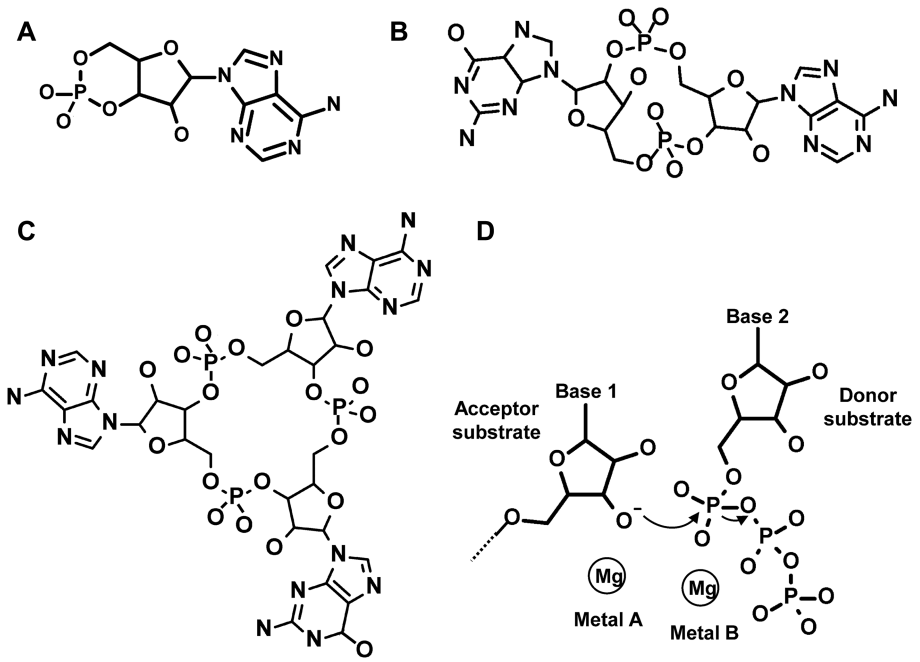

Figure 1. Cyclic nucleotides. The chemical structures of three cyclic nucleotides are presented as schematic diagrams in (**A**) cAMP, (**B**) 2′,3′-cGAMP and (**C**) cAAG. Hydrogen atoms and some double bonds are not shown. The shared mechanism of NTase-catalyzed reaction that forms CDN or CTN is depicted in (**D**). The Mg^{2+} ion bound to the acceptor ribose is designated metal A, and that bound to the donor triphosphate is metal B. The nucleophilic attack of O3′ (or O2′) at α-phosphate is indicated by an arrow.

to be detected by RECON (reductase controlling NF-κB), a PRR with aldo-keto reductase (AKR) activity for lipophilic metabolites (18). Binding of cAAG, as well as c-di-AMP and 3′,3′-cGAMP, to RECON directly blocks its active site and elicits STING-independent host responses (13).

In structure, bacterial CD-NTases are distinct from the c-di-GMP and c-di-AMP synthesizing enzymes, but are similar to mammalian cGAS and OAS (3). Each enzyme contains an N-terminal core domain, which is shared by non-templated and templated polymerases of the nucleotidyltransferase (NTase) superfamily, and a C-terminal helical domain that shows a higher degree of variation (13). Three conserved acidic residues in the central β-sheet of the NTase domain coordinate to two metal ions that are essential to the catalysis (19). In a shared mechanism for NTase, the ribose O3′ (or O2′) atom of the acceptor substrate is bound to one metal ion (termed metal A) and attacks the α-phosphate of the donor substrate, whose triphosphate associates with the other metal ion (termed metal B; Figure 1D) (20,21). Instead of base pairing, the substrate specificity of CD-NTase is determined by interactions with protein side chains from both N and C-terminal domains (22). The cage-like architecture may allow retention and reorientation of the intermediate molecule for the next step of catalysis (19). For example, the reaction intermediate of *Vc*DncV (actually an analogue) had the original acceptor nucleotide bound to the donor substrate binding site, and vice versa, facilitating the next reaction that would turn out a CDN product (10,22). CTN-synthesizing CD-NTases probably follow similar catalytic pathways.

On the other hand, CD-NTase signaling pathways are classified into Type I and Type II (19). Type I CD-NTases such as human cGAS and OAS1 are inactive without binding to the respective ligands of dsDNA and dsRNA. They require the ligand-induced conformational changes to become active, in which the N-terminal 'spine' helix connecting the NTase and helical domains is broken into two (5,6). The cAAA-synthesizing *Ec*CdnC and *Pa*CdnD are also Type I enzymes, which need activation by complex formation with HORMA (Hop1/Rev7/Mad2-like domain) as well as binding to DNA (14). In contrast, Type II CD-NTases such as *Vc*DncV and *Ec*CdnD02 are constitutively active (10,13). They do not need binding to other ligands to be catalytically functional. Along with several known Type I and Type II CDN-synthesizing enzymes, so far only two CTN-producing Type I CD-NTases have their crystal structures solved (14). To help elucidating the catalytic mechanism that turns out CTNs, we expressed, purified, crystallized and determined the structure of *Ec*CdnD02 (a CTN-producing Type II CD-NTase, denoted *Ec*CdnD hereafter) in apo and substrate-bound forms, including an active-site mutant. The enzyme was further characterized for affinity and activity by using different substrates and mutants.

MATERIALS AND METHODS

Expression and purification of EcCdnD

The full-length gene encoding the cAAG-producing *Ec*CdnD from *Enterobacter cloacae* UCI 50 (UniProt accession: P0DSP4; 381 amino acids) with *Escherichia coli* codon usage was synthesized chemically and subcloned into pET21 vector to be expressed as a C-terminal His_6-tagged protein. The plasmid was transformed into *E. coli* BL21 (DE3), which was grown in lysogeny broth (LB)

medium at 37°C until the OD_{600} reached 0.6–0.8. Overexpression of *Ec*CdnD was induced by addition of 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG), followed by incubation for 20 h at 15°C. The cells were harvested by centrifugation, resuspended in 50 mM Tris–HCl, pH 8.0, 500 mM NaCl, 10% glycerol, 1 mM tris(2-carboxyethyl) phosphine (TCEP), 1 mM phenylmethylsulfonyl fluoride (PMSF) (sonication buffer), and lysed by sonication on ice. After centrifugation at 28 000 g, 4°C, for 30 min, the supernatant was loaded onto a HisTrap FF column (GE Healthcare), washed with sonication buffer that contained 10 mM imidazole, and eluted with 20–200 mM imidazole gradient in the sonication buffer. Fractions containing the target protein were pooled and further purified by size-exclusion chromatography (HiLoad Superdex 200, Cytiva), using a buffer of 50 mM Tris–HCl, pH 8.0, 200 mM NaCl, 5% glycerol and 1 mM TCEP (gel filtration buffer). Substitutions of amino acid residues in *Ec*CdnD were performed by using the Quick Change site-directed mutagenesis kits with *Pfu*Ultra DNA polymerase and confirmed by DNA sequencing. The primers for making the mutants are listed in Supplementary Table S1. Expression and purification of the mutant proteins and Se-derivative followed the same protocols as above, except that the latter used an M9 minimal medium supplemented with 40 mg/l seleno-L-methionine (SeMet) instead of LB.

Crystallization of EcCdnD

All screening for crystallization conditions and subsequent optimization were carried out by using the sitting-drop vapor-diffusion method at 4°C. Each protein sample, typically 8 mg/ml in the gel filtration buffer, was mixed with a reservoir solution by a ratio of 1 μl to 1 μl, and equilibrated against 200 μl of the reservoir. The tetragonal *Ec*CdnD crystal was first obtained in 12 weeks by using a solution of 8% v/v Tacsimate, pH 6.0, and 20% (w/v) PEG 3350 as the reservoir. Subsequent crystallization of the SeMet-containing protein in a similar unit cell used a reservoir of 0.1 M ammonium tartrate dibasic, pH 7.0, and 20% (w/v) PEG 3350. Soaking experiments used crystals grown from a reservoir of 0.2 M K–Na–tartrate, pH 7.4 and 20% (w/v) PEG 3350, which were soaked in the reservoir containing 2 mM ATP or GTP for 30 s or 2 h before data collection. In co-crystallization experiments, the protein was incubated overnight at 4°C in solutions that contained various ligands before use in drop set-ups. The complex crystals of *Ec*CdnD with the ATP analogue 2′,3′-dideoxyadenosine-5′-triphosphate (ddATP) was obtained by incubation with 2 mM ddATP, 10 mM GTP and 10 mM $MgCl_2$ and by using a reservoir of 0.2 M ammonium tartrate dibasic, pH 7.0 and 26% (w/v) PEG 3350. The monoclinic crystals that contained ATP and ADP were grown by incubation with 5 mM ATP, 10 mM GTP and 10 mM $MgCl_2$, followed by the use of 0.2 M ammonium sulfate, 0.1 M Na-acetate, pH 4.6, and 25% (w/v) PEG 4000 as the reservoir. To cocrystallize *Ec*CdnD with the non-hydrolysable ATP analogue adenosine-5′-[(α,β)-methyleno] triphosphate (AMPcPP), in an orthorhombic unit cell, the protein was incubated with 5 mM AMPcPP, 10 mM GTP and 10 mM $MgCl_2$, and then used in drop set-ups with a reservoir of 0.2 M Na-malonate, pH 6.0, and 20% (w/v) PEG 3350. Complex crystals of the double mutant D69K/D71K (KK) with ATP were grown by incubation with 5 mM ATP, 10 mM GTP and 10 mM $MgCl_2$ and setup with a reservoir of 5% (w/v) 1-butyl-3-methylimidazolium chloride (BMIM-Cl), pH 7.4, and 20% (w/v) PEG 4000.

Data collection and structure determination

The diffraction data of all *Ec*CdnD crystals in this study were collected at the National Synchrotron Radiation Research Center (NSRRC) in Hsinchu, Taiwan. The native data set was collected by using the Taiwan Photon Source (TPS) beam line 05A. Other data sets including the multi-wavelength anomalous dispersion (MAD) data of the SeMet crystal used the Taiwan Light Source (TLS) beam line 15A1. Prior to flash cooling to cryogenic temperatures, each crystal was briefly washed in the reservoir or soaking solution that contained 20% glycerol as a cryoprotectant. The data were processed by using the HKL2000 software package (23).

The structure of EcCdnD was determined by using the three-wavelength MAD data collected from the tetragonal SeMet crystal (Supplementary Table S2). Initial phase angles were calculated by using PHENIX (24), which was also employed in subsequent model building and refinement. The software CNS (25) was used in solving the monoclinic and orthorhombic crystal structures by molecular replacement (MR). Manual adjustment of the protein models, placement of the ligands, and inclusion of the solvent molecules used the program Coot (26). Where applicable, local non-crystallographic symmetry (NCS) restraints and translation-libration-screw (TLS) tensors were used in the refinement. The models were validated by MolProbity for correct geometry (27). Figures that depict the 3D protein and ligand structures were prepared by using PyMOL (28).

Isothermal titration calorimetry (ITC)

The binding affinity of ATP, GTP or other nucleotide ligands to *Ec*CdnD was performed on a MicroCal ITC200 instrument at 25°C. Nucleotides and protein samples were dissolved in assay buffer (20 mM Tris pH 8.0, 100 mM NaCl, 1 mM $MgCl_2$). Aliquots of 2 μl nucleotides at a total concentration of 1–3 mM in the syringe were injected into cells containing 0.05–0.15 mM wildtype *Ec*CdnD or the mutant KK at 3-min intervals. Depending on the binding profiles, the data were fitted to one-site or two-site binding model using the commercial Origin 7.0 program to obtain ΔH, ΔS, K_A and K_D values.

Activity measurement

The enzymatic activity of *Ec*CdnD was measured using the EnzChek® Pyrophosphate Detection Kit (Invitrogen™, Cat. No. E6646). Briefly, nucleotide substrates with a total concentration of 0.2 mM were pre-incubated with 0.03 U pyrophosphatase, 1 U purine nucleoside phosphorylase, 0.2 mM 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG) in reaction buffer of 50 mM Tris–HCl, pH 7.5, and 1 mM $MgCl_2$ at 22°C for 10 min. The reaction

was started by adding 5 μM *Ec*CdnD. The product concentration at each time point was determined by measuring OD_{360} using a Synergy™ H1 hybrid multi-mode microplate reader (BioTek Instruments, Inc.) and comparison with a standard curve of inorganic phosphate. The phosphate concentrations of control reactions devoid of *Ec*CdnD were subtracted from those of reactions in the presence of the enzyme. Each reaction was carried out in triplicate. The first 2-min data of the reaction were used to calculate initial velocity (V_0) by linear regression.

RESULTS AND DISCUSSION

Structural variations of EcCdnD from other CD-NTases

A BLAST search (29) showed that the sequence of *Ec*CdnD has 20% identity to *Pa*CdnD with a large insertion of about 50 amino-acid residues in the middle region (Supplementary Figure S1). No other close homologue was found in the PDB. Both enzymes belong to Clade D of bacterial CD-NTases (13,14). *Ec*CdnD was first crystallized in a tetragonal unit cell with space group $P4_12_12$ and one protein molecule in its asymmetric unit. However, attempts to solve the crystal structure of *Ec*CdnD by MR using the *Pa*CdnD structures (in PDB 6P82, 6P8J and 6P8U) as search models failed, probably due to significant difference between the group D2 and D5 enzymes. In an alternative attempt to solve the structure, *Ec*CdnD was expressed as a SeMet-containing protein, from which an isomorphous crystal form was obtained. The crystal structure was determined by MAD using the SeMet-derivative (Supplementary Table S2), and subsequently refined to 2.14-Å resolution using the native diffraction data (Table 1). Similar to *Pa*CdnD and other CD-NTases, the *Ec*CdnD protein folds into an N-terminal catalytic domain and a C-terminal helical domain (Figure 2). The apo-form structures of *Ec*CdnD and *Pa*CdnD differ from each other by a root-mean-square deviation (RMSD) of 3.1–3.4 Å for 257–262 matched pairs of Cα atoms, which is much larger than the RMSD of 0.47–0.70 Å for 300–302 pairs of Cα in the different *Pa*CdnD monomers. The 50-residue insertion between strands βE and βF in *Ec*CdnD replaces the equivalent short turn in *Pa*CdnD and includes an additional helix α4 (Supplementary Figure S2) and Trp170, whose side chain was once mistaken as a purine base bound allosterically to a neighboring protein molecule (Supplementary Figure S3). The 26 C-terminal residues and the appended His-tag were not seen.

A DALI search (30) showed that *Ec*CdnD has >400 similar structures in the PDB with higher *Z* scores than 10. They include the mouse nuclear factors NF90/NF45 (31) and the human protein MAB21L1 (male abnormal 21 like 1) (32), as well as other NTase-family proteins. On top of the list is *Pa*CdnD, with a *Z* score of 25.5, followed by *Vc*DncV ($Z = 23.4$) and *Ec*CdnC ($Z = 23.1$). The latter structures have RMSD of 3.4 and 2.9 Å from *Ec*CdnD for 262 and 246 Cα pairs, respectively. Major structural differences between *Ec*CdnD and *Vc*DncV are seen in the regions of α2-βA and βE-βF (Supplementary Figure S4A). The helix α2 of *Vc*DncV is longer by three turns, which is followed by three additional short helices before strand βA. In the βE–βF insertion, *Vc*DncV contains an equivalent to helix α4 of *Ec*CdnD, but with a very different orientation. The βE–βF loop of *Ec*CdnC is not as prominent, but its equivalent to the α8-α9 loop of *Ec*CdnD is twice as long, forming a doubly wound lid structure (Supplementary Figure S4B). Despite these differences, the conserved two-domain architecture is seen in *Ec*CdnD as well as *Pa*CdnD, *Vc*DncV and *Ec*CdnC, and the cage-like substrate-binding pocket is lined with three catalytically important acidic residues, Asp69, Asp71 and Asp121 at the bottom (Figure 2). Interestingly, Pro16 of *Ec*CdnD replaces the conserved leucine residue in other CD-NTases, and its rigidity may render the enzyme constitutively active. The substrate binding mode and catalytic mechanism of *Vc*DncV and other CDN-synthesizing enzymes have been studied extensively (3,22), but so far the structures of CTN-synthesizing enzymes have no more than a bound ATP (14). Consequently, attempts were made to obtain complex crystals of *Ec*CdnD with substrates and substrate analogues.

Substrate binding modes and protein conformational changes

Soaking experiments of the native *Ec*CdnD crystals with GTP or ATP showed some electron densities in the active site, which however could only be modeled as phosphate and pyrophosphate ions (Supplementary Figure S5). It might be a result of the absence of divalent metal ion in the solution. Subsequent cocrystallization experiments all had Mg^{2+} included, and several complex crystals were obtained. One is isomorphous to the native crystal form (Table 1). The protein models in the two tetragonal crystals show an RMSD of 0.131 Å for 308 matched pairs of Cα atoms. The crystallization solution contained ddATP and GTP, but only ddATP was observed in the active site (Supplementary Figure S6A). Specifically, the ddATP occupied the binding site for the donor substrate. In the refined model, the triphosphate moiety of ddATP is bound to an Mg^{2+} ion, with an O atom from each of the α-, β- and γ-phosphate contributing a coordination bond (Figure 3A). This Mg^{2+} ion is equivalent to the metal B in Figure 1D. The other three ligands of the octahedral Mg^{2+} complex are served by two water molecules and the side chain of Asp71. The N3 and N6 atoms of the adenine base are hydrogen bonded to the sidechain NH_2 of Gln210 and the backbone O of Val304, respectively. The base is also in contact, at 3.7–3.8 Å distances, with the aromatic side chain of Tyr250 but does not show π- or T-stacking interaction. Another bound metal ion with octahedral configuration is located at the C-terminus of helix α7, where the six coordinates are served by four water molecules and the backbone O of Asn258 and Leu260 (Figure 3B). This metal ion was observed in all *Ec*CdnD crystals, either as an Na^+ in the native form or replaced by an Mg^{2+} in the substrate complexes, and its location away from the active site suggests a structural role.

Cocrystallization in the presence of ATP and GTP at lower pH yielded a new crystal form that belongs to the monoclinic space group of $P2_1$ and contains two molecules of *Ec*CdnD in its asymmetric unit. The structure was solved by MR and refined to 1.87-Å resolution (Table 1). In this crystal, the two protein molecules have virtually identical conformation, with an RMSD of 0.037 Å for 298 matched

Table 1. Data-collection and refinement statistics of *Ec*CdnD crystals. See below for some description about the listed items

	Native	ddATP	ATP+ADP	AMPcPP	KK-ATP
Data collection					
Space group	$P4_12_12$	$P4_12_12$	$P2_1$	$C222_1$	$C222_1$
Unit-cell *a*, *b*, *c* (Å)	52.9, 52.9, 344.4	53.2, 53.2, 345.8	65.8, 107.2, 65.9	66.3, 116.4, 107.4	68.9, 118.2, 102.4
Unit-cell α, β, γ (°)	90, 90, 90	90, 90, 90	90, 118.3, 90	90, 90, 90	90, 90, 90
Resolution range (Å)	30–2.14 (2.22–2.14)	50–2.09 (2.16–2.09)	30–1.87 (1.94–1.87)	50–1.93 (2.00–1.93)	30–2.70 (2.80–2.70)
Unique reflections	27 647 (2654)	30 653 (2873)	66 128 (6346)	31 398 (3124)	11 739 (1142)
Redundancy	8.7 (6.5)	9.7 (5.0)	3.8 (3.7)	7.0 (7.1)	6.9 (6.6)
Completeness (%)	98.9 (97.5)	99.1 (96.7)	99.5 (95.2)	99.9 (100.0)	99.7 (99.2)
Average $I/\sigma(I)$	17.0 (2.7)	25.0 (3.4)	29.0 (8.5)	25.7 (3.6)	17.1 (2.9)
Average $CC_{1/2}$	0.988 (0.969)	0.985 (0.943)	0.992 (0.974)	0.986 (0.950)	0.961 (0.838)
R_{merge} (%)	9.6 (40.3)	8.3 (37.2)	4.3 (15.4)	7.2 (51.4)	11.3 (63.3)
R_{pim} (%)	3.6 (16.1)	2.7 (16.0)	2.6 (9.3)	3.0 (20.5)	4.6 (26.0)
Structure refinement					
Resolution range (Å)	24–2.15 (2.23–2.15)	28–2.09 (2.17–2.09)	29–1.87 (1.94–1.87)	34–1.93 (2.00–1.93)	24–2.70 (2.80–2.70)
No. of reflections	24 213 (1206)	28 906 (1722)	66 096 (6464)	31 018 (2713)	11 641 (1042)
Completeness (%)	86.7 (44.5)	93.9 (58.3)	99.7 (97.3)	98.4 (87.0)	98.8 (90.7)
R_{work} for 95% data (%)	18.2 (21.5)	17.6 (20.9)	13.6 (15.7)	15.2 (19.7)	19.2 (24.1)
R_{free} for 5% data (%)	22.8 (30.4)	21.0 (24.2)	16.3 (19.6)	19.2 (22.2)	25.0 (33.0)
RMSD bond lengths (Å)	0.0029	0.0037	0.0120	0.0079	0.0019
RMSD bond angles (°)	0.593	0.749	1.216	0.897	0.648
B_{ave} (Å^2)/protein atoms	37.7/2877	29.4/2793	20.4/5619	22.0/2803	42.9/2773
B_{ave} (Å^2)/ligand atoms	30.9/1	41.6/31	22.5/170	19.4/66	35.6/62
B_{ave} (Å^2)/water molecules	46.1/371	40.8/373	33.7/935	36.6/454	43.4/206
Ramachandran favored (%)	98.0	99.1	98.1	98.2	97.9
Ramachandran allowed (%)	2.0	0.9	1.9	1.8	2.1
Ramachandran outliers (%)	0.0	0.0	0.0	0.0	0.0
Clashscore	2.64	2.52	3.41	1.94	4.65
MolProbity score	1.13	1.04	1.13	0.96	1.53
PDB code	**7D48**	**7D4J**	**7D4O**	**7D4S**	**7D4U**

Numbers in parentheses are for the highest resolution shells.
$I/\sigma(I)$ = ratio of intensity against background, or signal-noise level
$CC_{1/2}$ = correlation coefficient between intensity estimates of half data sets
R_{merge} = merging *R* factor = $\Sigma_h\ \Sigma_i|<I_h> - I_{h,i}|/\Sigma_h\ \Sigma_i I_{h,I}$ where *h* and *i* enumerate unique reflections and symmetry-equivalent contributors.
R_{pim} = precision-indicating merging *R* factor = $\Sigma_h\ [1/(n_h - 1)]^{1/2}\ \Sigma i|<I_h> - I_{h,i}|/\Sigma_h\ \Sigma_i I_{h,i}$ where n_h denotes multiplicity.
R_{work} = *R* value for the working data subset used in the refinement = $\Sigma_h\ |\ F_{h,obs} - F_{h,calc}\ |/\ \Sigma_h F_{h,obs}$ where $F_{h,obs}$ and $F_{h,calc}$ are the observed and calculated structure factors.
R_{free} = *R* value for the free data subset not used in the refinement to monitor the correctness of the model and the progress of refinement.
The RMSDs are root-mean-square deviations of the bond lengths and bond angles in the refined models from the ideal values in the dictionary of MolProbity (27)
B_{ave} = average isotropic temperature factors over the numbers of specified atoms.
The numbers of amino acid residues in the favored and allowed regions of Ramachandran plot for the peptide dihedral angles were determined by MolProbity, as well as the clash scores, which represent the numbers of serious clashes per 1000 atoms.
The MolProbity score serves as a single composite metric for model quality (lower numbers indicate better models).

Cα pairs, and so do the bound ligands. Besides an ATP in the donor site, there was also an ADP that apparently occupied the acceptor site (Supplementary Figure S6B). The presence of ADP was probably due to Mg^{2+}-assisted hydrolysis of ATP. It could also be a result of bound ATP with disordered γ-phosphate. The two nucleotide molecules are packed back-to-back and related by a dyad axis. The adenine bases are not stacked but each makes contacts to the ribose O4′ of another at 3.0–3.2 Å distances. The triphosphate part of ATP now binds to two Mg^{2+} ions. One Mg^{2+} ion, designated Mg-B, makes similar interactions with the protein and ligand as those observed in the above ddATP complex crystal, but the side chain of Asp69 replaces a coordinating water molecule (Figure 3C). The other Mg^{2+} ion, equivalent to the Metal A in Figure 1D and designated Mg-A, is coordinated to the α-phosphate, the side chains of Asp69, Asp71 and Asp121, and two water molecules, but with longer metal-oxygen distances of 2.2–2.5 Å than those of 2.0–2.1 Å for the Mg–B ion. The adenine base of ATP shows similar interactions with the protein as well, while the adenine N3 and the ribose 2′-OH are hydrogen bonded to the Gln210 side chain, and the 3′-OH is hydrogen bonded to the β-phosphate and the Arg56 side chain. On the other hand, the adenine base of ADP has its N1 and N3 atoms hydrogen bonded to the side chains of Gln51 and Thr205, and its ribose 2′-OH to the backbone O of Arg204 (Figure 3D). The diphosphate group makes seven hydrogen bonds with Arg109, Arg197 and Arg307.

Another crystal form obtained at pH 6.0 in the presence of AMPcPP and GTP turned out to belong to the orthorhombic space group of $C222_1$ with one protein molecule in its asymmetric unit (Table 1). Two AMPcPP

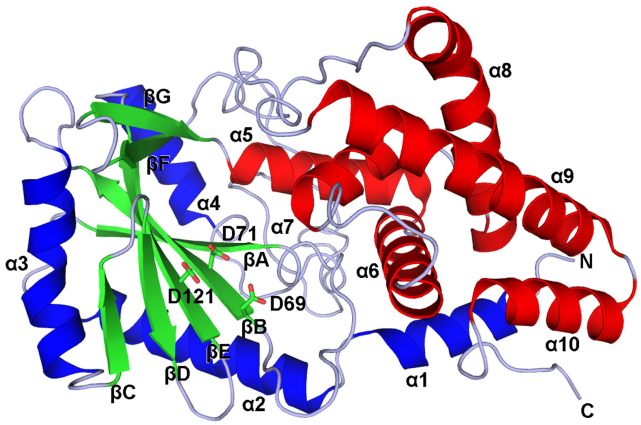

Figure 2. Overall structure of *Ec*CdnD. The protein is presented as a ribbon diagram, in which the N and C-terminal α-helices are colored blue and red. The β-strands are green. On the left side is the N-terminal NTase domain. It contains the three essential residues Asp69, Asp71 and Asp121 for catalysis, whose side chains are shown as stick models. On the right side is the C-terminal helical domain. The first two helices α1 and α2 correspond to the 'spine' in other NTases.

were bound to the enzyme by similar binding modes as the ATP and ADP in the monoclinic crystal. Most interactions with the donor AMPcPP and the two metal cofactors are conserved, but the Mg-A ion is bound to one water molecule instead of two (Figure 4A). The side chain of Arg307 now turns to form two hydrogen bonds with the γ-phosphate of the acceptor AMPcPP, which together with the β-phosphate may bind to a third Mg^{2+} ion (Figure 4B). The protein conformation in the orthorhombic crystal is similar to that in the monoclinic crystal, with RMSD of 0.107–0.108 Å for 325–328 pairs of matched Cα atoms, but both differ significantly from the tetragonal structures with RMSD of 0.62–0.64 Å for 297–300 Cα pairs (Supplementary Figure S7A). The largest shift is located in the βF-βG loop, where the Cα of Arg197 is moved by 6.8 Å and the side chain swings to the opposite side by nearly 18 Å (Supplementary Figure S7B). Other prominent shifts are seen in helices α2, α3, strands βB, βC, βD and the associated loops. Because a similar binding mode of the donor substrate, be it either ddATP, ATP or AMPcPP, is shared by all complex structures, the conformational difference is likely a result of the acceptor nucleotide binding. These structural rearrangements may be induced by the strong interactions between the positively charged arginine side chains of *Ec*CdnD and the negatively charged phosphate groups of ADP or AMPcPP, as consistently observed in the $P2_1$ and $C222_1$ crystals.

Furthermore, several mutants of Asp69, Asp71 and Asp121 were produced for activity analysis and crystallization. Mutants with two or three alanine substitutions precipitated easily and were difficult to purify and crystallize. Consequently, the aspartate residues were also replaced by lysine with a hope that its positive charged side chain would increase protein solubility and mimic Mg^{2+} in substrate binding. Among the mutants, only the double mutant KK crystallized, at pH 7.4 in a $C222_1$ unit cell similar to the wildtype (RMSD = 0.33 Å for 308 Cα pairs). Again, despite the presence of more GTP than ATP in the solution, only ATP was observed. In the refined model, two ATP molecules are packed back-to-back like the AMPcPP in the wildtype crystal. However, as expected, the sidechain NH_3^+ group of Lys71 replaces the Mg–B ion in association with the triphosphate moiety of the donor ATP, which has the 3′-OH hydrogen bonded to the γ-phosphate instead of the β-phosphate (Figure 4C). Ser53 and Arg56 now bind to the γ-phosphate but not the β-phosphate, which is hydrogen bonded to Tyr250. The binding mode of the acceptor ATP also varies considerably, with a different orientation of the γ-phosphate that binds to the 3′-OH and additional interactions between the β-phosphate and the Asp196 side chain (Figure 4D). In addition, with a hope to trap its reaction intermediates, we also crystallized the wildtype *Ec*CdnD with ATP and a non-hydrolysable GTP analogue guanosine-5′-[(α,β)-methyleno] triphosphate (GMPcPP). However, in three different crystals thus obtained, only ATP and ADP with the above back-to-back binding mode were seen (Supplementary Table S3 and Supplementary Figure S8).

Biophysical and biochemical characterization of EcCdnD

According to the recent study by Whiteley *et al.* (13), *Ec*CdnD catalyzes the formation of cAAG by condensing two molecules of ATP and one molecule of GTP, and the reaction may start with the attack at the α-phosphate of ATP because it was completely blocked by the non-hydrolysable analogue AMPcPP. On the other hand, some reaction intermediates were produced by using GMPcPP (13). The enzyme also turns out cAAA and cyclic AMP–GMP–GMP (cAGG) as minor products (15). However, as described above, only ATP and its analogues were seen in the crystal structures of *Ec*CdnD, and attempts to obtain complex crystals with GTP or GMPcPP were not successful. To find out the binding strengths of ATP and GTP, ITC experiments were carried out with each substrate as well as a mixture of both (Supplementary Figure S9). ATP showed strong binding but GTP showed no binding at all. However, use of both ATP and GTP complicated the binding pattern, probably due to on-going reactions catalyzed by *Ec*CdnD. To avoid interference of binding isotherm by the heat of reaction, further experiments were carried out by using the non-hydrolysable analogues AMPcPP and GMPcPP. The best fitting model for AMPcPP binding to *Ec*CdnD shows two independent binding sites with similar dissociation constants (K_D) of 39.5 ± 4.9 and 44.9 ± 6.0 μM (Figure 5A), which are consistent with the above crystallographic observation of two binding sites for ATP and comparable to the nearly 100 μM K_M of DNA-bound cGAS for nucleotide triphosphate substrates (33). Interestingly, the binding isotherm of AMPcPP does not fit to the model of two sequential binding sites, suggesting a lack of cooperative binding of the two nucleotide ligands (34). Binding of ATP to the donor site and GTP to the acceptor site of the mammalian cGAS also showed non-cooperative binding, which may be a conserved mechanism among the CD-NTases (35). In contrast, the ITC data of the double mutant KK and AMPcPP show only one binding site with a K_D of 34.3 ± 2.7 μM (Figure 5B). Although the hydrogen bonding interactions with the donor nucleotide may be conserved, the lack of two Mg^{2+}-coordinating aspartate side

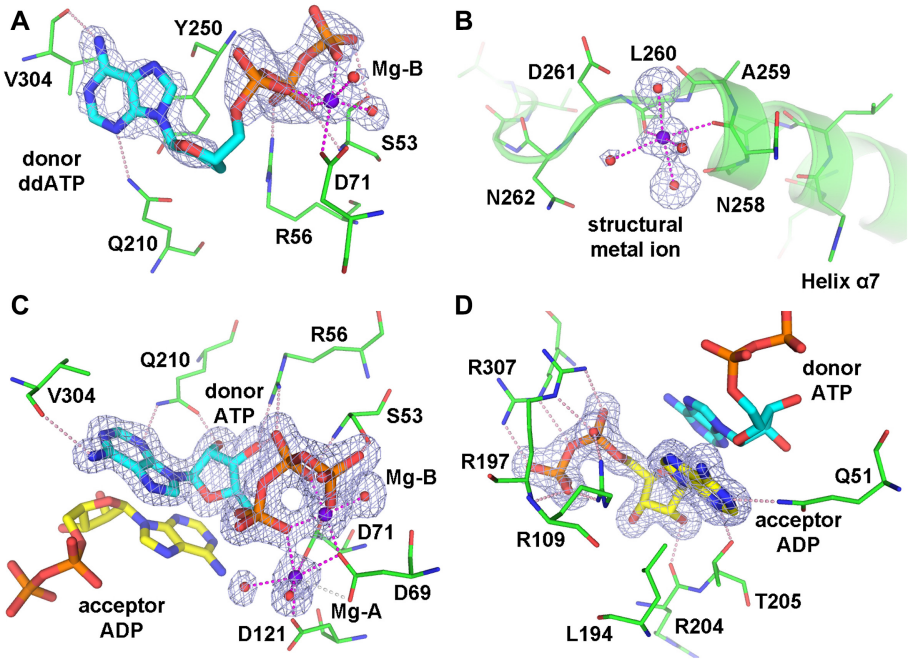

Figure 3. Ligand binding modes in the tetragonal and monoclinic *Ec*CdnD crystals. The bound nucleotides are shown as thick stick models, while the amino-acid residues involved in the binding are shown as thin sticks with the carbon atoms colored green. The metal ions and water molecules are shown as purple and red spheres. Hydrogen bonds and coordination bonds are indicated by dashed lines. The models of the bound ligands are superimposed on the *Fo* – *Fc* maps, each calculated by omitting the corresponding ligand atoms and contoured at 4σ level. (**A**) The bound ddATP (colored cyan) in the tetragonal crystal has an Mg^{2+} ion associated with its triphosphate moiety. (**B**) At the C-terminus of helix α7 a metal ion is bound to two protein backbone C=O groups and four water molecules. (**C**) The bound donor ATP in the monoclinic crystal shows more interactions with the protein, via the ribose 2′- and 3′-OH groups and an additional Mg^{2+} ion. (**D**) The bound acceptor nucleotide (colored yellow), here an ADP, makes several hydrogen bonds to three arginine residues via its phosphate groups.

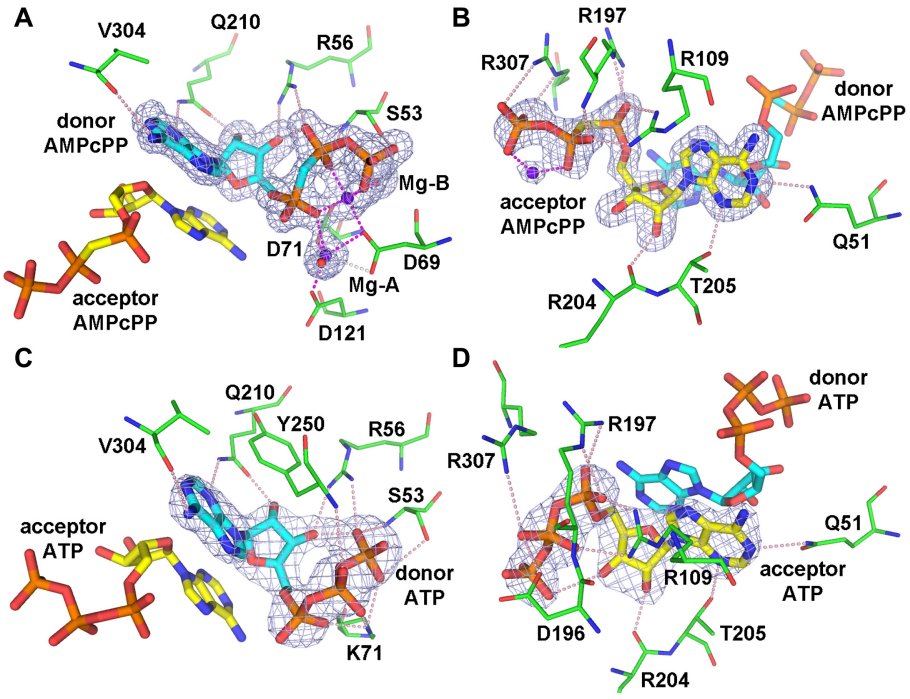

Figure 4. Ligand binding modes in the wildtype and mutant orthorhombic crystals. The models are shown in a similar way as in Figure 3, with the omit maps also contoured at 4σ level. (**A**) The binding mode of substrate analogue AMPcPP to the donor site is virtually identical to that of ATP in Figure 3C. (**B**) In addition to the three arginine residues, the PcPP group of the acceptor nucleotide is also bound to an Mg^{2+} ion. (**C**) The NH_3^+ group of Lys71 replaces Mg^{2+} ion for interacting with the donor ATP in the mutant protein. (**D**) The γ-phosphate of the acceptor ATP has a different disposition than its equivalent of AMPcPP in (B), and there is no bound Mg^{2+} ion.

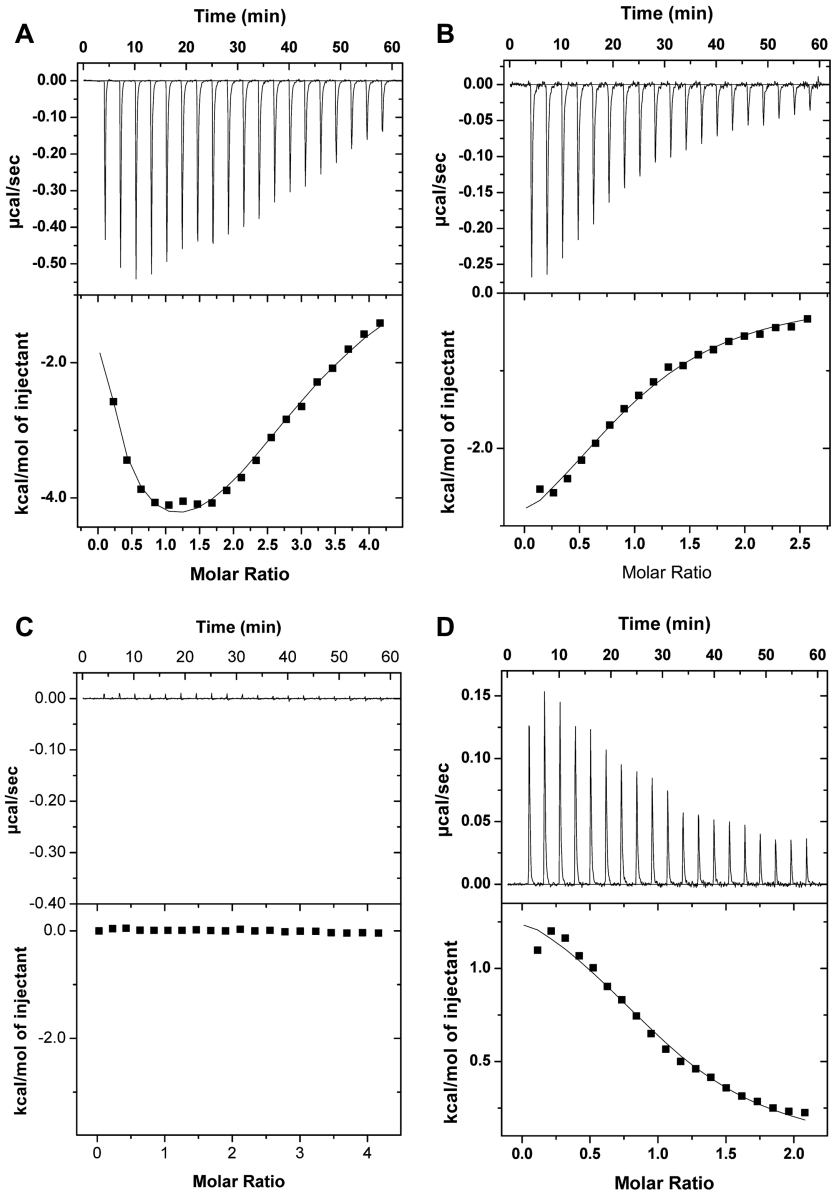

Figure 5. ITC measurements of nucleotide binding. (**A**) AMPcPP (1 mM) was used as the ligand and the binding isotherm to wildtype *Ec*CdnD (0.05 mM) turned out to be biphasic. Fitting to a model of two independent binding sites yields binding parameters of $K_A1 = 25.7 \pm 3.2$ mM^{-1}, $\Delta H1 = -2.9 \pm 0.3$ kcal/mol, $\Delta S1 = 10.3 \pm 1.4$ cal/mol/deg and $K_A2 = 22.7 \pm 3.0$ mM^{-1}, $\Delta H2 = -15.8 \pm 0.9$ kcal/mol, $\Delta S2 = -33.2 \pm 3.3$ cal/mol/deg. (**B**) The binding isotherm of AMPcPP (1 mM) to the double mutant KK (0.15 mM) shows a single site, with $N = 1.02 \pm 0.01$, $K_A = 29.4 \pm 2.3$ mM^{-1}, $\Delta H = -3.8 \pm 0.2$ kcal/mol and $\Delta S = 7.6 \pm 0.8$ cal/mol/deg. (**C**) No binding was observed when 0.05 mM wildtype *Ec*CdnD was titrated with 1.0 mM GMPcPP after pre-incubation with 1.0 mM AMPcPP. (**D**) The dinucleotide pppApG (1 mM) showed binding to wildtype *Ec*CdnD (0.1 mM) with $N = 1.02 \pm 0.06$, $K_A = 33.8 \pm 2.5$ mM^{-1}, $\Delta H = 2.2 \pm 0.6$ kcal/mol and $\Delta S = 28.1 \pm 2.0$ cal/mol/deg.

chains may significantly reduce the binding affinity to a level that cannot be detected by ITC. Consequently, the single AMPcPP binding site in KK probably corresponds to the acceptor site. The presence of the donor ATP in the KK mutant crystal might be a result of the high ATP concentration employed in crystallization. On the other hand, GMPcPP showed no binding to the wildtype *Ec*CdnD, either alone (Supplementary Figure S10A) or after preincubation of the enzyme with AMPcPP (Figure 5C), a reflection of the fact that no GTP was found in the crystals despite the two or five-fold higher concentration than ATP. The latter results suggest that GMPcPP (or GTP) cannot compete with AMPcPP (or ATP) for binding to either the donor site or the acceptor site. Further ITC experiments with the dinucleotides 5′-ApA-3′ and 5′-ApG-3′ showed no binding to *Ec*CdnD (Figure S10B, S10C), but 5′-pppApG-3′, a possible intermediate, showed a single binding site with a K_D of 29.7 ± 2.2 μM (Figure 5D), underscoring the importance of the 5′-triphosphate in the enzyme-substrate and enzyme-intermediate interactions.

Subsequently, the enzymatic activity of *Ec*CdnD was measured by using various combinations of ATP and GTP

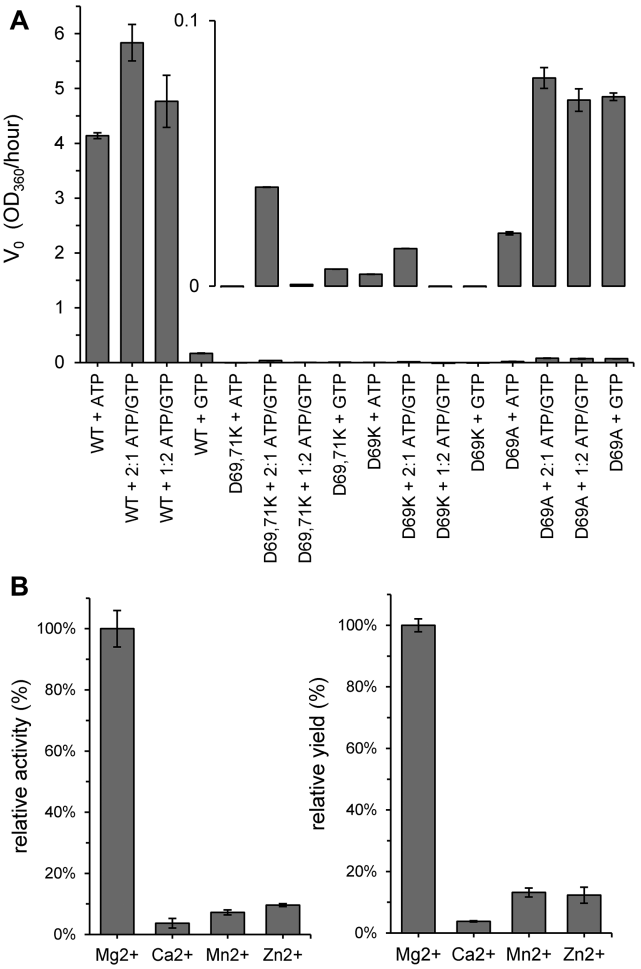

Figure 6. Catalytic activity for different substrates and cofactors. (**A**) ATP, GTP or two different combinations of ATP and GTP with ratio of 2:1 or 1:2 was used as the substrate. In addition to the wildtype *Ec*CdnD, the mutants KK (D69K/D71K), D69K, and D69A were also tested. (**B**) The substrate contained ATP and GTP with 2:1 ratio. The left chart used the first 2-min data, whereas the right chart used the first 10-min data.

as the substrates (Supplementary Figure S11). As expected, the substrate with an ATP:GTP ratio of 2:1 showed the highest turnover rate, and GTP alone showed the lowest (Figure 6A). Interestingly, using ATP alone as the substrate or a mixture with an ATP:GTP ratio of 1:2, *Ec*CdnD also showed comparable catalytic activity. In addition, the double mutant KK and two additional mutants D69K and D69A were tested for activity (Figure 6A, inset). Both KK and D69K lost most of the enzymatic activity, except for the small fraction that remained with KK when the substrate ATP:GTP ratio was 2:1. Although the double mutant KK does not retain the ability to form proper coordination bonds with Mg^{2+} ion, its two lysine side chains might still bind to and facilitate the dissociation of pyrophosphate from the donor substrate. However, the single mutant D69K might generate a strong interaction between the neighboring but oppositely charged side chains of Lys69 and Asp71, resulting in a complete loss of function. By contrast, D69A showed higher activity than the other two mutants. Presumably the effects of a single carboxylate group removal on Mg^{2+} binding could be more easily compensated, for example, by solvent molecules. The higher activity of D69A for GTP than ATP also suggests favored binding of GTP to the donor site, probably by a different mode from that of ATP.

When the active-site structures of *Ec*CdnD and DNA/RNA polymerase are compared, the Mg^{2+} ion that binds to the triphosphate moiety of the donor substrate in *Ec*CdnD corresponds to the lower affinity metal (Metal-B or MG2) in the A site (or the i+1 site; 'A' for addition) of DNA/RNA polymerase (20,21). The other Mg^{2+} ion that binds to the side chains of Asp69, Asp71 and Asp121 in *Ec*CdnD corresponds to the high affinity metal (Metal-A or MG1) in the P site (or the i site; 'P' for primer) of DNA/RNA polymerase. Interestingly, the triphosphate-bound metal ion, or Mg-B, seems to be more stable than the aspartate-bound Mg-A in CD-NTases, because it was clearly observed to associate with the bound donor substrate in the crystal structures of *Ec*CdnD as well as *Pa*CdnD and *Ec*CdnC (14). The octahedral configuration of the former is also better defined than the latter. Besides binding to the ligands with non-canonical distances and angles, this second metal ion even showed variation in coordination numbers by forming an additional coordination bond with the second sidechain oxygen of Asp69 at a distance of 2.8–2.9 Å (Figures 3C, 4A). To find out possible use of other divalent metal ions than Mg^{2+} as cofactor, the 1 mM $MgCl_2$ in the activity assay was first replaced by $MnCl_2$, $ZnCl_2$ or $CaCl_2$. Preliminary tests with ATP as the substrate showed that Mg^{2+} is the best cofactor, followed by Mn^{2+}, while Ca^{2+} and Zn^{2+} are poor (Supplementary Figure S12A). Further experiments by using both ATP and GTP as the substrate with a ratio of 2:1 showed similar results (Supplementary Figure S12B). Considering the requirement of Mg^{2+} by pyrophosphatase, the latter assays included 1 mM of the metal cofactor in addition to the original 1 mM $MgCl_2$. As shown in Figure 6B, despite the presence of Mg^{2+}, the addition of $MnCl_2$, $ZnCl_2$ or $CaCl_2$ all resulted in reduced activity. The divalent metal ions other than Mg^{2+} actually inhibited the enzyme, probably by occupying the metal binding sites for Mg^{2+}. The inhibition by Ca^{2+} was the strongest, while 5–10% activity was retained with Mn^{2+} and Zn^{2+}. Although the reaction appears to be slowed down, Mn^{2+} may itself participate in the catalysis, turning out more than 10% product (Figure 6B). Zn^{2+} may prefer a tetrahedral configuration, making it a weaker inhibitor for an octahedral metal binding site. Taken together, these results suggest that *Ec*CdnD uses Mg^{2+} as the cofactor in catalysis, and both metal ions in the active site should be Mg^{2+}.

Catalytic pathway based on other ligand-bound NTase structures

The catalytic reaction of a CD-NTase starts with the attack at the α-phosphate group of the donor substrate by the 3′-OH of the acceptor substrate. However, in the complex structures of *Ec*CdnD, the two bound nucleotides are disposed back-to-back, and the α-phosphate of one nucleotide is more than 10 Å away from the 3′-OH of another. In either case, the 3′-OH is directed away from the target phosphate, making it impossible for any nucleophilic

attack. Consequently, the observed acceptor ATP (ADP or AMPcPP) structure may indeed represent some inhibitory binding mode due to the high nucleotide concentrations used in crystallization, similar to the substrate inhibition of mammalian cGAS by high concentration of ATP and high ATP-to-GTP ratio (35). To investigate other possible binding modes of the substrates and reaction intermediates to *Ec*CdnD, the substrate-bound structures of three most similar CD-NTases are compared and used as a basis for rational modeling. In the complex structure of *Pa*CdnD and ATP (double mutant D62N/D64N; PDB 6P8J; Supplementary Figure S13A) (14), the nucleotide is located in the donor substrate binding site, as is the first ATP (or AMPcPP) in *Ec*CdnD. An Mg^{2+} ion is bound to the α- and β-phosphate groups of the ATP, the backbone O of Asn62, the side chains of Asn64 and Asp130, and a water molecule. The triphosphate also makes hydrogen bonds with two serine and two lysine residues. Interactions with the adenosine part are largely conserved in *Ec*CdnD and *Pa*CdnD, except that the N3 and 2′-OH are bound to separate protein side chains in the latter. In *Ec*CdnC (PDB 6P80; Supplementary Figure S13B) (14), the Mg^{2+} is bound to the β- and γ-phosphates, the Asp74 side chain, and three waters. The adenine base forms hydrogen bonds with the protein via its N1 and N7 atoms, but not N3. These bound ATP molecules in *Pa*CdnD and *Ec*CdnC, however, did not add much information to the reaction pathway of a CTN-synthesizing enzyme. In either case the cage-like active-site pocket seems only wide enough to accommodate the donor substrate (Supplementary Figure S14).

On the other hand, the complex structures of *Vc*DncV contain both donor and acceptor substrates (PDB 4XJ3) (22), as well as an analogue of reaction intermediate (PDB 4TY0) (10). In the substrate complex, the binding mode of the donor GTP is similar to that of the donor ATP (or AMPcPP) in *Ec*CdnD, with one Mg^{2+} ion bound to the α-, β- and γ-phosphate groups and another Mg^{2+} to the α-phosphate (Supplementary Figure S15A). The N1 and N2 atoms of the guanine base forms two hydrogen bonds with the side chain of Asp348. If GTP were bound to *Ec*CdnD by a similar mode, the N2 atom would clash with the Asp296 side chain and the O6 atom would not form a proper hydrogen bond with the backbone O of Val304. These unfavorable interactions may exclude GTP from its use as a donor substrate by *Ec*CdnD. The acceptor GTP in *Vc*DncV shows a very different binding mode from that of the acceptor ATP (ADP or AMPcPP) in *Ec*CdnD. However, this GTP binding mode is compatible with the geometry for nucleophilic attack at the donor substrate. The 3′-OH of one GTP is directed toward the α-phosphate of the other at a distance of 4.4 Å. The guanine base has each of its N1, N2 and N3 atoms hydrogen bonded to the side chains of Gln112, Tyr197 and Ser259. The N2 atom is also hydrogen bonded to the backbone O of Ser259. The protein makes identical interaction with the guanine base of the intermediate GMP-AMPcPP, which consists of a GMP linked to the 3′-OH of AMPcPP by a phosphodiester bond (Supplementary Figure S15B). In contrast, the adenine base of this intermediate does not make any specific bond with the protein. When compared with the donor GTP structure above, the adenosine ribose of GMP-AMPcPP is rotated by nearly 180° and the purine base is flipped over, while the PcPP part retains a similar disposition as that of the triphosphate, except for a rotation of the α-phosphate group (Supplementary Figure S16). This orientation still allows its attack by the 3′-OH of the acceptor nucleotide. Since GTP alone does not bind to *Ec*CdnD, it is likely that the guanosine part of some GTP-containing intermediate interacts with the enzyme by a similar binding mode in the donor site.

Although both nucleotide ligands in the substrate complex structure of *Vc*DncV are GTP, the donor has stronger electron densities than the acceptor, suggesting that GTP serves as the first donor in 3′,3′-cGAMP formation (Figure 7A). The intermediate complex structure contains GMP-AMPcPP also. Consequently, the *Vc*DncV-catalyzed reaction is likely to start with the ATP 3′-OH attacking the GTP α-phosphate (Figure 7B). This is in contrast to the production of 2′,3′-cGAMP by mammalian cGAS, which starts with the GTP 2′-OH attacking the ATP α-phosphate (Figure 7C). Regarding *Ec*CdnD, because the donor site is occupied by ATP in all complex structures, where bound GTP has never been observed, the cAAG-synthesizing reaction is likely to start with the GTP 3′-OH attacking the ATP α-phosphate (Figure 7D). The corresponding residues of Gln112, Tyr197 and Ser259 in *Vc*DncV that interact with the base of the acceptor substrate are Gln51, Thr125 and Thr205 in *Ec*CdnD. Both enzymes should be able to use either ATP or GTP, or the 3′-nucleoside of the reaction intermediate, as the acceptor. However, because GTP alone does not bind to *Ec*CdnD, it is unlikely to serve as a donor substrate, and the first two steps of reaction should use ATP as the donor, turning out 5′-pppGpApA-3′. The guanosine part of GTP in this second intermediate then flips over and fits into the donor substrate binding site in a similar way as does the adenosine part of the *Vc*DncV-GMP-AMPcPP complex (PDB 4TY0). If the first acceptor is an ATP, the enzyme will turn out cAAA following the same pathway (Figure 7E). It is also possible that the nascent intermediate 5′-pppGpA-3′ (or 5′-pppApA-3′) flips over and binds to the donor binding site. In this case, subsequent reaction with GTP as the acceptor will lead to cAGG (or cAAG) production (Figure 7F), but reaction with ATP will turn out cAAG (or cAAA) as well. It should be noted that the catalytic pathways of *Ec*CdnD as tentatively proposed above are derived indirectly from structure comparison and require further experiments to verify. In fact, although cAAG was shown to be the major product (13,15), it remains unclear which conditions are required for GTP binding to the enzyme.

Structural implications in function and regulation

The structural analyses above suggest that the catalytic pathway of *Ec*CdnD is similar to those of other CD-NTases such as *Vc*DncV and mammalian cGAS, following a common mechanism shared also by more distantly related NTases (20). The difference of CD-NTase from DNA/RNA polymerase is the use of protein interactions to discriminate the substrates, instead of base pairing with the DNA/RNA template (19). Moreover, the cage-like active site pocket is supposed to retain the product and put a limit to its size. In the active site of *Vc*DncV, the triphosphate moiety of

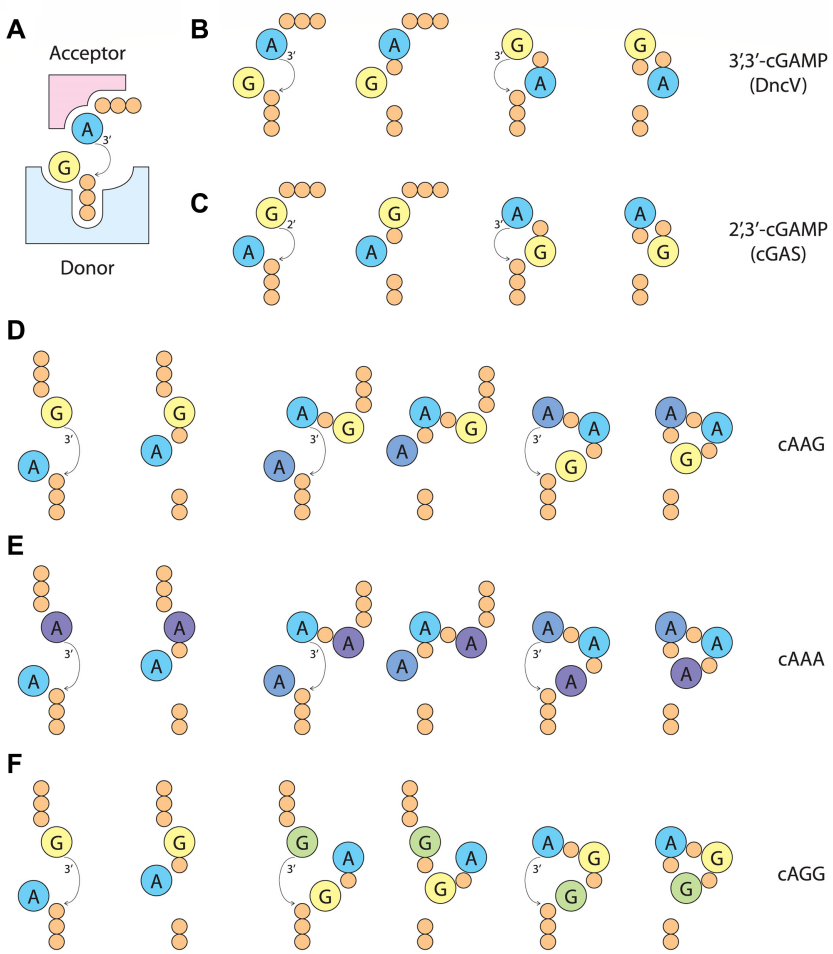

Figure 7. Reaction pathways of CD-NTase. (**A**) The nucleosides are represented by large circles, and the phosphate groups by small circles. The donor and acceptor substrate binding sites for are indicated by boxes. The nucleophilic attack by ribose OH group at the α-phosphate is indicated by an arrow. (**B**) and (**C**) show known pathways for CDN biosynthesis. (**D**), (**E**) and (**F**) show possible pathways that turn out three different CTNs by *Ec*CdnD.

the acceptor substrate interacts with an adjacent positively charged patch on the protein surface (Supplementary Figure S17A). After the first nucleophilic attack of the acceptor at the donor substrate, the resulting intermediate undergoes rearrangements for the second step, probably driven by the triphosphate rolling over into the nearby donor site. However, either as an apo-form or bound to ATP/ADP analogues, the active site of *Ec*CdnD is open on the distal end (Supplementary Figures S5, S6). The equivalent surface area for binding to the acceptor substrate triphosphate in *Vc*DncV is not positively charged in *Ec*CdnD. Instead, positively charged patches are seen on the opposite side of the pocket and a groove between the βF-βG and α8-α9 loops (the 'gateway'), where the triphosphate (or diphosphate) group of the bound acceptor nucleotide is located. Interestingly, if the primer strand of an RNA polymerase bound RNA, along with the A-site substrate (from PDB 2O5J) (36), is superimposed onto the bound substrates in *Vc*DncV and *Ec*CdnD, it will penetrate the 'cage' of *Vc*DncV but fit roughly into the 'gateway' of *Ec*CdnD (Supplementary Figure S17B), while the P-site and A-site nucleotides match well with the acceptor and donor substrates (Supplementary Figure S18A). Furthermore, if the nucleotide adjacent to the P-site is taken from the primer RNA strand, its 5′-phosphate can be rotated and connected to the 3′-OH of the donor nucleotide with an intermediate-like conformation.

Based on these comparative observations, a reasonable trinucleotide intermediate model with three bases stacked can be constructed (Supplementary Figure S18B-S18F). It is likely that the first two steps of *Ec*CdnD catalysis proceed in much the same way as that of RNA polymerase, and the 5′-triphophate moiety of the intermediate interacts with the positive charges at the 'gateway'. However, even with large conformational changes like that in the βF–βG loop that contains Arg197, the active site pocket can accommodate only three nucleotides, and therefore the polymerization reaction does not go further. Instead, the trinucleotide intermediate undergoes rearrangements, probably starting with the triphosphate relocation into the donor site. The CTN producing reaction then concludes with a third nucleophilic attack. On the other hand, premature rearrangement a dinucleotide intermediate should also be possible (for cAGG production), but the more open active site pocket would fa-

cilitate the intercalation of a third nucleotide as the new acceptor, rather than turning out a CDN product.

As a Type II CD-NTase, *Ec*CdnD is constitutively active, and so is *Vc*DncV (10,13,19). However, *Vc*DncV is inhibited by folic acid derivatives, which bind adjacent to the active site groove and prevent the enzyme from interacting with the substrates (37). Besides the ATP (or AMPcPP) in the donor substrate binding site, the complex crystal structures of *Ec*CdnD determined in this study show an additional dyad-related molecule of ATP (ADP or AMPcPP) that occupies the acceptor substrate binding site by a nonproductive mode. By comparison with the apo-form and donor-bound structures, binding of this 'fake acceptor' nucleotide induces significant local conformational changes in the enzyme, which are probably more inhibitory than activating. In this way, high ATP concentrations may prevent GTP from binding to *Ec*CdnD. On the other hand, it is tempting to speculate that some metabolic nucleotide cofactor like NAD or FAD, which contains an adenosine nucleotide moiety, might also be capable of occupying this site by a similar binding mode. However, preliminary tests with ITC showed that *Ec*CdnD does not bind to either NAD or FAD with high affinity (Supplementary Figure S19). Neither are they potent inhibitors of the CTN-producing reaction (Supplementary Figure S20). Perhaps ATP may itself serve as a bona fide regulator in inhibiting *Ec*CdnD as it inhibits the mammalian cGAS at a high concentration (35). Whether such nonproductive binding of the dyad-related nucleotide in place of a real acceptor substrate might play a role in regulating *Ec*CdnD activity, as well as the precise catalytic pathway for synthesizing CTNs, remains to be further investigated. Since Type II CD-NTases do not require additional activation mechanism of Type I enzymes such as conformational change induction by binding to effector proteins or dsDNA/RNA, the crystal structures of *Ec*CdnD presented above may provide an easily accessible route for further biochemical and biophysical studies of CTN-producing CD-NTases.

DATA AVAILABLITY

Atomic coordinates and structure factors for the reported *Ec*CdnD crystals have been deposited with the Protein Data Bank under accession numbers 7D48 7D4J, 7D4O, 7D4S and 7D4U.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to NSRRC for beam time allocations and data collection assistance.

FUNDING

Academia Sinica and the Taiwan Protein Project [AS-KPQ-109-TPP2]; Ministry of Science and Technology of Taiwan [108-2311-B241-001 and 109-2311-B241-001]; China Medical University in Taiwan [CMU109-MF-18 and CMU109-S-07]. Funding for open access charge: Ministry of Science and Technology in Taiwan.

Conflict of interest statement. None declared.

REFERENCES

1. Walsh,D.A., Brostrom,C.O., Brostrom,M.A., Chen,L., Corbin,J.D., Reimann,E., Soderling,T.R. and Krebs,E.G. (1972) Cyclic AMP-dependent protein kinases from skeletal muscle and liver. *Adv. Cyclic Nucleotide Res.*, **1**, 33–45.
2. Zubay,G., Schwartz,D. and Beckwith,J. (1970) Mechanism of activation of catabolite-sensitive genes: a positive control system. *Proc. Natl. Acad. Sci. U.S.A.*, **66**, 104–110.
3. Aline Dias da,P., Nathalia Marins de,A., Gabriel Guarany de,A., Robson Francisco de,S. and Cristiane Rodrigues,G. (2020) The world of cyclic dinucleotides in bacterial behavior. *Molecules*, **25**, 2462.
4. He,J., Yin,W., Galperin,M.Y. and Chou,S.H. (2020) Cyclic di-AMP, a second messenger of primary importance: tertiary structures and binding mechanisms. *Nucleic Acids Res.*, **48**, 2807–2829.
5. Zhou,W., Whiteley,A.T., de Oliveira Mann,C.C., Morehouse,B.R., Nowak,R.P., Fischer,E.S., Gray,N.S., Mekalanos,J.J. and Kranzusch,P.J. (2018) Structure of the human cGAS-DNA complex reveals enhanced control of immune surveillance. *Cell*, **174**, 300–311.
6. Lohöfener,J., Steinke,N., Kay-Fedorov,P., Baruch,P., Nikulin,A., Tishchenko,S., Manstein,D.J. and Fedorov,R. (2015) The activation mechanism of 2′-5′-oligoadenylate synthetase gives new insights into OAS/cGAS triggers of innate immunity. *Structure*, **23**, 851–862.
7. Hopfner,K.P. and Hornung,V. (2020) Molecular mechanisms and cellular functions of cGAS-STING signalling. *Nat. Rev. Mol. Cell Biol.*, **21**, 501–521.
8. Tanaka,N., Nakanishi,M., Kusakabe,Y., Goto,Y., Kitade,Y. and Nakamura,K.T. (2004) Structural basis for recognition of 2′,5′-linked oligoadenylates by human ribonuclease L. *EMBO J.*, **23**, 3929–3938.
9. Kato,K., Omura,H., Ishitani,R. and Nureki,O. (2017) Cyclic GMP-AMP as an endogenous second messenger in innate immune signaling by cytosolic DNA. *Annu. Rev. Biochem.*, **86**, 541–566.
10. Kranzusch,P.J., Lee,A.S.Y., Wilson,S.C., Solovykh,M.S., Vance,R.E., Berger,J.M. and Doudna,J.A. (2014) Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. *Cell*, **158**, 1011–1021.
11. Severin,G.B., Ramliden,M.S., Hawver,L.A., Wang,K., Pell,M.E., Kieninger,A.K., Khataokar,A., O'Hara,B.J., Behrmann,L.V., Neiditch,M.B. *et al.* (2018) Direct activation of a phospholipase by cyclic GMP-AMP in El Tor Vibrio cholerae. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E6048–E6055.
12. Burroughs,A.M., Zhang,D., Schäffer,D.E., Iyer,L.M. and Aravind,L. (2015) Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res.*, **43**, 10633–10654.
13. Whiteley,A.T., Eaglesham,J.B., de Oliveira Mann,C.C., Morehouse,B.R., Lowey,B., Nieminen,E.A., Danilchanka,O., King,D.S., Lee,A.S.Y., Mekalanos,J.J. *et al.* (2019) Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature*, **567**, 194–199.
14. Ye,Q., Lau,R.K., Mathews,I.T., Birkholz,E.A., Watrous,J.D., Azimi,C.S., Pogliano,J., Jain,M. and Corbett,K.D. (2020) HORMA domain proteins and a Trip13-like ATPase regulate bacterial cGAS-like enzymes to mediate bacteriophage immunity. *Mol. Cell*, **77**, 709–722.
15. Lau,R.K., Ye,Q., Birkholz,E.A., Berg,K.R., Patel,L., Mathews,I.T., Watrous,J.D., Ego,K., Whiteley,A.T., Lowey,B. *et al.* (2020) Structure and mechanism of a cyclic trinucleotide-activated bacterial endonuclease mediating bacteriophage immunity. *Mol. Cell*, **77**, 723–733.
16. Cohen,D., Melamed,S., Millman,A., Shulman,G., Oppenheimer-Shaanan,Y., Kacen,A., Doron,S., Amitai,G. and Sorek,R. (2019) Cyclic GMP-AMP signalling protects bacteria against viral infection. *Nature*, **574**, 691–695.
17. Lowey,B., Whiteley,A.T., Keszei,A.F.A., Morehouse,B.R., Mathews,I.T., Antine,S.P., Cabrera,V.J., Kashin,D., Niemann,P., Jain,M. *et al.* (2020) CBASS immunity uses CARF-related effectors to sense 3′-5′- and 2′-5′-linked cyclic oligonucleotide signals and protect bacteria from phage infection. *Cell*, **182**, 38–49.
18. McFarland,A.P., Luo,S., Ahmed-Qadri,F., Zuck,M., Thayer,E.F., Goo,Y.A., Hybiske,K., Tong,L. and Woodward,J.J. (2017) Sensing of bacterial cyclic dinucleotides by the oxidoreductase RECON

promotes NF-κB activation and shapes a proinflammatory antibacterial state. *Immunity*, **46**, 433–445.
19. Kranzusch,P.J. (2019) cGAS and CD-NTase enzymes: structure, mechanism, and evolution. *Curr. Opin. Struct. Biol.*, **59**, 178–187.
20. Steitz,T.A. (1998) A mechanism for all polymerases. *Nature*, **391**, 231–232.
21. Belogurov,G.A. and Artsimovitch,I. (2019) The mechanisms of substrate selection, catalysis, and translocation by the elongating RNA polymerase. *J. Mol. Biol.*, **431**, 3975–4006.
22. Kato,K., Ishii,R., Hirano,S., Ishitani,R. and Nureki,O. (2015) Structural basis for the catalytic mechanism of DncV, bacterial homolog of cyclic GMP-AMP synthase. *Structure*, **23**, 843–850.
23. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
24. Liebschner,D., Afonine,P.V., Baker,M.L., Bunkóczi,G., Chen,V.B., Croll,T.I., Hintze,B., Hung,L.W., Jain,S., McCoy,A.J. *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Biol Crystallogr.*, **75**, 861–877.
25. Brünger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
26. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 486–501.
27. Chen,V.B., Arendall,W.B., Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
28. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
29. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
30. Holm,L. (2020) DALI and the persistence of protein shape. *Protein Sci.*, **29**, 128–140.
31. Wolkowicz,U.M. and Cook,A.G. (2012) NF45 dimerizes with NF90, Zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold. *Nucleic Acids Res.*, **40**, 9356–9368.
32. de Oliveira Mann,C.C., Kiefersauer,R., Witte,G. and Hopfner,K.P. (2016) Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein MAB21L1. *Sci. Rep.*, **6**, 27498.
33. Hooy,R.M. and Sohn,J. (2018) The allosteric activation of cGAS underpins its dynamic signaling landscape. *Elife*, **7**, e39984.
34. Brown,A. (2009) Analysis of cooperativity by isothermal titration calorimetry. *Int. J. Mol. Sci.*, **10**, 3457–3477.
35. Hall,J., Ralph,E.C., Shanker,S., Wang,H., Byrnes,L.J., Horst,R., Wong,J., Brault,A., Dumlao,D., Smith,J.F. *et al.* (2017) The catalytic mechanism of cyclic GMP-AMP synthase (cGAS) and implications for innate immunity and inhibition. *Protein Sci.*, **26**, 2367–2380.
36. Vassylyev,D.G., Vassylyeva,M.N., Zhang,J., Palangat,M., Artsimovitch,I. and Landick,R. (2007) Structural basis for substrate loading in bacterial RNA polymerase. *Nature*, **448**, 163–168.
37. Zhu,D., Wang,L., Shang,G., Liu,X., Zhu,J., Lu,D., Wang,L., Kan,B., Zhang,J.R. and Xiang,Y. (2014) Structural biochemistry of a Vibrio cholerae dinucleotide cyclase reveals cyclase activity regulation by folates. *Mol. Cell*, **55**, 931–937.