A bacterial E2 enzyme regulates an antiviral cGAS through a peptide exchange–ligation–cleavage mechanism

Yan Yan[1,#], Fengtao Huang[1,2,#,*], Bingbing Yu[1], Rui Cheng[1], Hui Wu[1], Xueling Lu[1], Xionglue Wang[1], Jun Xiao[3], Longfei Wang[3,*], Bin Zhu[1,2,*]

[1]Key Laboratory of Molecular Biophysics, the Ministry of Education, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

[2]Shenzhen Huazhong University of Science and Technology Research Institute, Shenzhen 518063, China

[3]School of Pharmaceutical Sciences, Wuhan University, Wuhan, China

[#]These authors contribute equally to this work.

[*]To whom correspondence should be addressed. Email: huang_fengtao@126.com; wanglf@whu.edu.cn; bin_zhu@hust.edu.cn

ABSTRACT

The cyclic oligonucleotide based anti-phage signaling systems (CBASSs) are a family of defense systems in prokaryotes. Composed of a cyclic GMP-AMP synthase (cGAS) and CBASS-associated proteins, CBASSs utilize cyclic oligonucleotides to activate antiviral immunity. One major group of CBASS-associated proteins are homologs of eukaryotic E2 ubiquitin-conjugating enzymes. However, the function of E2 in CBASSs remains elusive. Here, we report that a bacterial E2 forms a covalent bond with cGAS, between the active site of E2 and a glycine near the C-terminus of cGAS. Unlike the eukaryotic ubiquitination

system where E1 catalyzes covalent bond formation between E2 and ubiquitin, this covalent bond is directly formed between the "glycinated" cysteine at the E2 active site and the cGAS glycine residue, thereby releasing the C-terminal peptide of the cGAS. We further found that this link stabilizes cGAS, while it inhibits its synthetase activity. In the presence of danger signals such as ADP and PPi, the covalent E2–cGAS complex undergoes auto-proteolysis to release cGAS, which then produces 3′,2′-cGAMP to activate antiviral defense. Our findings reveal a novel form of E2-mediated covalent modification, a negative regulatory role of E2 in CBASS, and an unexpected interplay between cGAS signaling and the ubiquitin system in bacteria.

Cyclic dinucleotides such as c-di-GMP or c-di-AMP are key secondary messenger molecules in many organisms[1-5]. The first hybrid cyclic AMP-GMP molecule 3′,3′-cGAMP, which is synthesized by the cyclic dinucleotide synthase DncV, was discovered in *Vibrio cholerae* [6,7]. It was later found that 2′,3′-cGAMP synthesized by mammalian cyclic GMP-AMP synthase (cGAS) plays a key role in mammalian cGAS-STING innate immunity[8-10]. DncV and cGAS share a conserved catalytic domain and were named cGAS/DncV-like nucleotidyltransferases (CD-NTases)[11,12]. Recently, bioinformatic and functional studies revealed a large family of prokaryotic CD-NTases with diverse cyclic oligonucleotide molecules and signaling pathways[13]. CD-NTases and ancillary proteins involved in bacterial anti-phage defense, which are encoded by operons, from the cyclic-oligonucleotide-based anti-phage signaling system

(CBASS)[14]. Certain CBASSs share ancient evolutionary origins with the eukaryotic cGAS-STING innate immunity pathway[14-16], indicating similar underlying principles between bacterial anti-phage defense and mammalian innate immunity. Over 6,000 distinct CBASSs have been identified in more than 10% of bacteria with known genome sequences[13,14,17,18]. About half of CBASS operons encode the minimal version, which consists of CD-NTase generating signal molecules, and an effector that induces programed cell death as a defense against infection[17,18]. The remaining CBASS operons encode ancillary genes in addition to the effector[17,18]. Some of these ancillary components have been reported to function as a threat sensor or a regulator of bacterial defenses[15,19-21]; however, most of their functions and mechanisms are unclear. Among them are analogs of eukaryotic proteins involved in the ubiquitin protein modification systems[17,18]. Eukaryotic ubiquitin systems are mainly composed of ubiquitin, E1 ubiquitin-activating enzymes, E2 ubiquitin-conjugating enzymes, E3 ubiquitin-ligases and deubiquitinating enzymes (DUBs)[22,23]. Ubiquitin-protein-modification-like systems with conjugation and deconjugation apparatuses have also been observed in prokaryotes[24-27], which are mostly related to the sulfur incorporation reactions of the biosynthesis processes for thiamine and molybdenum/tungsten cofactors. Among over 5,000 predicted CBASSs, 2,199 CBASSs (39%) encode analogs of ubiquitin systems[18], which can be divided into two main categories: one category contains an E1 domain, an E2 domain, and a JAB deubiquitinating peptidase domain, but no E3 domain[18], while the other category consisting of 616 CBASSs only contains an E2 domain protein without E1 or JAB domains[17,18]. The role of such E2-like proteins has not yet been

described.

_Sm_CdnG is a 3′,2′-cGAMP synthase

In E2-containing CBASSs, E2 proteins were mainly associated with clade G CD-NTases (here termed CdnG). We first focused on a CdnG+E2 operon from *Serratia marcescens* to elucidate the role of the CBASS E2 protein (Fig. 1a). The *S. marcescens* operon is found in several patient-derived *S. marcescens* strains and the operon also encodes a putative exonuclease domain and a nucleotide sensor domain (termed SAVED) fused to two transmembrane (TM) helices (termed Saf-2TM) (Fig. 1a)[17].

To characterize the CD-NTase activity of *Sm*CdnG, we purified the recombinant protein with an N-terminal 6×His-tag (Fig. 1b). The potential CD-NTase activity of *Sm*CdnG was tested on various nucleotide substrates with different divalent ions as cofactors. *Sm*CdnG synthesizes a product only when supplied with ATP, GTP, and Mn^{2+} (Fig. 1c, Extended Data Fig. 1a). This product resisted calf intestinal phosphatase (CIP) digestion, indicating that it is a cyclic oligonucleotide without a 5′-terminal phosphate. One phosphodiester bond was cleaved by nuclease P1, which specifically hydrolyzes 3′-5′ but not 2′-5′ phosphodiester bonds (Extended Data Fig. 1b), indicating the presence of a 2′-5′ and a 3′-5′ phosphodiester bond. Mutation of the Dx[D/E] motif at the catalytic center (D84A and D86A) completely abolished the activity of *Sm*CdnG (Extended Data Fig. 1c, d)[28], confirming that the product is synthesized by *Sm*CdnG. Together, these results suggested that *Sm*CdnG is a CD-NTase and the product is a cyclic oligonucleotide with a 2′-5′ and a 3′-5′ phosphodiester bond.

To determine the chemical nature of the *Sm*CdnG product, we compared it with

several available cyclic oligonucleotides derived from ATP and GTP. Reversed-phase liquid chromatography- ultraviolet (RPLC-UV) detection showed that the *Sm*CdnG product had the same retention time as the commercial 3′,2′-cGAMP standard (Fig. 1d). Further, liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis showed that the *Sm*CdnG product and 3′,2′-cGAMP have the same molecular mass and fragmentation fingerprint (Fig. 1e), confirming that *Sm*CdnG synthesizes 3′,2′-cGAMP (Fig. 1f). Consistently, three recent publications also reported enzymatic synthesis of 3′,2′-cGAMP by CD-NTases from *Drosophila*[29,30] and *Asticcacaulis* sp.[31].

We next tested the effect of divalent cations on the catalytic activity of *Sm*CdnG (Extended Data Fig. 1e). The product was detected in the presence of Mn^{2+} but not Mg^{2+} (Extended Data Fig. 1e), confirming that *Sm*CdnG is a Mn^{2+}-dependent CD-NTase. The optimal concentration of Mn^{2+} for *Sm*CdnG activity is 2.5 mM (Extended Data Fig. 1f).

Covalent linkage between *Sm*E2 and *Sm*CdnG

Previous bioinformatic analyses have identified a few CBASS operons encoding a fusion protein of a CD-NTase domain and an E2-like domain, which implies potential interaction between the CD-NTase and E2[17]. To test this hypothesis, we constructed three forms for *Sm*CdnG–*Sm*E2 co-expression: (i) *Sm*CdnG and *Sm*E2 arranged as the native polycistron form in the genome and cloned into pET28a, with an N-terminal 6×His-tag on *Sm*CdnG; (ii) the gene encoding *Sm*CdnG with N-terminal 6×His-tag cloned into the pET28a vector and the gene encoding *Sm*E2 without tag cloned into the pQE82L vector (the T5 promoter in pQE82L replaced by a T7 promoter); and (iii) the gene encoding *Sm*CdnG without tag cloned into the pQE82L vector (the T5 promoter

in pQE82L replaced by a T7 promoter) and the gene encoding *Sm*E2 with N-terminal 6×His-tag cloned into the pET28a vector. We tested whether one protein with His-tag could pull down the other without His-tag through Ni-NTA purification (Fig. 2a). Interestingly, an additional band appeared on the SDS-PAGE gel for all three forms of co-expression and its molecular weight (64 kDa) suggested a fusion protein of *Sm*CdnG (46 kDa) and *Sm*E2 (18 kDa) (Fig. 2b). *Sm*CdnG and *Sm*E2 fusion were not separated by SDS and dithiothreitol (DTT), suggesting a covalent linkage.

To validate the covalent linkage between *Sm*CdnG and *Sm*E2, we purified *Sm*CdnG and *Sm*E2 separately and mixed them at a 1:1 molar ratio *in vitro*. Surprisingly, the covalent link was still formed and increased over time without any cofactors except for ingredients in the enzyme storage buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 0.1% Triton X-100, and 50% glycerol) (Fig. 2c). At a 1:1 molar ratio, the turnover of *Sm*CdnG and *Sm*E2 to the covalently linked *Sm*CdnG-*Sm*E2 was not complete even after long incubation times. Excess *Sm*E2 efficiently promoted the covalent linkage process and turned most free *Sm*CdnG into covalently linked *Sm*CdnG-*Sm*E2 (Extended Data Fig. 2).

Covalent *Sm*E2-*Sm*CdnG complex resembles ubiquitin-charged E2

To elucidate the structural basis of the covalent linkage between *Sm*E2 and *Sm*CdnG, we predicted the structure of the *Sm*CdnG-*Sm*E2 complex using AlphaFold and ColabFold[32,33] (Fig. 2d). All five predicted structures of *Sm*CdnG-*Sm*E2 were nearly identical and most per-residue IDDT scores were over 80, with the exceptions of the terminal disordered regions and internal loop regions (Extended Data Fig. 3a). In the

AlphaFold prediction, *Sm*CdnG resembles CD-NTases and has a root mean square deviation (RMSD) of 2.7 when overlaid to the CD-NTase from *Bacteroides fragilis* (Extended Data Fig. 3b)[8,11,12,34,35]. The predicted structure of *Sm*E2 is very similar to the structure of the eukaryotic E2 with an RMSD of 2.5 (Extended Data Fig. 3c)[36-38]. Interestingly, the C-terminal region (amino acids 374–407) of *Sm*CdnG consists of a three-stranded antiparallel β-sheet, which is in contact with *Sm*E2 and mimics the eukaryotic ubiquitin-E2 complex (Extended Data Fig. 4a, b).

Similar to the ubiquitin-E2 complex, the C-terminal tail of *Sm*CdnG is inserted into the catalytic pocket of *Sm*E2 (Fig. 2d). As eukaryotic E2 ubiquitin-conjugating enzymes use a cysteine residue to form a thioester bond with the C-terminus of ubiquitin, we hypothesized that the thiol group of *Sm*E2 may be covalently linked to the C-terminus of *Sm*CdnG. Among all three cysteine residues in *Sm*E2, only C101 is in contact with the C-terminus of *Sm*CdnG in the predicted *Sm*CdnG-*Sm*E2 heterodimer (Fig. 2d). We also aligned amino acid sequences of various prokaryotic E2 homologs and identified two conserved cysteine residues (C69 and C101 in *Sm*E2) (Fig. 2e). We constructed the three mutants C69A, C101A, and C69A/C101A and co-purified them with *Sm*CdnG. The mutant C69A, but not mutants C101A and C69A/C101A, formed a covalent link with *Sm*CdnG, confirming that residue C101 of *Sm*E2 is critical for the covalent link (Fig. 2f).

_Sm_E2 is covalently linked to _Sm_CdnG through a thioester bond between a cysteine and a glycine

Alignment of the C-terminal sequences of CD-NTase homologs associated with E2 in

various CBASSs showed that the glycine corresponding to *Sm*CdnG G406 is highly conserved (Fig. 2g). To verify the key residue in *Sm*CdnG for covalent linkage between *Sm*CdnG and *Sm*E2, we constructed five *Sm*CdnG mutants: K401A, Y405A, G406/E407-deletion (ΔGE), E407-deletion (ΔE), and a mutant lacking the last eight residues at the C-terminus (ΔC). We co-expressed and co-purified them with *Sm*E2. Among these mutants, only ΔGE and ΔC no longer form a covalent link to *Sm*E2 (Fig. 2h, Extended Data Fig. 5a). In contrast, K401A, Y405A, and ΔE still formed the covalent link, suggesting that G406 is the critical residue for covalent linkage in *Sm*CdnG. Then, we mutated G406 to A, V, and L and found that only glycine and alanine at this position can support bond formation (Fig. 2h), indicating that the size of the residue at this position is critical for the covalent linkage.

To further identify the site of the covalent link formed between *Sm*CdnG and *Sm*E2, we performed LC-MS/MS experiments on covalently linked *Sm*CdnG-*Sm*E2, which showed that C101 in *Sm*E2 and G406 in *Sm*CdnG carry a glycine (57.060 Da) and a cysteine (103.140 Da) modification, respectively (Fig. 2i), indicating that *Sm*E2 is covalently linked to *Sm*CdnG through a thioester bond between C101 in *Sm*E2 and G406 in *Sm*CdnG, the same bond as that between ubiquitin and E2.

Cleavage of the *Sm*CdnG C-terminus

The formation of the thioester bond between *Sm*CdnG G406 and *Sm*E2 C101 requires the removal of the last C-terminal residue E407 of *Sm*CdnG in order to expose the carboxyl group of G406. In eukaryotes, the removal of C-terminal residues of ubiquitin is carried out by accessory proteases[22,23]. To confirm the C-terminal processing of

*Sm*CdnG, we constructed and purified *Sm*CdnG-MBP, which has a linker (SSSSLVPRSH) and a C-terminal maltose binding protein (MBP), to monitor the processing of the *Sm*CdnG C-terminus during bond formation. We first co-expressed and co-purified His-tagged *Sm*CdnG-MBP with non-tagged *Sm*E2. The presence of a protein with a higher molecular weight indicated that the covalent link between *Sm*CdnG-MBP and *Sm*E2 is still formed, and the molecular weight of the covalently linked protein was consistent with that of the covalently linked *Sm*CdnG-*Sm*E2, suggesting that the C-terminus of *Sm*CdnG-MBP was removed (Fig. 3a, lane 5). The covalently linked *Sm*CdnG-MBP-*Sm*E2 protein was also reconstituted *in vitro* by mixing the two proteins at a 1:1 stoichiometry. Since His-tagged *Sm*E2 was used for the *in vitro* experiments and the non-tagged *Sm*E2 was used for the *in vivo* experiments, the resulting covalently linked *Sm*CdnG-*Sm*E2 protein was slightly larger than that obtained *in vivo* (Fig. 3a, lane 7). The presence of a single protein with the same molecular weight as MBP in addition to *Sm*CdnG-MBP, *Sm*E2, and the covalently linked *Sm*CdnG-MBP-*Sm*E2 further confirmed that the C-terminus of *Sm*CdnG was indeed removed during the covalent linking between *Sm*CdnG-MBP and *Sm*E2 (Fig. 3a, lane 7). To determine the position where the C-terminus of *Sm*CdnG-MBP was processed, N-terminal sequencing was performed on the MBP-tagged product generated during the covalent linkage using the Edman degradation method. Surprisingly, the sequencing results showed that the N-terminal sequence of the MBP-containing peptide is GESSSSLVPRSH (Fig. 3b, colored in green). We also constructed the mutant G406A of *Sm*CdnG-MBP, which should also support the formation of a

covalent link with *Sm*E2 (Fig. 2h). N-terminal sequencing of the C-terminal peptide of *Sm*CdnG(G406A)-MBP released after *Sm*CdnG-*Sm*E2 formation showed that the N-terminal sequence of the MBP-containing peptide is AESSSSLVPRSH (Fig. 3b, colored in purple). These results indicated that the cleavage occurs between Y405 and G406 of *Sm*CdnG and G406 is removed from the covalently linked proteins, which apparently contradicted our previous MS results that the covalently linked *Sm*CdnG-*Sm*E2 was formed through a thioester bond between *Sm*E2 C101 and *Sm*CdnG G406. How can the G406 be removed during the covalent linkage between *Sm*CdnG G406 and *Sm*E2 C101? Although no similar mechanism is known for E2 enzymes, we speculate that *Sm*E2 is activated by a "glycine modification," by which the thioester bond between C101 and a freestanding glycine is pre-formed. This glycine-modified cysteine then replaces the G406 residue of *Sm*CdnG to link the glycine-*Sm*E2 to the Y405 residue of *Sm*CdnG. To test this hypothesis, LC-MS/MS analysis was performed on purified *Sm*E2 protein alone to identify protein binding sites. The results showed that C101 in *Sm*E2 indeed carries a glycine (57.060 Da) modification (Fig. 3c). The formation of a high-energy thioester bond between ubiquitin and E1 requires Mg^{2+} and ATP[22,39], while the *in vitro* formation of a bond between *Sm*CdnG and *Sm*E2 requires neither ATP nor Mg^{2+}. Even in the presence of 10 mM EDTA, the covalent link between *Sm*CdnG and *Sm*E2 is still formed (Fig. 3d). These puzzling observations can partially be explained by the pre-formation of a thioester bond between a glycine residue and *Sm*E2. A schematic diagram of the peptide exchange and ligation between *Sm*CdnG-MBP and glycine-*Sm*E2 as described above is shown in Fig. 3e.

To validate that the cleavage of the *Sm*CdnG C-terminus is mediated by *Sm*E2 in the absence of any protease, we co-expressed and co-purified *Sm*CdnG-C10 (where the C-terminus of *Sm*CdnG is extended by a 10-residue-peptide SSSSLVPRSH) with the C101A mutant of *Sm*E2 (Extended Data Fig. 6a), as well as mixing individual proteins at a 1:1 stoichiometry *in vitro* (Extended Data Fig. 6b). We then compared the molecular weight of *Sm*CdnG-C10 with or without *Sm*E2-C101A by checking the small difference in size between processed and unprocessed *Sm*CdnG-C10 and found that the molecular weight of *Sm*CdnG-C10 was not changed in the presence of *Sm*E2-C101A (Extended Data Fig. 6c, d). Together, our data suggest that the covalent linkage and C-terminal cleavage are coupled and *Sm*E2 is responsible for the protease activity to process the C-terminus of *Sm*CdnG. Residue C101 plays a pivotal role in both linkage and cleavage.

Separation of the covalently linked *Sm*CdnG-*Sm*E2 induced by ADP or PPi

To further explore whether this bond formation could be affected by provided nucleotides. ATP, ADP, or AMP (5 mM each) was added to the 1:1 mixture of *Sm*CdnG and *Sm*E2. GTP, GDP, GMP, NTPs, or dNTPs were used as controls. Unexpectedly, the formation of the covalently linked *Sm*CdnG-*Sm*E2 protein was not improved by any of the nucleotides tested but was significantly inhibited in the presence of ADP without any other cofactors except for ingredients in the enzyme storage buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.1mM EDTA, 1 mM DTT, 0.1% Triton X-100, and 50% glycerol) (Fig. 4a). We further tested effects of the related phosphoric acid (Pi) and pyrophosphoric acid (PPi). Individual ADP, Pi, and PPi and mixtures of ADP/Pi or ADP/PPi were added to the 1:1 mixture of *Sm*CdnG and *Sm*E2, and ATP and AMP were

used as controls. The results clearly showed that the covalent linkage was inhibited in the presence of ADP or PPi (Fig. 4b).

The apparent inhibition of the formation of covalently linked *Sm*CdnG-*Sm*E2 by ADP or PPi could be caused by inhibition of the formation of the covalent link or by stimulation of cleavage of the covalent link. To verify these possibilities, we added individual ADP, Pi, or PPi, ADP/Pi mixture, or ADP/PPi mixture to the pre-formed covalently linked *Sm*CdnG-*Sm*E2 protein. ATP and AMP were used as controls. We found that after overnight incubation at 4°C, the covalent link can be cleaved (Fig. 4c). Similar to the results shown in Fig. 4a and b, the covalent link was only cleaved in the presence of ADP or PPi (Fig. 4c). GDP or CDP also showed no effect (Extended Data Fig. 7a). We further tested the optimal concentrations of ADP and PPi for inducing cleavage of the covalent link and determined that the optimal concentration was 2.5 mM for both ADP and PPi (Extended Data Fig. 7b, c). Additionally, we also incubated the covalently linked *Sm*CdnG-*Sm*E2 protein with ADP or PPi at 4°C for various time periods, and found that free *Sm*CdnG and *Sm*E2 levels increased over time (Extended Data Fig. 7d, e). Even in the presence of 10 mM EDTA, the covalent link was cleaved in the presence of ADP or PPi (Fig. 4d), indicating that metal ions are not required for this process.

To understand how the covalently linked *Sm*CdnG-*Sm*E2 was separated in the presence of ADP or PPi, we first co-expressed and co-purified His-tagged *Sm*CdnG-C10 with non-tagged *Sm*E2, and the covalent link between *Sm*CdnG-C10 and *Sm*E2 was formed (Fig. 4e, lane 5). The covalently linked *Sm*CdnG-(C10)-*Sm*E2 protein was

also reconstituted *in vitro* by mixing the two proteins at a 1:1 stoichiometry (Fig. 4e, lane 7). Then we added ADP to the pre-formed covalently linked *Sm*CdnG-C10-*Sm*E2 protein to break the covalent link (Fig. 4f). Since the C-terminus of *Sm*CdnG was processed, the molecular weight (estimated by SDS-PAGE) of *Sm*CdnG-C10 produced after cleavage of the covalent link is lower than the molecular weight of the original *Sm*CdnG-C10 expressed and purified alone (Fig. 4g). We also used ADP to induce cleavage of the covalent link of the *Sm*CdnG-C10-*Sm*E2 protein formed *in vitro* and observed two bands of similar but distinct size (marked by arrows), indicating the processed (red arrow) and unprocessed (black arrow) *Sm*CdnG-C10 (Fig. 4h). To determine the cleavage position of the covalent linkage, protein C-terminal sequencing was performed on the *Sm*CdnG-C10 produced after cleavage of the covalent link. The results showed that the C-terminal sequence of *Sm*CdnG-C10 produced after cleavage of the covalent link is AGTAAQAGVPKNTFY (Fig. 4i), revealing that ADP or PPi induced cleavage at the peptide bond between G406 and Y405 in *Sm*CdnG, not between G406 and C101 in *Sm*E2. A schematic diagram of the cleavage of the covalently linked *Sm*CdnG-C10-*Sm*E2 protein induced by ADP or PPi is shown in Fig. 4j.

_Sm_CdnG-_Sm_E2 shows a lower 3′,2′-cGAMP synthesis activity but a higher stability compared to _Sm_CdnG

To explore the function of *Sm*E2 in the antiviral system, we first compared the catalytic activity of the covalently linked *Sm*CdnG-*Sm*E2 protein with *Sm*CdnG alone. The results showed that the yield of 3′,2′-cGAMP synthesized by the covalently linked *Sm*CdnG-*Sm*E2 protein obtained from three forms of co-expression of *Sm*CdnG and

*Sm*E2 significantly decreased compared to *Sm*CdnG (Fig. 5a, Extended Data Fig. 8a). Additionally, when *Sm*E2 was added to the *Sm*CdnG reaction system at a ratio of 1:1 (*Sm*CdnG+*Sm*E2), the yield of 3′,2′-cGAMP synthesized was significantly decreased compared to *Sm*CdnG alone (Fig. 5a, Extended Data Fig. 8a). These results suggest that *Sm*E2 negatively regulates the catalytic activity of *Sm*CdnG. To further explore the differences in catalytic activity between *Sm*CdnG and the covalently linked *Sm*CdnG-*Sm*E2 protein, we compared the steady-state kinetics of *Sm*CdnG and the *Sm*CdnG-*Sm*E2 complex by quantifying the 3′,2′-cGAMP synthesis after HPLC analysis. Because both ATP and GTP are used as substrates, we evaluated the steady-state kinetics of both of them (Extended Data Fig. 8b, c). For *Sm*CdnG, K_m and k_{cat} for ATP are 393.0 ± 15.3 μM and 2.4 ± 0.3 min^{-1}, respectively. For *Sm*CdnG-*Sm*E2, K_m for ATP is 573.9 ± 19.7 μM and k_{cat} is 0.32 ± 0.02 min^{-1} (Fig. 5b). The k_{cat}/K_{mATP} ratio of *Sm*CdnG (5.9 × 10^{-3} $min^{-1}μM^{-1}$) is 10 times that of *Sm*CdnG-*Sm*E2 (6.5 × 10^{-4} $min^{-1}μM^{-1}$). For GTP, although the k_{cat} value is similar to that of ATP (2.6 ± 0.2 min^{-1} for *Sm*CdnG and 0.25 ± 0.01 min^{-1} for *Sm*CdnG-*Sm*E2), K_m is 4-fold lower for *Sm*CdnG (94.2 ± 11.1 μM) and 7-fold lower for *Sm*CdnG-*Sm*E2 (83.0 ± 12.5 μM) compared to those of ATP, respectively (Fig. 5b), indicating a higher affinity to GTP over ATP. The catalytic efficiency of *Sm*CdnG is also 10 times higher than that of *Sm*CdnG-*Sm*E2 (k_{cat}/K_{mGTP} values of *Sm*CdnG and *Sm*CdnG-*Sm*E2 are 2.8 × 10^{-2} and 3.0 × 10^{-3}, respectively, Fig. 5b). These results indicate that *Sm*CdnG has a higher activity than *Sm*CdnG-*Sm*E2.

We next purified G406V and G406L mutants of *Sm*CdnG (Extended Data Fig. 5b)

and analyzed their 3′,2′-cGAMP synthesis activities. The results showed that G406V and G406L mutations, which abolish the covalent linkage, did not affect *Sm*CdnG activity (Fig. 5c, Extended Data Fig. 8d). When *Sm*CdnG G406V and G406L mutants were co-expressed and co-purified with *Sm*E2, we obtained non-covalent *Sm*CdnG (G406V)–*Sm*E2 and *Sm*CdnG (G406L)–*Sm*E2 complexes (Fig. 2h). The yield of 3′,2′-cGAMP synthesized by *Sm*CdnG (G406V)–*Sm*E2 and *Sm*CdnG (G406L)–*Sm*E2 complexes actually increased compared to the covalently linked *Sm*CdnG-*Sm*E2 protein (Fig. 5c, Extended Data Fig. 8d). We also measured 3′,2′-cGAMP synthesis activities of the co-purified E2 mutants with *Sm*CdnG (Fig. 5d, Extended Data Fig. 8e). The protein complexes *Sm*CdnG-*Sm*E2 (C101A) and *Sm*CdnG-*Sm*E2 (C101A/C69A) also showed higher 3′,2′-cGAMP synthesis activities than the covalently linked *Sm*CdnG-*Sm*E2 protein, while the covalently linked *Sm*CdnG-*Sm*E2 (C69A) protein had comparable 3′,2′-cGAMP synthesis activity compared to *Sm*CdnG-*Sm*E2 (Fig. 5d, Extended Data Fig. 8e). These results confirmed that *Sm*E2 negatively regulates the signal molecule synthesis activity of *Sm*CdnG and the regulation depends on covalent linkage.

The covalently linked *Sm*CdnG-*Sm*E2 showed a lower 3′,2′-cGAMP yield compared to *Sm*CdnG during a short period of time (Fig. 5e). However, as the reaction time was extended, the 3′,2′-cGAMP yield of *Sm*CdnG-*Sm*E2 exceeded that of *Sm*CdnG (Fig. 5e), indicating that the stability of *Sm*CdnG was increased after covalently linking with *Sm*E2. The stability of *Sm*CdnG and *Sm*CdnG-*Sm*E2 was further investigated by a two-step assay. *Sm*CdnG or *Sm*CdnG-*Sm*E2 was first pre-incubated with various ions and

nucleotides at 37°C for 30 min; then, the full reaction condition (Mn^{2+}, ATP, and/or GTP) was provided or added to start the reaction, and the yield of 3′,2′-cGAMP was quantified after 4 h. After pre-incubation with Mn^{2+} alone, or ATP plus GTP without Mn^{2+}, ATP plus GTP plus Mg^{2+} at 37°C for 30 min, both *Sm*CdnG and *Sm*CdnG-*Sm*E2 lost the 3′,2′-cGAMP synthesis activity even after the missing component was added (Fig. 5f), indicating a low stability of *Sm*CdnG. When Mn^{2+} and one of the substrates ATP or GTP were present during the pre-incubation, both *Sm*CdnG and *Sm*CdnG-*Sm*E2 retained 3′,2′-cGAMP synthesis activity when the missing component was added. However, the 3′,2′-cGAMP yield of *Sm*CdnG-*Sm*E2 was significantly higher than that of *Sm*CdnG (Fig. 5f). If Mn^{2+} and one of the non-substrate nucleotides CTP or UTP were present during the pre-incubation, only *Sm*CdnG-*Sm*E2 retained the activity to synthesize 3′,2′-cGAMP when ATP and GTP were added (Fig. 5f). We also compared the time course of 3′,2′-cGAMP synthesis of *Sm*CdnG and *Sm*CdnG-*Sm*E2 after pre-incubation with Mn^{2+}/ATP or Mn^{2+}/GTP at 37°C for 30 min. In contrast to the results obtained from directly initiated reactions (Fig. 5e), after pre-incubation, *Sm*CdnG-*Sm*E2 consistently showed a higher 3′,2′-cGAMP yield compared to *Sm*CdnG at various reaction times in both cases (Fig. 5g, h). These results suggest a model in which *Sm*E2 is covalently linked to *Sm*CdnG to suppress its activity while protecting its stability.

Similar but simplified regulation mode by related bacterial E2

Bacterial CBASSs systems also encode natural fusion proteins consisting of a CD-NTase domain and an E2-like domain[17], and the action mode of the E2 domain in such systems also attracted our interest. We first purified the N-terminal His-tagged protein

consisting of a clade G CD-NTase domain and an E2 domain from *Alphaproteobacteria bacterium HGW-Alphaproteobacteria-1* (*Ha*CdnG-E2) and found that in addition to the fusion protein with a molecular weight of 66.5 kDa, there were also a smaller protein with a molecular weight of 40–50 kDa and several large proteins in the form of polymers (Fig. 6a). Interestingly, we found that the amount of the 66.5-kDa-fusion protein decreased, while the 40–50-kDa-protein increased as time extended, indicating that the 40~50 kDa-protein was formed by self-cleavage of the fusion protein. To test this hypothesis, protein C-terminal sequencing was performed on the 40–50-kDa-protein. Sequencing results showed that the C-terminal sequence is SLSAQAK (Fig. 6b), revealing that *Ha*CdnG-E2 was indeed self-cleaved at K432.

To investigate the self-cleavage mechanism of *Ha*CdnG-E2, we predicted the structure of *Ha*CdnG-E2 using AlphaFold and ColabFold (Fig. 6c). Strikingly, in the predicted structures, we found that K432 lies at the interface between the CdnG and E2 domains, and G411 in the CdnG domain is very close to C523 in the E2 domain (Fig. 6c). Therefore, three mutants, G411L, K432A, and C523A, of *Ha*CdnG-E2 were constructed and purified. Consistently, *Ha*CdnG-E2 did not self-cleave after G411 or C523 was mutated, while mutation of K432 residue did not affect its self-cleavage (Fig. 6d). These results suggest that *Ha*CdnG-E2 may adopt a similar mechanism to that of *Sm*CdnG-E2 to regulate the activity of the cyclic oligonucleotide synthase through the covalent link (in this case, pre-formed) and cleavage. The cleavage between domains in *Ha*CdnG-E2 also depends on the interaction between a glycine and a cysteine. The cleavage site K432 of *Ha*CdnG-E2, although not adjacent to G411 and C523 in

sequence, is also close to these two key residues in the structure. The schematic diagram of the presumed regulation mode of *Ha*CdnG-E2 is shown in Fig. 6e.

Discussion

A novel covalent modification of a cGAS by a ubiquitin-conjugating enzyme in prokaryotes

The secondary messenger system and the ubiquitin system are two well-known but unrelated systems in higher eukaryotes. As the core components of the two systems, a eukaryotic cyclic-nucleotide synthase and an E2 ubiquitin-conjugating enzyme had not been reported to regulate each other. In this work, our data support a peptide exchange–ligation mechanism between an E2 enzyme and a CD-NTase in bacteria that is distantly reminiscent of the roles of E2 enzymes in the eukaryotic ubiquitin conjugation system in which the E2 ubiquitin-conjugating enzyme receives a ubiquitin-like (Ubl) protein from its cognate E1 and transfers it to a substrate with the aid of an E3 enzyme[22]. In this pathway, precursor Ubls are processed by deubiquitinating enzymes to expose the carboxylate group of a C-terminal glycine and to be adenylated with ATP by E1 and form a Ubl-AMP intermediate. The AMP of the intermediate is immediately attacked and replaced by the sulfhydryl group of the E1 active-site cysteine, forming a thioester bond between E1 and the Ubl for Ubl's substrate conjugation[40]. The Ubl is then transferred to the catalytic cysteine of the E2 enzyme and covalently linked to it through a thioester bond as well. Similar processes with the apparatuses for conjugation and deconjugation have also been described in prokaryotes, while most processes are

related to the sulfur incorporation reactions. In these pathways, a sulfur carrier protein, ThiS or MoaD, carries the sulfur in the form of a thiocarboxylate of a terminal glycine, just like the thioester links of the eukaryotic ubiquitin-conjugation system[23,24,26,27]. In the thiamine pathway, the covalent link between ThiS and ThiF is an acyl persulfide link[41-43]. Although the prokaryotic sulfur transfer systems are considered as the prokaryotic antecedents of the eukaryotic ubiquitin-conjugation system, there is no direct proof of an evolutionary connection between them.

Here, we demonstrated that the a bacterial E2 enzyme is covalently linked to a CD-NTase through a thioester bond between a glycine and a cysteine (Fig. 2), resembling the bond between eukaryotic E2 enzymes and ubiquitin. However, the formation of a high-energy thioester bond in eukaryotes requires ATP and Mg^{2+}, while the *in vitro* bond formation between *Sm*CdnG and *Sm*E2 does not need any supplied energy or metal ions (Fig. 3e). The requirement or divalent ions as in the ubiquitin pathway can also be excluded since the bond is still formed in the presence of 10 mM EDTA (Fig. 3e). Another question is how the carboxyl group of the tyrosine is exposed for bond formation. The tyrosine is the third last residue in the C-terminus of *Sm*CdnG; so, to form a bond with the glycine in *Sm*E2, the last two C-terminal residues (glycine and glutamate) must be removed. In the ubiquitin pathway, the removal of extra C-terminal residues of ubiquitin is carried out by certain proteases[22,23]. In contrast, neither protease nor protease-like domains have been identified in the *Sm*CdnG-*Sm*E2 system. Also, removal of the C-terminal peptide of *Sm*CdnG and ligation between *Sm*CdnG and *Sm*E2 occur in a single step, apparently behaving as an exchange between the glycine carried

by *Sm*E2 and G406 of *Sm*CdnG. The pre-formed thioester bond between *Sm*E2 and a glycine partially explains why there is no requirement for energy, while the peptide exchange mechanism between the two glycine residues is still intriguing. To the best of our knowledge, such a glycine exchange mechanism is unique among known peptide ligation mechanisms. Future structural studies are needed to elucidate the details of such mechanism.

Signals from nucleotide metabolism induce cleavage of the covalently linked *Sm*CdnG-*Sm*E2

In mammals, CD-NTases produce different cyclic dinucleotides in response to different types of nucleic acids in the cytoplasm, resulting in an antiviral immune response[8,9,29,30]. The ancillary proteins in CBASSs may play important roles in sensing virus invasion signals. Previous studies have shown that specific bacteriophage proteins can activate CBASSs as a signal for phage invasion that trigger cAAA signaling in the CD-NTase+HORMA+Trip13 system[20]. Here, we also propose that PPi and ADP may be signals of phage invasion. We observed that the covalently linked *Sm*CdnG-*Sm*E2 protein can be cleaved in the presence of PPi or ADP (Fig. 4j). The normal cellular ADP concentration is maintained low (between undetectable and 0.12 mM)[44,45], and the intracellular PPi concentration is approximately 0.5 mM[46]. Phages usually use the host's nucleotides to synthesize their own genetic material after invasion. In this scenario, the intracellular PPi concentrations resulting from phage DNA replication and transcription and the ADP levels resulting from energy-consuming processes such as DNA unwinding by helicases could be temporarily high. If reaching a concentration of

2.5 mM, which is the optimal concentration for separation of *Sm*CdnG and *Sm*E2 (Extended Data Fig. 7b, c), ADP and PPi are ideal danger signals to activate *Sm*CdnG by releasing the covalently linked *Sm*E2. The messenger 3′,2′-cGAMP synthesized by *Sm*CdnG can in turn activate the downstream effector proteins in the CBASS. After cleavage, the C-terminal residue of *Sm*CdnG is Y405 (Fig 4i); the lack of glycine G406 eliminates its ability to form the covalent link with *Sm*E2 (Fig. 2h). The instability of *Sm*CdnG without protection from *Sm*E2 ensures a shut-off mechanism of the signaling molecule synthesis. How PPi and ADP induce cleavage between *Sm*CdnG Y405 and G406 in the covalently linked *Sm*CdnG-*Sm*E2 remains to be elucidated in detail; it could resemble the cleavage between *Sm*CdnG Y405 and G406 during glycine exchange and peptide ligation.

Regulation of *Sm*CdnG by *Sm*E2

Based on our results, we propose a model on the regulation of *Sm*CdnG by *Sm*E2 through covalent linkage and cleavage (Fig. 5i). *Sm*CdnG is a structurally unstable protein, which is protected and maintained stable by covalently attached *Sm*E2. The covalently linked complex inhibits the cGAMP synthesis activity of *Sm*CdnG, preventing the production of signaling molecules to interfere with normal cell growth. After phage invasion, high nucleotide metabolism, including phage DNA replication and transcription releases PPi and ADP, which induce *Sm*CdnG-*Sm*E2 cleavage. The free-standing *Sm*CdnG thus produces cGAMP to activate the effector protein Saf-2TM-SAVED in the CBASS to combat phage infection[47]. The free-standing *Sm*CdnG quickly loses its activity due to its instability, ensuring a shut-off mechanism for the signaling

pathway. In addition to CdnG, the E2 enzyme, and the Saf-2TM-SAVED effector protein, the *S. marcescens* CBASS operon also encodes an unknown component predicted to be an RNaseH 3'-5' exonuclease (Fig. 1a). We also purified the protein and investigated its function *in vitro*. Neither RNase H nor any nuclease activity was detected, and no effect was observed when this protein was added to the reactions including cGAMP synthesis, covalent linking, and cleavage of *Sm*CdnG-*Sm*E2. However, this element is not always associated with E2 enzymes and only exists in a few CBASSs[17], indicating its regulatory but not key role in the E2-featured CBASS.

Additionally, we investigated the regulatory mode of a naturally occurring simplified version—a single protein consisting of an N-terminal CD-NTase domain and a C-terminal E2 domain. Interestingly, self-cleavage between the fused CD-NTase and E2 domains was observed (Fig. 6), which is also mediated by a glycine in the CD-NTase domain and a structurally close cysteine in the E2 domain. These results suggest that covalent linkage and cleavage between the bacterial CD-NTase and E2 is a common mechanism in the prokaryotic antiviral CBASSs.

Evolutionary origin and relationship of bacterial E2 protein and CD-NTase

The CBASS operon in this study only has an E2 domain protein without E1 or JAB (Fig. 1a). Systems like this account for about 11% of all CBASSs[18]. Although the secondary structure of the bacterial E2 protein in this study is predicted to be similar to but different from the secondary structure of eukaryotic E2 proteins[38], the conserved cysteine residue of the bacterial proteins precisely matched that of the eukaryotic proteins. Besides, we have shown that a conserved cysteine of bacterial E2 binds to a

glycine residue of the C-terminal of its cognate CD-NTase through a thioester bond (Fig. 2e–i), which is also consistent with the linkage between eukaryotic E2 and ubiquitin. These observations imply a direct evolutionary connection between the eukaryotic ubiquitin-signaling system and the prokaryotic E2-featured antiviral system. However, the functional connections between them are still unclear. Here, the bacterial E2 protein is covalently linked to its cognate CD-NTase, thereby stabilizing the CD-NTase but inhibiting its cGAMP signaling molecule synthesis. This is quite different from eukaryotic E2 proteins, which serve as linkers and transport machines. Moreover, *Sm*CdnG is also the prokaryotic homolog of cGAS in mammalian innate immune pathways. How did these two unrelated signaling pathways evolve is also an interesting question.

Here, we have defined the regulatory mechanism and biochemical characteristics of a subset of a class of bacterial CBASSs containing only an E2-like protein. However, other widely available CBASSs encode a fusion protein (Cap2) with an E1-like domain and an E2-like domain and a JAB-like domain peptidase (Cap3), which usually removes ubiquitin from target proteins in eukaryotes[17,18]. Previous research has demonstrated that E1/E2 domain proteins or JAB domain proteins are necessary for defense against some phages but not all[14]. Interestingly, a recent study showed that Cap2 ligates the cGAS C-terminus to a target molecule in cells through a process called cGASylation and increases the synthesis of cGAMP by cGAS. Cap2 activity is attenuated by Cap3, which deconjugates cGAS and antagonizes antiviral signaling[48]. Although both can achieve the goal of cGAS activity regulation, this mechanism depending on the

activities of three enzymes[48] is more closely related to the eukaryotic ubiquitin pathway than the system that we report here, which relies on a single E2 protein. More studies are required to reveal the divergence and common features of different CBASSs, which will provide clues with respect to the evolutionary relationship between prokaryotic CBASSs and eukaryotic innate immune pathways.

Precise information regarding the interaction between the conserved glycine and cysteine residues was provided by both predicted structures in this work, for both the single *Ha*CdnG-E2 protein (Fig. 6c) and the *Sm*CdnG-*Sm*E2 complex (Fig. 2d), which demonstrates the value of the latest AI-based structure prediction[32,33] as a powerful tool to elucidate novel protein–protein interactions, especially in bacteria, since abundant homologous sequences are available.

REFERENCES

1 Witte, G., Hartung, S., Buttner, K. & Hopfner, K. P. Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Molecular cell* **30**, 167-178, doi:10.1016/j.molcel.2008.02.020 (2008).

2 Schirmer, T. & Jenal, U. Structural and mechanistic determinants of c-di-GMP signalling. *Nature reviews. Microbiology* **7**, 724-735, doi:10.1038/nrmicro2203 (2009).

3 Corrigan, R. M. & Grundling, A. Cyclic di-AMP: another second messenger enters the fray. *Nature reviews. Microbiology* **11**, 513-524, doi:10.1038/nrmicro3069 (2013).

4 Jenal, U., Reinders, A. & Lori, C. Cyclic di-GMP: second messenger extraordinaire. *Nature reviews. Microbiology* **15**, 271-284, doi:10.1038/nrmicro.2016.190 (2017).

5 He, J., Yin, W., Galperin, M. Y. & Chou, S. H. Cyclic di-AMP, a second messenger of primary importance: tertiary structures and binding mechanisms. *Nucleic acids research* **48**, 2807-2829, doi:10.1093/nar/gkaa112 (2020).

6 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, doi:10.1016/j.cell.2012.01.053 (2012).

7 Zhu, D. *et al.* Structural biochemistry of a Vibrio cholerae dinucleotide cyclase reveals cyclase activity regulation by folates. *Molecular cell* **55**, 931-937, doi:10.1016/j.molcel.2014.08.001 (2014).

8 Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786-791, doi:10.1126/science.1232458 (2013).

9 Wu, J. *et al.* Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339**, 826-830, doi:10.1126/science.1229963 (2013).

10 Diner, E. J. *et al.* The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell reports* **3**, 1355-1361, doi:10.1016/j.celrep.2013.05.009 (2013).

11 Kranzusch, P. J., Lee, A. S., Berger, J. M. & Doudna, J. A. Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. *Cell reports* **3**, 1362-1368, doi:10.1016/j.celrep.2013.05.008 (2013).

12 Kranzusch, P. J. *et al.* Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. *Cell* **158**, 1011-1021, doi:10.1016/j.cell.2014.07.028 (2014).

13 Whiteley, A. T. *et al.* Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194-199, doi:10.1038/s41586-019-0953-5 (2019).

14 Cohen, D. *et al.* Cyclic GMP-AMP signalling protects bacteria against viral infection. *Nature* **574**, 691-695, doi:10.1038/s41586-019-1605-5 (2019).

15 Morehouse, B. R. *et al.* STING cyclic dinucleotide sensing originated in bacteria. *Nature* **586**, 429-433, doi:10.1038/s41586-020-2719-5 (2020).

16 Ko, T. P. *et al.* Crystal structure and functional implication of a bacterial cyclic AMP-AMP-GMP synthetase. *Nucleic acids research* **49**, 4725-4737, doi:10.1093/nar/gkab165 (2021).

17 Burroughs, A. M., Zhang, D., Schaffer, D. E., Iyer, L. M. & Aravind, L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic acids research* **43**, 10633-10654, doi:10.1093/nar/gkv1267 (2015).

18 Millman, A., Melamed, S., Amitai, G. & Sorek, R. Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. *Nature microbiology* **5**, 1608-1615, doi:10.1038/s41564-020-0777-y (2020).

19 Lowey, B. *et al.* 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 e17, doi:10.1016/j.cell.2020.05.019 (2020).

20 Ye, Q. *et al.* HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial cGAS-like Enzymes to Mediate Bacteriophage Immunity. *Molecular cell* **77**, 709-722 e707, doi:10.1016/j.molcel.2019.12.009 (2020).

21 Lau, R. K. *et al.* Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial

Endonuclease Mediating Bacteriophage Immunity. *Molecular cell* **77**, 723-733 e726, doi:10.1016/j.molcel.2019.12.010 (2020).

22 Kerscher, O., Felberbaum, R. & Hochstrasser, M. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annual review of cell and developmental biology* **22**, 159-180, doi:10.1146/annurev.cellbio.22.010605.093503 (2006).

23 Iyer, L. M., Burroughs, A. M. & Aravind, L. The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like beta-grasp domains. *Genome biology* **7**, R60, doi:10.1186/gb-2006-7-7-r60 (2006).

24 Goehring, A. S., Rivers, D. M. & Sprague, G. F., Jr. Attachment of the ubiquitin-related protein Urm1p to the antioxidant protein Ahp1p. *Eukaryotic cell* **2**, 930-936, doi:10.1128/EC.2.5.930-936.2003 (2003).

25 Rudolph, M. J., Wuebbens, M. M., Rajagopalan, K. V. & Schindelin, H. Crystal structure of molybdopterin synthase and its evolutionary relationship to ubiquitin activation. *Nature Structural Biology* **8**, 42 (2001).

26 Lake, M. W., Wuebbens, M. M., Rajagopalan, K. V. & Schindelin, H. Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex. *Nature* **414**, 325-329 (2001).

27 Furukawa, K., Mizushima, N., Noda, T. & Ohsumi, Y. A protein conjugation system in yeast with homology to biosynthetic enzyme reaction of prokaryotes. *The Journal of biological chemistry* **275**, 7462-7465, doi:10.1074/jbc.275.11.7462 (2000).

28 Holm, L. & Sander, C. DNA polymerase beta belongs to an ancient nucleotidyltransferase superfamily. *Trends Biochem Sci* **20**, 345-347, doi:10.1016/s0968-0004(00)89071-4 (1995).

29 Slavik, K. M. *et al.* cGAS-like receptors sense RNA and control 3'2'-cGAMP signalling in Drosophila. *Nature* **597**, 109-113, doi:10.1038/s41586-021-03743-5 (2021).

30 Holleufer, A. *et al.* Two cGAS-like receptors induce antiviral immunity in Drosophila. *Nature* **597**, 114-118, doi:10.1038/s41586-021-03800-z (2021).

31 Fatma, S., Chakravarti, A., Zeng, X. & Huang, R. H. Molecular mechanisms of the CdnG-Cap5 antiphage defense system employing 3',2'-cGAMP as the second messenger. *Nature communications* **12**, 6381, doi:10.1038/s41467-021-26738-2 (2021).

32 Mirdita, M., Ovchinnikov, S. & Steinegger, M. ColabFold - Making protein folding accessible to all. *bioRxiv*, 2021.2008.2015.456425, doi:10.1101/2021.08.15.456425 (2021).

33 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

34 Kato, K. *et al.* Structural and functional analyses of DNA-sensing and immune activation by human cGAS. *PloS one* **8**, e76983, doi:10.1371/journal.pone.0076983 (2013).

35 Li, X. D. *et al.* Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects. *Science* **341**, 1390-1394, doi:10.1126/science.1244040 (2013).

36 Im, Y. J. *et al.* Crystallographic and functional analysis of the ESCRT-I /HIV-1 Gag PTAP interaction. *Structure* **18**, 1536-1547, doi:10.1016/j.str.2010.08.010 (2010).

37 Sheng, Y. *et al.* A human ubiquitin conjugating enzyme (E2)-HECT E3 ligase structure-function screen. *Molecular & cellular proteomics : MCP* **11**, 329-341, doi:10.1074/mcp.O111.013706 (2012).

38 Nameki, N. *et al.* Solution structure of the RWD domain of the mouse GCN2 protein. *Protein science : a publication of the Protein Society* **13**, 2089-2100,

doi:10.1110/ps.04751804 (2004).

39 Ciechanover, A., Elias, S., Heller, H. & Hershko, A. "Covalent affinity" purification of ubiquitin-activating enzyme. *Journal of Biological Chemistry* **257**, 2537-2542, doi:10.1016/s0021-9258(18)34957-3 (1982).

40 Haas, A. L., Warms, J. V. B., Hershko, A. & Rose, I. A. Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation. *Journal of Biological Chemistry* **257**, 2543-2548 (1982).

41 Xi, J., Ge, Y., Kinsland, C., Mclafferty, F. W. & Begley, T. P. Biosynthesis of the thiazole moiety of thiamin in Escherichia coli: Identification of an acyldisulfide-linked protein-protein conjugate that is functionally analogous to the ubiquitin/E1 complex. *Proceedings of the National Academy of Sciences* (2001).

42 Duda, D. M., Walden, H., Sfondouris, J. & Schulman, B. A. Structural analysis of Escherichia coli ThiF. *Journal of molecular biology* **349**, 774-786, doi:10.1016/j.jmb.2005.04.011 (2005).

43 Lehmann, C., Begley, T. P. & Ealick, S. E. Structure of the Escherichia coli ThiS-ThIF complex, a key component of the sulfur transfer system in thiamin biosynthesis. *Biochemistry* **45**, 11 (2006).

44 Zimmerman, J. J., Arnim, A. V. & Mclaughlin, J. Cellular Respiration. *Pediatric Critical Care Medicine*, 1058-1072 (2011).

45 Meyrat, A. & von Ballmoos, C. ATP synthesis at physiological nucleotide concentrations. *Scientific reports* **9**, 3070, doi:10.1038/s41598-019-38564-0 (2019).

46 KUKKO, E. & HEINONEN, J. The intracellular concentration of pyrophosphate in the

batch culture of Escherichia coli. *European Journal of Biochemistry* **127**, 347-349 (1982).

47 Duncan-Lowey, B., McNamara-Bordewick, N. K., Tal, N., Sorek, R. & Kranzusch, P. J. Effector-mediated membrane disruption controls cell death in CBASS antiphage defense. *Molecular cell* **81**, 5039-5051.e5035, doi:10.1016/j.molcel.2021.10.020 (2021).

48 Ledvina, H. E. *et al.* cGASylation by a bacterial E1-E2 fusion protein primes antiviral immune signaling. *bioRxiv*, 2022.2003.2031.486616, doi:10.1101/2022.03.31.486616 (2022).

Figure legends

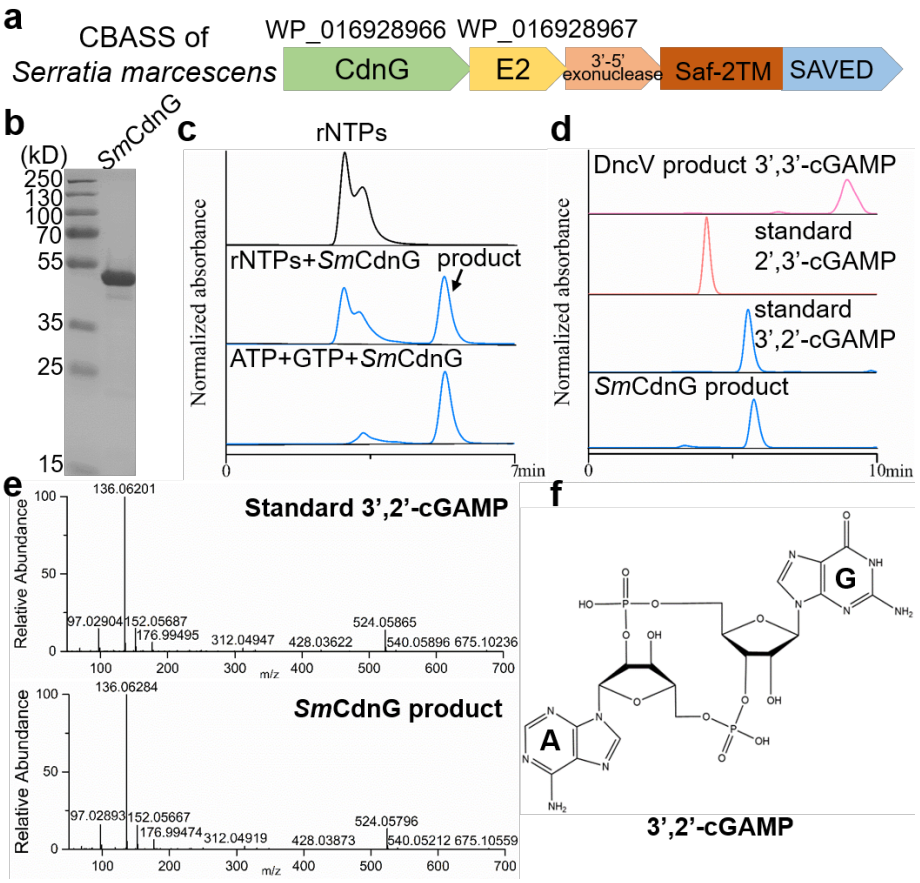


Fig. 1 *Sm*CdnG synthesizes 3′,2′-cGAMP. a, Schematic diagram showing the CBASS operon from *S. marcescens*. The operon contains four genes that encode a CD-NTase,

an E2-like protein, a potential 3′-5′ exonuclease, and an effector protein (Saf-2TM-SAVED). **b**, SDS-PAGE analysis of the purified *Sm*CdnG. **c**, HPLC analysis of the *Sm*CdnG product. *Sm*CdnG synthesizes a product only in the presence of both ATP and GTP. Results are representative of three independent experiments. **d**, Comparison of the retention time of possible cGAMP variants (3′,3′-cGAMP, 2′,3′-cGAMP, and 3′,2′-cGAMP) and the *Sm*CdnG product. **e**, MS/MS fragmentation spectra of the 3′,2′-cGAMP standard (top) and the *Sm*CdnG product (bottom). **f**, Chemical structure of 3′,2′-cGAMP.

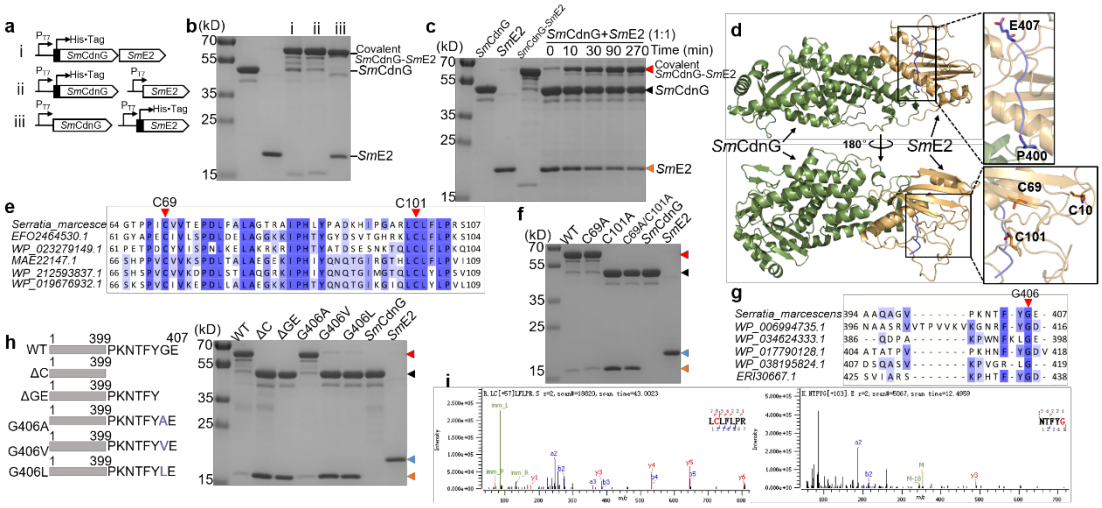


Fig. 2 *Sm*CdnG and *Sm*E2 are covalently linked through a thioester bond between the G406 and C101 residues. a, Schematic diagram showing the constructs for co-expression of *Sm*CdnG and *Sm*E2. *Sm*CdnG and *Sm*E2 were co-expressed in different forms (i-iii). **b**, SDS-PAGE analysis of the purified proteins from various constructs as shown in (**a**). **c**, SDS-PAGE analysis of the *in vitro* assembly of the covalently linked *Sm*CdnG-*Sm*E2 protein. The *in vivo* formation of the covalently linked *Sm*CdnG-*Sm*E2 protein was achieved by the co-expression and co-purification of His-tagged *Sm*CdnG

and non-tagged *Sm*E2 (*Sm*CdnG-*Sm*E2). For the *in vitro* assembly, two proteins were mixed at a molar ratio of 1:1 and incubated on ice for various periods as indicated (*Sm*CdnG+*Sm*E2). Since both *Sm*CdnG and *Sm*E2 were His-tagged during the *in vitro* reconstruction, the molecular weight of the covalently linked *Sm*CdnG-*Sm*E2 protein was slightly higher than that obtained *in vivo* from His-tagged *Sm*CdnG and non-tagged *Sm*E2. The red triangle on the right of the gel indicates the covalently linked *Sm*CdnG-*Sm*E2 protein. The black and orange triangles indicate the *Sm*CdnG and *Sm*E2, respectively. **d**, Ribbon diagram of the predicted *Sm*CdnG-*Sm*E2 complex structure highlighting the C-terminal region (purple) of *Sm*CdnG (green) that is inserted into the *Sm*E2 domain (orange). The potential interaction region (boxed) was enlarged. **e**, Sequence alignment of the C-terminal residues of *Sm*CdnG and other E2-associated CD-NTase homologs. A conserved C-terminal glycine is marked by a red arrow. **f**, SDS-PAGE analysis of the co-purified proteins consisting of *Sm*CdnG with *Sm*E2 or *Sm*E2 mutants. The red triangle on the right of the gel indicates the covalently linked *Sm*CdnG-*Sm*E2 protein. The black, blue, and orange triangles on the right of the gel indicate His-tagged *Sm*CdnG, His-tagged *Sm*E2, and non-tagged *Sm*E2 or *Sm*E2 mutants, respectively. The C101 residue of *Sm*E2 was identified to be the key residue for the covalent linkage. **g**, Sequence alignment of *Sm*E2 and its orthologs shows two conserved cysteine residues (marked by red arrows). **h**, Left: schematic diagram of the *Sm*CdnG mutants. Right: SDS-PAGE analysis of the co-purified proteins of corresponding *Sm*CdnG mutants with *Sm*E2. The red triangle on the right of the gel indicates the covalently linked *Sm*CdnG-*Sm*E2 protein. The black, blue, and orange

triangles on the right of the gel indicate His-tagged *Sm*CdnG or *Sm*CdnG mutants, His-tagged *Sm*E2, and non-tagged *Sm*E2, respectively. The G406 residue of *Sm*CdnG was identified to be the critical residue for the covalent linkage. **i**, Top: MS/MS of the C101-containing peptide from *Sm*E2 in the covalently linked *Sm*CdnG-*Sm*E2 protein showing that the C101 residue of *Sm*E2 carries a glycine (57.060 Da) modification. Bottom: MS/MS of the G406-containing peptide from *Sm*CdnG in the covalently linked *Sm*CdnG-*Sm*E2 protein showing that the G406 residue of *Sm*CdnG carries a cystine (103.140 Da) modification.

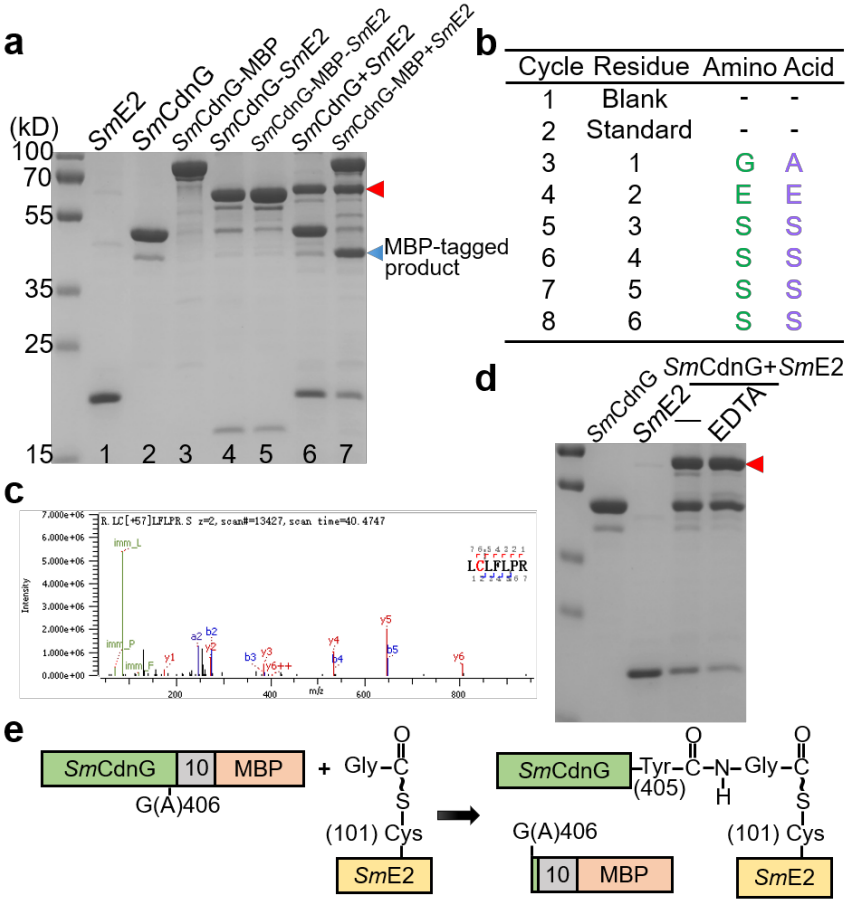


Fig. 3 C-terminus of *Sm*CdnG is cleaved for the formation of a covalent link. a, SDS-PAGE analysis showing the *in vivo* and *in vitro* formation of the covalently linked

*Sm*CdnG-MBP-*Sm*E2 protein. The *in vivo* formation of the covalently linked *Sm*CdnG-MBP-*Sm*E2 protein was achieved by the co-expression and co-purification of His-tagged *Sm*CdnG-MBP and non-tagged *Sm*E2 (lane 5). For the *in vitro* assembly, two proteins were mixed at a molar ratio of 1:1 overnight (lane 7). Since both *Sm*CdnG-MBP and *Sm*E2 were His-tagged during the *in vitro* reconstruction, the molecular weight of the covalently linked *Sm*CdnG-MBP-*Sm*E2 protein was slightly higher than that obtained *in vivo* from His-tagged *Sm*CdnG-MBP and non-tagged *Sm*E2. The red and blue triangles indicate the covalently linked *Sm*CdnG-MBP-*Sm*E2 protein and the MBP-tagged protein which is cleaved during the ligation of *Sm*CdnG-MBP and *Sm*E2, respectively. **b**, Protein N-terminal sequencing based on the Edman degradation method showed that the N-terminal sequence of the cleaved C-terminal MBP-containing peptide of *Sm*CdnG-MBP is GESSSS (in green), starting from G406. Protein N-terminal sequencing based on MS/MS showed that the N-terminal sequence of the cleaved C-terminal MBP-containing peptide of *Sm*CdnG(G406A)-MBP is AESSSS (in purple), starting from A406. **c**, MS/MS of the C101-containing peptide from *Sm*E2 purified alone showing that its C101 residue carries a glycine (57.060 Da) modification. **d**, SDS-PAGE analysis showing the *in vitro* formation of the covalently linked *Sm*CdnG-*Sm*E2 protein in the presence of EDTA. EDTA (10 mM) was added to the 1:1 mixture of *Sm*CdnG and *Sm*E2 and samples were incubated on ice overnight. **e**, Schematic diagram showing the exchange between the *Sm*E2-bound glycine and G406 (or A406) of *Sm*CdnG-MBP, the ligation between *Sm*CdnG-MBP (MBP cleaved) and *Sm*E2, and the release of MBP with G406 (or A406).

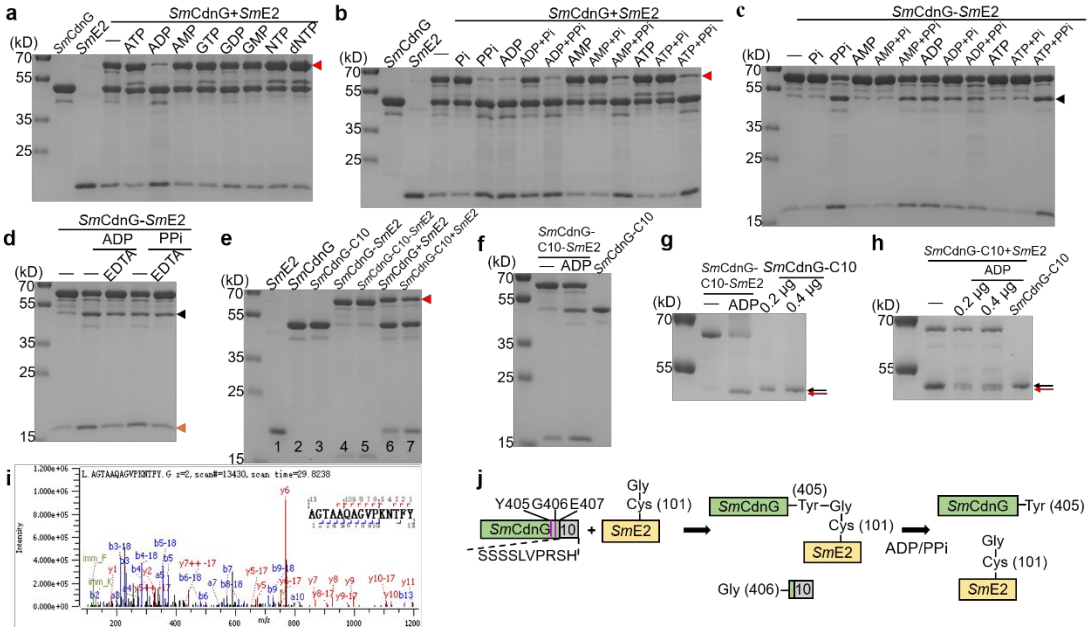
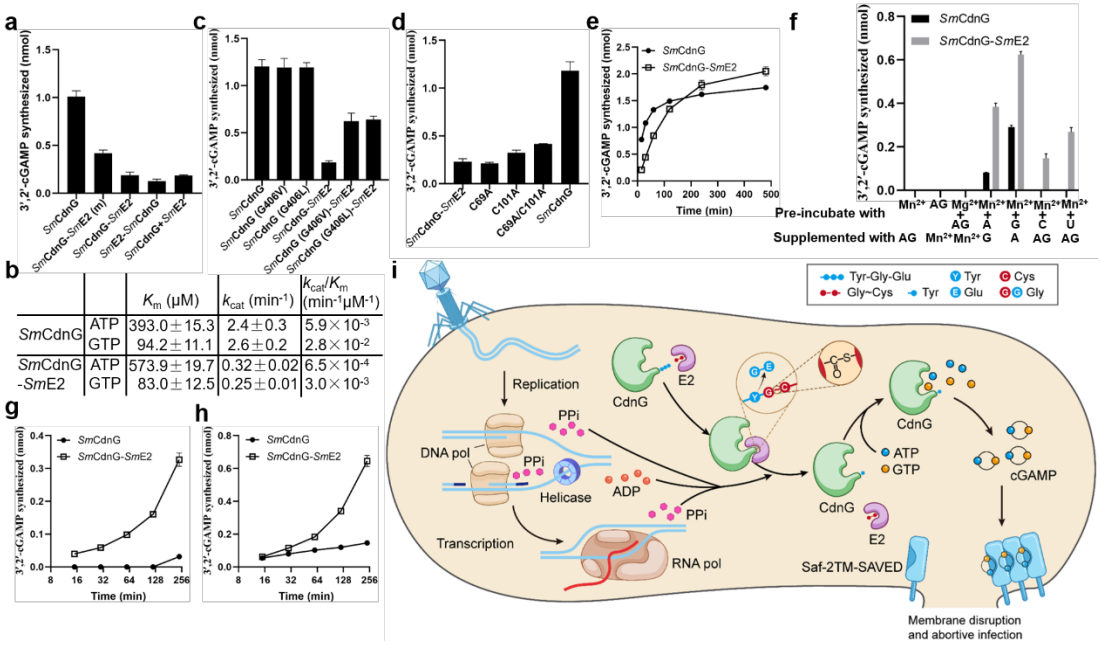


Fig. 4 Covalently linked *Sm*CdnG-*Sm*E2 protein was cleaved in the presence of ADP or PPi. a, SDS-PAGE analysis showed that the *in vitro* formation of the covalently linked *Sm*CdnG-*Sm*E2 protein was significantly inhibited in the presence of ADP. Different nucleotides (5 mM each) were added to the 1:1 mixture of *Sm*CdnG and *Sm*E2. The red triangle on the right of the gel indicates the covalently linked *Sm*CdnG-*Sm*E2 protein. **b**, SDS-PAGE analysis showed that the *in vitro* formation of the covalently linked *Sm*CdnG-*Sm*E2 protein was inhibited in the presence of ADP or PPi. Individual ADP (or AMP or ATP), Pi or PPi, ADP (or AMP or ATP)/Pi mixture, or ADP (or AMP or ATP)/PPi mixture was added to the 1:1 mixture of *Sm*CdnG and *Sm*E2. **c**, SDS-PAGE analysis showed that the covalently linked *Sm*CdnG-*Sm*E2 protein was cleaved in the presence of ADP or PPi. Individual ADP (or AMP or ATP), Pi or PPi, ADP (or AMP or ATP)/Pi mixture or ADP (or AMP or ATP)/PPi mixture was added to the covalently linked *Sm*CdnG-*Sm*E2 protein. The red triangle on the right of the gel indicates the

covalently linked *Sm*CdnG-*Sm*E2 protein. The black and orange triangles indicate the *Sm*CdnG and *Sm*E2, respectively. **d**, SDS-PAGE analysis showed that cleavage of the covalently linked *Sm*CdnG-*Sm*E2 protein can be induced by ADP or PPi in the presence of 10 mM EDTA. The black and orange triangles indicate the *Sm*CdnG and *Sm*E2, respectively. **e**, SDS-PAGE analysis showing the *in vivo* and *in vitro* formation of covalently linked *Sm*CdnG-C10-*Sm*E2 protein. The *in vivo* formation of the covalently linked *Sm*CdnG-C10-*Sm*E2 protein was achieved by the co-expression and co-purification of His-tagged *Sm*CdnG-C10 and non-tagged *Sm*E2 (lane 5). For the *in vitro* formation, two proteins were mixed at a molar ratio of 1:1 and incubated overnight (lane 7). The red triangle indicates the covalently linked *Sm*CdnG-C10-*Sm*E2 protein. Since both *Sm*CdnG-C10 and *Sm*E2 were His-tagged during the *in vitro* reconstruction, the molecular weight of the covalently linked *Sm*CdnG-C10-*Sm*E2 protein was slightly higher than that obtained *in vivo* from His-tagged *Sm*CdnG-C10 and non-tagged *Sm*E2. **f**, SDS-PAGE analysis showing the cleavage of the covalently linked *Sm*CdnG-C10-*Sm*E2 protein by ADP. ADP was added to the covalently linked *Sm*CdnG-C10-*Sm*E2 protein to break the covalent link. **g**, Comparison of the molecular weight of *Sm*CdnG-C10 released from the covalent complex and original *Sm*CdnG-C10. Electrophoresis time was extended to show the small difference in size between processed (after cleavage of the C10 peptide following G406 cleavage, red arrow) and unprocessed *Sm*CdnG-C10 (black arrow). **h**, SDS-PAGE analysis showing the cleavage of the covalently linked *Sm*CdnG-C10-*Sm*E2 protein formed *in vitro*. The black and red arrows indicate unprocessed and processed (after cleavage of the C10 peptide following

G406 cleavage) *Sm*CdnG-C10, respectively. **i**, Protein C-terminal sequencing by MS/MS showed that the C-terminal sequence of *Sm*CdnG-C10 cleaved by ADP is AGTAAQAGVPKNTFY, terminated by Y405. **j**, Schematic diagram showing the covalent link between *Sm*CdnG-C10 and *Sm*E2 and their cleavage induced by ADP or PPi.



		K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m (min^{-1}μM^{-1})
*Sm*CdnG	ATP	393.0±15.3	2.4±0.3	5.9×10^{-3}
	GTP	94.2±11.1	2.6±0.2	2.8×10^{-2}
*Sm*CdnG-*Sm*E2	ATP	573.9±19.7	0.32±0.02	6.5×10^{-4}
	GTP	83.0±12.5	0.25±0.01	3.0×10^{-3}

Fig. 5 Regulation of activity and stability of *Sm*CdnG by *Sm*E2. a, Comparison of the 3′,2′-cGAMP synthesis activities of *Sm*CdnG and the covalently linked *Sm*CdnG-*Sm*E2 proteins expressed in different forms. *Sm*CdnG-*Sm*E2 (m), *Sm*CdnG-*Sm*E2 and *Sm*E2-*Sm*CdnG correspond to the covalently linked *Sm*CdnG-*Sm*E2 protein shown in **Fig. 2a** (i–iii), respectively. *Sm*CdnG+*Sm*E2 represents the addition of *Sm*E2 to the *Sm*CdnG reaction at a molar ratio of 1:1. **b**, Comparison of enzyme kinetics between *Sm*CdnG and *Sm*CdnG-*Sm*E2. The catalytic efficiency (k_{cat}/K_m) of *Sm*CdnG-*Sm*E2 is ~10 times lower than that of *Sm*CdnG for both ATP and GTP. **c**, HPLC analysis and

quantification of the 3′,2′-cGAMP synthesized by *Sm*CdnG and *Sm*CdnG mutants co-purified with *Sm*E2. **d**, HPLC analysis and quantification of the 3′,2′-cGAMP synthesized by *Sm*CdnG co-purified with *Sm*E2 or *Sm*E2 mutants. **e**, Time-course analysis of the efficiency of 3′,2′-cGAMP synthesized by *Sm*CdnG and *Sm*CdnG-*Sm*E2. *Sm*CdnG-*Sm*E2 showed a lower activity compared to *Sm*CdnG at the early stage of the reaction, while it retained a higher activity during prolonged incubation. **f**, Two-step assay was conducted to compare the stability of *Sm*CdnG and *Sm*CdnG-*Sm*E2. For each reaction, *Sm*CdnG and *Sm*CdnG-*Sm*E2 were pre-incubated with Mn^{2+}, ATP/GTP (AG), Mg^{2+}/ATP/GTP (Mg^{2+}+AG), Mn^{2+}/ATP (Mn^{2+}+A), Mn^{2+}/GTP (Mn^{2+}+G), Mn^{2+}/CTP (Mn^{2+}+C), or Mn^{2+}/UTP (Mn^{2+}+U) at 37°C for 30 min, and then, the missing components were added (to a full set including Mn^{2+}, ATP, and GTP) to start 3′,2′-cGAMP synthesis. The yield of 3′,2′-cGAMP after an additional incubation step of 4 h for each reaction was quantified by HPLC. **g**, Time-course analysis of 3′,2′-cGAMP synthesized by *Sm*CdnG and *Sm*CdnG-*Sm*E2 after pre-incubation with Mn^{2+}/ATP at 37°C for 30 min. **h**, Time-course analysis of 3′,2′-cGAMP synthesized by *Sm*CdnG and *Sm*CdnG-*Sm*E2 after pre-incubation with Mn^{2+}/GTP at 37°C for 30 min. All the data represent the mean ± SD of three independent experiments. **i**, Model for bacteriophage sensing and immunity by CD-NTase+E2+Saf-2TM-SAVED operons. In the absence of bacteriophage infection, *Sm*E2 is activated by forming a high-energy thioester bond between its active cysteine and free glycine, and then, the *Sm*E2-bound glycine reacts with and replaces the C-terminal glycine of *Sm*CdnG, thereby covalently ligating *Sm*E2 to *Sm*CdnG. *Sm*E2 modifies *Sm*CdnG to enhance its stability while inhibiting its

cGAMP synthesis activity. After phage invasion, the high nucleotide metabolism including phage DNA replication and transcription releases PPi and ADP, which is sensed by the ATP binding site of *Sm*E2 to induce the cleavage of the peptidyl bond between Y405 and G406 in *Sm*CdnG. The free-standing *Sm*CdnG thus produces cGAMP to activate Saf-2TM-SAVED, which could mediate bacterial membrane disruption and abortive infection.

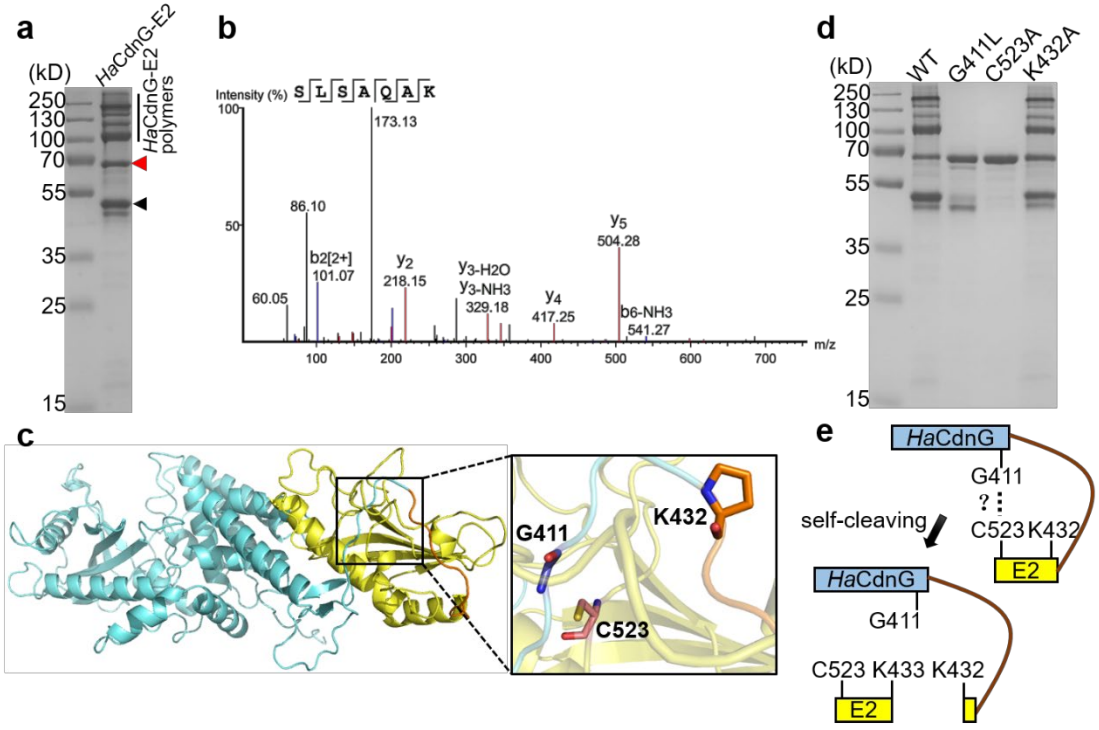


Fig. 6 Bacterial E2 domain mediates self-cleavage of the peptidyl bond between the CdnG and E2 domains in a single protein. a, SDS-PAGE analysis of the purified *Ha*CdnG-E2. The red and black triangles on the right of the gel indicate the uncleaved and cleaved *Ha*CdnG-E2 proteins, respectively. **b**, MS/MS of protein C-terminal sequencing showing that the C-terminal sequence of the 40–50-kDa protein is SLSAQAK, terminated by K432. **c**, Predicted *Ha*CdnG-E2 structure, with the CD-

NTase domain core in blue, the ubiquitin E2 variant (UEV) domain in yellow, and their linking region in orange. The detailed diagram shows the relative positions of residues G411, C523, and K432. **d**, SDS-PAGE analysis showing the purified mutants of *Ha*CdnG-E2. **e**, Schematic diagram showing the self-cleavage between the CD-NTase domain and the E2 domain of *Ha*CdnG-E2.

Methods

Materials

The DNA fragments encoding *Sm*CdnG (WP_016928966), *Sm*E2 (WP_016928967), and *Ha*CdnG-E2 (PKQ12786.1) were synthesized by Genscript. Oligonucleotides and primers were ordered from Genecreate Company. The ClonExpress II One Step Cloning Kit (C112-01) was purchased from Vazyme Biotech. NTPs (N0450L), CIP (M0525V), and nuclease P1 (M0660S) were purchased from New England BioLabs. ADP (A0180), GDP (G8730), CDP (YZ-140675), and AMP (SA9510) were purchased from Solarbio Life Sciences. Pi (106586) and PPi (433314) were purchased from Sigma-Aldrich. PrimeSTAR Max DNA Polymerase (R022A) was purchased from TaKaRa, and 3′,2′-cGAMP (C 238-005) and 2′,3′-cGAMP (C 161-005) were purchased from Biolog Life Science Institute. The DNA purification kit (AP-PCR-50) was purchased from Axygen. Ni-NTA Agarose (30230) was purchased from Qiagen. All other chemical reagents used were from Sigma-Aldrich unless otherwise noted.

Cloning and plasmid construction

Plasmids generated in this study were constructed by Gibson from synthesized genes

with ≥18 base pairs of homology flanking the insert sequence using the ClonExpress II One Step Cloning Kit (Vazyme Biotech). Plasmids were transformed into competent DH5α cells and confirmed by Sanger sequencing. For the co-expression of *Sm*CdnG and *Sm*E2, an N-terminal 6×His-tag sequence was constructed in the pET28a vector. Non-tagged expression vectors were constructed from the pQE82L vector, with the T5 promoter replaced by a T7 promoter for *Sm*E2. For the expression and purification of individual *Sm*CdnG and *Sm*E2, their coding sequences were codon-optimized. Meanwhile, a vector with the native sequence of *Sm*CdnG and *Sm*E2 in a polycistronic form derived from *S. marcescens* was also constructed. Details about cloning of the native sequences of *Sm*CdnG and *Sm*E2 are provided in Table S1.

Protein expression and purification

Plasmids for protein expression were transformed into chemically competent *E. coli* BL21 (DE3). Bacterial cells were cultivated as a 4 mL starter culture in LB liquid medium overnight at 37°C with shaking at 220 rpm. The cultures were then transferred into 200 mL LB medium and cultivated for approximately 3 h at 37°C with shaking at 220 rpm until the optical density at 600 nm (OD_{600}) reached 0.8–1.0. After cooling down to room temperature, 0.25 mM IPTG was added to the cultures, and cultivation was continued overnight at 16°C with shaking at 220 rpm.

Cultures were collected and lysed by sonication in Lysis Buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole), and the lysates were clarified by centrifugation at 21,000 × *g*, 4°C for 1 h and filtered through a 0.45-μm filter. Recombinant protein was purified by affinity chromatography using Ni-NTA resin and

a gravity column. Ni-NTA resin was pre-equilibrated with Lysis Buffer, bound to target proteins, washed with Wash Buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, and 50 mM imidazole for *Sm*CdnG, its mutants and *Sm*E2; 20 mM Tris-HCl, pH 8.0, 300 mM NaCl, and 50, 80, or 100 mM imidazole for the co-expression proteins) and eluted with Elution Buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, and 100 mM imidazole for *Sm*CdnG and its mutants; 20 mM Tris-HCl, pH 8.0, 300 mM NaCl, and 200 mM imidazole for *Sm*E2 and the co-expression proteins). Proteins were filter-concentrated using centrifugation and a 10-kDa or 30-kDa cut-off column (Millipore Sigma) and dialyzed at 4°C for ~24 h against Dialysis Buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 0.1% Triton X-100 and 50% glycerol). Proteins were analyzed by SDS-PAGE with Coomassie blue (Bio-Rad) staining.

HPLC analysis of the enzymatic reactions

CD-NTase reactions were performed essentially as previously described [13]. Briefly, 20-μL reactions contained 2 μM enzyme and 250 μM NTPs in reaction buffer with 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 2.5 mM $MnCl_2$/$MgCl_2$, and 1 mM DTT. Reactions were incubated at 37°C for the indicated periods, inactivated at 80°C for 10 min, and centrifuged for 15 min at 21,000 × *g* to remove precipitated protein. Reaction products were analyzed by HPLC with a ZORBAX Eclipse XDB-C18 (4.6 × 250 mm) column and an Agilent 1260 Infinity II Series LC system. Next, 10 μL of the reaction product was injected into the column and eluted with solvent A (methanol) and solvent B (20 mM ammonium acetate) at a flow rate of 1 mL/min using the following linear gradient: 0–5 min, 5% A; 5–15 min, 5–100% A; 15–20 min, 100% A. The column was re-

equilibrated for 5 min at 5% A. For analyzing the reaction products treated with CIP or nuclease P1, 20 µL of the reaction product was treated with 1 µL of CIP or nuclease P1 (New England BioLabs) at 37°C for 1 h and inactivated at 80°C for 10 min prior to centrifugation and analysis by HPLC.

Mass spectrometry

To characterize the product of *Sm*CdnG, we performed LC-MS/MS on a Thermo Scientific™ UltiMate™ 3000 system coupled to a Thermo Scientific Orbitrap LC/MS (Q Exactive), using a Thermo Hypersil GOLD C18 column (100 mm × 2.1 mm, 3 µm) maintained at 25°C with a flow rate of 0.25 mL/min. The mobile phase consisted of methanol (A) and 20 mM ammonium acetate (B). The HPLC gradient was as follows: 0–7 min, 2% A; 7–12 min, 2–30% A. The column was re-equilibrated for 7 min at 2% A. Detection was performed in positive ionization mode using an electrospray ionization (ESI) source under the following parameters: spray voltage, 3.2 kV; sheath and auxiliary gas flow rates, 40 and 15 arbitrary units, respectively; max spray current, 100.00 µA; S-Lens RF Level, 50%; capillary temperature, 300°C; probe heater temperature, 350°C. Profile MS1 spectra were acquired with the following settings: mass resolution, 70,000; AGC volume, 3×10^6, maximum IT, 100 ms, scan range, 300–1,000 *m*/*z*. Acquisition of data-dependent MS/MS spectra was performed using collision-induced dissociation (CID) with the following settings: mass resolution, 17,500; AGC volume, 1×10^5; maximum IT, 50 ms; loop count, 10; isolation window, 4.0 *m*/*z*; normalized collision energy, 20, 40, and 60 eV. Data are reported for the $z = 1$ acquisition for each indicated cyclic oligonucleotide. The chemical structures were

drawn using ChemDraw 19.0.

Co-expression of *Sm*CdnG and *Sm*E2 in *E. coli*

To determine the interaction between *Sm*CdnG and *Sm*E2, they were co-expressed in *E. coli* BL21 (DE3). Three forms of co-expression were assessed for *Sm*CdnG and *Sm*E2: (1) the gene encoding *Sm*CdnG with an N-terminal 6×His-tag cloned into the pET28a vector and the gene encoding *Sm*E2 gene without tag cloned into the pQE82L vector and the T5 promoter in pQE82L replaced with a T7 promoter; (2) the gene encoding *Sm*CdnG without tag cloned into the pQE82L vector and the gene encoding *Sm*E2 with N-terminal 6×His-tag cloned into the pET28a vector; (3) *Sm*CdnG and *Sm*E2 arranged in the native polycistron form in the genome and cloned into the pET28a vector. Plasmids were co-transformed into competent BL21 (DE3) and selected with relevant antibiotics. The expression and purification of co-expressed proteins were performed as described above for the co-expression proteins.

Covalent linkage and cleavage

The covalent linkage and cleavage for *Sm*CdnG and *Sm*E2 were analyzed by SDS-PAGE. Covalent linkage experiments were performed in 16 μL containing 4 μM *Sm*CdnG, 4 μM *Sm*E2, and 50 mM Tris-HCl (pH 7.5). Samples were incubated on ice for 10, 30, 90, or 270 min. Then, the samples were mixed with 4 μL 5× SDS-PAGE loading buffer (with DTT) (Solarbio life sciences), heated at 100°C for 10 min, and loaded (10 μL per lane) onto a 12.5% SDS-PAGE gel, which was stained with Coomassie blue.

Cleavage experiments were performed in 16 μL containing 4 μM *Sm*CdnG-*Sm*E2,

various concentrations of PPi and/or other nucleotides, and 50 mM Tris-HCl (pH 7.5), in the presence or absence of 5 mM $MgCl_2$. Samples were incubated overnight at 4°C. The results were analyzed as above.

Protein structure prediction

ColabFold was used to predict the structure of the *Sm*CdnG-*Sm*E2 complex and the *Ha*CdnG-E2 protein. Multiple sequence alignment (MSA) were performed using MMseqs2 (UniRef + Environmental). Structural prediction was performed using AlphaFold on a virtual cluster node with a Nvidia Tesla P100 GPU.

Steady-state kinetic measurement

For steady-state kinetic measurements, the standard 20 µL-reaction mixture included 50 mM Tris-HCl (pH 7.5), 2.5 mM $MnCl_2$, 100 mM NaCl, 1 mM DTT, 1 µM enzyme, 1 mM ATP (or GTP), and various concentrations (15.6 µM, 31.3 µM, 62.5 µM, 125 µM, 250 µM, 500 µM and 1 mM) of GTP (or ATP). For each nucleotide concentration, samples were incubated at 37°C for 15 min for *Sm*CdnG and for 30 min for *Sm*CdnG-*Sm*E2. Reactions were immediately heat-inactivated at 80°C for 10 min and the samples were centrifuged for 15 min at 17,000 × *g* to remove precipitated protein. Of each sample, 10 µL was injected and reactions were analyzed by HPLC, as detailed above. Absorbance units were converted to µmol/L by comparing to a standard curve from 7.8 µM to 1 mM of chemically synthesized 3′,2′-cGAMP (Biolog Life Sciences). Data were fitted by linear regression, and non-linear curve fitting Michalis–Menten kinetics and allosteric sigmoidal were calculated using GraphPad Prism version 8.0.0.

Stability measurement

For stability measurements, the standard reaction mixture included 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM DTT, 2 μM enzyme, and one or more components of 2.5 mM $MnCl_2$ or $MgCl_2$ and 250 μM ATP, GTP, CTP, or UTP, with one component missing to prevent reaction initiation. The reaction mixtures were first incubated at 37°C for 30 min, and then supplemented with the missing component to initiate reactions, followed by incubation at 37°C for 4 h. For time course measurements, reactions were incubated at 37°C for 15, 30, 60, 120, and 240 minutes, respectively, then immediately heat inactivated at 80°C for 10 min, and centrifuged for 15 min at 17,000 × *g* to remove precipitated protein. Of each sample, 10 μL was injected and analyzed by HPLC as described above.

Data analysis

Data analysis was performed using Microsoft Excel 2019, OriginPro 9.1 and GraphPad Prism 8.0.0. Graph plotting and statistical analysis was performed using GraphPad Prism 8.0.0.

Acknowledgements We thank all lab members for helpful discussion and LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript. The work was supported by National Natural Science Foundation of China (grant 32150009 and 31870165 to B.Z., 31900032 to F.T.H.) and Fund from Science,Technology and Innovation Commission of Shenzhen Municipality (grant JCYJ20210324115811032 to B.Z.). Funding for open access charge: National Natural

Science Foundation of China.

Author contributions F.T.H. and B.Z. conceived the project. Y.Y., F.T.H. and B.Z. designed the experiments. Y.Y. carried out the experiments. J.X. and L.F.W. analyzed structures. Y.Y., F.T.H., L.F.W. and B.Z. analyzed the data and wrote the manuscript. All authors discussed the results and contributed to the final manuscript.

Competing interests The authors declared that they have no conflicts of interest to this work.

Additional Information

Supplementary Information is available for this paper.

Correspondence and requests for materials should be addressed to Bin Zhu**.**

Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.

Extended data

Extended Data Table. 1 The whole sequence of *Sm*CdnG and *Sm*E2 genes in their native form

ATGTACGGTTCCACTACCGCCAGAAACCTGCCTTCAGGGAAAAAACAGC
GTATCGCCGATTTATTATCGCAAATTATTGAAACGCTGGATCTCACCAAAA

CCCAATACGCCAACATCGAAAGCGCCTATAACGGCGTCGGCACCTTTCTG

TCCGAAGGCGACGATCCGCTATTGCAAGATGCCGTTATTTACCCGCAGGG

CAGCGTGCGGCTCAACACCACCGTTAAACCCAAAAATGAAGAGCAATAC

GATATTGACCTGATTTGTTATCTGCCCCATGCCACCCAGGCGGATTATACC

GGCGTGATATCGGCCATTCGCCAGCGGCTGGAATCACATAAAACCTACAA

AACGCTATTAAGCGAGTTACCGCGCGGGTTTCGCATCAATTACGCCGGGG

ATTACCATCTGGATATCACGCCGGGCCGCGATCACACCGGCACAGCACAT

CCGGGCCAGCCGCTGTGGGTCGTGGATGCACAAACCGCCTGGAAGGAG

TCCAACCCCAGCGGCTACGCCGAGTGGTTCGAGAGCAGCGCCAGCGTGC

AACCCCTGCGCACCATTCTGGTCATGGATTCCGCCAGCCGCGTGGGTACC

GAGGCGTTGCTCCCGTTGCCGGACAGCACCGACAAGAAATTGCTTAATC

ACATCGTACAAATTCTCAAACGCCACCGTGACGAATGGGCCGCAGAGCA

GGATGAGGTCCGGCAGCGCTGCCGCCCGATTTCGGTCATCATTACCACGC

TGGCGTGCCATGCCTACAACCACATCATTGCGGACAGGCGCGCCTACGA

CAACGACCTGGATATCCTGTTAGACGTCCTGGAACTGATGCCGGATTTCA

TCGTGTCGACACAGGGAGCAATTCACGTCAACAACCCGCACATGCCGGA

GGAGAACTTCGCCGAGAAGTGGAACCGTTCAGAGCAGGATGAGGGCCC

TCAGCGCAGCGAAGCCTTTTACCAGTGGCATGCCGCGGCCCAGGCGACG

TTTAACACCATCGCCGCCAGCGTGGGGGAAGATAATCTGTTCCTGAGCCT

CGAAGACAGCTTCGGCAAAACGCCGGTCGATGTCGTCAGGCAACGTCTG

ATGGAGCATATGCAGTCAGCCAGAGAACAAGGCAGCCTGCATCTGGATA

AGAAAACCGGCGGGCTGATTGCCACCGGCCTCGCCGGTACGGCAGCCCA

GGCCGGCGTGCCTAAAAACACCTTCTACGGTGAATAACGTGGTTATCAG

GCATCACTGCAAACCGCTGACCATCGCTCAGCAATATCGCGCGTTGAAA

GCCGGCGGCCCGTATGAACGGCTGCGCATTATCCATCACGACCGCACCCT

GCTGTGGGAAGGGTGGTTGCAACCCTCGCTGTTTAGCCGCCGCTACAAG

GTCGCGGTCCGGTACAGCCTCGGCACCCCACCGATTTGCGTGGTGACAG

AGCCTGACCTTTTCGCGCTGGCCGGCACGCGGGCCATCCCGCACCTGTAT

CCCGCCGATAAGCATATTCCCGGCGCGCGGCTGTGCCTGTTTTTGCCCCG

TTCACAGGCCGACGACGGGCTCAGCGAATGGCGCGCCCAGTTAAAAATC

AGCGACACCCTCATTCCCTGGGCGTCGCTCTGGCTGTTTTATTTTGAACA

GTGGCTGCACACCGGCCACTGGGAAGGCGGCGGTAAGCACCCACGCCC

CAGCGAGGTTAAAAATGAGCGTTGA[1]

[1]Color-coding: Red: CdnG / Blue: E2 / Purple: CdnG/E2 overlap

Extended Data Table. 2 Primers for construction of *Sm*CdnG mutations.

Name	**Sequence**
*Sm*CdnG-D84A/D864-F	GCAATACGCGATCGCGCTGATTTGTTATCTGC
*Sm*CdnG-D84A/D864-R	CAAATCAGCGCGATCGCGTATTGCTCTTCGTTT
*Sm*CdnG-ΔC-F	AGGCTGGCGTGTAAGAATTCGAGCTCCGTC
*Sm*CdnG-ΔC-R	CTCGAATTCTTACACGCCAGCCTGGGCCGCCGT
*Sm*CdnG-ΔGE-F	CACCTTCTACTAAGAATTCGAGCTCCGTCGAC
*Sm*CdnG-ΔGE-R	CGAATTCTTAGTAGAAGGTGTTCTTCGGCACGC
*Sm*CdnG-ΔE-F	CTTCTACGGCTAAGAATTCGAGCTCCGTCGAC

*Sm*CdnG-ΔE-R	CGAATTCTTAGCCGTAGAAGGTGTTCTTCG
*Sm*CdnG-K401A-F	TGGCGTGCCGGCGAACACCTTCTACGGCGA
*Sm*CdnG-K401A-R	AGAAGGTGTTCGCCGGCACGCCAGCCTGGGCC
*Sm*CdnG-Y405A-F	ACACCTTCGCGGGCGAGTAAGAATTCG
*Sm*CdnG-Y405A-R	TACTCGCCCGCGAAGGTGTTCTTCGGC
*Sm*CdnG-G406A-F	CACCTTCTACGCGGAGTAAGAATTCGAGCTCC
*Sm*CdnG-G406A-R	ATTCTTACTCCGCGTAGAAGGTGTTCTTCGGC
*Sm*CdnG-G406V-F	CACCTTCTACGTGGAGTAAGAATTCGAGCTCC
*Sm*CdnG-G406V-R	ATTCTTACTCCACGTAGAAGGTGTTCTTCGG
*Sm*CdnG-G406L-F	CACCTTCTACCTGGAGTAAGAATTCGAGCTCC
*Sm*CdnG-G406L-R	ATTCTTACTCCAGGTAGAAGGTGTTCTTCGGC

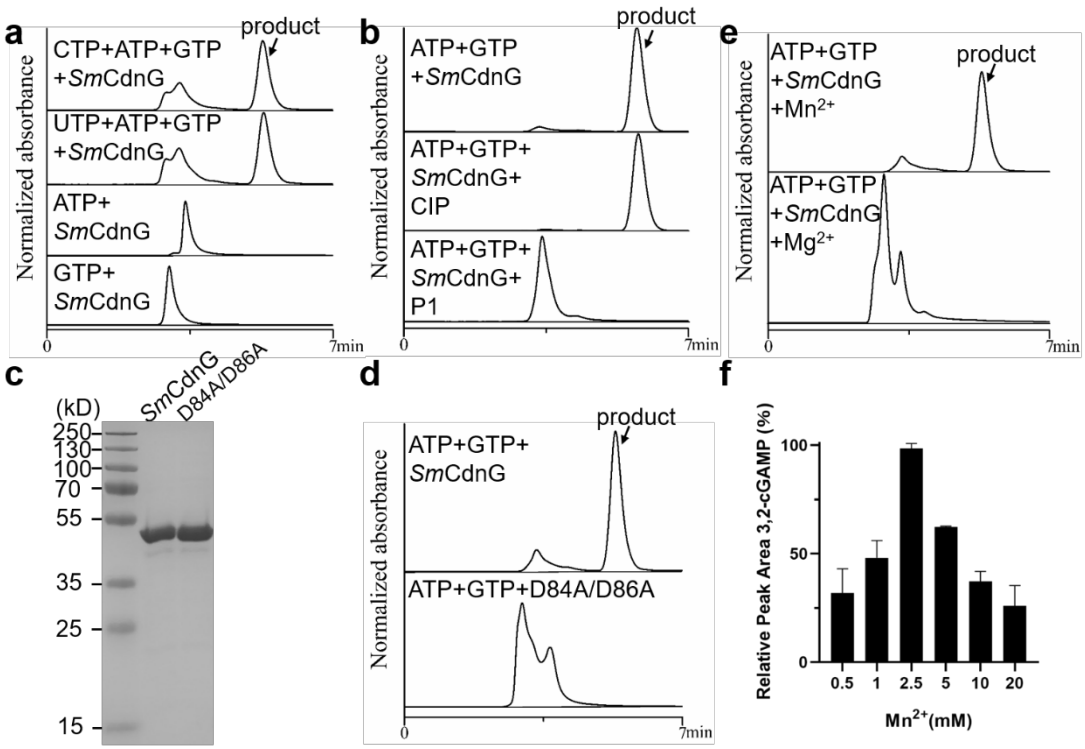
Extended Data Table. 3 Primers for construction of *Sm*E2 mutations.

Name	**Sequence**
*Sm*E2-C69A-F	ACCCCGCCCATCGCGGTTGTTACCGAACCG
*Sm*E2-C69A-R	TCGGTAACAACCGCGATGGGCGGGGTGCCG
*Sm*E2-C101A-F	GGTGCGCGTCTGGCGCTGTTCCTGCCGAGAT
*Sm*E2-C101A-R	CGGCAGGAACAGCGCCAGACGCGCACCCGGAAT

Extended Data Table. 4 Primers for construction of *Ha*CdnG-E2 mutations.

Name	**Sequence**
*Ha*CdnG-E2-G411L-F	CCTTTTTCCTGTCTAAGCGTCCCGTCT

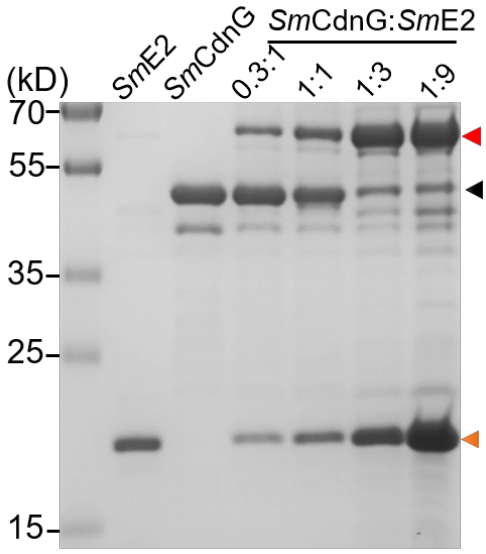
*Ha*CdnG-E2-G411L-R	CGCTTAGACAGGAAAAAGGTGTTCGTC
*Ha*CdnG-E2-K432A-F	CCCAGGCCGCGGCCATGGCCCGGAGGT
*Ha*CdnG-E2-K432A-R	GCCATGGCCGCGGCCTGGGCCGAGAGC
*Ha*CdnG-E2-C523A-F	CGCCCCTTGCGCTGTTCGATCCCCAGG
*Ha*CdnG-E2-C523A-R	TCGAACAGCGCAAGGGGCGACAGGCGG



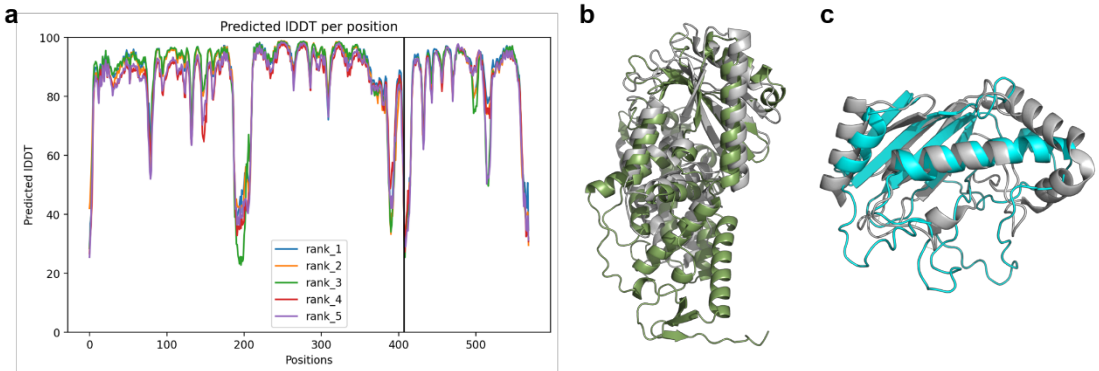
Extended Data Fig. 1 *In vitro* characterization of the enzymatic activity of CdnG.

a, HPLC analysis of nucleotide second messenger synthesis by *Sm*CdnG using different combinations of four kinds of nucleotide substrates. *Sm*CdnG synthesizes a product only in the presence of both ATP and GTP. **b**, *Sm*CdnG product in the presence of ATP and GTP was degraded by nuclease P1 but not CIP, as analyzed by HPLC. **c**, SDS-PAGE analysis of the purified catalytic center mutant of *Sm*CdnG. **d**, Comparison of the activities of *Sm*CdnG and its D84A/D86A mutant. **e**, *Sm*CdnG synthesis activity is

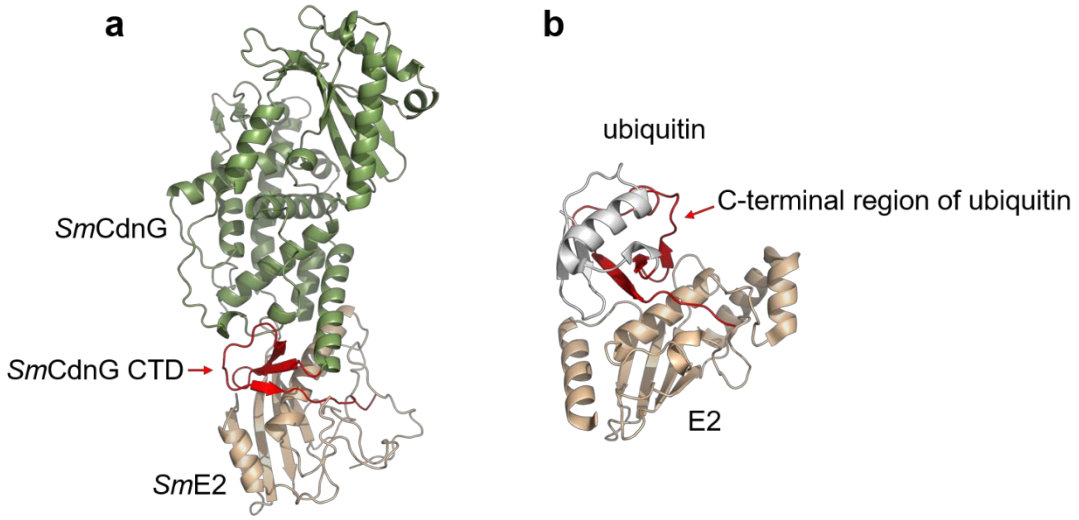
dependent on Mn^{2+}. **f**, Optimal Mn^{2+} concentration for *Sm*CdnG. The products synthesized at various Mn^{2+} concentrations were quantified by HPLC. Data represent the mean ± SD of three independent experiments.



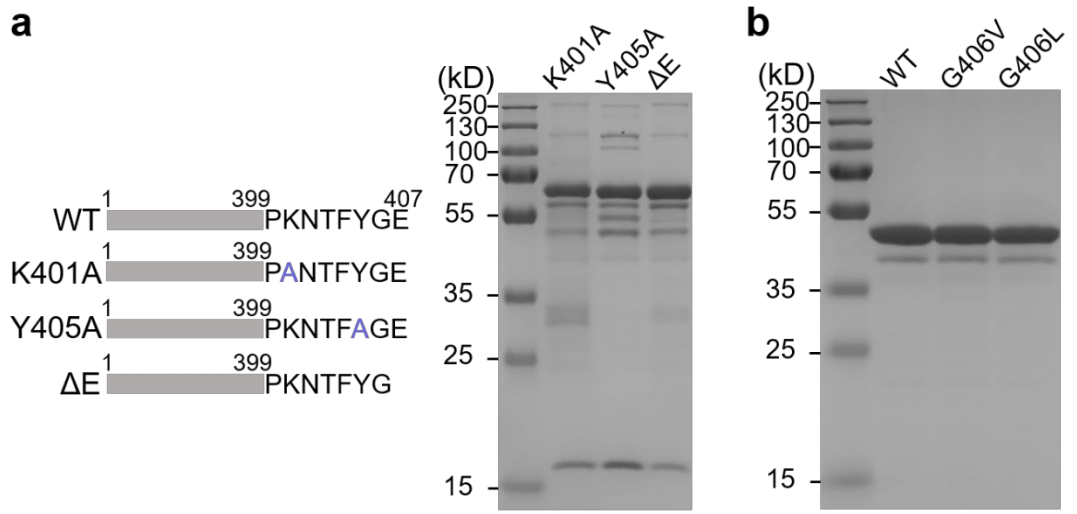
Extended Data Fig. 2 SDS-PAGE analysis of the *in vitro* assembly of the covalently linked *Sm*CdnG-*Sm*E2 protein. *Sm*CdnG and *Sm*E2 were mixed at different molar ratios as indicated on top of the gel, and incubated at 4°C overnight. The red triangle on the right of the gel indicates the covalently linked *Sm*CdnG-*Sm*E2 protein. The black and orange triangles indicate the *Sm*CdnG and *Sm*E2, respectively.



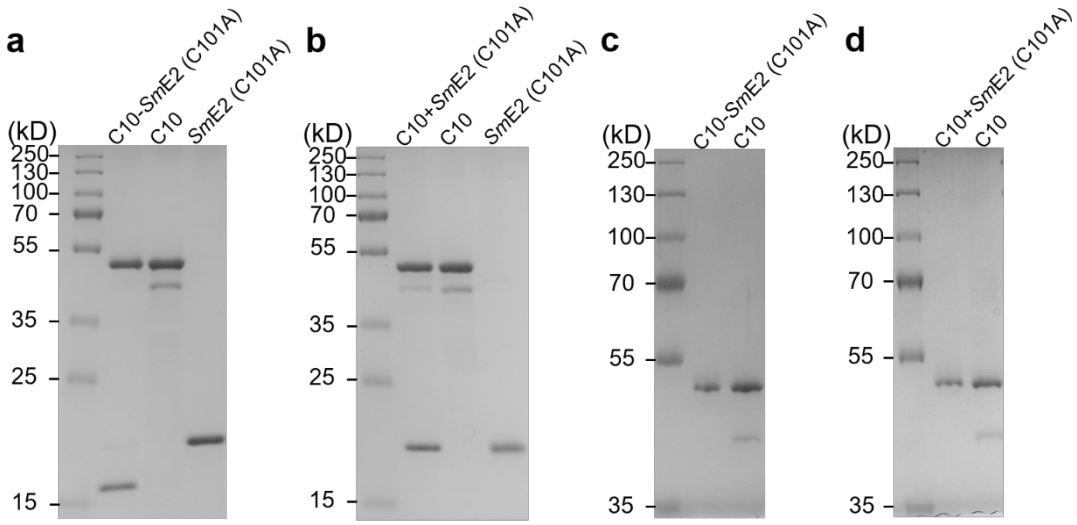
Extended Data Fig. 3 Predicted structure of the *Sm*CdnG-*Sm*E2 complex. a, Per-residue IDDT of five predicted structures of the covalently linked *Sm*CdnG-*Sm*E2 protein. All five predicted structures were nearly identical, and the per-residue IDDT scores are mostly over 80, with the exceptions of the terminal disordered regions and internal loop regions. **b**, Overlaid structures of *Sm*CdnG (dark green) and *B. fragilis* CD-NTase (gray, PDB ID 7lju). **c**, Overlaid structures of *Sm*E2 (cyan) and eukaryotic E2 (gray, PDB ID 4auq).



Extended Data Fig. 4 C-terminal domain of *Sm*CdnG interacts with *Sm*E2 and mimics ubiquitin-E2 complex. a, Predicted structure of the *Sm*CdnG-*Sm*E2 complex, with the C-terminal domain (CID) of *Sm*CdnG (amino acids 374–407) colored in red. **b**, Structures of the E2-ubiquitin complex (PDB ID 4auq), with the C-terminal region of ubiquitin colored in red.

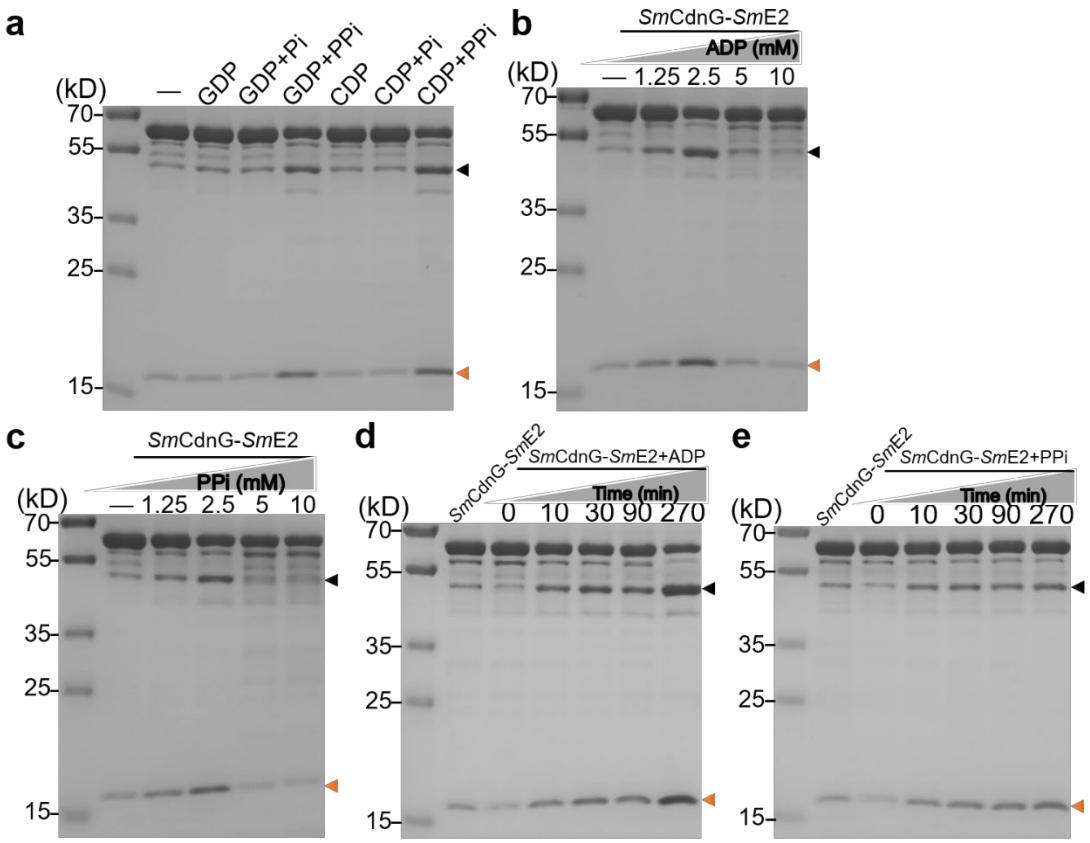


Extended Data Fig. 5 *Sm*CdnG C-terminal mutants not shown in the text. a, Left: Schematic diagram of the *Sm*CdnG mutants. Right: SDS-PAGE analysis of the indicated (on top of the gel) His-tagged *Sm*CdnG mutants co-purified with non-tagged *Sm*E2. The K401, Y405 and E407 residues of *Sm*CdnG are not critical for the covalent linkage between *Sm*CdnG and *Sm*E2. **b**, SDS-PAGE analysis of *Sm*CdnG and its G406V and G406L mutants.



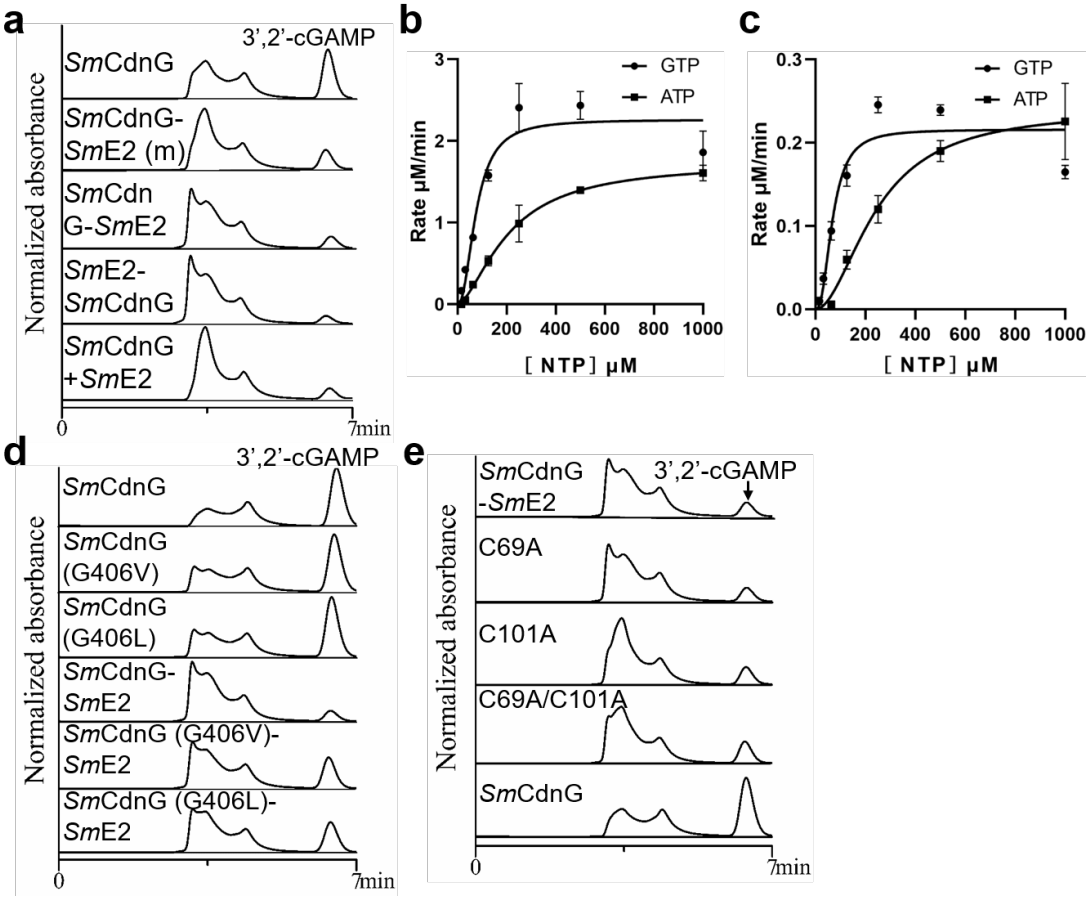
Extended Data Fig. 6 C-terminus of *Sm*CdnG is not processed in the presence of *Sm*E2 C101A mutant. a, SDS-PAGE analysis of the co-purified proteins of His-tagged

*Sm*CdnG-C10 with non-tagged *Sm*E2 C101 mutant (C10-*Sm*E2 (C101A)). The molecular weight of non-tagged *Sm*E2 is lower than that of the His-tagged *Sm*E2 expressed and purified alone. **b**, SDS-PAGE analysis of the *in vitro* mixture of *Sm*CdnG-C10 and the *Sm*E2 C101A mutant. **c**, Comparison of the molecular weight of *Sm*CdnG-C10 co-expressed and co-purified with *Sm*E2-C101A and *Sm*CdnG-C10 expressed and purified alone. **d**, Comparison of the molecular weight of *Sm*CdnG-C10 mixed with *Sm*E2-C101A and *Sm*CdnG-C10 alone. For (**c**) and (**d**), the electrophoresis time was extended to show the small size difference between processed and unprocessed *Sm*CdnG-C10.



Extended Data Fig. 7 *Sm*CdnG-*Sm*E2 is cleaved in the presence of ADP or PPi. a, SDS-PAGE analysis showed that the cleavage of *Sm*CdnG-*Sm*E2 cannot be induced by

GDP or CDP. Individual GDP (or CDP), GDP (or CDP)/Pi mixture and GDP (or CDP)/PPi mixture were added to the covalently linked *Sm*CdnG-*Sm*E2 protein. **b**, SDS-PAGE analysis showed that the optimal ADP concentration for inducing cleavage of the covalently linked *Sm*CdnG-*Sm*E2 protein was 2.5 mM. **c**, SDS-PAGE analysis showed that the optimal PPi concentration for inducing cleavage of the covalently linked *Sm*CdnG-*Sm*E2 protein was 2.5 mM. **d**, SDS-PAGE analysis showing that the *in vitro* cleavage of the covalently linked *Sm*CdnG-*Sm*E2 protein was induced by ADP. ADP (2.5 mM) was mixed with *Sm*CdnG-*Sm*E2 and incubated on ice for various time periods as indicated. **e**, SDS-PAGE analysis showing the *in vitro* cleavage of the covalently linked *Sm*CdnG-*Sm*E2 protein induced by PPi. PPi (2.5 mM) was mixed with *Sm*CdnG-*Sm*E2 and incubated on ice for various time periods as indicated. *Sm*CdnG and *Sm*E2 are indicated by black and orange triangles, respectively.



Extended Data Fig. 8 *Sm*E2 negatively regulates the catalytic activity of *Sm*CdnG.

a, HPLC analysis of the 3′,2′-cGAMP synthesis activities of *Sm*CdnG and the covalently linked *Sm*CdnG-*Sm*E2 proteins expressed in different forms, related to **Fig. 5a**. *Sm*CdnG-*Sm*E2 (m), *Sm*CdnG-*Sm*E2, and *Sm*E2-*Sm*CdnG correspond to the covalently linked *Sm*CdnG-*Sm*E2 protein shown in **Fig. 2a** (i–iii), respectively. *Sm*CdnG+*Sm*E2 represents the addition of *Sm*E2 to the *Sm*CdnG reaction at a molar ratio of 1:1. **b**, Michaelis–Menten kinetics plot of ATP and GTP for *Sm*CdnG. Data points are the mean of three independently calculated rates of product formation at each concentration, and error bars indicate standard deviation. **c**, Michaelis–Menten kinetics plot of ATP and GTP for the covalently linked *Sm*CdnG-*Sm*E2 protein. Data points are the mean of three independently calculated rates of product formation at each

concentration and error bars indicate standard deviation. **d**, HPLC analysis of the 3′,2′-cGAMP synthesized by *Sm*CdnG and *Sm*CdnG mutants co-purified with *Sm*E2. **e**, HPLC analysis of the 3′,2′-cGAMP synthesized by *Sm*CdnG co-purified with *Sm*E2 or *Sm*E2 mutants.

Extended Data Methods

MS

MS-related experiments and analyses were performed by Beijing Bio-Tech Pack Technology Company Ltd. Briefly, for LC-MS/MS identification of the modified peptides, samples were separated on an SDS-PAGE gel, and the protein bands to be identified were reduced and alkylated using dithiothreitol and iodoacetamide, respectively. The protein samples were digested overnight with trypsin and chymotrypsin or Glu-C. The resulting peptides were dried on a concentrator and a vacuum concentrator and then analyzed by LC-MS/MS. For nano-LC, a 150 μm × 15 cm in-house made column packed with Acclaim PepMap RPLC C18 (1.9 μm, 100 Å, Dr. Maisch GmbH, Germany) was used. The LC mobile phase comprised solvent A ((0.1% formic acid in water) and solvent B (20% 0.1% formic acid in water and 80% acetonitrile). An EASY-nLC 1200 HPLC system (Thermo Fisher Scientific) was used to generate the following HPLC gradient: 0–2 min, 4–8% B; 2–45 min, 8–28% B; 45–55 min, 28–40% B; 55–56 min, 40–95%; 56–66 min, 95%. For MS, a Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer (Thermo Fisher Scientific, USA) was used to identify and analyze the modified peptides under the

following parameters: spray voltage, 2.2 kV; capillary temperature of 270°C. Profile MS1 spectra were acquired with the following settings: mass resolution, 70000 at 400 *m*/*z*; scan rang, 300–1800 *m*/*z*. MS/MS spectra were acquired with the following settings: activation type, HCD; normalized collision energy, 28.0 eV; activation time, 66.0 ms. For data-dependent MS/MS, the top 20 most intense peptide ions from the preview scan in the Orbitrap were used. The raw MS files were analyzed and searched against a target protein database based on the species of the samples using Byonic. The parameters were as follows: the protein modifications were carbamidomethylation (C) (variable), acetyl (protein N-terminus), oxidation (M) (variable), ATP (H, R, F, A, C, G, Q, D, E, K, L, M, N, S, Y, T, I, W, P, V) (variable), NTFYGE (C) (variable), Y (C) (variable), and G (C) (variable); the enzyme specificity was set to chymotrypsin or trypsin&Asp-N; the maximum missed cleavages were set to 3; the precursor ion mass tolerance was set to 20 ppm; and MS/MS tolerance was 0.02 Da. Only high confidence identified peptides were chosen for downstream protein identification analysis.

Protein N-terminal sequencing based on the Edman degradation method

Protein N-terminal sequencing based on the Edman degradation method was performed by Beijing Bio-Tech Pack Technology Company Ltd. Briefly, protein samples were first separated by SDS-PAGE and then transferred to a PVDF membrane (GE). The protein bands were identified by staining and cut out. A PPSQ-33A Edman sequencer (SHIMADZU) was used to analyze the protein samples on the PVDF membrane. The amino acid species were identified one by one starting from the N-terminus of the protein by a cyclic reaction, determining the N-terminal sequence. The original data

and spectra generated by PPSQ-33A were identified by PPSQ-30 data processing software and the corresponding spectrum was derived.