













Phage defence system CBASS is regulated by a prokaryotic E2 enzyme that imitates the ubiquitin pathway

Received: 15 October 2023

Accepted: 21 March 2024

Published online: 22 April 2024

 Check for updates

Yan Yan ^{1,10}, Jun Xiao ^{2,10}, Fengtao Huang ^{1,3,10} ✉, Wei Xian ^{4,5,10}, Bingbing Yu¹, Rui Cheng ¹, Hui Wu¹, Xueling Lu¹, Xionglue Wang¹, Wenjing Huang², Jing Li ², Greater Kayode Oyejobi ², Carol V. Robinson ^{6,7}, Hao Wu^{8,9}, Di Wu ^{6,7} ✉, Xiaoyun Liu ^{4,5} ✉, Longfei Wang ² ✉ & Bin Zhu ^{1,3} ✉

The cyclic-oligonucleotide-based anti-phage signalling system (CBASS) is a type of innate prokaryotic immune system. Composed of a cyclic GMP–AMP synthase (cGAS) and CBASS-associated proteins, CBASS uses cyclic oligonucleotides to activate antiviral immunity. One major class of CBASS contains a homologue of eukaryotic ubiquitin-conjugating enzymes, which is either an E1–E2 fusion or a single E2. However, the functions of single E2s in CBASS remain elusive. Here, using biochemical, genetic, cryo-electron microscopy and mass spectrometry investigations, we discover that the E2 enzyme from *Serratia marcescens* regulates cGAS by imitating the ubiquitination cascade. This includes the processing of the cGAS C terminus, conjugation of cGAS to a cysteine residue, ligation of cGAS to a lysine residue, cleavage of the isopeptide bond and poly-cGASylation. The poly-cGASylation activates cGAS to produce cGAMP, which acts as an antiviral signal and leads to cell death. Thus, our findings reveal a unique regulatory role of E2 in CBASS.

Cyclic GMP–AMP synthase (cGAS) has a key role in mammalian cGAS–STING innate immunity^{1,2}. Coincidentally, cGAS shares a conserved catalytic domain with the *Vibrio cholerae* enzyme dinucleotide cyclase (DncV)³; thus, this family of enzymes has been classified as cGAS/DncV-like nucleotidyltransferases (CD-NTases)^{4,5}. The prokaryotic CD-NTases and ancillary proteins, which are encoded by operons and involved in bacterial anti-phage defence, constitute the cyclic-oligonucleotide-based anti-phage signalling system (CBASS)⁶.

Interestingly, certain CBASSs share evolutionary origins with the eukaryotic cGAS–STING innate immunity pathway^{6–8}, indicating similar underlying principles between bacterial anti-phage defence and mammalian innate immunity.

Over 5,000 distinct CBASSs have been identified in more than 10% of bacteria with known genome sequences^{5,9,10}. Furthermore, about half of all CBASS operons encode a CD-NTase that can generate signal molecules in response to phage infection, and an effector that is

¹Key Laboratory of Molecular Biophysics, the Ministry of Education, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, China. ²Department of Cardiovascular Surgery, Taikang Center for Life and Medical Sciences Zhongnan Hospital of Wuhan University, School of Pharmaceutical Sciences, Wuhan University, Wuhan, China. ³Shenzhen Huazhong University of Science and Technology Research Institute, Shenzhen, China. ⁴Microbiology and Infectious Disease Center, School of Basic Medical Sciences, Peking University Health Science Center, Beijing, China. ⁵NHC Key Laboratory of Medical Immunology, Peking University, Beijing, China. ⁶Department of Chemistry, University of Oxford, Oxford, UK. ⁷Kavli Institute for Nanoscience Discovery, University of Oxford, Oxford, UK. ⁸Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA. ⁹Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA, USA. ¹⁰These authors contributed equally: Yan Yan, Jun Xiao, Fengtao Huang, Wei Xian. ✉ e-mail: huang_fengtao@126.com; di.wu2@chem.ox.ac.uk; xiaoyun.liu@bjmu.edu.cn; wanglf@whu.edu.cn; bin_zhu@hust.edu.cn

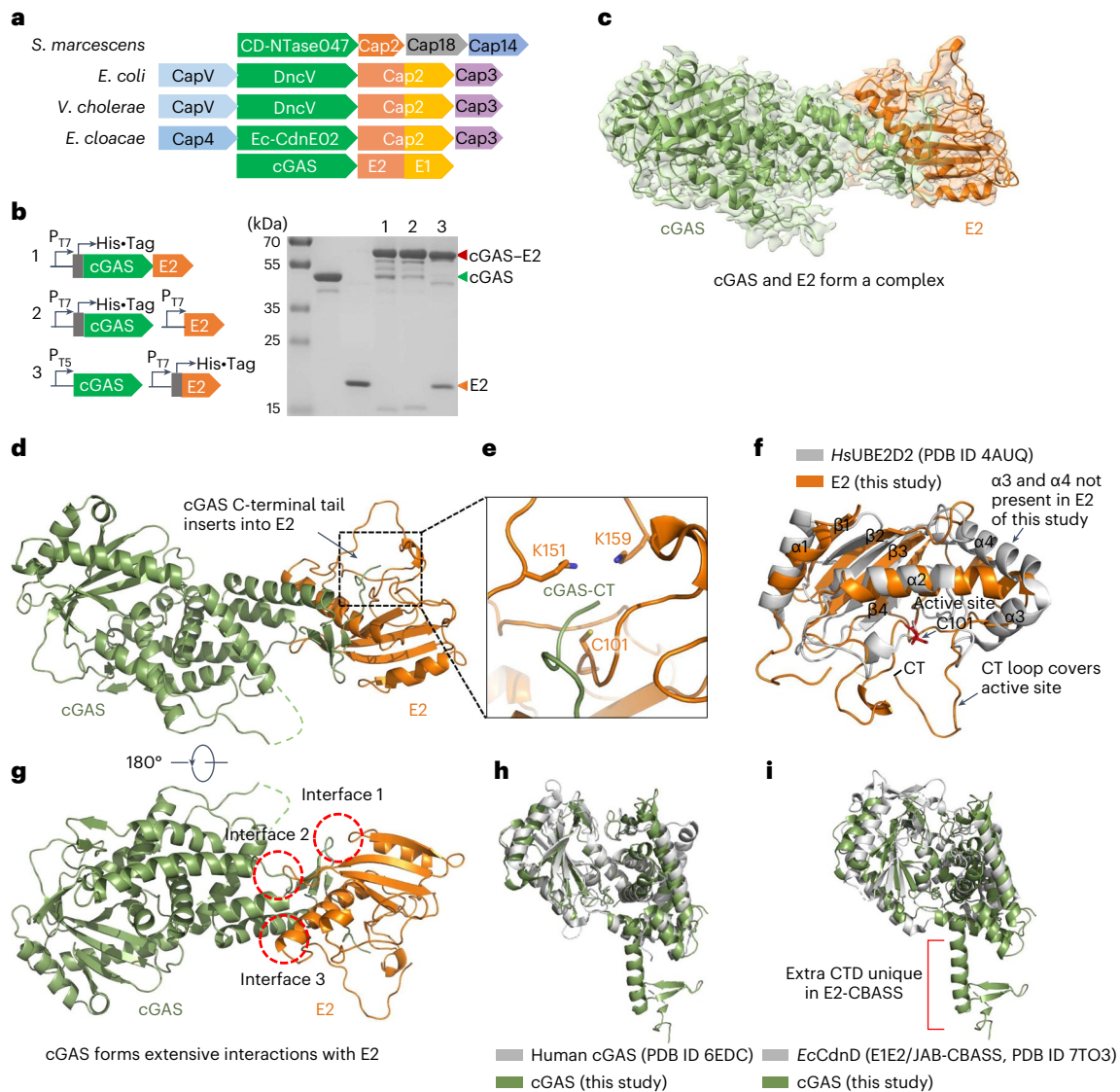


Fig. 1 | Cryo-EM structure of a cGAS-E2 complex. **a**, Domain organization of the CBASS operons from *S. marcescens*, *E. coli*¹⁵, *V. cholerae*¹⁴. **b**, Left, schematic diagram showing the constructs for co-expression of cGAS and E2 in different forms (1-3). Right, SDS-PAGE analysis of the purified proteins from various constructs as shown on the left. Data are representative of three independent experiments. **c**, Overlay of cryo-EM density and the atomic model of the cGAS-E2 complex. **d, e**, Ribbon diagram of the cGAS-E2 complex highlighting the C-terminal tail of cGAS (green) that is inserted into the E2 domain (orange).

with the potential interaction region (boxed) enlarged (**e**). **f**, Superimposed structures of E2 used in this study (orange) and human E2 (grey, PDB identifier (ID) 4AUQ). CT, C terminus. **g**, Ribbon diagram of the cGAS-E2 complex highlighting the interface between cGAS and E2. **h**, Superimposed structures of cGAS used in this study (green) and human cGAS (grey, PDB ID 6EDC). **i**, Superimposed structures of cGAS used in this study from E2-CBASS (green) and *EcCdnD*¹⁴ from E1E2/JAB-CBASS (grey, PDB ID 7TO3). CTD, C-terminal domain.

activated by the cyclic oligonucleotides and induces programmed cell death; the remaining CBASS operons encode ancillary genes^{9,10}. Some of these ancillary components have been reported to function as a threat sensor or a regulator of bacterial defences^{7,8,11}; however, most of their functions and mechanisms are unclear.

Among the over 5,000 predicted CBASSs, 2,199 CBASSs encode homologues of ubiquitin systems, which are typical protein post-translational modification systems in eukaryotes that use an activation, transfer and ligation cascade¹⁰. The ubiquitin system is highly conserved in eukaryotes and has a critical role in a variety of cellular processes, including protein degradation, DNA repair and signal transduction^{12,13}. The ubiquitination cascade consists of the ubiquitin-activating enzyme E1, ubiquitin-conjugating enzyme E2 and ubiquitin-ligating enzyme E3 to covalently link ubiquitin to a substrate; conversely, the deubiquitinating enzyme can remove ubiquitin from a substrate¹². CBASS systems encoding domains homologous to the

eukaryotic ubiquitin machinery can be divided into two main classes. The first is a type II CBASS consisting of 1,583 systems and containing an E1-E2 fusion protein and a JAB deubiquitinating peptidase, with no E3 ligase^{14,15} (hereafter referred to as E1E2/JAB-CBASS; Fig. 1a). The second is a short type II CBASS consisting of 616 systems and containing only an E2 protein^{9,10} (hereafter referred to as E2-CBASS; Fig. 1a), the function of which is elusive. In this work, we reveal that E2-CBASS uses a more concise and specific mechanism, with a distinct protein chemistry, to regulate anti-phage signalling, compared with the reported mechanism of E1E2/JAB-CBASS^{14,15}.

Results

Structure of the E2-cGAS complex reveals unique interactions
To elucidate the role of the CBASS-related E2 protein, we focused on an E2-CBASS operon from *Serratia marcescens* (Fig. 1a). The E2-CBASS operon encodes a cGAS (CD-NTase047; Extended Data Fig. 1), an E2

(Cap2), a predicted 3'–5' exonuclease (Cap18) and a nucleotide sensor domain fused to two transmembrane helices⁹ (2TM-SAVED, Cap14; Fig. 1a). Reversed-phase liquid chromatography with ultraviolet detection and mass spectrometry revealed that the product of cGAS is 3',2'-cGAMP (Extended Data Fig. 1). Interestingly, when cGAS and E2 were co-expressed and co-purified, an additional band with a molecular weight of 64 kDa appears on the SDS–PAGE gel, suggesting that cGAS (46 kDa) and E2 (18 kDa) were fused (Fig. 1b). When separately purified cGAS and E2 were mixed at a 1:1 molar ratio, the linkage still formed without cofactors in the storage buffer (50 mM Tris–HCl (pH 7.5), 100 mM NaCl, 0.1 mM EDTA, 0.1% Triton X-100 and 50% glycerol) (Extended Data Fig. 2a). This fusion of E2 and cGAS could not be separated by SDS, confirming a covalent linkage between the two.

To elucidate the structural basis of the cGAS–E2 complex, the complex was purified and structural studies were performed using cryo-electron microscopy (cryo-EM) (Extended Data Fig. 2b–d). Two cryo-EM datasets of the cGAS–E2 complex were collected and processed (Extended Data Fig. 3a,b and Supplementary Table 1), and the three-dimensional (3D) reconstruction was refined to 3.3 Å and 3.1 Å, respectively. An initial model of the cGAS–E2 complex was generated using AlphaFold and ColabFold^{16,17} and rigid-body fitted into the electron density map (Extended Data Fig. 2e,f). Although the cryo-EM density map of cGAS–E2 was anisotropic as a result of preferred orientations, it was of sufficient quality to perform model refinement guided by the initial model (Extended Data Fig. 3c–e).

The refined cryo-EM structure of cGAS–E2 is consistent with the AlphaFold prediction (Extended Data Fig. 3f), with E2 binding to the C-terminal region of cGAS (Fig. 1c,d). The C terminus of cGAS is threaded into a pocket of E2 (Fig. 1e) and sits in close proximity with the side chains of one Cys and two Lys residues (C101, K151 and K159). We performed a structure similarity search of *S. marcescens* E2 against published structures in the Protein Data Bank (PDB) using the Dali server¹⁸. One of the top hits was a human E2 (UBE2D2)¹⁹ with a root mean square deviation of 2.5 (Fig. 1f). Interestingly, C101 of *S. marcescens* E2 is located in close proximity to the active site of human E2, indicating that it is the active-site cysteine. Notably, *S. marcescens* E2 has a C-terminal loop (residue 150–160) that covers the active site, unlike the C-terminal α -helices (α 3 and α 4) found in human E2. Furthermore, in contrast to the E1E2/JAB–CBASS system from *Enterobacter cloacae*¹⁴, where cGAS interacts with dimerized E1–E2 fusion, cGAS of E2–CBASS forms extensive interactions with E2 at three interfaces in addition to the C terminus (Fig. 1g). Thus, although the core region of cGAS from E2–CBASS resembles human cGAS and cGAS from the E1E2/JAB–CBASS system, the C-terminal region that mediates E2 interactions is completely unique (Fig. 1h,i).

Covalent linkage between E2 and cGAS

In eukaryotes, ubiquitin forms a thioester bond between its C-terminal Gly and a Cys in E1, E2 and E3, or forms an isopeptide bond between its C-terminal Gly and a Lys in a target protein or another ubiquitin. Similarly, the C-terminal tail of *S. marcescens* cGAS (residues 400–407) is inserted into the catalytic pocket of E2 (Fig. 1e), as previously discussed. The C-terminal residue of cGAS in the complex is probably either G406 or Y405 based on densities of the cryo-EM structure. Interestingly, the C-terminal residue of cGAS is in close proximity to C101, K151 and K159 of E2, suggesting that the covalent bond between cGAS and E2 is either a thioester bond between cGAS and C101 of E2, or an isopeptide bond between cGAS and K151 or K159 of E2.

Alignment of the C-terminal sequences of cGAS (CD-NTase) homologues associated with E2 in various CBASSs showed a conserved glycine near their C termini (Extended Data Fig. 4a and Supplementary Fig. 1). Deletion of the C-terminal eight residues (Δ P400–E407) or two residues (Δ G406–E407) of cGAS abolished the covalent linkage between cGAS and E2, while Δ E407 had no effect. These results indicate that G406 participates in the covalent linkage with E2 (Fig. 2a).

Consistently, mutation of G406 to Val or Leu also abolished the covalent linkage, while mutation of G406 to Ala, which has a shorter side chain, still allowed the linkage (Fig. 2a). In addition, we mutated two residues in the C terminus of cGAS with potential for covalent interactions, finding that both K401A and Y405A mutations showed no effect (Fig. 2a).

As previously mentioned, C101 is the active-site cysteine (Fig. 1f) and is conserved among prokaryotic E2 homologues (Extended Data Fig. 4b and Supplementary Fig. 2). To confirm its importance, we made a C101A mutant, which, unsurprisingly, failed to form the covalent complex with cGAS (Fig. 2a), in contrast to C69A that showed no effect. These results indicate a thioester bond may form between cGAS G406 and E2 C101.

To validate the thioester bond, we incubated cGAS–E2 fusion with 10 mM dithiothreitol (DTT), as thioester bonds are unstable in the presence of DTT²⁰. Interestingly, the cGAS–E2 complexes are mostly resistant to DTT treatment and only a small portion was separated by 10 mM DTT (Fig. 2b). Therefore, the majority of cGAS–E2 complexes are probably not linked by a thioester bond, but rather by an isopeptide bond.

To test whether cGAS and E2 are linked by an isopeptide bond, we mutated K151 and K159 of E2, which are close to the C terminus of cGAS (Fig. 1e). Mutation of K159A but not K151A significantly reduced the cGAS–E2 complex in the presence of DTT (Fig. 2b), indicating that the DTT-insensitive bond between cGAS and E2 is an isopeptide bond, which mirrors the isopeptide bond formation between ubiquitin and its target protein. Interestingly, E2(K159A) mutation not only reduced the DTT-insensitive isopeptide bond, but also increased the sensitivity of the cGAS–E2 complex to DTT (Fig. 2b), indicating that both types of covalent bonds exist and the E2 studied could catalyse the formation of both thioester and isopeptide bonds.

Consistently, cryo-EM structures of cGAS–E2 exhibit heterogeneity of E2, resulting in its relatively low resolution (Extended Data Fig. 3e). Nevertheless, continuous density was observed connecting Y405 and K159 in the 3.1 Å dataset (Extended Data Fig. 5a). Therefore, bacterial E2 mimics the function of both eukaryotic E2 and E3. To directly show the isopeptide bond between cGAS and E2, we performed mass spectrometry analysis of the cGAS–E2 complex. Interestingly, the results showed that an NTFY peptide of cGAS attached to the peptide HPRPSEVKNER of E2 at K159, indicating the formation of an isopeptide bond between cGAS Y405 and E2 K159 (Fig. 2c and Extended Data Fig. 5b). Therefore, G406 of cGAS is recognized but removed during protein ligation, partially resembling the sortase-mediated protein ligation in which a C-terminal glycine of the donor peptide is recognized but cleaved off^{21,22}. This distinguishes the E2–CBASS from E1E2/JAB–CBASS, in which a C-terminal glycine of cGAS directly participates in the covalent bonding^{14,15}. Consistently, shown by the sequence alignments of cGAS from E1E2/JAB–CBASSs and E2–CBASSs, cGAS from E1E2/JAB–CBASSs possesses a very conserved C-terminal glycine or alanine residue, while cGAS from E2–CBASSs usually has a glycine near their C termini (Extended Data Fig. 6a and Supplementary Figs. 1 and 3).

The formation of either the Y405–C101 thioester or the Y405–K159 isopeptide bond requires the removal of the last two C-terminal residues G406 and E407 of cGAS. In eukaryotes, the removal of C-terminal residues of ubiquitin is carried out by specific proteases¹². As the G406–E407 peptide is too small to observe, we constructed a cGAS–maltose-binding protein (MBP) by attaching an MBP to the C terminus of cGAS to monitor the cleavage. When cGAS–MBP was mixed with E2 in vitro, SDS–PAGE revealed the presence of a protein with the same molecular weight as cGAS–E2 (64 kDa), suggesting that the covalent linkage between cGAS–MBP and E2 was formed and the C terminus of cGAS–MBP was removed (Fig. 2d). The presence of a protein with the same molecular weight as MBP further confirmed that the C terminus of cGAS was indeed removed during the covalent bonding (Fig. 2d). Furthermore, Edman sequencing of the cleaved C-terminal fragment revealed an N-terminal sequence of GESSSS (Fig. 2e), confirming that G406 is removed during the ligation of cGAS and E2. In addition, we constructed the mutant G406A of cGAS–MBP, which could also form a covalent linkage with E2 (Fig. 2a).

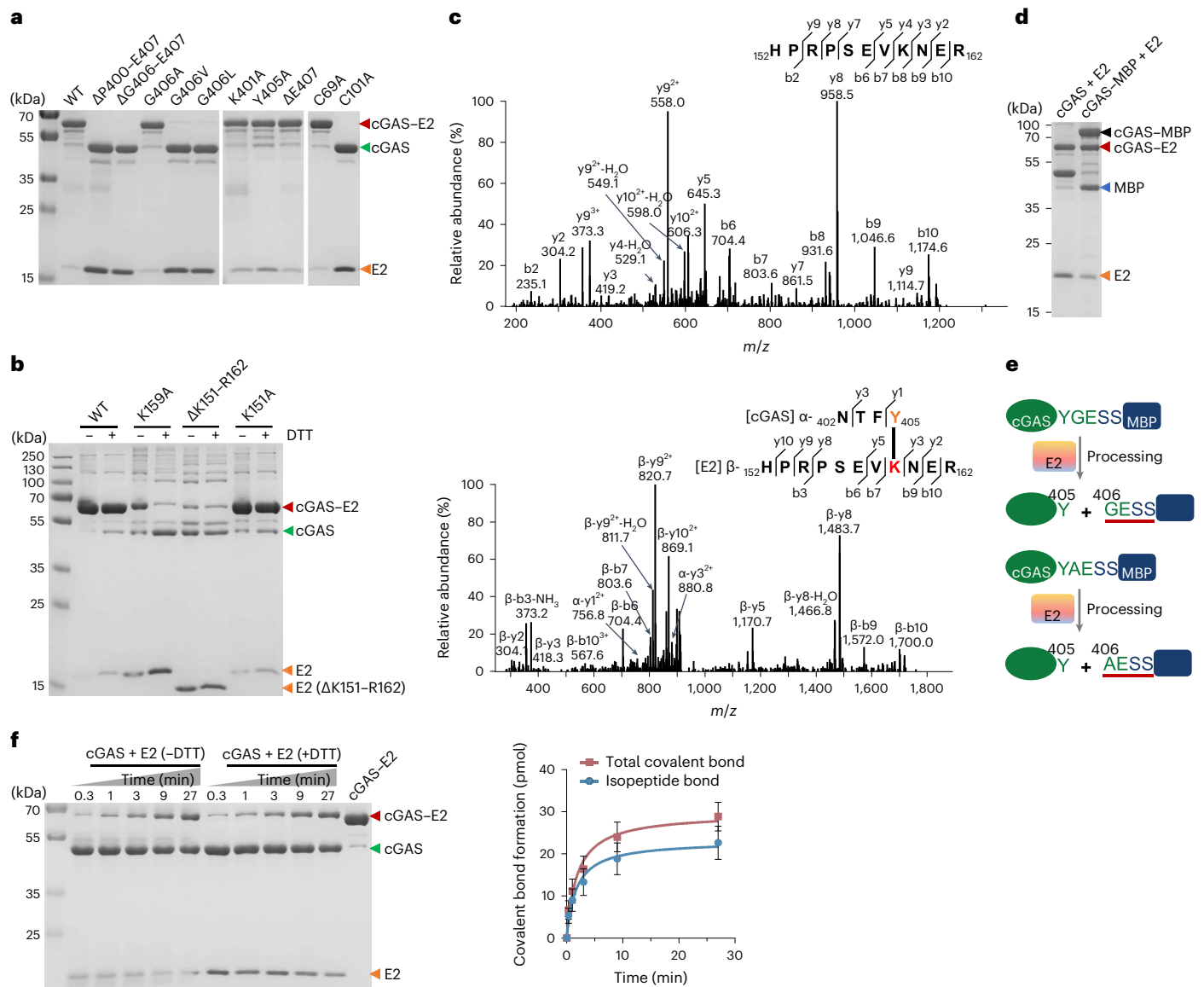


Fig. 2 | Covalent linkage between E2 and cGAS. **a**, SDS-PAGE analysis of cGAS mutants co-purified with E2, or E2 mutants co-purified with cGAS. The G406 residue of cGAS and the C101 residue of E2 were identified as critical residues for the covalent linkage. Data are representative of three independent experiments. WT, wild type. **b**, SDS-PAGE analysis of the co-purified proteins consisting of cGAS with E2 or E2 mutants in the presence (+DTT) or absence (-DTT) of 10 mM DTT (10 mM DTT was used in all succeeding analyses). The K159 residue of E2 was identified as the key residue for the isopeptide bonding. Data are representative of three independent experiments. **c**, Top, MS/MS analysis of peptide HPRPSEVKNIE₁₆₂ derived from E2. Bottom, MS/MS analysis of the same peptide bearing an NTFY remnant at K159 from the cGAS-E2 complex. The fragment ions β -b9, β -b10, β -y5 and β -y8- β -y10 from the β -chain (E2) have a mass increase of 525 Da corresponding to the addition of NTFY, whereas the fragment ions β -b3, β -b6, β -b7, β -y2 and β -y3 do not exhibit this mass shift, confirming the

formation of an isopeptide bond between cGAS Y405 (orange) and E2 K159 (red). Data are representative of three independent experiments. **d**, SDS-PAGE analysis showing the *in vitro* processing and bonding of cGAS-MBP by E2. cGAS-MBP or cGAS and E2 were mixed at a molar ratio of 1:1 overnight. Data are representative of three independent experiments. **e**, Cartoon diagram showing the C-terminal processing of cGAS-MBP or cGAS(G406A)-MBP by E2. The N-terminal sequences (underlined) of the released MBP-containing fragments indicate that either G406 or A406 is removed from cGAS. **f**, Left, SDS-PAGE analysis showing the *in vitro* formation of the cGAS-E2 complex in the presence (+DTT) or absence (-DTT) of 10 mM DTT. Right, rate of the total bond (thioester + isopeptide, -DTT) and isopeptide bond (+DTT) formation between cGAS and E2, as calculated based on the gels in the left panel. Data are the mean \pm s.d. of $n = 3$ independent replicates and are representative of three independent experiments.

Consequently, the N-terminal sequence of the released MBP-containing fragment was found to be AESSSS (Fig. 2e). These results confirmed that the cleavage occurs between Y405 and G406 of cGAS, and G406 is removed to form the Y405-K159 isopeptide bond. Thus, there is a distinguishing chemistry of the cGAS-E2 protein ligation to those in the ubiquitin¹² and E1E2/JAB-CBASS systems^{12,13}.

As evident from the ubiquitin cascade, it is likely that the unstable thioester bond between cGAS and E2 is transient and will be converted to a more stable isopeptide bond. The rates of total bond

(thioester + isopeptide, -DTT, $0.63 \pm 0.17 \text{ min}^{-1}$) and isopeptide bond (+DTT, $0.50 \pm 0.14 \text{ min}^{-1}$) formation *in vitro* were measured and are shown in Fig. 2f.

Cysteine protease activity of E2

The eukaryotic ubiquitin system¹² and the E1E2/JAB-CBASS system^{14,15} both rely on deubiquitinating enzymes such as DUBs or JAB, respectively, which are mostly cysteine proteases. However, the processing of cGAS requires only E2, suggesting that E2 itself may function as a

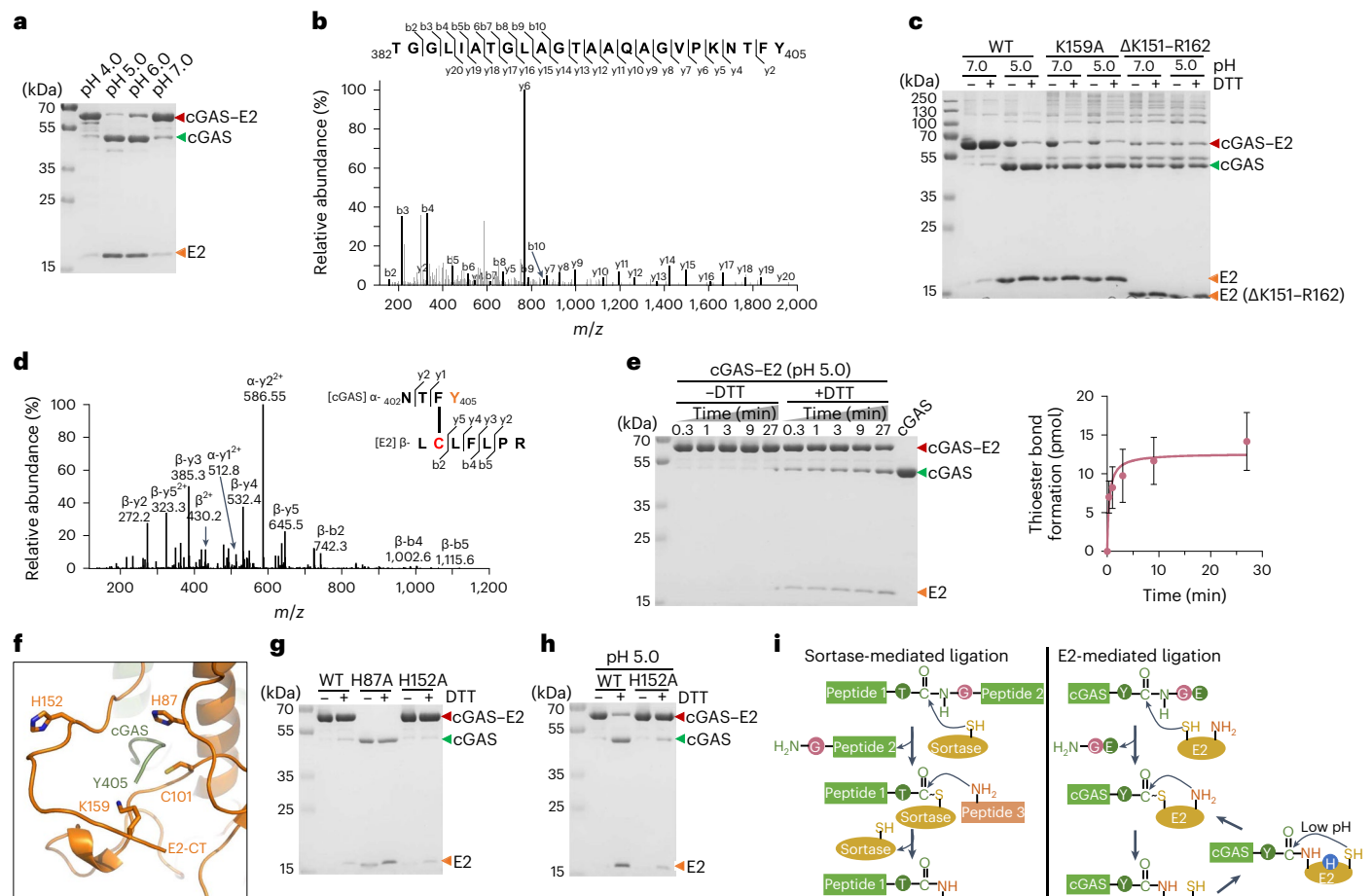


Fig. 3 | The function of E2 as a cysteine protease. a, SDS-PAGE analysis showing the low-pH-induced autocleavage of the covalent cGAS-E2 complex. Data are representative of three independent experiments. **b**, MS/MS analysis shows that the C terminus of released cGAS after DTT and low-pH treatment is terminated by Y405. Data are representative of three independent experiments. **c**, SDS-PAGE analysis of the low-pH-induced bond conversion in cGAS-E2 or cGAS-E2(K159A) or cGAS-E2(ΔK151-R162) in the presence (+DTT) or absence of DTT (-DTT). Data are representative of three independent experiments. **d**, Mass spectrometry analysis of the ligation site in cGAS-E2(K159A); the same spectrum was also detected in cGAS-E2 (pH 5.0). The fragment ions β-b2, β-b4 and β-b5 from the β-chain (E2) have a mass increase of 525 Da corresponding to the addition of NTFY, whereas the fragment ions β-y2, β-y3, β-y4 and β-y5 do not exhibit this mass shift, confirming the formation of a thioester bond between cGAS Y405 (orange) and E2 C101 (red). Data are representative of three independent experiments.

e, Rate of the thioester bond formation in cGAS-E2 at pH 5.0 (right) calculated based on gel analysis (left). Data are the mean ± s.d. of $n = 3$ independent replicates and are representative of three independent experiments. **f**, Magnified view of E2 showing C101, H87 and H152 and other key residues in the active site. E2-CT, the C terminus of E2. **g**, SDS-PAGE analysis of cGAS co-purified with E2 or E2 mutants (as indicated on the above gel) in the presence (+DTT) or absence (-DTT) of DTT. The H87 residue of E2 was identified to be crucial for the covalent linkage. Data are representative of three independent experiments. **h**, SDS-PAGE analysis of the low-pH-induced autocleavage of the covalent cGAS-E2 complex in the presence (+DTT) or absence (-DTT) of DTT. The H152 residue of E2 is crucial to mediate the low-pH-induced cleavage of the isopeptide bond between cGAS and E2. Data are representative of three independent experiments. **i**, Schematic showing the sortase-mediated (left) and E2-mediated (right) protein ligation and the low-pH-induced bond transition in the cGAS-E2 complex.

cysteine protease. Coincidentally, the presence of C101 provides a structural basis for an active-site cysteine. As low-pH-induced autocleavage is well known for classical cysteine proteases such as asparaginyl endopeptidase and cathepsin L^{23,24}, we examined the effect of low pH on the cGAS-E2 complex. Interestingly, pH 5.0 induced the cleavage of the isopeptide bond between cGAS and E2, as shown by the reduction of the cGAS-E2 complex in the presence of DTT (Fig. 3a). This was also confirmed by mass spectrometry data showing that the cGAS C terminus ends at Y405 after treatment by low pH and DTT (Fig. 3b). However, in the absence of DTT, we found that, like the K159A mutation, pH 5.0 also increased the thioester bond; pH 5.0 had no effect on E2(K159A) or E2(ΔK151-R162) mutants (Fig. 3c). With K159A mutation in E2 or pH 5.0, we were able to capture the thioester bond between cGAS Y405 and E2 C101 by mass spectrometry (Fig. 3d and Extended Data Fig. 7a). When the pH was adjusted from 5.0 to 8.0, the thioester bond (-DTT) between cGAS and E2 was converted back to an isopeptide

bond (+DTT; Extended Data Fig. 7b). Our results confirm that the two types of covalent bond are interconvertible. The rate of thioester bond formation at pH 5.0 ($0.50 \pm 0.07 \text{ min}^{-1}$) is shown in Fig. 3e.

Another typical feature of cysteine proteases is a His residue near the catalytic Cys for proton transfer during proteolytic catalysis^{25,26}. Interestingly, our structure of the cGAS-E2 complex reveals two His (H87 and H152) residues that are in close proximity to C101 (Fig. 3f). The E2(H87A) mutation, such as C101A (Fig. 2a), abolished the covalent linkage between cGAS and E2 (Fig. 3g). Both mutants failed to cleave the MBP of cGAS-MBP (Extended Data Fig. 7c), confirming the catalytic roles of C101 and H87 in E2 protease-like activity. The E2(H152A) mutation showed no effect on the complex formation (Fig. 3g); however, it significantly reduced the conversion from the isopeptide bond to the thioester bond at pH 5.0 (Fig. 3h). These results suggest that H152 is a pH sensor at the protease active site, similar to those reported previously^{27,28}.

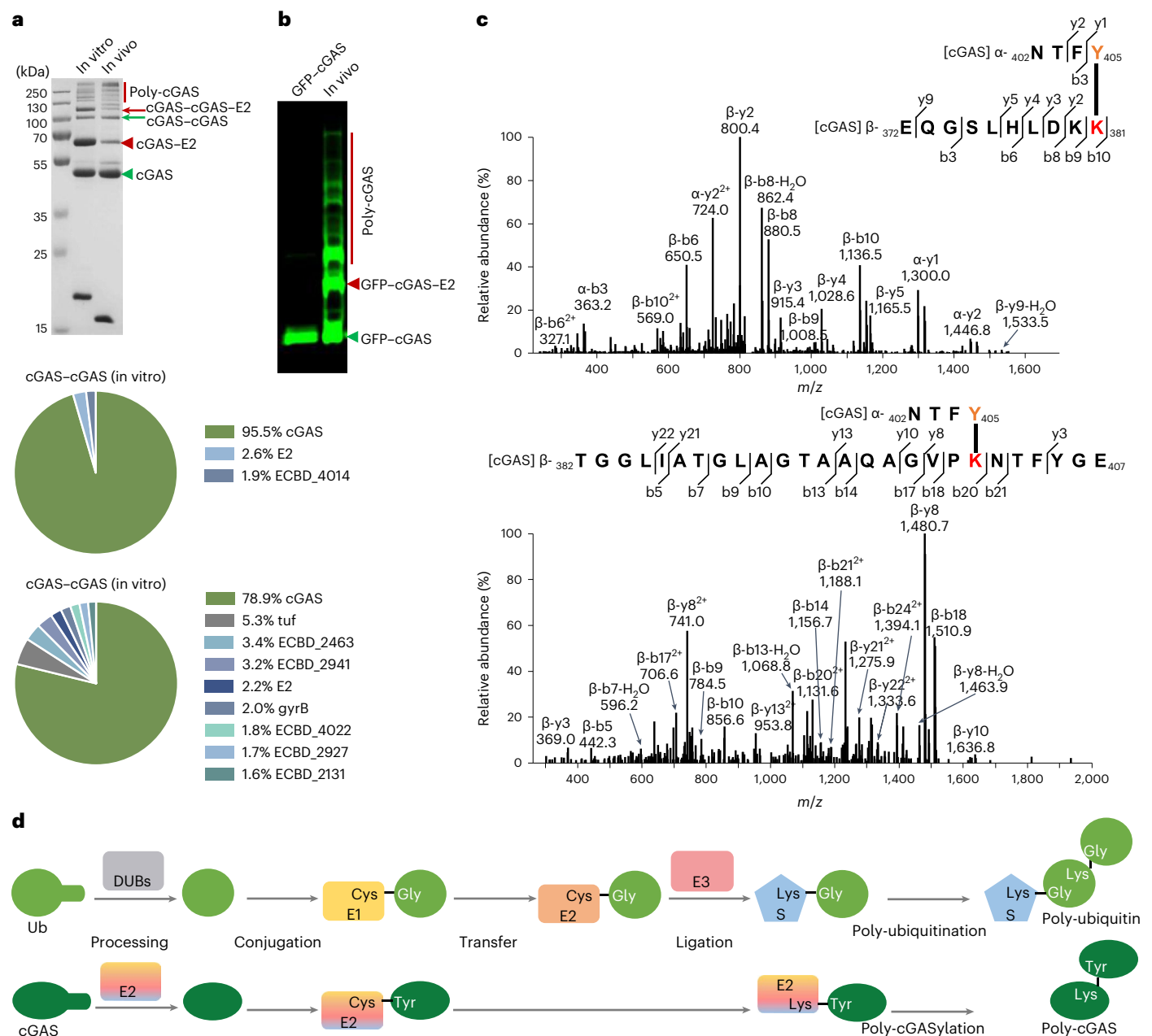


Fig. 4 | E2 catalyses poly-cGASylation of cGAS. **a**, Top, SDS-PAGE analysis showing the in vivo and in vitro formation of the covalent cGAS-E2(K159A) protein. The in vivo formation of the covalently linked cGAS-E2(K159A) protein was achieved by the co-purification of His-tagged cGAS and non-tagged E2(K159A) mutant. For the in vitro assembly, two proteins were mixed at a molar ratio of 1:1 overnight. Bottom, mass spectrometry showing the score of proteins contained in the potential cGAS-cGAS extracted from the gel (marked by the green arrow). Mass spectrometry signals corresponding to *E. coli* protein contaminants in the cGAS-cGAS dimer are detailed in Supplementary Table 2. Data are representative of two independent experiments. ECBD, *Escherichia coli* (strain B/BL21-DE3); tuf, elongation factor Tu. **b**, SDS-PAGE showing the in vivo formation of poly-cGAS. GFP-tagged cGAS was co-expressed and co-purified with E2(K159A). Data are representative of three independent experiments. **c**, Mass spectrometry analysis of the ligation sites in cGAS-cGAS. Top, tandem mass spectrum of the cGAS

peptide EQGSLHLDKK bearing an NTFY remnant at K381 (red). The fragment ion β -b10 and all β -y ions from the β -chain (cGAS) have a mass increase of 525 Da corresponding to the addition of NTFY, whereas the fragment ions β -b3, β -b6, β -b8 and β -b9 do not exhibit this mass shift, confirming the formation of an isopeptide bond between cGAS Y405 (orange) and K381 of another cGAS. Bottom, tandem mass spectrum of the cGAS peptide TGGLIATGLAGTAAQAGVVPKNTFYGE bearing an NTFY remnant at K401 (red). The fragment ions β -b20, β -b21, β -y8, β -y10, β -y13, β -y21 and β -y22 from the β -chain (cGAS) have a mass increase of 525 Da corresponding to the addition of NTFY, whereas the fragment ions β -b5, β -b7, β -b9, β -y10, β -b13, β -b14, β -b17, β -b18 and β -y3 do not exhibit this mass shift, confirming the formation of an isopeptide bond between cGAS Y405 (orange) and K401 of another cGAS. Data are representative of three independent experiments. **d**, Schematic diagrams showing the eukaryotic ubiquitin pathway (top) and E2-mediated cGAS regulation (bottom).

Thus, we have found that E2 possesses both the catalytic^{29,30} and regulatory elements of proteases. The protease activity of E2 appears unique, with sequence alignments of E2 from E2-CBASSs and the E2 domain from E1E2/JAB-CBASSs showing that the histidine residues corresponding to H87 and H152 are conserved only in E2 from the

E2-CBASSs, but not in E2 from the E1E2/JAB-CBASSs (Extended Data Fig. 6b,c and Supplementary Figs. 2, 4 and 5). In addition, no similar histidine residue was observed in eukaryotic E2 (Extended Data Fig. 6d). Together, the E2-mediated protein ligation is comparable to that by a sortase and depicted in Fig. 3i.

Poly-cGASylation of cGAS by E2

Notably, the disruption of the Y405–K159 isopeptide bond by E2(K159A) mutation not only increased the Y405–C101 thioester bond but also generated larger species insensitive to DTT (Fig. 2b). The molecular weight of these proteins indicates that they are polymers containing various numbers of cGAS. Either co-expressed with cGAS or mixed with cGAS in vitro, the E2(K159A) mutant generated the cGAS polymers (Fig. 4a). We excised the gel bands corresponding to the cGAS–cGAS dimer from both co-expressed and in-vitro-assembled cGAS–E2(K159A) and analysed them by mass spectrometry. The results showed that the majority of detected peptides were from cGAS, with minimal signals from E2 and *Escherichia coli* protein contaminants (Fig. 4a and Supplementary Table 2), supporting the formation of the cGAS–cGAS dimer. The poly-cGASylation was also directly visualized by tagging cGAS with GFP (Fig. 4b). These cGAS polymers are insensitive to DTT, suggesting that they are linked by isopeptide bonds similar to poly-ubiquitin.

It is noteworthy that the high molecular weight protein conjugates that were formed in vivo and in vitro were not identical (Fig. 4a, top). In addition, more peptide signals corresponding to *E. coli* proteins were detected from the in vivo conjugations by mass spectrometry (Fig. 4a, bottom). Although E2 may have a strong preference to catalyse poly-cGAS formation, it may also conjugate cGAS to other proteins, as reported for the E1E2/JAB-CBASS^{14,15}. Tandem mass spectrometry (MS/MS) revealed that K381 or K401, both near the C terminus of cGAS, are the conjugation sites between cGASs (Fig. 4c and Extended Data Fig. 8a,b). Like those in the ubiquitin system¹³ and E1E2/JAB-CBASS^{14,15}, it is likely that the isopeptide linkages between cGASs are heterogeneous. In addition to the K381 and K401 that were revealed by mass spectrometry as linkage sites (Fig. 4c), we mutated all 16 individual lysines in cGAS to identify the 'preferred' target lysine. We found that none of the alanine mutants prevented or significantly affected the poly-cGAS formation (Extended Data Fig. 8c), confirming that the linking sites between cGASs are heterogeneous. Deletion of the C terminus of E2, including K151 and K159 (Δ K151–R162), resulted in more significant poly-cGASylation and less thioester bond formation (Fig. 2b). Altogether, the E2 functionally resembles the eukaryotic ubiquitination system (Fig. 4d).

cGAS is activated by E2-mediated poly-cGASylation

Human cGAS is activated through oligomerization^{31,32}. The poly-cGAS observed in this study is probably the active form of cGAS. We compared the cGAMP synthesis activities of free cGAS, cGAS–E2, cGAS–E2(K159A) and cGAS–E2(Δ K151–R162). The concentrations of cGAS in the four samples were kept the same (800 nM); nevertheless, the latter two samples contained more poly-cGAS (Fig. 5a). Interestingly, although the cGAS–E2 linkage inhibited the activity of cGAS, cGAS–E2(K159A) and cGAS–E2(Δ K151–R162) significantly increased the synthesis activity of cGAS when poly-cGAS was formed (Fig. 5a and Extended Data Fig. 9a). Furthermore, when the concentration of ATP substrate was lowered by half, there was a more significant increase of synthesis activity by poly-cGAS (Fig. 5b and Extended Data Fig. 9b), suggesting that the poly-cGASylation increased the affinity of cGAS to ATP.

Immunoblotting using a cGAS-specific antibody detected the poly-cGAS in the CBASS-expressing *E. coli* upon T4 phage infection (Fig. 5c), confirming that the phage-induced poly-cGASylation occurs in vivo. In addition, recent work has shown that CD-NTase + 2TM-*SAVED* operons use antiviral nucleotide signals to activate CBASS TM effectors, which in turn induce inner-membrane disruption and cause cell death to limit phage replication³³. With this effector, the antiviral signal producer cGAS, and a regulatory E2, the E2-CBASS operon confers strong immunity on *E. coli* against bacteriophage T4, as judged by plaque formation (Fig. 5d). This immunity was abolished upon (1) deletion of either cGAS, E2 or the effector 2TM-*SAVED* gene; (2) disruption of the second messenger synthesis by the cGAS D84A/D86A mutation;

(3) disruption of the covalent linkage by the cGAS(G406R), E2(C101A) or E2(H87A) mutation; and (4) disruption of the low-pH-dependent bond conversion by E2(H152A) (Fig. 5d). Although K159 is the preferred residue for stable isopeptide bond formation, its mutation did not show significant effect on the antiviral immunity (Fig. 5d), indicating that the partial releasing of poly-cGAS by K159A does not impede the antiviral immunity.

Discussion

The ubiquitin machinery consists of at least five proteins: a protease, an E1, an E2, an E3 and a target protein. We show that the bacterial E2 in our study regulates cGAS by imitating the ubiquitination cascade, functioning like a protease, an E1 and E2, and an E3 and a target protein. This is evident in the ability of E2 to cleave the C terminus of cGAS or the isopeptide bond, conjugate cGAS to a cysteine residue, and ligate cGAS to a lysine residue (Fig. 4d). This multifunctional E2 provides compelling evidence for the central role of E2 in ubiquitin systems^{34,35}.

Although the E2 family was considered evolutionarily diverse³⁶, the E2 reported in this work is unique in its catalytic mechanism. Despite the E2-CBASS and the E1E2/JAB-CBASS^{14,15} both belonging to the type II CBASS, the E2 from E1E2/JAB-CBASS functions more similarly to the E2 from the eukaryotic ubiquitin system, both dependent on E1 to activate cGAS or ubiquitin in the presence of ATP^{12,14,15}. Conversely, the E2 from E2-CBASS processes the cGAS directly without any cofactors. Furthermore, the E2 from E1E2/JAB-CBASS and the E2 from the eukaryotic ubiquitin system both specifically link to a glycine at the C terminus of cGAS or ubiquitin, respectively, while the E2 from E2-CBASS recognizes a glycine in the cGAS but removes it for ligation. Moreover, the protein chemistry of the latter is more closely related to the reactions catalysed by a sortase^{21,22}. However, the most interesting features of the E2 from E2-CBASS are its cysteine protease activity and the related structural elements, such as the regulatory histidine (Extended Data Fig. 6b,c and Supplementary Fig. 5), which appear to be unique among all known E2s to date³⁶. Dop and PafA that catalyse ubiquitination-like protein modification had been found in *Mycobacterium tuberculosis*^{37,38}. However, these prokaryotic enzymes are structurally similar to the carboxylate-amine/ammonia ligase superfamily of glutamine synthetases, which are evolutionarily distinct from E2 (refs. 38,39). The evolutionary relationship of the eukaryotic ubiquitin system, the E1E2/JAB-CBASS, the E2-CBASS and the cysteine protease family, needs to be further explored.

One interesting question that remains unsolved is how a single E2 transfers a cGAS that is covalently attached to itself to another cGAS. Although our current results are not adequate to fully address the complex mechanism, we identified a small population of dimerized cGAS–E2 in the two-dimensional (2D) class averages of the cGAS–E2 complex at pH 5.0 (Extended Data Fig. 10a). The cGAS–E2 dimer suggests that, in addition to the covalent *cis*-interactions, non-covalent *trans*-interactions can be formed between E2 and cGAS, providing clues to how E2 conjugates two cGASs.

The poly-cGASylation for cGAS activation revealed in this study is unique. However, it shares similarity with human cGAS and cGAS in the E1E2/JAB-CBASS system. Such similarities include the human cGAS dependence on oligomerization for activation^{31,32} and bacterial cGAS, in the E1E2/JAB-CBASS system, activation by conjugation to other proteins^{14,15}. Furthermore, activation through inter-protein interaction is probably a common feature of cGAS. However, the poly-cGASylation in this study was triggered by mutations of the E2 C terminus. In addition, although the poly-cGASylation was also observed in vivo upon phage infection (Fig. 5c), the exact signalling from phages to E2 remains to be determined. To this end, we have tested the effect of phage infection on bacterial cellular pH and found that the cellular pH is decreased during phage infection (Extended Data Fig. 10b). Thus, the cellular pH of bacteria may serve as part of the E2 activation signal. Finally, upon the phage signal disrupting the C terminus of E2, the sequential

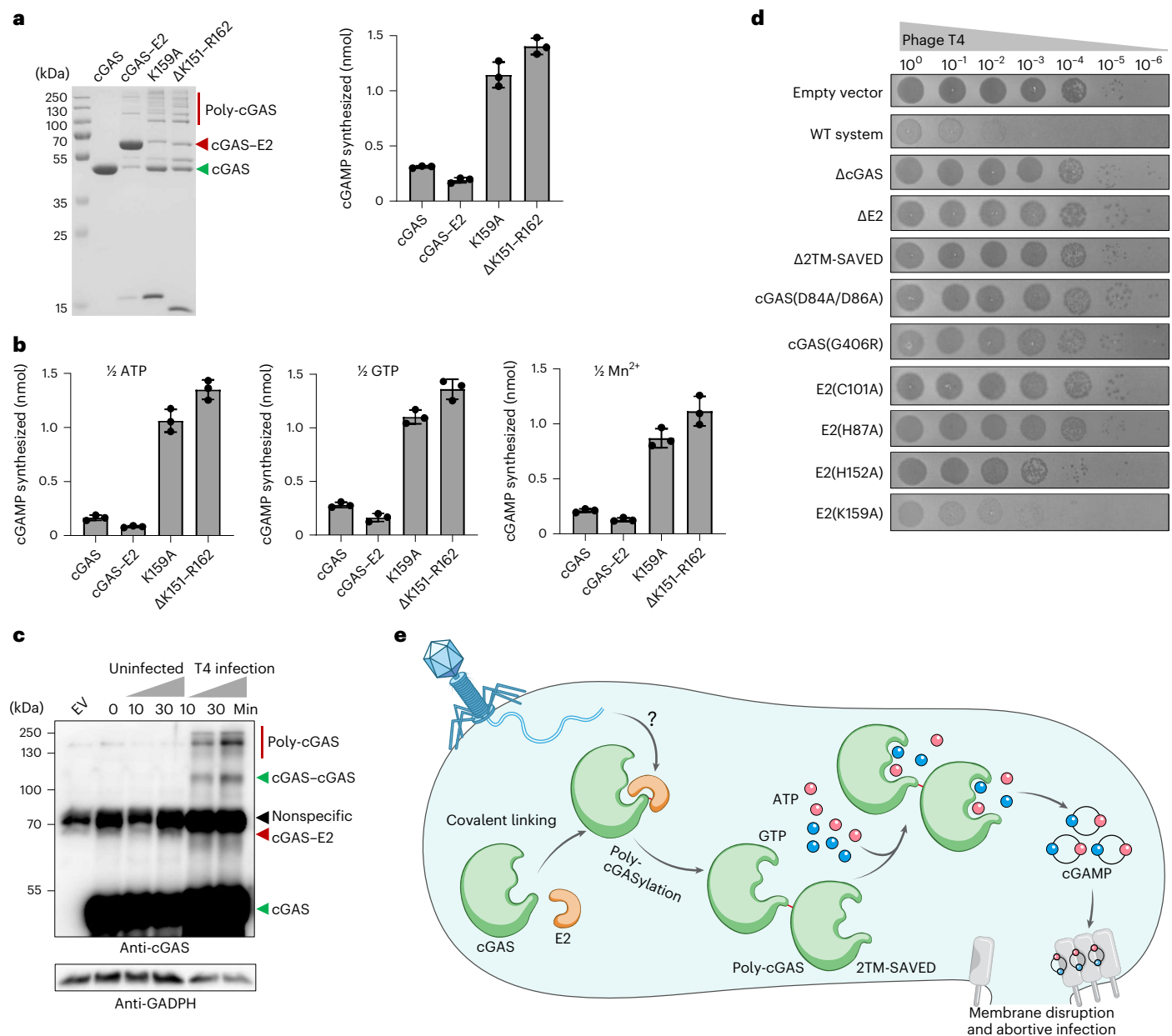


Fig. 5 | Poly-cGASylation stimulates cGAMP production and anti-phage immunity. a, Left, SDS-PAGE analysis of cGAS and the co-purified proteins consisting of cGAS with E2 or E2 mutants (indicated on the gel). Right, quantification of 3',2'-cGAMP (based on HPLC analysis shown in Extended Data Fig. 9a) synthesized by cGAS and variants. Data are the mean \pm s.d. of $n = 3$ independent replicates and are representative of three independent experiments. **b**, Quantification of 3',2'-cGAMP (based on the HPLC analysis shown in Extended Data Fig. 9b) synthesized by cGAS and variants, with the concentration of ATP, GTP or Mn²⁺ in the reaction reduced by half compared with those in **a**. Data are the mean \pm s.d. of $n = 3$ independent replicates and are

representative of three independent experiments. **c**, Western blot analysis with cGAS-specific antibody showing the poly-cGAS in *E. coli* BL21 expressing the E2-CBASS operon during T4 phage infection. Phage T4 was applied at an MOI of 5. Data are the mean \pm s.d. of $n = 3$ biological replicates. EV, empty vector. **d**, Phage T4 infection of cells expressing the empty vector, the full native E2-CBASS operon (wild-type system), the wild-type system with deletion or mutations of cGAS, the wild-type system with deletion or mutations of E2, or the wild-type system with deletion of the effector 2TM-SAVED. Data are representative of three independent experiments. **e**, Model for the E2-CBASS antiviral mechanism.

poly-cGASylation and cGAMP production trigger membrane disruption and abortive infection (Fig. 5e).

Methods

Plasmid construction

All experiments were performed with the CBASS system from *S. marcescens* (NCBI RefSeq NZ_KES60335.1). For the expression and purification of cGAS (NCBI RefSeq WP_016928966.1) and E2 (NCBI RefSeq WP_016928967.1), their coding sequences were codon optimized and commercially synthesized by Genscript. For the co-expression of

cGAS and E2, an N-terminal 6 \times 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 E2. Three forms of co-expression present in Fig. 1b were assessed for cGAS and E2: (1) the gene encoding cGAS with an N-terminal 6 \times His-tag cloned into the pET28a vector, the gene encoding E2 without tag cloned into the pQE82L vector and the T5 promoter in pQE82L replaced with a T7 promoter; (2) the gene encoding cGAS without tag cloned into the pQE82L vector and the gene encoding E2 with an N-terminal 6 \times His-tag cloned into the pET28a vector; and (3) the native coding sequence of

cGAS and E2 in a polycistronic form derived from *S. marcescens* was cloned into the pET28a vector. To monitor the C-terminal cleavage of cGAS, DNA encoding a C-terminal MBP-tagged cGAS was constructed in the pET28a vector by attaching the synthetic construct MBP gene (NCBI GenBank MN433888.1) to the 3' of the cGAS via a linker sequence (5'-AGCAGCAGCAGCCTGGTCCGCGCAGCCAT-3'). For phage infections, the four-gene CBASS system operon from *S. marcescens* spanning a nucleotide range of 26,427–29,783 was cloned into plasmid pQE82L. Details on the cloning of cGAS and E2 mutants are listed in Supplementary Tables 3 and 4, respectively. Plasmids generated in this study were constructed by Gibson from synthesized genes with ≥ 18 base pairs of homologies flanking the insert sequence using a ClonExpress II One Step Cloning Kit (Vazyme Biotech). Plasmids were transformed into *E. coli* DH5 α cells and confirmed by Sanger sequencing.

Protein expression, co-expression and purification

Plasmids for protein expression were transformed into *E. coli* BL21(DE3). To determine the interaction between cGAS and E2, both were co-transformed into *E. coli* BL21(DE3) and selected with their respective antibiotic marker. Bacterial cells were cultivated as a 4 ml starter culture in LB overnight at 37 °C with shaking at 220 rpm. The cultures were then transferred into 200 ml LB and cultivated for approximately 3 h until the optical density at 600 nm (OD₆₀₀) 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.

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 $\times g$, 4 °C for 1 h, and filtered through a 0.45 μ m filter. Proteins were purified by affinity chromatography using Ni-NTA resin (Qiagen) and a gravity column (Biorad). 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 cGAS and its mutants; 20 mM Tris–HCl (pH 8.0), 300 mM NaCl, and 50, 80 or 100 mM imidazole for E2, its mutants and the co-expressed proteins, respectively) and eluted with elution buffer (20 mM Tris–HCl (pH 8.0), 300 mM NaCl and 100 mM imidazole for cGAS, its mutants and the co-expressed proteins; 20 mM Tris–HCl (pH 8.0), 300 mM NaCl and 200 mM imidazole for E2 and its mutants). Proteins were filter concentrated using centrifugation and a 10 kDa or 30 kDa cut-off column (Millipore Sigma) and dialysed at 4 °C for -24 h against dialysis buffer (50 mM Tris–HCl (pH 7.5), 100 mM NaCl, 0.1 mM EDTA, 0.1% Triton X-100 and 50% glycerol). Proteins were analysed by SDS–PAGE with Coomassie blue (Bio-Rad) staining.

Structure prediction

ColabFold was used to predict the structure of the cGAS–E2 complex. Multiple sequence alignment was performed using MMseqs2 (UniRef + Environmental). Structural prediction was performed using AlphaFold on a virtual cluster node with a Nvidia Tesla P100 GPU.

Cryo-EM sample preparation and data collection

The cGAS–E2 complex was further purified by size-exclusion chromatography on a HiLoad 16/600 Superdex 200 pg column (Cytiva) in TBS buffer (20 mM Tris–HCl (pH 8.0), 100 mM NaCl). Proteins were concentrated to 3–7 mg ml⁻¹, flash-frozen in liquid nitrogen and stored at -80 °C.

A 3 μ l drop of cGAS–E2 sample at 0.7 mg ml⁻¹ was applied to glow-discharged copper Quantifoil R1.2/R1.3 grids, blotted for 3 s in 100% humidity at 8 °C and then plunged into liquid ethane using a Vitrobot Mark IV. All grids were screened using a Glacios microscope. The data were collected on a 300 keV Titan Krios G4 microscope equipped with a Gatan K3 direct electron detector and a Gatan Quantum energy filter at the Cryo-EM Unit of the Core Facility of Wuhan University.

For the initial dataset, 5,883 videos were collected in super-resolution mode, with 40 frames per video, 1.61 s exposure time, 50 e⁻ Å⁻² accumulated dose and 0.335 Å pixel size. For the final dataset,

the cGAS–E2 sample concentration was increased to 3.2 mg ml⁻¹ and UltrAuFoil grids were used instead for plunging. A total of 4,752 videos were collected in super-resolution mode, with 40 frames per video, 2.16 s exposure time, 70 e⁻ Å⁻² accumulated dose and 0.335 Å pixel size.

Cryo-EM data processing and model building

For dataset 1, videos were binned 2 \times (pixel size, 0.67 Å) and motion corrected using patch motion correction in CryoSPARC^{40,41}, followed by patch contrast transfer function estimation performed in the same software. Particles were picked using a blob picker, binned 2 \times (pixel size, 1.34 Å) and subjected to two rounds of 2D classifications in CryoSPARC with the number of online-EM iterations set to 40 and batch size per class set to 400. A total of 869,610 particles representing the binary complex were selected and were used for ab initio 3D reconstruction and heterogeneous refinements (Extended Data Fig. 3a). One class out of four was selected for 3D refinement using homogeneous refinement in C1 symmetry, resulting in a 3.8 Å cryo-EM density map of cGAS–E2. After global contrast transfer function refinement, a soft mask excluding the flexible terminal loops of the cGAS–E2 complex was generated and used for local refinement in CryoSPARC. Particles were unbinned and used for another round of local refinement, and the final map resolution of the cGAS–E2 complex was 3.35 Å.

The data processing procedures of dataset 2 were similar to those of dataset 1, with a few differences (Extended Data Fig. 3b). (1) A general Topaz^{42–45} model was trained with the particles from the initial dataset and used for particle picking. A blob picker was also used for picking, and after the first round of 2D classification, particles picked from the two pickers were merged and duplicate particles were removed. (2) Heterogeneous refinement (3D classification) was performed on 512,003 particles from 2D classification using the low-passed 3.35 Å map as the initial model. (3) The final cryo-EM map of the cGAS–E2 complex reached a resolution of 3.0 Å. Some degree of preferred orientation was present in both of the datasets, resulting in anisotropic maps (Extended Data Fig. 3c,d). DeepEMhancer⁴⁶ was used for post-processing to improve the anisotropic map, which helped with the interpretation of the map. The initial model of cGAS–E2 was generated by rigid body fitting of the predicted cGAS–E2 structure into the cryo-EM maps using Chimera⁴⁷. Regions with no electron densities were removed from the model, including residues 1–4, 191–210 and 389–396 from cGAS and 1–3 from E2. Linkage between cGAS Y405 and E2 K159 was modelled based on the densities with local adjustments. Initial B factors for all atoms of the model were set to 94.7 based on the Guinier plot from the final local refinement job. Inspection, model building and manual adjustments were performed in Coot⁴⁸, and real-space refinements were carried out in Phenix^{49,50}. Representations of cryo-EM densities and structural models were generated using ChimeraX⁴⁷ and PyMOL⁵¹.

In vitro bonding assay

To reconstitute the covalent attachment of cGAS to E2 in vitro, cGAS and E2 or their mutants were mixed in 50 mM Tris–HCl (pH 7.0), unless otherwise noted. The mixture was incubated at 4 °C for the indicated periods. The reaction was quenched by adding SDS sample buffer with DTT (Solarbio Life Sciences) or without DTT (Biosharp Life Sciences) and heating at 95 °C for 5 min. Proteins in the solutions were separated by SDS–PAGE in the absence of reducing agent and stained with Coomassie blue. The rate of the thioester bond formation in cGAS–E2 at pH 5.0 was calculated based on gel analysis. At each time interval, the amount of thioester bonds was calculated by subtracting the amount of isopeptide bonds (as quantified by cGAS–E2 in the presence of DTT) from the amount of total bonds (thioester + isopeptide, as quantified by cGAS–E2 in the absence of DTT). GraphPad Prism was used for data analysis.

In vitro cleavage assay

To examine the effect of low pH on the cGAS–E2 complex, the cGAS–E2 complex or its mutant was incubated in citrate buffer with a different

pH (Shyuanye) at 4 °C for the indicated periods. The reaction was quenched by adding SDS sample buffer with DTT or without DTT and heating at 95 °C for 5 min. Proteins in the solutions were separated by SDS–PAGE in the absence of a reducing agent and stained with Coomassie blue. GraphPad Prism was used for data analysis.

LC–MS/MS analysis

For liquid chromatography (LC)–MS/MS identification of the modified proteins, samples were separated on SDS–PAGE and protein bands of interest were excised and subjected to in-gel trypsin digestion as previously described⁵². Peptide identification was conducted using a hybrid ion trap–Orbitrap mass spectrometer (LTQ Orbitrap Velos, Thermo Fisher Scientific) coupled with nanoflow reversed-phase liquid chromatography (EASY-nLC 1200, Thermo Scientific). A capillary column (75 $\mu\text{m} \times 150\text{ mm}$) packed with 4 μm , 100 Å silica-based Magic C18AQ silica particles (Michrom BioResources) was prepared with a laser-pulled electrospray tip (model P-2000, Sutter Instrument). The column operated at a flow rate of 300 nl min^{-1} with the mobile phases consisting of solvent A (97% water, 3% acetonitrile and 0.1% formic acid) and solvent B (20% water, 80% acetonitrile and 0.1% formic acid). The LC gradient started at 7% solvent B for 3 min and linearly increased to 40% solvent B in 40 min. Within 2 min, the gradient was ramped up to 90% solvent B and maintained for 10 min. In a data-dependent acquisition mode, the eluted peptides from the capillary column were electrosprayed directly into the mass spectrometer. A full MS scan (m/z range 400–1,500, resolution 60,000) was performed using the Orbitrap mass analyser, and the 10 most intense ions were selected for fragmentation under collision-induced dissociation. Dynamic exclusion was applied with a repeat duration of 30 s and an exclusion duration of 12 s. MS/MS spectra were searched against cGAS and E2 protein sequences using Mascot software (version 2.3.02, Matrix Science). Key search parameters included mass tolerance of 20 ppm for precursor ions and 0.8 Da for MS/MS ions, a maximum of two missed cleavages, carbamidomethylation of cysteine (+57.02146 Da) as static modification and oxidation (+15.99492 Da) of methionine as dynamic modification. Identified peptides were filtered to achieve a false discovery rate < 1%. The corresponding peptide peaks were obtained from Thermo Xcalibur 2.2 (Thermo Fisher Scientific).

Proteomic analysis of the cGAS–cGAS bands was performed using a commercial company vendor (Gencreate). The protein samples were mixed with NuPAGE LDS sample buffer (Invitrogen) and separated using NuPAGE 4 to 12% Bis–Tris gel. The corresponding gel bands were then cut and washed with 40% H_2O , 60% acetonitrile and 50 mM NH_4HCO_3 (pH 8.0), and the protein bands to be identified were reduced and alkylated using TCEP and chloroacetamide, respectively. The gel pieces were digested with trypsin (20 ng ml^{-1}) in 50 mM NH_4HCO_3 (pH 8.0) at 37 °C overnight. The digested peptides were extracted with 40% H_2O , 60% acetonitrile and 1% formic acid, dried and reconstituted with 1% formic acid for LC–MS/MS analysis. The tryptic peptides (100 ng) were analysed on an UltiMate 3000 RSLCnano coupled to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). The peptides were first loaded onto a 75 $\mu\text{m} \times 2\text{ cm}$ trap column and separated on a 75 $\mu\text{m} \times 25\text{ cm}$ Pepmap C18 analytical column (Thermo Fisher Scientific) with a binary buffer system. Buffer A was 0.1% formic acid and 3% DMSO in 100% H_2O , and buffer B was 0.1% formic acid and 3% DMSO in 80% acetonitrile with 20% H_2O . The Eclipse mass spectrometer was operated in data-dependent acquisition mode with one full MS scan followed by MS/MS scans with higher-energy collision-induced dissociation fragmentation. The typical MS settings were spray voltage of 2.2 kV, heated capillary temperature of 320 °C and higher-energy collision-induced dissociation energy of 27. The LC–MS/MS raw data were processed with Maxquant for protein identification using a tailored database containing the sequences of E2 and cGAS and the *E. coli* strain B/BL21(DE3) proteome database in Uniprot. The typical search parameters were enzymatic specificity of Trypsin/P with a maximum

of two missed cleavages and variable modifications of oxidation on methionine. Carbamidomethyl (Cys) was set as the fixed modification. All results were screened at 1% protein and peptide level false discovery rate. Only peptides identified with high confidence were chosen for downstream protein identification analysis.

To characterize the product of cGAS, 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 $\text{mm} \times 2.1\text{ mm}$, 3 μm) maintained at 25 °C with a flow rate of 0.25 ml min^{-1} . The mobile phase consisted of methanol (A) and 20 mM ammonium acetate (B). The high-performance liquid chromatography (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 source under the following parameters: spray voltage, 3.2 kV; sheath and auxiliary gas flow rates, 40 and 15 arbitrary units, respectively; maximum spray current, 100.00 μA ; S-lens RF level, 50%; capillary temperature, 300 °C; and probe heater temperature, 350 °C. Profile MS1 spectra were acquired with the following settings: mass resolution, 70,000; AGC volume, 3×10^6 ; maximum injection time, 100 ms; and scan range, 300–1,000 m/z . Acquisition of data-dependent MS/MS spectra was performed using collision-induced dissociation 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 ; and normalized collision energy, 20 eV, 40 eV and 60 eV. Data are reported for the $z = 1$ acquisition for each indicated cyclic nucleotide. The chemical structures were drawn using ChemDraw version 19.0.

Peptide N-terminal sequencing

N-terminal sequencing of the MBP-containing peptides based on the Edman degradation method was performed by Beijing Bio-Tech Pack Technology. Peptide samples were separated by SDS–PAGE, transferred to a PVDF membrane (GE), identified by staining and cut out. A PPSQ-33A Edman sequencer (SHIMADZU) was used to analyse the protein samples on the PVDF membrane. Amino acids were identified one by one starting from the N terminus of the peptides by a cyclic reaction to determine the N-terminal sequences. The original data and spectra generated by PPSQ-33A were identified using PPSQ-30 data processing software, and the corresponding spectra were derived.

Fluorescence analysis

For the fluorescence imaging, we constructed an N-terminal GFP-tagged cGAS by attaching a synthetic construct GFP+ gene (NCBI GenBank KF410615.1) to the coding sequence of cGAS to visualize the poly-cGAS. GFP–cGAS was co-expressed and co-purified with the E2(K159A) mutant, and proteins eluted were analysed by SDS–PAGE and detected with the Fluorescence/Chemiluminescence imaging system (ChemiScope 6000 Touch).

HPLC analysis of enzymatic reactions

CD-NTase reactions were performed essentially as previously described⁵. In brief, 20 μl reactions contained 2 μM enzyme and 250 μM all four ribonucleotides (New England BioLabs) in reaction buffer with 50 mM Tris–HCl (pH 7.5), 100 mM NaCl, 2.5 mM $\text{MnCl}_2/\text{MgCl}_2$ and 1 mM DTT, unless otherwise noted. Reactions were incubated at 37 °C for the indicated periods, inactivated at 80 °C for 10 min and centrifuged for 15 min at $21,000 \times g$ to remove precipitated protein. Reaction products were analysed by HPLC with a ZORBAX Eclipse XDB-C18 (4.6 \times 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, Sigma) and solvent B (20 mM ammonium acetate, Sigma) at a flow rate of 1 ml min^{-1} using the following linear gradient: 0–5 min, 5% A; 5–15 min, 5–100% A; and 15–20 min, 100% A. The column was re-equilibrated for 5 min at 5% A. To analyse the reaction products treated with CIP or nuclease P1, 20 μl of the reaction product

was treated with 1 μ l of calf intestinal alkaline phosphatase (CIP) or nuclease P1 (New England BioLabs) at 37 °C for 1 h and inactivated at 80 °C for 10 min before centrifugation and analysis by HPLC.

Western blots

Rabbit cGAS polyclonal antibody was generated with a commercial vendor (AtaGenix) using a purified cGAS antigen. Serum was used at 1:1,000 for cGAS immunoblot detection. GAPDH antibody (ab125247, Abcam) was used at 1:1,000 as a loading control.

For whole-cell lysate analysis, *E. coli* BL21 carrying the indicated plasmid were grown to mid-logarithmic phase and protein expression was induced using 0.2 mM IPTG. Cells were infected with phage T4 for the indicated periods at a multiplicity of infection (MOI) of 5. Cells were lysed using FastBreak cell lysis reagent (Promega) at room temperature for 15 min, followed by centrifugation at 4 °C to remove cellular debris. Subsequently, 80 μ l soluble lysates were mixed with 20 μ l 5 \times SDS sample buffer with DTT, heated and concentrated at 100 °C for 30 min followed by a 5 min centrifugation at 21,000g. Proteins in the solutions were separated by SDS–PAGE, then transferred to PVDF membranes (Millipore Sigma) charged in methanol. Membranes were blocked in nonfat dry milk for 1 h at 24 °C, followed by incubation with primary antibodies diluted in antibody dilution buffer at 1:1,000 overnight at 4 °C. PVDF membranes were then incubated with the HRP-conjugated goat anti-rabbit IgG (SA00001-2, Proteintech) or HRP-conjugated goat anti-mouse IgG (SA00001-1, Proteintech) at 1:5,000 dilution in TBST for 1 h at 24 °C and detected with an ECL Chemiluminescent Substrate Kit (YEASEN) and Amersham Imager 680 (GE Health).

Plaque assays

Phages were propagated by infecting exponentially growing *E. coli* B culture in LB medium at an MOI of 1:100–1:1,000, incubated at 37 °C with shaking (200 rpm) until clearing of the culture. The lysate was centrifuged and the supernatant was filter sterilized through a 0.45 μ m filter.

Bacteriophage infection assay was determined using the small drop plaque assay method⁵³. The E2-CBASS operon and its mutants, as indicated, were cloned into the pQE82L vector and transformed into *E. coli* BL21. Cells containing an empty vector (pQE82L) were used as control. A single bacterial colony was picked and grown in LB at 37 °C to an OD₆₀₀ of ~0.2. E2-CBASS operon expression was induced using 50 μ M IPTG, followed by further growth for 1 h to an OD₆₀₀ of 0.6–0.7. Cells at a volume of 800 μ l were mixed with 25 ml of LB with 0.75% agar containing the appropriate antibiotic selection and IPTG (50 μ M), and the entire sample was poured onto plates. The phage stock was tenfold serially diluted, and 4 μ l drops of diluted phage lysate were placed on the solidified agar. The plates were then incubated at 37 °C for 16–18 h before imaging.

Steady-state kinetic measurement

For steady-state kinetic measurements of cGAS, the standard 20 μ l reaction mixture included 50 mM Tris–HCl (pH 7.5), 2.5 mM MnCl₂, 100 mM NaCl, 1 mM DTT, 1 μ M cGAS, 1 mM ATP (or GTP) and various concentrations (15.6 μ M, 31.3 μ M, 62.5 μ M, 125 μ M, 250 μ M, 500 μ M and 1,000 μ M) of GTP (or ATP). For each nucleotide concentration, samples were incubated at 37 °C for 15 min. Reactions were immediately heat inactivated at 80 °C for 10 min, and the samples were centrifuged for 15 min at 17,000 \times g to remove precipitated protein. Of each sample, 10 μ l was injected and reactions were analysed by HPLC, as detailed above. Absorbance units were converted to μ mol l⁻¹ by comparing with a standard curve from 7.8 μ M to 1 mM of chemically synthesized 3',2'-cGAMP (Biolog Life Sciences). Data were fitted by linear regression, and nonlinear curve fitting Michaelis–Menten kinetics and allosteric sigmoidal curves were calculated using GraphPad Prism version 8.0.2.

Bioinformatic analyses

Bioinformatic analysis of E2s and CD-NTases (cGASs) was performed by a commercial company vendor (Abiostar). The alignments in Extended Data Fig. 6a,c and Supplementary Figs. 1–5 were adapted from previously published datasets¹⁴. More than 2,200 CD-NTase and Cap2 protein sequences from different strains were downloaded from the Integrated Microbial Genomes & Microbiomes database. These protein sequences were analysed using Clustal Omega (version 1.2.4) for multiple sequence alignment, with default parameters selected. For each subset of CD-NTases indicated in Extended Data Fig. 6a, ten amino acids at the C terminus were extracted and aligned to create a sequence logo using the R language ggseqlogo (version 0.1). For each subset of E2s indicated in Extended Data Fig. 6c, 20 amino acids upstream and downstream of the active site cysteine were extracted and aligned to create a sequence logo. In addition, five amino acids upstream and downstream of the H152 residue of E2 in this study were also extracted and aligned to create a sequence logo using the R language ggseqlogo.

Measurement of intracellular pH during phage T4 infection

For the intracellular pH measurement during phage T4 infection using the intracellular pH probe BCECF-AM⁵⁴, *E. coli* BL21 were grown until OD₆₀₀ reached 1.0, and 10 μ M BCECF-AM was added to the cultures for 60 min at 37 °C in the dark. Then, the medium containing the probe was eliminated, and the cells were washed twice and resuspended with the elution buffer (20 mM HEPES (pH 7.4), 153 mM NaCl, 5 mM KCl and 5 mM glucose). Cells were then infected by phage T4 at an MOI of 5, and uninfected cells were used as controls. Fluorescence intensities were measured at excitation wavelengths of 490 nm and 430 nm, and the emission wavelength was recorded at 530 nm in a Synergy HI microplate reader (BioTek) at various postinfection times. A standard curve was determined using the free acid form of BCECF suspended in elution buffer ranging from pH 4.0 to pH 10.0. The intracellular pH was calculated from the ratios of emission fluorescence intensities upon excitations at 490 nm and 430 nm using the standard curve.

For the intracellular pH measurement using the fluorescence resonance energy transfer (FRET)-based protein sensor⁵⁵, a plasmid of pH-Lemon was transformed into chemically competent *E. coli* BL21, and cells were grown until OD₆₀₀ reached 1.0. Then, the culture medium was eliminated, and the cells were washed twice and resuspended with the elution buffer. Cells were infected by phage T4 at an MOI of 5, and uninfected cells were used as controls. Fluorescence intensities were measured at excitation wavelengths of 435 nm and emission wavelengths of 480 nm and 530 nm in the microplate reader at various postinfection times. A standard curve was determined using purified pH-Lemon at different pH values (elution buffer as above) in vitro. The intracellular pH was calculated from the ratios of emission fluorescence intensities upon excitations at 480 nm and 530 nm using the standard curve.

Statistics and reproducibility

All statistical analyses were performed using Microsoft Excel 2019, OriginPro version 9.1 and GraphPad Prism version 8.0.2. Graph plotting and statistical analysis were performed using GraphPad Prism version 8.0.2. Error bars and the number of replicates for each experiment are defined in the figure legends. No statistical methods were used to pre-determine sample sizes used for experiments, but sample sizes are in line with field standards. No data were excluded from the analyses. Randomization was used for sample injection order in high-performance liquid chromatography mass spectrometry measurements, and other experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data supporting the findings of this study are available in this Article and its Supplementary Information. Genomes of *S. marcescens* were sequenced previously, and sequences are publicly available. GenBank accessions for the CBASS operons used in this study are found in Methods. Primers used in this study are provided in Supplementary Tables 3 and 4. The *E. coli* strain B/BL21-DE3 proteome database that was used in this study can be accessed at Uniprot (UP000002032). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<https://proteomecentral.proteomexchange.org>) via the iProX partner repository^{56,57} with the dataset identifier PXD050178. For the cGAS–E2 complex, coordinates are available at the RCSB PDB (<http://www.rcsb.org>) under accession codes 8HSB and 8YJY, and electron microscopy maps are available at the Electron Microscopy Data Bank (EMDB; <https://www.ebi.ac.uk/emdb/>) under accession codes EMD-34992 and EMD-39353. Source data are provided with this paper.

Code availability

No custom code was used for data analyses.

References

- 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 (2013).
- Wu, J. et al. Cyclic GMP–AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339**, 826–830 (2013).
- Davies, B. W., Bogard, R. W., Young, T. S. & Mekalanos, J. J. Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for *V. cholerae* virulence. *Cell* **149**, 358–370 (2012).
- 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 Rep.* **3**, 1362–1368 (2013).
- Whiteley, A. T. et al. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194–199 (2019).
- Cohen, D. et al. Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* **574**, 691–695 (2019).
- Morehouse, B. R. et al. STING cyclic dinucleotide sensing originated in bacteria. *Nature* **586**, 429–433 (2020).
- Ye, Q. et al. HORMA domain proteins and a Trip13-like ATPase regulate bacterial cGAS-like enzymes to mediate bacteriophage immunity. *Mol. Cell* **77**, 709–722.e7 (2020).
- Burroughs, A. M., Zhang, D., Schaffer, D. E., Iyer, L. M. & Aravind, L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res.* **43**, 10633–10654 (2015).
- Millman, A., Melamed, S., Amitai, G. & Sorek, R. Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. *Nat. Microbiol.* **5**, 1608–1615 (2020).
- 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 (2020).
- Kerscher, O., Felberbaum, R. & Hochstrasser, M. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu. Rev. Cell Dev. Biol.* **22**, 159–180 (2006).
- Li, W. & Ye, Y. Polyubiquitin chains: functions, structures, and mechanisms. *Cell. Mol. Life Sci.* **65**, 2397–2406 (2008).
- Ledvina, H. E. et al. An E1–E2 fusion protein primes antiviral immune signalling in bacteria. *Nature* **616**, 319–325 (2023).
- Jenson, J. M., Li, T., Du, F., Ea, C. K. & Chen, Z. J. Ubiquitin-like conjugation by bacterial cGAS enhances anti-phage defence. *Nature* **616**, 326–331 (2023).
- Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- Mirdita, M. & Schütze, K. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
- Holm, L., Laiho, A., Törönen, P. & Salgado, M. DALI shines a light on remote homologs: one hundred discoveries. *Protein Sci.* **32**, e4519 (2023).
- Dou, H., Buetow, L., Sibbet, G. J., Cameron, K. & Huang, D. T. BIRC7–E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nat. Struct. Mol. Biol.* **19**, 876–883 (2012).
- Termathe, M. & Leidel, S. A. The Uba4 domain interplay is mediated via a thioester that is critical for tRNA thiolation through Urm1 thiocarboxylation. *Nucleic Acids Res.* **46**, 5171–5181 (2018).
- Hendrickx, A. P., Budzik, J. M., Oh, S. Y. & Schneewind, O. Architects at the bacterial surface—sortases and the assembly of pili with isopeptide bonds. *Nat. Rev. Microbiol.* **9**, 166–176 (2011).
- Jacobitz, A. W., Kattke, M. D., Wereszczynski, J. & Clubb, R. T. Sortase transpeptidases: structural biology and catalytic mechanism. *Adv. Protein Chem. Struct. Biol.* **109**, 223–264 (2017).
- Lowther, J. et al. The importance of pH in regulating the function of the *Fasciola hepatica* cathepsin L1 cysteine protease. *PLoS Negl. Trop. Dis.* **3**, e369 (2009).
- Zhao, L. et al. Structural analysis of asparaginyl endopeptidase reveals the activation mechanism and a reversible intermediate maturation stage. *Cell Res.* **24**, 344–358 (2014).
- Verma, S., Dixit, R. & Pandey, K. C. Cysteine proteases: modes of activation and future prospects as pharmacological targets. *Front. Pharmacol.* **7**, 107 (2016).
- Elsässer, B. et al. Distinct roles of catalytic cysteine and histidine in the protease and ligase mechanisms of human legumain as revealed by DFT-based QM/MM simulations. *ACS Catal.* **7**, 5585–5593 (2017).
- Feliciangeli, S. F. et al. Identification of a pH sensor in the furin propeptide that regulates enzyme activation. *J. Biol. Chem.* **281**, 16108–16116 (2006).
- Williamson, D. M., Elferich, J., Ramakrishnan, P., Thomas, G. & Shinde, U. The mechanism by which a propeptide-encoded pH sensor regulates spatiotemporal activation of furin. *J. Biol. Chem.* **288**, 19154–19165 (2013).
- Vernet, T. et al. Processing of the papain precursor. The ionization state of a conserved amino acid motif within the Pro region participates in the regulation of intramolecular processing. *J. Biol. Chem.* **270**, 10838–10846 (1995).
- Fuentes-Prior, P. & Salvesen, G. S. The protein structures that shape caspase activity, specificity, activation and inhibition. *Biochem. J.* **384**, 201–232 (2004).
- Li, X. et al. Cyclic GMP–AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity* **39**, 1019–1031 (2013).
- Yu, L. & Liu, P. Cytosolic DNA sensing by cGAS: regulation, function, and human diseases. *Signal Transduct. Target. Ther.* **6**, 170 (2021).
- 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. *Mol. Cell* **81**, 5039–5051.e5035 (2021).
- David, Y., Ziv, T., Admon, A. & Navon, A. The E2 ubiquitin-conjugating enzymes direct polyubiquitination to preferred lysines. *J. Biol. Chem.* **285**, 8595–8604 (2010).
- Stewart, M. D., Ritterhoff, T., Klevit, R. E. & Brzovic, P. S. E2 enzymes: more than just middle men. *Cell Res.* **26**, 423–440 (2016).

36. Burroughs, A. M., Jaffee, M., Iyer, L. M. & Aravind, L. Anatomy of the E2 ligase fold: implications for enzymology and evolution of ubiquitin/Ub-like protein conjugation. *J. Struct. Biol.* **162**, 205–218 (2008).
37. Pearce, M. J., Mintseris, J., Ferreyra, J., Gygi, S. P. & Darwin, K. H. Ubiquitin-like protein involved in the proteasome pathway of *Mycobacterium tuberculosis*. *Science* **322**, 1104–1107 (2008).
38. Burns, K. E. et al. “Depupylation” of prokaryotic ubiquitin-like protein from mycobacterial proteasome substrates. *Mol. Cell* **39**, 821–827 (2010).
39. Iyer, L. M., Burroughs, A. M. & Aravind, L. Unraveling the biochemistry and provenance of pupylation: a prokaryotic analog of ubiquitination. *Biol. Direct* **3**, 45 (2008).
40. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
41. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
42. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
43. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCr J* **6**, 5–17 (2019).
44. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
45. Punjani, A. & Fleet, D. J. 3D variability analysis: resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol.* **213**, 107702 (2021).
46. Sanchez-Garcia, R. et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874 (2021).
47. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
48. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D.* **66**, 486–501 (2010).
49. Wang Ray, Y.-R. et al. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *eLife* **5**, e17219 (2016).
50. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G. & Adams, P. D. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D.* **75**, 861–877 (2019).
51. Leman, J. K. et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods* **17**, 665–680 (2020).
52. Hu, M., Liu, Y., Yu, K. & Liu, X. Decreasing the amount of trypsin in in-gel digestion leads to diminished chemical noise and improved protein identifications. *J. Proteomics* **109**, 16–25 (2014).
53. Mazzocco, A., Waddell, T. E., Lingohr, E. & Johnson, R. P. Enumeration of bacteriophages using the small drop plaque assay system. *Methods Mol. Biol.* **501**, 81–85 (2009).
54. Chakraborty, S., Mizusaki, H. & Kenney, L. J. A FRET-based DNA biosensor tracks OmpR-dependent acidification of *Salmonella* during macrophage infection. *PLoS Biol.* **13**, e1002116 (2015).
55. Burgstaller, S. et al. pH-Lemon, a fluorescent protein-based pH reporter for acidic compartments. *ACS Sens.* **4**, 883–891 (2019).
56. Ma, J. et al. iProX: an integrated proteome resource. *Nucleic Acids Res.* **47**, D1211–d1217 (2019).
57. Chen, T. & Ma, J. iProX in 2021: connecting proteomics data sharing with big data. *Nucleic Acids Res.* **50**, D1522–d1527 (2022).

Acknowledgements

We would like to express our special gratitude to L. Liu of Tsinghua University for his invaluable guidance in the field of protein chemistry. We thank E.-D. Wang, W. Yan and H. Hu for their advice, P. Xu for providing materials, X. Li for technical support and all lab members for helpful discussions. We thank D. Li, X. Li and Y. Zeng of the Core Facility of Wuhan University for their assistance with cryo-EM grid screening and data collection. This work was supported by the National Natural Science Foundation of China (grants 32150009 and 31870165 to B.Z., 22174003 and 21974002 to X. Liu, and 31900032 to F.H.), the National Key Research and Development Program of China (2022YFA0912200 to L.W. and 2022YFA1304500 to X. Liu) and the Fund from Science, Technology and Innovation Commission of Shenzhen Municipality (grant JCYJ20210324115811032 to B.Z.).

Author contributions

F.H. and B.Z. conceived the project. Y.Y., J.X., W.X., F.H., D.W., X. Liu, L.W. and B.Z. designed the experiments. Y.Y., J.X., W.X. and D.W. performed the experiments. J.X. and L.W. determined and analysed structures. Y.Y., F.H., W.X., G.K.O., C.V.R., Hao Wu, D.W., X. Liu, L.W. and B.Z. analysed the data and wrote the paper. All authors discussed the results and contributed to the final paper.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41564-024-01684-z>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41564-024-01684-z>.

Correspondence and requests for materials should be addressed to Fengtao Huang, Di Wu, Xiaoyun Liu, Longfei Wang or Bin Zhu.

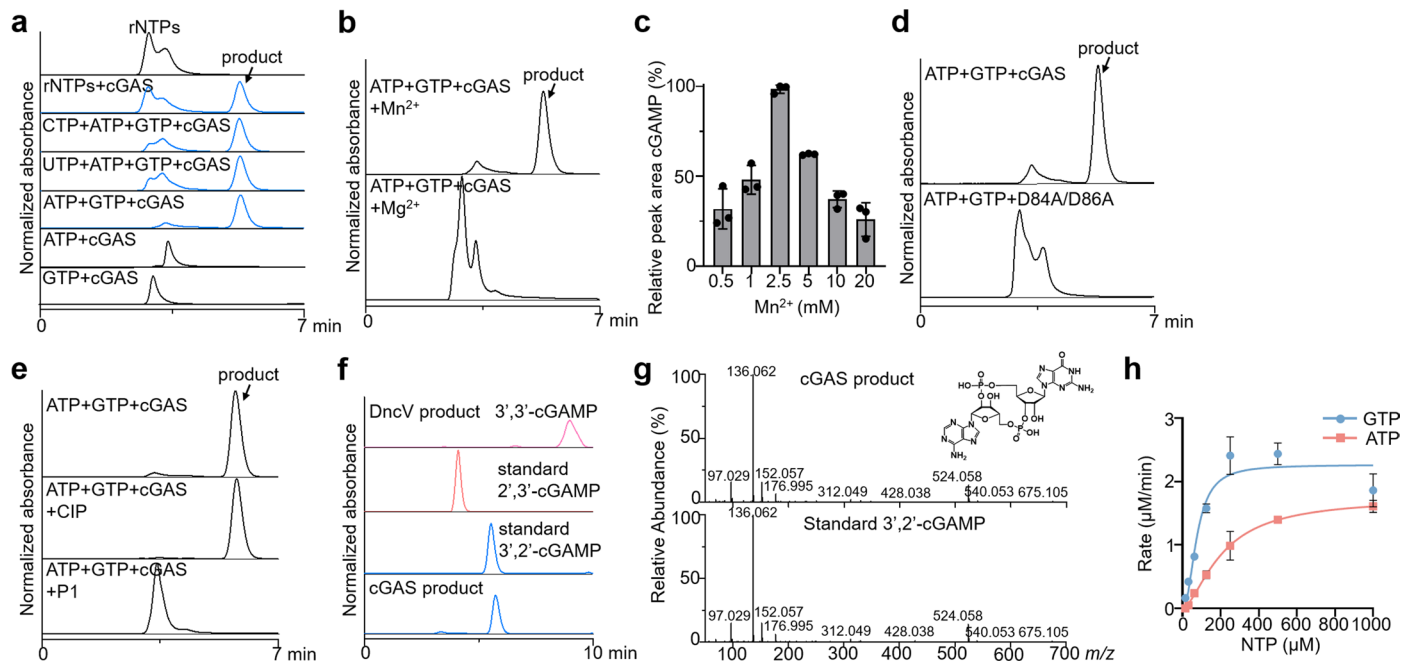
Peer review information *Nature Microbiology* thanks Karl-Peter Hopfner and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

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

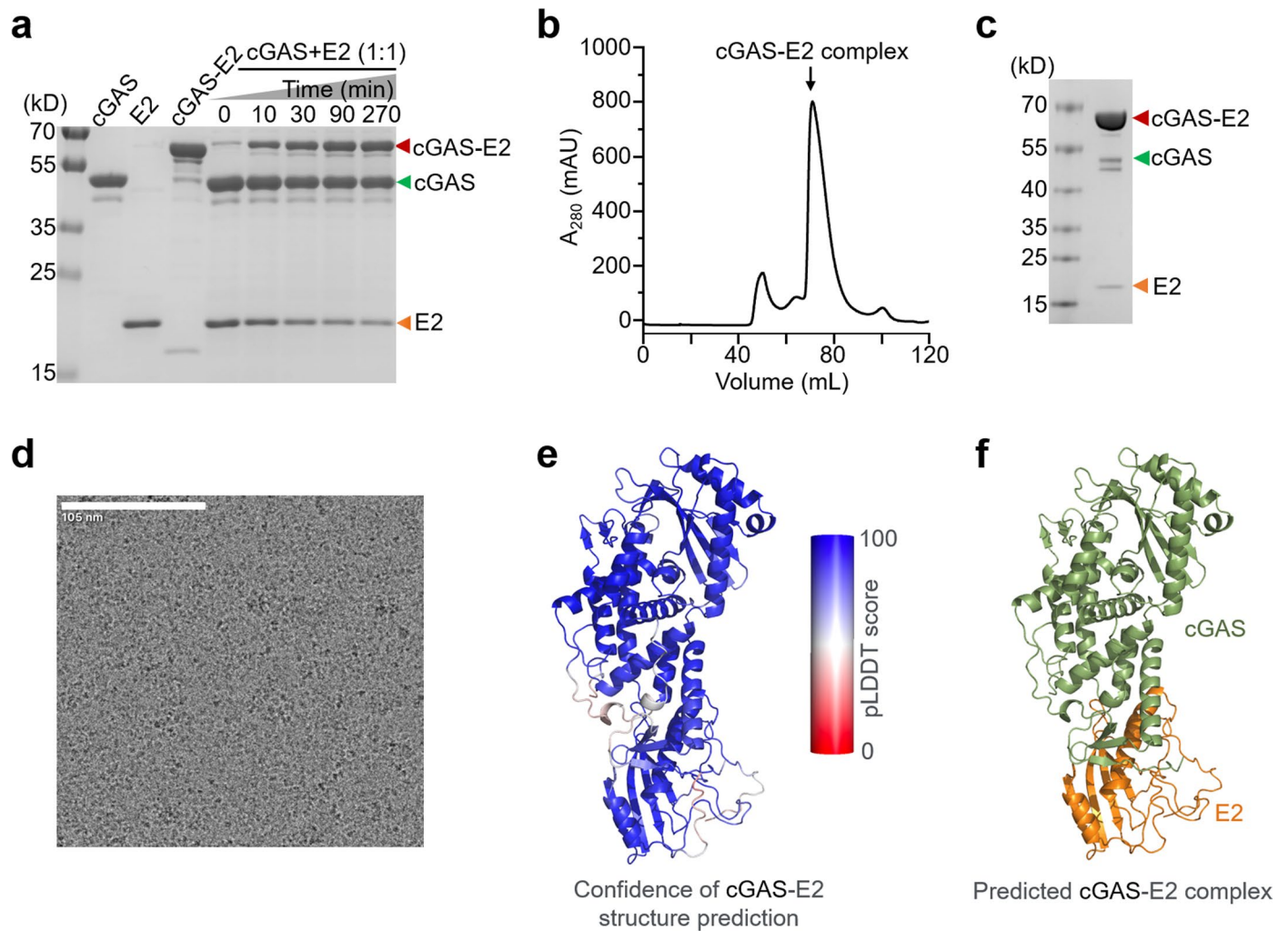
Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2024



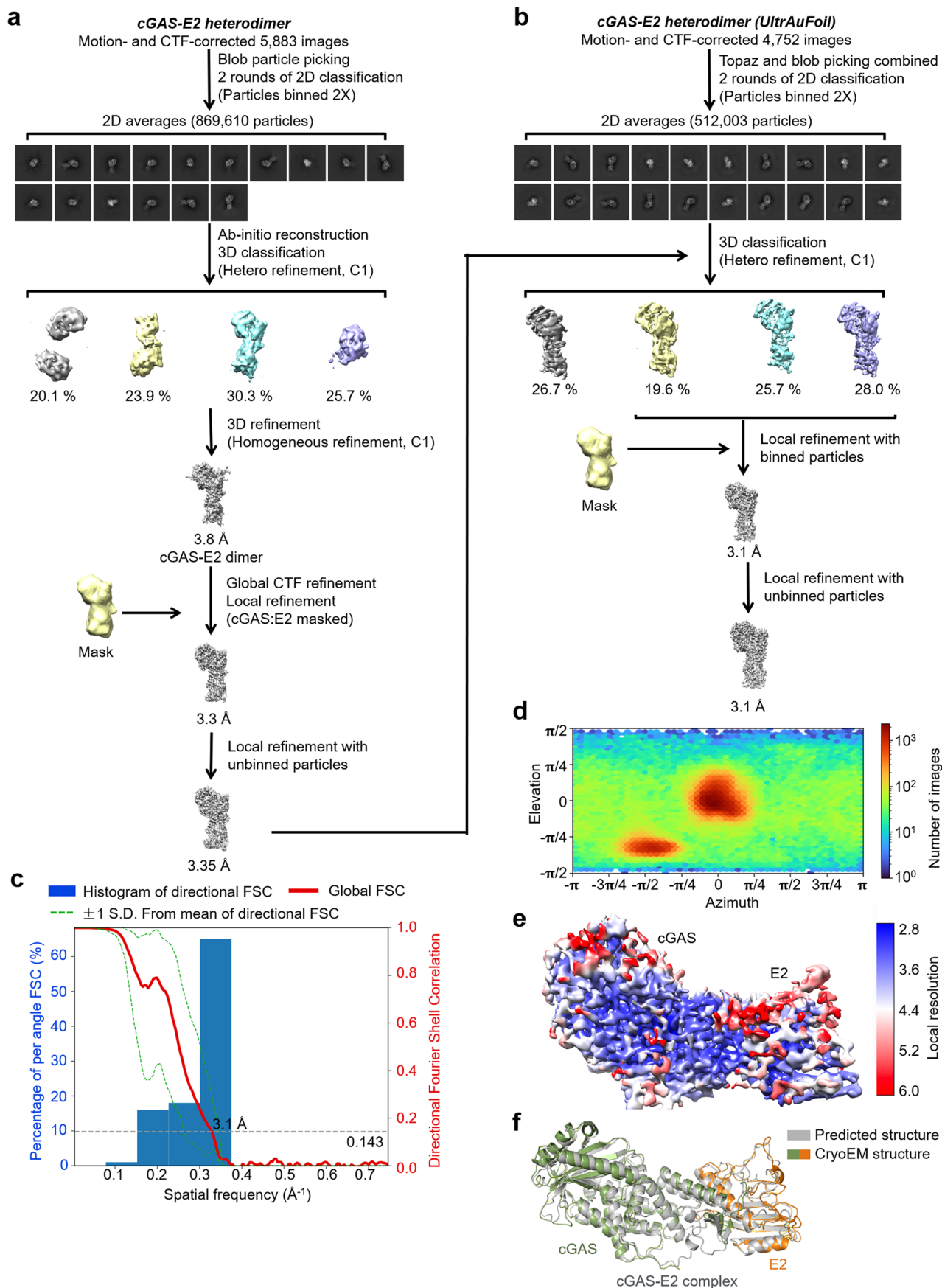
Extended Data Fig. 1 | *In vitro* characterization of the enzymatic activity of cGAS. **a**, HPLC analysis of nucleotide second messenger synthesis by cGAS using various nucleotide substrates. cGAS synthesizes a product only in the presence of both ATP and GTP. **b**, cGAS synthesis activity is dependent on Mn^{2+} . **c**, Optimal Mn^{2+} concentration for cGAS. The products synthesized at various Mn^{2+} concentrations were quantified by HPLC. Data are mean \pm s.d. for $n = 3$ independent replicates. **d**, Comparison of the activities of cGAS and its D84A/D86A mutant. **e**, cGAS product was degraded by nuclease P1 but not CIP, as analyzed by HPLC. **f**, Comparison of the retention time of possible cGAMP

variants (3',3'-cGAMP, 2',3'-cGAMP, and 3',2'-cGAMP) and the cGAS product. **g**, MS/MS fragmentation spectra of the cGAS product (top) and the standard 3',2'-cGAMP (bottom), and the chemical structure of 3',2'-cGAMP. **h**, Michaelis-Menten kinetics plot of ATP and GTP for cGAS. K_m , k_{cat} and k_{cat}/K_m for ATP are $393.0 \pm 15.3 \mu M$, $2.4 \pm 0.3 \text{ min}^{-1}$ and 5.9×10^{-3} , respectively. For GTP, K_m , k_{cat} and k_{cat}/K_m are $94.2 \pm 11.1 \mu M$, $2.6 \pm 0.2 \text{ min}^{-1}$ and 2.8×10^{-2} , respectively. Data are mean \pm s.d. for $n = 3$ independent replicates. All data are representative of three independent experiments.



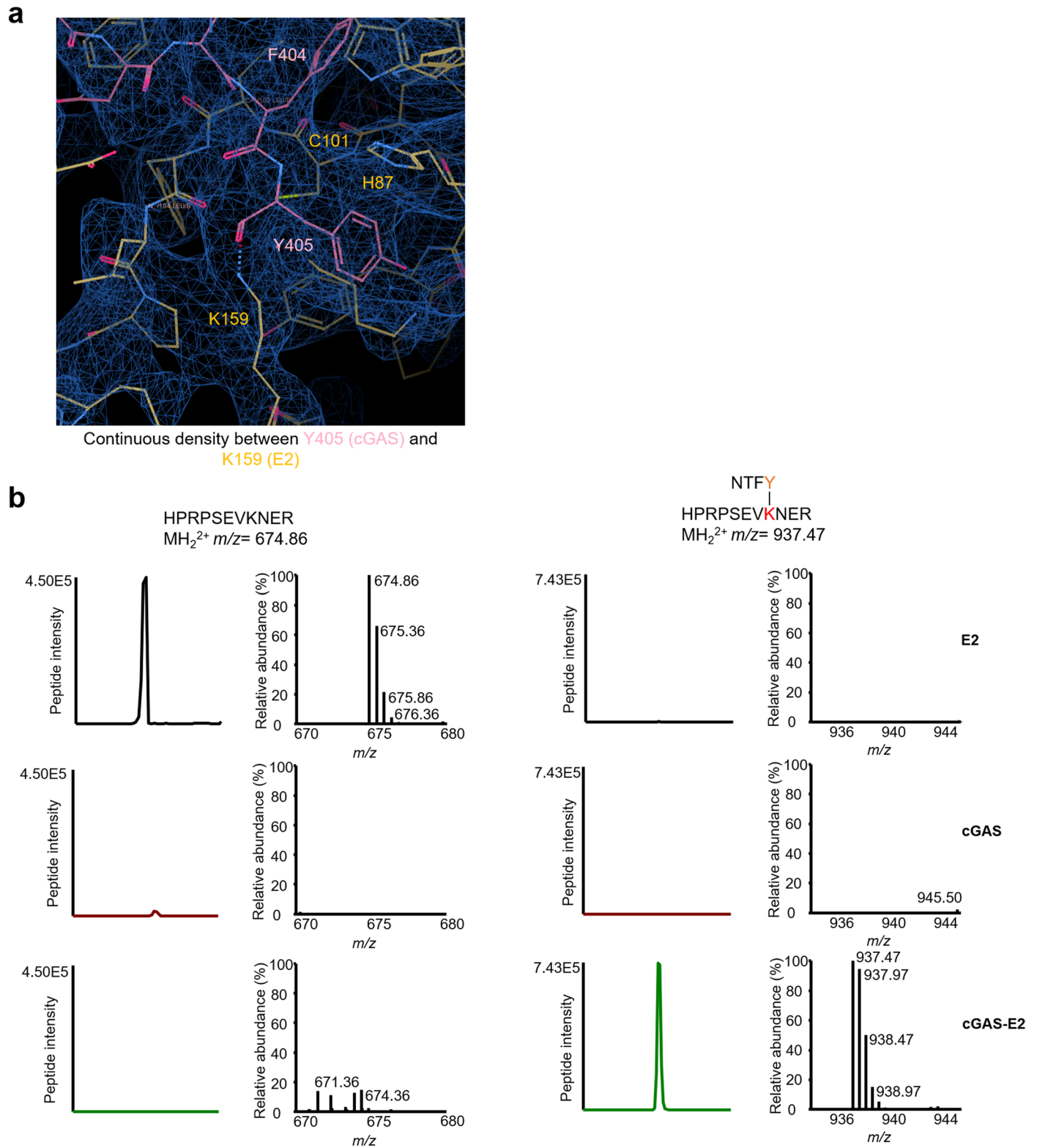
Extended Data Fig. 2 | Predicted structure of the cGAS-E2 complex. **a**, SDS-PAGE analysis of the *in vitro* assembly of the covalent cGAS-E2 protein. Two proteins were mixed at a molar ratio of 1:1 and incubated at 4 °C for the indicated periods (cGAS+E2). Since both cGAS and E2 were His-tagged during the *in vitro* reconstruction, the molecular weight of cGAS-E2 protein was slightly higher than that obtained *in vivo* from His-tagged cGAS and non-tagged E2. Data are representative of three independent experiments. **b,c**, Chromatography traces

and SDS-PAGE analysis of purified cGAS-E2 complex. Data are representative of three independent experiments. **d**, Representative cryo-EM micrograph of cGAS-E2 complex. Data are representative of three independent experiments. **e**, Ribbon diagram of the top ranked structure prediction of cGAS-E2 complex colored by the per-residue LDDT scores. **f**, Ribbon diagram of the top ranked structure prediction of cGAS-E2 complex with cGAS and E2 colored in green and orange, respectively.



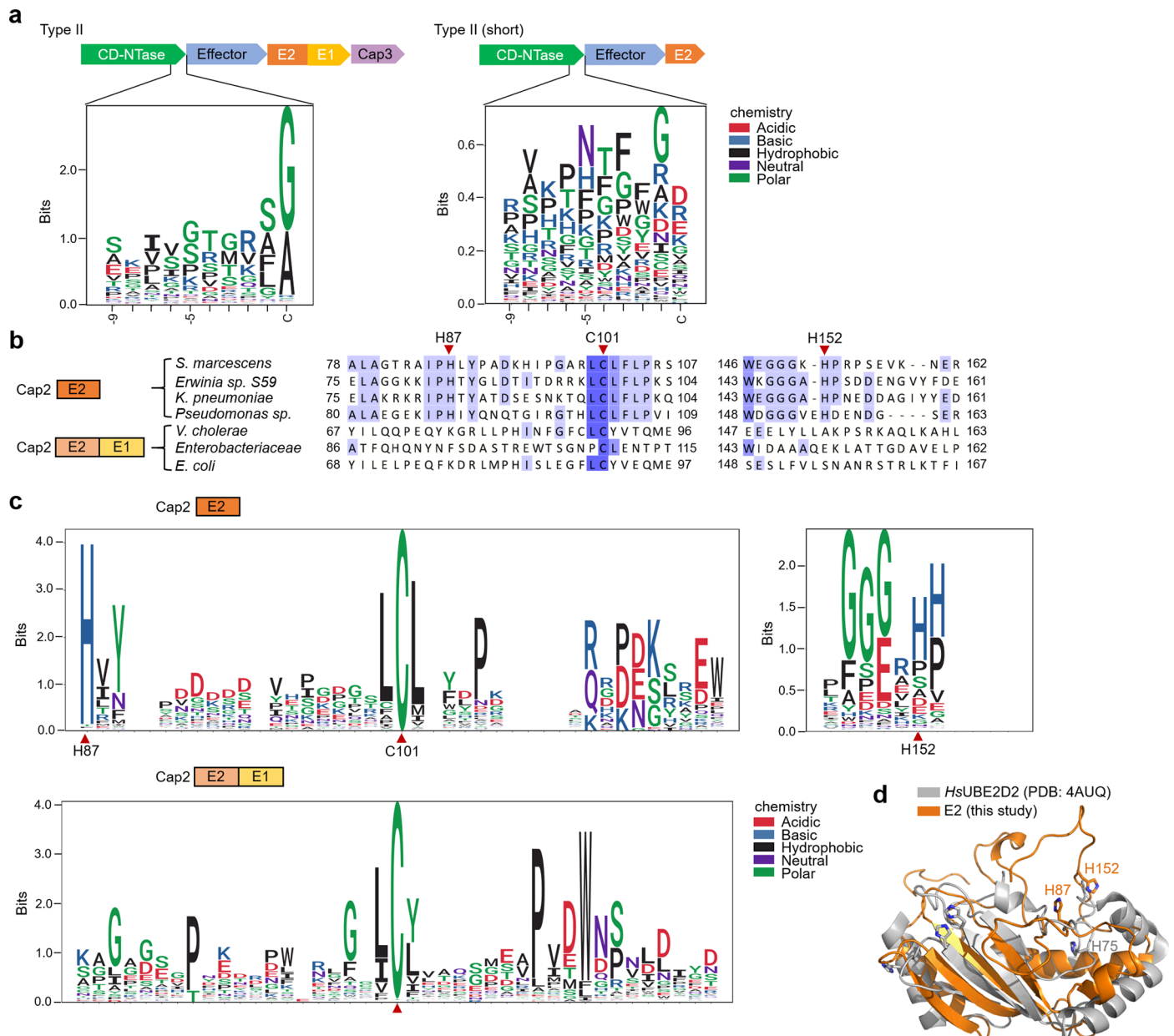
Extended Data Fig. 3 | Cryo-EM workflows and the quality of reconstructed cryo-EM maps. a, b, Workflow of 3D reconstruction of the initial cGAS-E2 dataset (a) and the UltrAuFoil cGAS-E2 dataset (b). **c,** Directional FSC of the cGAS-E2

cryo-EM density map. **d,** Orientation distributions of cGAS-E2 3D reconstruction. **e,** Local resolutions of the cGAS-E2 cryo-EM map. **f,** Superimposed predicted and experimental structures of cGAS-E2.



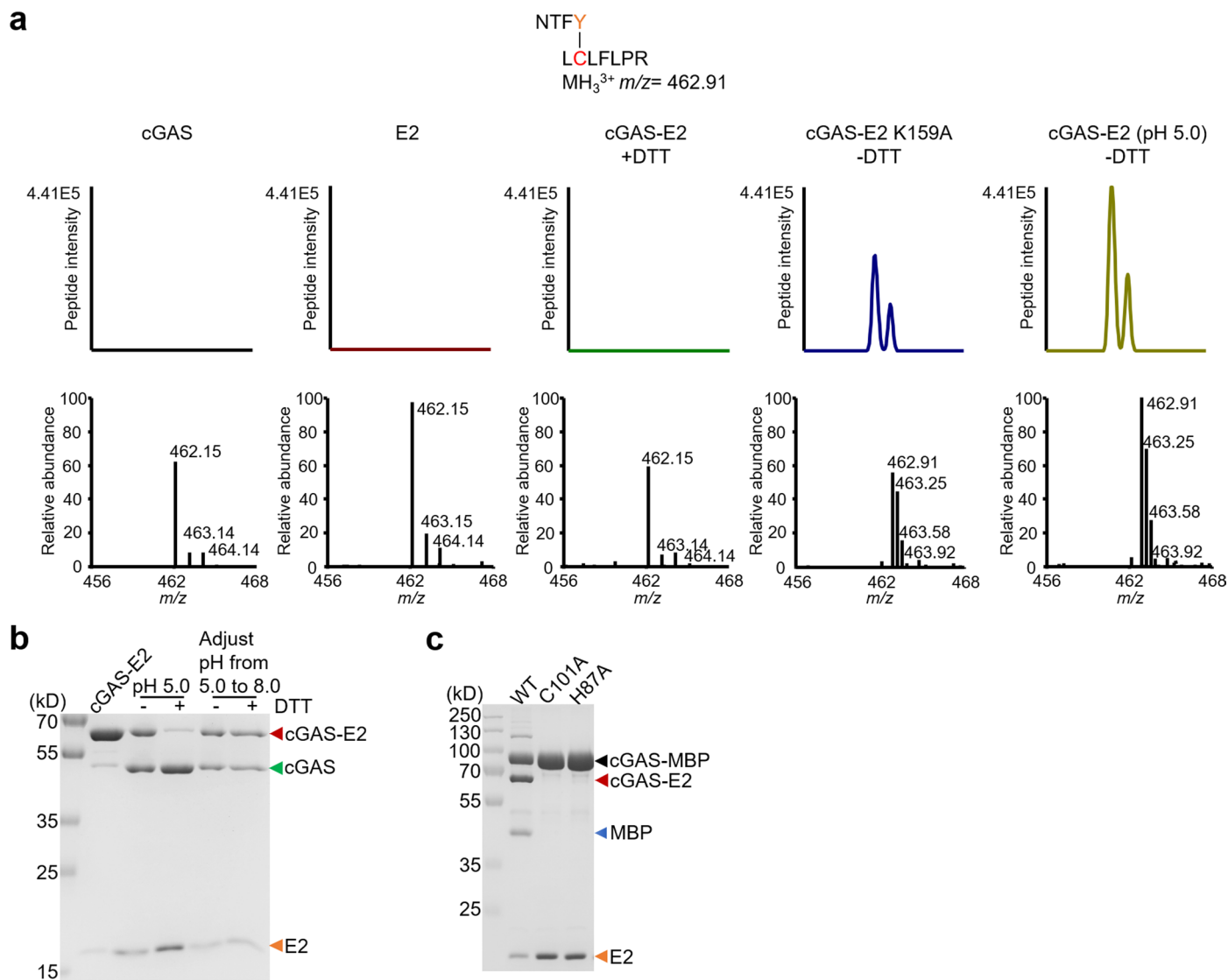
Extended Data Fig. 5 | Characterization of the bond formed between cGAS Y405 and E2 K159. a, Magnified view of the E2 active site in cGAS-E2 complex showing continuous density between Y405 of cGAS and K159 of E2. **b**, Extracted ion chromatograms (left panel) and corresponding mass spectra of precursor ions (right panel) of the control peptide HPRPSEVKNER and the modified peptide

HPRPSEVKNER bearing an NTFY remnant at K159 (red). MS signals of the control peptide were only detected in E2 but not cGAS or cGAS-E2 samples. In contrast, MS signals of the modified peptide were only detected in cGAS-E2 but not E2 or cGAS samples. All data are representative of three independent experiments.



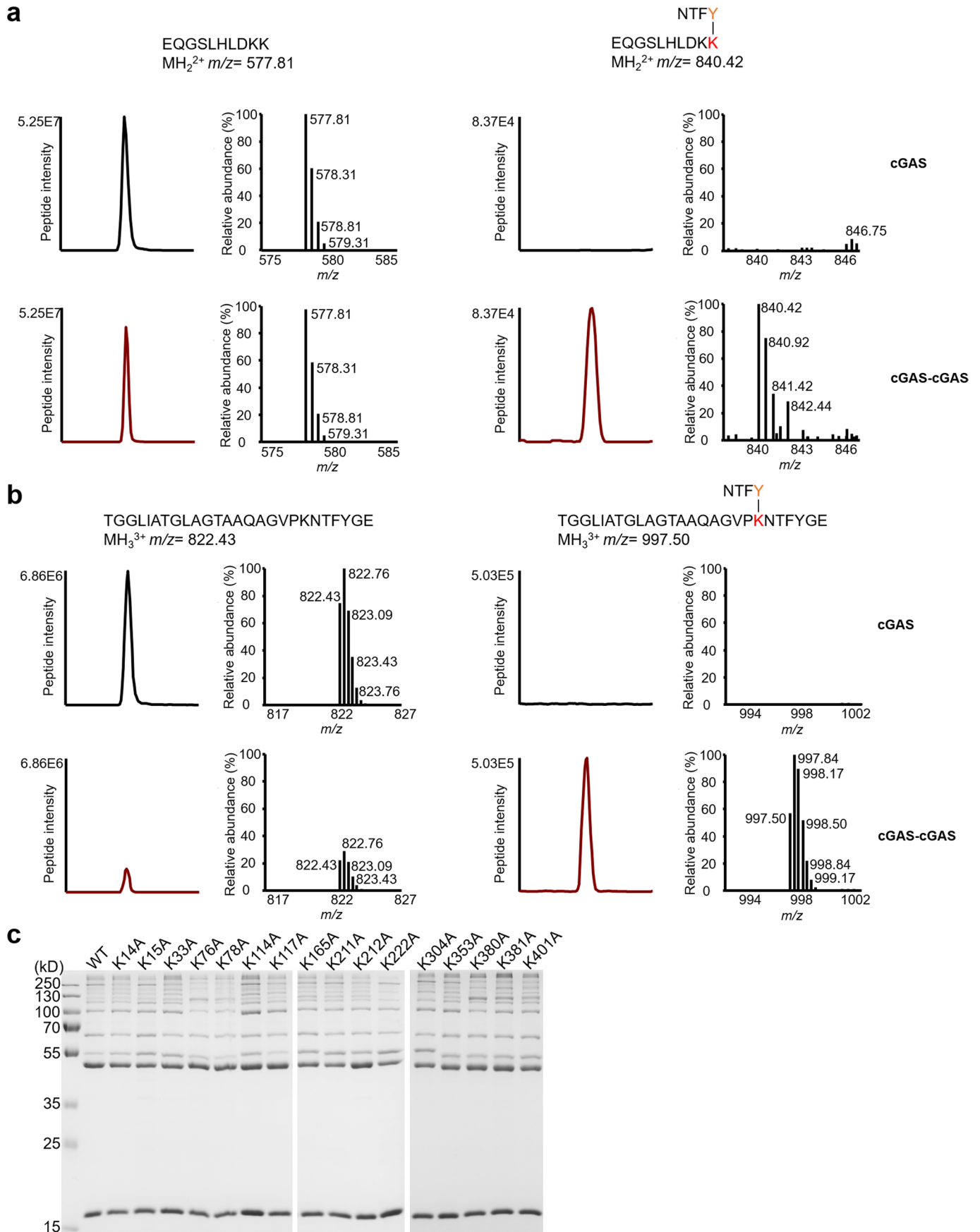
Extended Data Fig. 6 | Comparison of CD-NTase and E2 homologs. **a**, Sequence logos for CD-NTase (cGAS) from type II CBASSs. Left and right show sequence logos for the C-terminal 10 residues of CD-NTase from E1E2/JAB-CBASSs and E2-CBASSs, respectively. Data are depicted as bits and signified by the height of each residue. Different colors indicate the chemical properties of amino acids. See also Supplementary Fig. 1, 3. **b**, Sequence alignment of E2 homologs from E2-CBASSs and E1E2/JAB-CBASSs show that they share low homology. The two crucial histidine residues (marked by red triangles) are conservative in E2 from E2-CBASSs but not E1E2/JAB-CBASSs. See also Supplementary Fig. 2, 4, 5. **c**, Sequence logos for the active site sequences of Cap2 (E2) from E2-CBASSs. Top left and right show 20 amino acids upstream and downstream of the active

site cysteine, and 5 amino acids upstream and downstream of the residues corresponding to H152, respectively. The positions corresponding to H87, C101, and H152 are marked by red triangles. Bottom, Sequence logo for the active site sequences of E2 domain of Cap2 (E1E2) from E1E2/JAB-CBASSs. 20 amino acids upstream and downstream of the active site cysteine are shown. The active site cysteine is marked by a red triangle. Data are depicted as bits and signified by the height of each residue. Different colors indicate the chemical properties of different amino acids. See also Supplementary Figs. 2, 4. **d**, Superimposed structures of E2 (orange) and human E2 (gray, PDB ID 4AUQ) show that no similar catalytic site is found in human E2.



Extended Data Fig. 7 | Characterization of the thioester bond in cGAS-E2.
a, Extracted ion chromatograms (upper panel) and corresponding mass spectra of precursor ions (lower panel) of the modified peptide LCLFLPR bearing an NTFY remnant at C101 (red). MS signals of the modified peptide were only detected in cGAS-E2 samples without DTT treatment but not samples of E2, cGAS

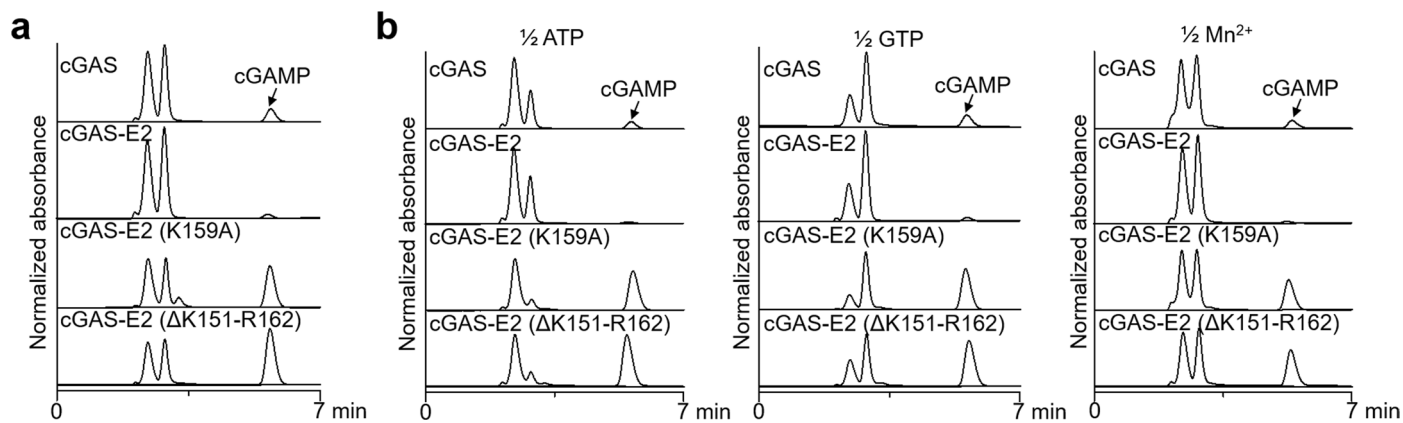
or cGAS-E2 treated with DTT. **b**, SDS-PAGE analysis showed that the thioester bond (-DTT) and isopeptide bond (+DTT) in cGAS-E2 are interconvertible at various pH. **c**, SDS-PAGE analysis of the *in vitro* processing of cGAS-MBP by E2, E2 (C101A), or E2 (H87A). All data are representative of three independent experiments.



Extended Data Fig. 8 | See next page for caption.

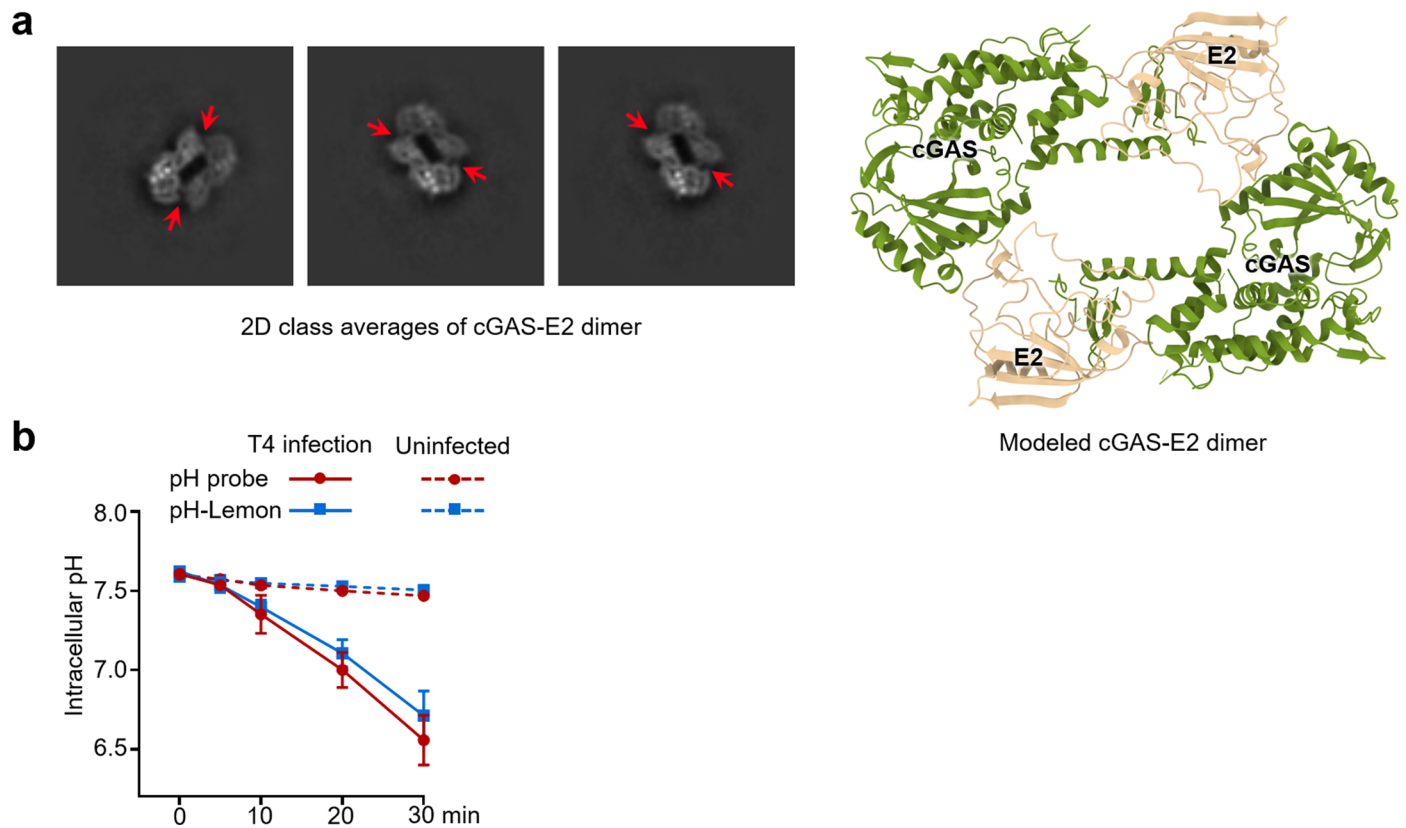
Extended Data Fig. 8 | Characterization of the isopeptide bond in poly-cGAS. **a**, Extracted ion chromatograms (left panel) and corresponding mass spectra of precursor ions (right panel) of the control peptide EQGSLHLDKK and the modified peptide EQGSLHLDKK bearing an NTFY remnant at K381 (red). **b**, Extracted ion chromatograms (left panel) and corresponding mass spectra of precursor ions (right panel) of the control

peptide TGGLIATGLAGTAAQAGVPKNTFYGE and the modified peptide TGGLIATGLAGTAAQAGVPKNTFYGE bearing an NTFY remnant at K401 (red). In both cases, MS signals of the modified peptides were only detected in cGAS-cGAS samples but not cGAS alone. **c**, SDS-PAGE analysis of cGAS mutants (indicated on above gel) co-purified with E2 K159A. All data are representative of three independent experiments.



Extended Data Fig. 9 | Regulation of cGAS activity by E2-mediated poly-cGASylation. **a**, HPLC analysis and quantification of the 3',2'-cGAMP synthesized by cGAS and variants as shown in Fig. 5a, with concentrations of ATP, GTP, and Mn being 250 μ M, 250 μ M, and 2.5 mM, respectively. **b**, HPLC analysis and

quantification of the 3',2'-cGAMP synthesized by cGAS and variants as shown in Fig. 5b, with the concentration of ATP, GTP, or Mn²⁺ in the reaction reduced by half compared to those in **a**. All data are representative of three independent experiments.



Extended Data Fig. 10 | Mechanisms underlying poly-cGAS formation. **a**, 2D class averages of cGAS-E2 dimer (left) and the modeled structure based on the 2D averages (right). **b**, Intracellular pH change of *E. coli* BL21 during T4 phage

infection. pH was determined using the intracellular pH probe BCECF-AM (red) or the FRET-based protein sensor (blue), respectively. Data are mean \pm s.d. for $n = 3$ biological replicates.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a | Confirmed |
|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection
Absorbance unit data was collected using Agilent OpenLAB CDS ChemStation Edition vC.01.10.
Mass spectrometry data was collected using Xcalibur v2.2.
Protein homologs identified using NCBI PSI-BLAST (web-based: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>).
Cryo-EM data was collected using SerialEM v3.6.

Data analysis
Cryo-EM data analysis: cryoSPARC v3.3.2.
Structure Refinement: Coot v0.9.6 and PHENIX v1.19.
Visualization: Chimera v1.12 and PyMOL v2.5.4.
Statistical analysis: GraphPad Prism v8.0.2.
Structure prediction: AlphaFold v2.2.1, ColabFold v1.3.0.
LC-MS/MS analysis: Mascot software v2.3.02, Maxquant v2.1.3.0, Xcalibur v2.2.
Edman sequencing analysis: PPSQ-30 DataProcessing v1.02.
Multiple Sequence Alignments: Clustal Omega v1.2.4, R language ggseqlogo v0.1.
HPLC analysis: OriginPro v9.1.0.
Chemical structure drawing: ChemDraw v19.0.
Other data analysis was performed using Microsoft Excel 2019

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

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data supporting the findings of this study are available in this article and its Supplementary Information files. Genomes of *Serratia marcescens* were sequenced previously, and sequences are publicly available. GenBank accessions for the CBASS operons used in this study are found in the Methods. Primers used in this study are provided in Supplementary Table 3 and 4. The *E. coli* strain B/BL21-DE3 proteome database that was used in this study can be accessed at Uniprot (UP000002032). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<https://proteomecentral.proteomexchange.org>) via the iProX partner repository with the dataset identifier PXD050178. For the cGAS-E2 complex, coordinates are available at the RCSB Protein Data Bank (PDB) (<http://www.rcsb.org>) under accession code 8HSB and 8YJY, and electron microscopy maps are available at the Electron Microscopy Data Bank (EMDB) (<https://www.ebi.ac.uk/emdb/>) under accession code EMD-34992 and EMD-39353. Minimum datasets that are necessary to interpret, verify, and extend the research in the article are provided as source data with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Reporting on race, ethnicity, or other socially relevant groupings

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size

Data exclusions

Replication

Randomization

Blinding

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|--|
| n/a | Involves in the study |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Plants |

Methods

- | | |
|-------------------------------------|---|
| n/a | Involves in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Antibodies

Antibodies used

Anti-GAPDH antibody produced in mouse (ab125247, Abcam), Custom rabbit anti-cGAS antibody generated by AtaGenix, HRP-conjugated goat anti-rabbit IgG (SA00001-2, Proteintech), and HRP-conjugated goat anti-mouse IgG (SA00001-1, Proteintech).

Validation

Anti-GAPDH was validated by the manufacturer and validation information can be found at the manufacturers' website (<https://www.abcam.cn/products/primary-antibodies/gapdh-antibody-ga1r-loading-control-ab125247.html>); Custom rabbit anti-cGAS antibody was validated in Figure 5c using cells not expressing cGAS; HRP-conjugated goat anti-rabbit IgG (<https://www.ptgcn.com/products/HRP-conjugated-Affinipure-Goat-Anti-Rabbit-IgG-H-L-secondary-antibody.htm>) and HRP-conjugated goat anti-mouse IgG (<https://www.ptgcn.com/products/HRP-conjugated-Affinipure-Goat-Anti-Mouse-IgG-H-L-secondary-antibody.htm>) were validated by the manufacturer and validation information can be found at the manufacturers' website.

Plants

Seed stocks

No plants were used.

Novel plant genotypes

No plants were used.

Authentication

No plants were used.