cGAS-like receptors sense RNA and control 3'2'-cGAMP signalling in Drosophila

Kailey M. Slavik, Benjamin R. Morehouse, Adelyn E. Ragucci, Wen Zhou, Xianlong Ai, Yuqiang Chen, Lihua Li, Ziming Wei, Heike Bähre, Martin König, et al.

cGAS-like receptors control RNA sensing and 3′2′-cGAMP antiviral signaling in Drosophila

Kailey M. Slavik[1,2], Benjamin R. Morehouse[1,2], Adelyn E. Ragucci[1,2], Wen Zhou[1,2,9], Xianlong Ai[3], Yuqiang Chen[3], Lihua Li[3], Ziming Wei[3], Heike Bähre[4], Martin König[4], Roland Seifert[4,5], Amy S. Y. Lee[2,6], Hua Cai[3], Jean-Luc Imler[3,7], Philip J. Kranzusch[1,2,8*]

[1] Department of Microbiology, Harvard Medical School, Boston, MA 02115, USA

[2] Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

[3] Sino-French Hoffmann Institute, State Key Laboratory of Respiratory Disease, School of Basic Medical Science, Guangzhou Medical University, Guangzhou, 511436, China.

[4] Research Core Unit Metabolomics, Hannover Medical School, 30625 Hannover, Germany

[5] Hannover Medical School, Institute of Pharmacology, Carl-Neuberg-Str. 1, 30625 Hannover, Germany

[6] Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

[7] Université de Strasbourg, CNRS UPR9022, 67084 Strasbourg, France

[8] Parker Institute for Cancer Immunotherapy at Dana-Farber Cancer Institute, Boston, MA 02115, USA

[9] Present address: School of Life Sciences, Southern University of Science and Technology, Shenzhen, Guangdong, China

*Correspondence: philip_kranzusch@dfci.harvard.edu

Cyclic GMP–AMP synthase (cGAS) is a cytosolic DNA sensor that produces the second messenger 2′3′-cGAMP and controls activation of innate immunity in mammalian cells[1-5]. Animal genomes typically encode multiple proteins with predicted homology to cGAS[6-12], but the function of these uncharacterized enzymes is unknown. Here we show that cGAS-like receptors (cGLRs) are innate immune sensors capable of recognizing divergent molecular patterns and catalyzing synthesis of distinct nucleotide second messenger signals. Crystal structures of human and insect cGLRs reveal a nucleotidyltransferase signaling core shared with cGAS and a diversified primary ligand-binding surface modified with significant insertions and deletions. We demonstrate that cGLR surface remodeling enables altered ligand specificity and use a forward biochemical screen to identify cGLR1 as a double-stranded RNA sensor in the model organism *Drosophila melanogaster*. Surprisingly, RNA recognition activates *Drosophila* cGLR1 to synthesize the novel product cG[3′–5′]pA[2′–5′]p (3′2′-cGAMP). A crystal structure of *Drosophila* Stimulator of Interferon Genes (STING) in complex with 3′2′-cGAMP explains selective isomer recognition and we demonstrate that 3′2′-cGAMP induces an enhanced antiviral state *in vivo* that protects from viral infection. Similar to radiation of Toll-like receptors in pathogen immunity, our results establish cGLRs as a diverse family of metazoan pattern recognition receptors.

To define the function of cGAS-like enzymes in animals, we screened predicted cGAS-homologs for suitability in structural analysis and determined a 2.4 Å crystal structure of the human protein MB21D2 (hMB21D2; C3orf59) and a 1.6 Å crystal structure of a protein from the beetle species *Tribolium castaneum* (Genbank XP_969398.1) (Supplementary Table 1). Despite primary sequence divergence, the hMB21D2 and *T. castaneum* XP_969398.1 structures each reveal close homology to human cGAS with a shared bi-lobed architecture, caged nucleotidyltransferase core, Gly-[Gly/Ser]-activation loop, and putative catalytic triad (Fig. 1a, Extended Data Fig. 1). In human cGAS, the primary ligand binding surface is a long groove on

the back of the enzyme formed by the α-helix spine and a Zn-ribbon motif that is essential for recognition of double-stranded DNA[3,13-17]. A conserved groove is present in both the hMB21D2 and *T. castaneum* XP_969398.1 structures (Fig. 1a) but is notably distinguished by the absence of a Zn-ribbon and the insertion of a C-terminal α-helix in hMB21D2 (Fig. 1b). We hypothesized that the remodeling of this groove controls the detection of distinct activating ligands. The hMB21D2 surface is overall neutral with no obvious capacity to coordinate nucleic acid and no enzyme activity was detected in the presence of potential activating ligands (Extended Data Fig. 1d,e). In contrast to hMB21D2, the surface of *T. castaneum* XP_969398.1 shares highly conserved basic residues with human cGAS (Fig. 1a) and we therefore tested this enzyme with candidate DNA and RNA ligands. Remarkably, we observed that *T. castaneum* XP_969398.1 is activated to synthesize a nucleotide product in response to recognition of double-stranded RNA (dsRNA) (Fig. 1c). Despite exhibiting a clear difference in ligand-specificity, analysis of all related structures in the Protein Data Bank confirms that *T. castaneum* XP_969398.1 is a close homolog of mammalian cGAS and is distinct from previously characterized RNA sensors including oligoadenylate synthase 1 (Extended Data Fig. 1f)[18,19]. Together, these results establish the existence of cGAS-like Receptors (cGLRs) in animals and demonstrate that remodeling of a primary ligand-binding surface enables the recognition of divergent molecular patterns.

To identify additional cGLRs that respond to dsRNA, we used the *T. castaneum* cGLR (*Tc*-cGLR) sequence to search for predicted cGAS homologs in species related to the model organism *Drosophila melanogaster*. We identified 153 cGLR genes across 42 species in the order *Diptera*, which cluster into distinct clades designated 1–5 (Fig. 2a; Supplementary Table 2). *Drosophila* exhibit a remarkable radiation of cGLR genes with individual species encoding between three and seven predicted enzymes (Extended Data Fig. 2a). In a systematic biochemical screen, we purified and tested 53 recombinant cGLR proteins and identified active enzymes from the species *Lucilia cuprina*, *Drosophila eugracilis*, *Drosophila erecta*, and *Drosophila simulans* (Extended Data Figs. 2b–f, 3a). Similar to *Tc*-cGLR, each active *Diptera*

enzyme specifically responds to dsRNA indicating that cGLR-based recognition of RNA is conserved across diverse insect species (Fig. 2b; Extended Data Fig. 4a).

The *D. simulans* enzyme identified in our screen shares 91% sequence identity with the protein product of the uncharacterized *D. melanogaster* gene *CG12970*. We purified recombinant *D. melanogaster* CG12970 and found that it synthesizes a nucleotide product specifically in the presence of dsRNA and we therefore named this gene *cGAS-like Receptor 1* (*Dm-cGLR1*) (Fig. 2c). To understand how dsRNA activates *Drosophila* cGLR1, we analyzed the molecular determinants for enzymatic activity *in vitro.* We observed that *D. simulans* cGLR1 (*Ds*-cGLR1) and *Dm*-cGLR1 recognize long >30 bp dsRNA with no preference for 5′ RNA phosphorylation (Fig. 2d; Extended Data Fig. 4b,c). Notably, activation of *Dm*-cGLR1 and *Ds*-cGLR1 requires dsRNA ligands that exceed the length of 21–23 bp RNA molecules commonly produced during RNA interference (RNAi) in *Drosophila*, suggesting specific avoidance of self-recognition[20-22]. *Ds*-cGLR1 selectively binds dsRNA and forms a higher-order complex that is dependent on dsRNA length, similar to condensate formation previously observed with hcGAS recognition of dsDNA (Fig. 2e; Extended Data Fig. 5)[23-25]. Using the *Tc*-cGLR and human cGAS–DNA structures as a template[17], we modeled interactions between *Drosophila* cGLR1 and dsRNA (Fig. 2f). Single glutamate substitutions to the conserved basic residues R41, R259, and K269 within the predicted *Ds*-cGLR1 ligand binding surface significantly diminish dsRNA-stimulated activity (Fig. 2g; Extended Data Fig. 3). We ectopically expressed *Dm*-cGLR1 or *Ds*-cGLR1 in human cells and observed that expression is sufficient to enable dsRNA sensing and drive activation of a STING-dependent immune response (Fig. 2h; Extended Data Fig. 4e). *Dm*-cGLR1 and *Ds*-cGLR1 signaling in cells is strictly dependent on dsRNA stimulation and mutations to the enzyme catalytic site or conserved ligand-binding surface disrupt signaling and prevent downstream STING activation (Fig. 2h; Extended Data Fig. 3f). Together, these data demonstrate that insect cGLRs and human cGAS use a shared mechanism of ligand detection and reveal that *Dm*-cGLR1 can function as a foreign RNA sensor.

A role in sensing foreign RNA suggests that the function of *Dm*-cGLR1 is to control activation of a downstream immune response in *Drosophila*. In human cells, cGAS catalyzes production of the nucleotide second messenger 2′3′-cGAMP (cG[2′–5′]pA[3′–5′]p) that contains a non-canonical 2′–5′ phosphodiester linkage required for potent activation of immune signaling[2-5]. To determine how *Dm*-cGLR1 controls cellular signaling, we purified the nucleotide reaction product for direct comparison to 2′3′-cGAMP. Surprisingly, *Dm*-cGLR1 synthesizes a nucleotide product that exhibits a C18 chromatography migration profile distinct from 2′3′-cGAMP and all previously known naturally occurring cyclic dinucleotide (CDN) signals (Fig. 3a; Extended Data Fig. 6). Production of this nucleotide signal is conserved in *Diptera* with *Ds*-cGLR1, *Lc*-cGLR, and *Deu*-cGLR reactions each synthesizing the same major reaction product (Extended Data Fig. 6a). Using nucleobase-specific labeling and nuclease digestion of the *Dm*-cGLR1 product we observed a 3′−5′-linkage connected to an adenosine phosphate and a protected 2′−5′-linkage connected to a guanosine phosphate suggesting the existence of a mixed-linkage cyclic GMP–AMP species (Fig. 3b). We verified these findings with comparative high performance liquid chromatography and tandem mass-spectrometry profiling against a chemically synthesized standard and confirmed that the shared *Diptera* cGLR product is the novel isomer 3′2′-cGAMP (cG[3′–5′]pA[2′–5′]p) (Figs. 3a,b; Extended Data Figs. 6a,b).

Dm-cGLR1 synthesizes 3′2′-cGAMP in a two-step reaction through production of the linear intermediate pppA[2′–5′]pG and uses an opposite nucleobase reaction order compared to human cGAS (Extended Data Fig. 7a)[2,3,26]. We next used mass-spectrometry to further analyze each lysate sample from our screen of recombinant dipteran cGLR proteins. 3′2′-cGAMP was detected as a main reaction product from 15 *Diptera* cGLRs including active enzymes from each sub-group within Clade 5 of the insect cGLR phylogeny (Fig. 3c, Extended Data Fig. 6c). cGLRs clustered within Clade 5 collectively represent 41 species suggesting widespread conservation of 3′2′-cGAMP-sigaling in *Diptera*. Interestingly, the beetle enzyme *Tc*-cGLR synthesizes 2′3′-cGAMP, supporting that 2′3′-cGAMP is an ancestral signaling molecule in metazoans and that 3′2′-

cGAMP-signaling is a recent adaptation in flies (Fig. 3d; Extended Data Fig. 6a)[10,27,28]. Insect and mammalian viruses encode 2′3′-cGAMP-specific nucleases named poxins that allow evasion of cGAS-STING immune responses[29,30]. Remarkably, 3′2′-cGAMP is protected from poxin cleavage (Extended Data Fig. 7b–d), indicating that an isomeric switch in phosphodiester linkage specificity endows *Drosophila* with a signaling pathway resistant to a major form of viral immune evasion.

Drosophila STING (dSTING) is known to function as a cyclic dinucleotide receptor *in vivo*[31-34], but an endogenous nucleotide second messenger has not previously been identified. We therefore developed an *in vitro* thermo-fluor binding assay to analyze the ability of dSTING to recognize specific CDNs. dSTING preferentially forms a thermo-stable complex with 3′2′-cGAMP and exhibits no detectable complex formation with 2′3′-cGAMP or other CDNs *in vitro* (Fig. 4a; Extended Data Fig. 8b,c). Using direct delivery of CDNs to permeabilized cells, we confirmed that dSTING preferentially responds to 3′2′-cGAMP in the cellular environment (Extended Data Fig. 8d). To define the mechanism of selective 3′2′-cGAMP recognition, we next determined a 2.0 Å crystal structure of the *D. eugracilis* STING CDN-binding domain in complex with 3′2′-cGAMP (Fig. 4b). dSTING adopts a highly conserved V-shaped homodimeric architecture with a deep central pocket that binds 3′2′-cGAMP. The dSTING–3′2′-cGAMP structure reveals a tightly "closed" conformation with dSTING protomers positioned 36 Å apart, similar to the closed conformation of human STING bound to 2′3′-cGAMP (Extended Data Fig. 8e)[5,35]. Each nucleobase of 3′2′-cGAMP is stacked between dSTING Y164 and R234, and E257 specifically coordinates the 3′2′-cGAMP guanosine N2 position (Fig. 4c,d). In human STING, high-affinity recognition of 2′3′-cGAMP requires readout of the 2′–5′ phosphodiester linkage by R232 in the β-strand lid[5,35]. In dSTING, the equivalent R229 makes no contact with either phosphodiester bond. Instead, R229 is repositioned to extend outward from the ligand binding pocket by the deletion of a single lid residue and the formation of a salt bridge with E267 on the opposing protomer, explaining the diminished affinity of dSTING for 2′3′-cGAMP (Fig. 4c,e). Additionally, a key asparagine substitution N159 in dSTING extends across the binding pocket to coordinate the

adenosine 3′ OH in 3′2′-cGAMP and directly replaces the human STING S162 residue that contacts the guanosine 3′ OH in 2′3′-cGAMP (Fig. 4f). We tested a panel of dSTING mutant proteins and confirmed that mutations to each coordinating residue disrupt dSTING–3′2′-cGAMP complex formation (Extended Data Fig. 8i). The unique ligand-binding pocket adaptations observed in the dSTING–3′2′-cGAMP structure are widely conserved in *Diptera* and together explain a mechanism for how specific 3′2′-cGAMP-dependent signaling drives STING activation.

To determine how *Dm*-cGLR1–3′2′-cGAMP–STING signaling controls immune responses *in vivo*, we next injected 3′2′-cGAMP into *D. melanogaster* to directly monitor the dSTING response. 3′2′-cGAMP potently induces the expression of *dSTING* and three other STING-regulated genes in a dose-dependent manner (Fig 5a; Extended Data Fig. 9). Notably, 3′2′-cGAMP-dependent signaling through dSTING is significantly more potent than the response triggered by injection of the bacterial CDN signal 3′3′-c-di-GMP (Fig 5a; Extended Data Fig. 9e–k). Genetic mutations to *dSTING* and the NF-κB homolog *Relish* ablate 3′2′-cGAMP-induced responses, demonstrating that signaling is dependent upon each downstream pathway component (Fig. 5a, Extended Data Fig. 9e–k). We challenged flies with viral infection and observed that 3′2′-cGAMP markedly suppresses the replication of two unrelated RNA viruses, *Drosophila* C virus (*Dicistroviridae*), a natural *Drosophila* pathogen, and vesicular stomatitis virus (*Rhabdoviridae*) (Figs. 5b,c; Extended Data Fig. 10a,b). 3′2′-cGAMP activation of antiviral immunity is strictly dependent on dSTING and results in a response that significantly delays pathogen-mediated mortality (Figs. 5b,c; Extended Data Fig. 10a,b). Direct comparison of the protective effects against DCV infection demonstrates that the endogenous signal 3′2′-cGAMP exhibits greater antiviral potency than 2′3′-cGAMP. 3′2′-cGAMP more robustly suppresses RNA viral loads and extends animal survival (Fig. 5d; Extended data Fig. 10c,d), revealing that the dSTING antiviral signaling axis is preferentially activated by 3′2′-cGAMP *in vivo*. Together, these results demonstrate that 3′2′-cGAMP is an antiviral nucleotide second messenger in *D.*

melanogaster and establish a novel cGLR–STING–NF-κB axis that protects animals from viral replication.

Along with cGAS recognition of dsDNA, the discovery of animal cGLR dsRNA sensors establishes a diverse class of pattern recognition receptors conserved throughout metazoans. Divergent structural homologs of cGAS in humans and insects demonstrate that cGLRs constitute a rapidly evolving family of proteins in which remodeling of a shared primary binding surface enables the detection of diverse ligands. Our characterization of the mechanism of *Drosophila* cGLR1 activation shows that cGLRs function as direct sensors of pathogen-associated molecular patterns and synthesize distinct second messengers to control a conserved downstream signaling axis. *Drosophila* were previously thought to respond to foreign nucleic acid exclusively through RNAi and direct cleavage of pathogen RNA[21,22]. *Drosophila* cGLR1 reveals a parallel signaling system for sensing foreign RNA and directing an inducible immune response through STING. Synthesis of the novel second messenger 3′2′-cGAMP by *Drosophila* cGLR1 and selective recognition by dSTING provides the first evidence that metazoans use CDNs beyond 2′3′-cGAMP as endogenous second messengers and highlights the evolutionary plasticity of cGLR signaling. Our structural analysis also reveals that the human cGLR MB21D2 is competent for nucleotide second messenger synthesis and has a remodeled ligand-binding groove likely adapted for detection of an unknown stimulus. Together with the known high frequency of hMB21D2 mutation in cancer[37,38], these results support a more extensive role for cGLR signaling in human biology. The existence of multiple unique cGLRs encoded within a single species (Extended Data Fig. 2a) suggests a model in which the cGLR signaling scaffold is harnessed to detect a number of distinct stimuli (Fig. 5e). In support of this conclusion, Hartmann, Imler, Cai and colleagues identify cGLR2 as a second functional cGLR in *Drosophila* and demonstrate *in vivo* that cGLR1 and cGLR2 have distinct functions in *Drosophlia* immunity[36]*.* Together, our results define cGLRs as receptors in animal cells capable of detecting diverse pathogen-associated molecular patterns and dictating response to the foreign environment.

References

1 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* (2013) **339**, 786-791. PMC3863629.

2 Ablasser, A., Goldeck, M., Cavlar, T., Deimling, T., Witte, G., Rohl, I., Hopfner, K. P., Ludwig, J. & Hornung, V. cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. *Nature* (2013) **498**, 380-384. PMC4143541.

3 Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B. L., Zillinger, T., Serganov, A. A., Liu, Y., Jones, R. A., Hartmann, G., Tuschl, T. & Patel, D. J. Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* (2013) **153**, 1094-1107. PMC4382009.

4 Diner, E. J., Burdette, D. L., Wilson, S. C., Monroe, K. M., Kellenberger, C. A., Hyodo, M., Hayakawa, Y., Hammond, M. C. & Vance, R. E. The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell Rep* (2013) **3**, 1355-1361. PMC3706192.

5 Zhang, X., Shi, H., Wu, J., Zhang, X., Sun, L., Chen, C. & Chen, Z. J. Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol Cell* (2013) **51**, 226-235. PMC3808999.

6 Kranzusch, P. J. cGAS and CD-NTase enzymes: structure, mechanism, and evolution. *Curr Opin Struct Biol* (2019) **59**, 178-187. PMC7127440.

7 Wu, J. & Chen, Z. J. Innate immune sensing and signaling of cytosolic nucleic acids. *Annu Rev Immunol* (2014) **32**, 461-488

8 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* (2015) **43**, 10633-10654. PMC4678834.

9 Wu, X., Wu, F. H., Wang, X., Wang, L., Siedow, J. N., Zhang, W. & Pei, Z. M. Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic Acids Res* (2014) **42**, 8243-8257. PMC4117786.

10 Kranzusch, P. J., Wilson, S. C., Lee, A. S., Berger, J. M., Doudna, J. A. & Vance, R. E. Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2',3' cGAMP Signaling. *Mol Cell* (2015) **59**, 891-903. PMC4575873.

11 Whiteley, A. T., Eaglesham, J. B., de Oliveira Mann, C. C., Morehouse, B. R., Lowey, B., Nieminen, E. A., Danilchanka, O., King, D. S., Lee, A. S. Y., Mekalanos, J. J. & Kranzusch, P. J. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* (2019) **567**, 194-199. PMC6544370.

12 de Oliveira Mann, C. C., Kiefersauer, R., Witte, G. & Hopfner, K. P. Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein MAB21L1. *Sci Rep* (2016) **6**, 27498. PMC4897736.

13 Civril, F., Deimling, T., de Oliveira Mann, C. C., Ablasser, A., Moldt, M., Witte, G., Hornung, V. & Hopfner, K. P. Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* (2013) **498**, 332-337. PMC3768140.

14 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* (2013) **3**, 1362-1368. PMC3800681.

15 Li, X., Shu, C., Yi, G., Chaton, C. T., Shelton, C. L., Diao, J., Zuo, X., Kao, C. C., Herr, A. B. & Li, P. Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity* (2013) **39**, 1019-1031. PMC3886715.

16 Zhang, X., Wu, J., Du, F., Xu, H., Sun, L., Chen, Z., Brautigam, C. A., Zhang, X. & Chen, Z. J. The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes switch-like conformational changes in the activation loop. *Cell Rep* (2014) **6**, 421-430. PMC3969844.

17 Zhou, W., Whiteley, A. T., de Oliveira Mann, C. C., Morehouse, B. R., Nowak, R. P., Fischer, E. S., Gray, N. S., Mekalanos, J. J. & Kranzusch, P. J. Structure of the Human cGAS-DNA Complex Reveals Enhanced Control of Immune Surveillance. *Cell* (2018) **174**, 300-311 e311. PMC6084792.

18 Lohofener, J., Steinke, N., Kay-Fedorov, P., Baruch, P., Nikulin, A., Tishchenko, S., Manstein, D. J. & Fedorov, R. The Activation Mechanism of 2'-5'-Oligoadenylate Synthetase Gives New Insights Into OAS/cGAS Triggers of Innate Immunity. *Structure* (2015) **23**, 851-862

19 Donovan, J., Dufner, M. & Korennykh, A. Structural basis for cytosolic double-stranded RNA surveillance by human oligoadenylate synthetase 1. *Proc Natl Acad Sci U S A* (2013) **110**, 1652-1657. PMC3562804.

20 Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nat Rev Genet* (2009) **10**, 94-108. PMC2724769.

21 Guo, Z., Li, Y. & Ding, S. W. Small RNA-based antimicrobial immunity. *Nat Rev Immunol* (2019) **19**, 31-44

22 Schneider, J. & Imler, J. L. Sensing and signalling viral infection in drosophila. *Dev Comp Immunol* (2021) **117**, 103985

23 Du, M. & Chen, Z. J. DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science* (2018) **361**, 704-709

24 Xie, W., Lama, L., Adura, C., Tomita, D., Glickman, J. F., Tuschl, T. & Patel, D. J. Human cGAS catalytic domain has an additional DNA-binding interface that enhances enzymatic activity and liquid-phase condensation. *Proc Natl Acad Sci U S A* (2019) **116**, 11946-11955. PMC6575157.
25 Zhou, W., Mohr, L., Maciejowski, J. & Kranzusch, P. J. cGAS phase separation inhibits TREX1-mediated DNA degradation and enhances cytosolic DNA sensing. *Mol Cell* (2021) **81**, DOI 10.1016/j.molcel.2021.1001.1024.
26 Kranzusch, P. J., Lee, A. S. Y., Wilson, S. C., Solovykh, M. S., Vance, R. E., Berger, J. M. & Doudna, J. A. Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. *Cell* (2014) **158**, 1011-1021. PMC4157622.
27 Gui, X., Yang, H., Li, T., Tan, X., Shi, P., Li, M., Du, F. & Chen, Z. J. Autophagy induction via STING trafficking is a primordial function of the cGAS pathway. *Nature* (2019) **567**, 262-266
28 Morehouse, B. R., Govande, A. A., Millman, A., Keszei, A. F. A., Lowey, B., Ofir, G., Shao, S., Sorek, R. & Kranzusch, P. J. STING cyclic dinucleotide sensing originated in bacteria. *Nature* (2020) **586**, 429-433. PMC7572726.
29 Eaglesham, J. B., Pan, Y., Kupper, T. S. & Kranzusch, P. J. Viral and metazoan poxins are cGAMP-specific nucleases that restrict cGAS-STING signalling. *Nature* (2019) **566**, 259-263. PMC6640140.
30 Eaglesham, J. B., McCarty, K. L. & Kranzusch, P. J. Structures of diverse poxin cGAMP nucleases reveal a widespread role for cGAS-STING evasion in host-pathogen conflict. *Elife* (2020) **9**. PMC7688311.
31 Goto, A., Okado, K., Martins, N., Cai, H., Barbier, V., Lamiable, O., Troxler, L., Santiago, E., Kuhn, L., Paik, D., Silverman, N., Holleufer, A., Hartmann, R., Liu, J., Peng, T., Hoffmann, J. A., Meignin, C., Daeffler, L. & Imler, J. L. The Kinase IKKbeta Regulates a STING- and NF-kappaB-Dependent Antiviral Response Pathway in Drosophila. *Immunity* (2018) **49**, 225-234 e224. PMC6267954.
32 Liu, Y., Gordesky-Gold, B., Leney-Greene, M., Weinbren, N. L., Tudor, M. & Cherry, S. Inflammation-Induced, STING-Dependent Autophagy Restricts Zika Virus Infection in the Drosophila Brain. *Cell Host Microbe* (2018) **24**, 57-68 e53. PMC6173519.
33 Martin, M., Hiroyasu, A., Guzman, R. M., Roberts, S. A. & Goodman, A. G. Analysis of Drosophila STING Reveals an Evolutionarily Conserved Antimicrobial Function. *Cell Rep* (2018) **23**, 3537-3550 e3536. PMC6114933.
34 Cai, H., Holleufer, A., Simonsen, B., Schneider, J., Lemoine, A., Gad, H. H., Huang, J., Huang, J., Chen, D., Peng, T., Marques, J. T., Hartmann, R., Martins, N. E. & Imler, J. L. 2'3'-cGAMP triggers a STING- and NF-kappaB-dependent broad antiviral response in Drosophila. *Sci Signal* (2020) **13**
35 Gao, P., Ascano, M., Zillinger, T., Wang, W., Dai, P., Serganov, A. A., Gaffney, B. L., Shuman, S., Jones, R. A., Deng, L., Hartmann, G., Barchet, W., Tuschl, T. & Patel, D. J. Structure-function analysis of STING activation by c[G(2',5')pA(3',5')p] and targeting by antiviral DMXAA. *Cell* (2013) **154**, 748-762. PMC4386733.
36 Holleufer, A., Winther, K. G., Gad, H. H., Ai, X., Chen, Y., Li, L., Wei, Z., Deng, H., Liu, J., Frederiksen, N. A., Simonsen, B., Andersen, L. L., Kleigrewe, K., Pichlmair, A., Cai, H., Imler, J. L. & Hartmann, R. Two cGAS-like receptors induce a Sting-dependent antiviral immune response in Drosophila melanogaster. *Submitted Manuscript* (2021)

Acknowledgements: The authors are grateful to P.M. Devant, K. Chat, A. Holleufer, R. Hartmann, T. IV. L., and members of the Kranzusch lab for helpful comments and discussion, C. de Oliveira Mann for assistance developing mass spectrometry analysis of insect cGLRs, and M. Burroughs and A. Iyer for bioinformatics analysis of cGAS-like enzymes. The work was funded by grants to P.J.K. from the Pew Biomedical Scholars program, the Burroughs Wellcome Fund PATH program, The Richard and Susan Smith Family Foundation, The Mathers Foundation, The Mark Foundation for Cancer Research, a Cancer Research Institute CLIP grant, a V Foundation V Scholar Award, and the Parker Institute for Cancer Immunotherapy; grants to J.L.I and H.C. from the Agence Nationale de la Recherche (ANR-17-CE15-0014), the Investissement d'Avenir Programme (ANR-10-LABX-0036 and ANR-11-EQPX-0022), the Institut Universitaire de France,

the Chinese National Overseas expertise Introduction Center for Discipline Innovation (Project "111" [D18010]), the Foreign Experts Program (2020A1414010306) and The Natural Science Foundation (32000662); and a grant from the Deutsche Forschungsgemeinschaft (DFG) within the Priority Program SPP1879 and INST 192/524-1 FUGG. W.Z. is supported as a Benacerraf Fellow in Immunology and through a Charles A. King Trust Postdoctoral Fellowship. B.R.M. is supported as a Ruth L. Kirschstein NRSA Postdoctoral Fellow NIH F32GM133063. X-ray data were collected at the Northeastern Collaborative Access Team beamlines 24-ID-C and 24-ID-E (P30 GM124165), and used a Pilatus detector (S10RR029205), an Eiger detector (S10OD021527) and the Argonne National Laboratory Advanced Photon Source (DE-AC02-06CH11357), and at beamlines 5.0.1 and 8.2.2 of the Advanced Light Source, a U.S. DOE Office of Science User Facility under Contract No. DE-AC02-05CH11231 and supported in part by the ALS-ENABLE program and NIGMS grant P30 GM124169-01.

Author Contributions: Experiments were designed and conceived by K.M.S. and P.J.K. Gene identification and phylogenetic analyses were performed by K.M.S., B.R.M., and P.J.K. *Tc*-cGLR structural experiments were performed by K.M.S., hMB21D2 structural experiments were performed by B.R.M, and dSTING structural experiments were performed by K.M.S and A.E.R. hMB21D2, cGLR and dSTING biochemical experiments were performed by K.M.S. and A.E.R. Phase separation and RNA-binding analysis were performed by W.Z and A.E.R. Cell biology experiments were designed by K.M.S. and A.S.Y.L., and performed by K.M.S. Nucleotide purification and mass-spectrometry experiments were performed by K.M.S., H.B., M.K, and R.S. *In vivo* drosophila experiments were designed and performed by X.A., Y.C., L.L., Z.W., H.C., and J.L.I. The manuscript was written by K.M.S. and P.J.K. All authors contributed to editing the manuscript and support the conclusions.

Author Information: The authors have no competing financial interests to declare. Correspondence and requests for materials should be addressed to P.J.K.

Figure Legends

Figure 1 | Structural remodeling in animal cGLRs enables divergent pattern recognition.

a, Crystal structures and surface electrostatic views of human MB21D2 and a cGLR from the beetle *T. castaneum* (*Tc*-cGLR). Structural comparison with the human cGAS–DNA complex (PDB 6CTA) reveals that cGLRs adopt a conserved overall architecture with a nucleotidyltransferase signaling core and shared primary ligand-binding surface (indicated by dashed lines). **b**, Zoom-in cutaways highlighting structural insertions and deletions unique to each cGLR. hMB21D2 and *Tc*-cGLR lack the Zn-ribbon motif present in cGAS (left) and hMB21D2 contains a C-terminal α-helix extension that contacts the central "spine" helix (right). Specific alterations in the predicted ligand-binding surfaces suggest individual cGLRs are remodeled for recognition of different molecular patterns. **c**, Thin-layer chromatography (TLC) analysis and quantification of *Tc*-cGLR enzymatic reactions in the presence of candidate 40 nt or bp nucleic acid ligands. In contrast to mouse cGAS recognition of dsDNA, *Tc*-cGLR synthesizes a nucleotide product specifically in response to recognition of dsRNA confirming altered ligand specificity. Data are relative to maximum enzymatic activity and represent the mean ± SEM for n = 3 independent experiments.

Figure 2 | *Drosophila* cGLR1 senses long double-stranded RNA.

a, Summary of a forward biochemical screen of 53 cGLR proteins in the order *Diptera*. cGLR genes were selected based on predicted structural homology to *Tc*-cGLR and human cGAS, and conservation of putative nucleotidyltransferase catalytic residues. Screened enzymes are denoted with a red dot, and active dsRNA-sensors are denoted with a blue circle. The phylogeny represents 153 proteins clustered into Clades 1–5, with <30% sequence identity between clades. 41/42 *Diptera* species are represented in Clade 5, which contains *D. melanogaster* CG12970 (cGLR1) and CG30424. **b,** Functional *Diptera* cGLRs identified in the biochemical screen are activated to form a nucleotide product by the dsRNA mimic poly I:C. **c,** A single mutation to the

Dm-cGLR1 active site disrupts all enzymatic activity. **d,** *Dm*-cGLR1 *in vitro* activity was monitored in the presence of a panel of synthetic dsRNAs from 10–40 bp and quantified relative to 40 bp reactions. **e,** Electrophoretic mobility shift assay (EMSA) shows binding between *Ds*-cGLR1 and a 40 bp dsRNA. Formation of a higher order protein–nucleic acid complex which does not migrate through the gel is also observed with human cGAS and a 45 bp dsDNA ligand, in contrast to 2:2 binding between mouse cGAS and dsDNA. **f,** Model of the *Tc*-cGLR–dsRNA complex based on the hcGAS–dsDNA structure (PDB 6CTA). Predicted *Dm*-cGLR1 (*Ds*-cGLR1) ligand-binding residues R23 (41), K42 (60), K52 (70), R 241 (259) and K251 (269) are analogous to *Tc*-cGLR residues R22, K40, R51, R249, and K259 respectively. **g,** Analysis and quantification of *in vitro* *Ds*-cGLR1 activity demonstrates that charge-swap mutations to ligand-binding residues significantly impair enzyme activation (see also **Extended Data Fig. 3**) **h**, Analysis of *Dm*-cGLR1 activity in human cells using a mammalian STING-controlled IFN-β luciferase reporter. IFN-β is quantified relative to vector control and shown here relative to wild-type activity. *Dm*-cGLR1 signaling to STING is strictly dependent on poly I:C stimulation and mutation of the catalytic site or predicted ligand-binding residues ablates this activity. Data in **h** are mean ± SEM of n = 3 technical replicates and representative of n = 3 independent experiments; all other data are the mean ± SEM of n = 3 independent experiments.

Figure 3 | Discovery of 3′2′-cGAMP as a metazoan nucleotide second messenger.

a, HPLC chromatogram of the *Dm*-cGLR1 reaction (orange) and comparison with synthetic nucleotide standards. *Dm*-cGLR1 synthesizes 3′2′-cGAMP as a major product that migrates identical to a synthetic standard (black). The retention time of other standards are indicated with a dotted line, demonstrating that a minor *Dm*-cGLR1 reaction product is 2′3′-c-di-AMP (see also **Extended Data Fig. 6a**). **b,** TLC analysis of mouse cGAS and *Dm*-cGLR1 reactions labeled with either α-^{32}P-ATP or α-^{32}P-GTP and treated as indicated. Schematic on right demonstrates how pairwise labeling and Nuclease P1 digestion verifies that cGAS and *Dm*-cGLR1 synthesize

368 distinct cGAMP isomers with opposite phosphodiester linkage specificities. High-resolution mass
369 spectrometry confirms the major *Diptera* cGLR product as 3′2′-cGAMP (see also **Extended Data**
370 **Fig. 6b**). **c,** Inset of Clade 5 in the *Diptera* cGLR phylogeny from **Fig. 2a** annotated to denote all
371 enzymes identified to synthesize 3′2′-cGAMP by bacterial lysate analysis. **d,** HPLC analysis and
372 quantification of product formation by insect cGLRs. 3′2′-cGAMP is the dominant product of each
373 identified *Diptera* cGLR, and 2′3′-cGAMP is the dominant product of cGAS and *Tc*-cGLR. Data
374 are mean ± SEM for n = 3 independent experiments.
375
376 **Figure 4 | Structural basis for 3′2′-cGAMP recognition by *Drosophila* STING**
377 **a**, Thermal denaturation assay showing selective recognition of 3′2′-cGAMP *Drosophila* STING
378 (see also **Extended Data Fig. 8b,c**). Data shown are representative of n = 3 independent
379 experiments. **b**, Crystal structure of the dSTING–3′2′-cGAMP complex reveals a tightly closed
380 homodimer conformation and an ordered β-strand lid, indicating high-affinity engagement with the
381 endogenous *Drosophila* second messenger 3′2′-cGAMP. **c**, Phylogenetic alignment of the stem
382 helix and β-strand lid in human and insect STING proteins colored by amino-acid conservation.
383 Critical ligand binding residues are denoted with a navy circle and adaptations specific to *Diptera*
384 are highlighted in red outline. **d**, Highlight of 3′2′-cGAMP in the CDN-binding pocket of dSTING
385 showing key ligand contacts. **e**, Superposition of the dSTING–3′2′-cGAMP (blue-orange) and
386 human STING–2′3′-cGAMP (gray-pink) complexes reveals human STING readout of the 2′–5′
387 phosphodiester bond by the conserved R229 is absent in dSTING. dSTING R229 is stabilized in
388 a conformation extended outward from the ligand-binding pocket interaction with E267 on the
389 opposite dimer. **f**, Human STING S162 (gray) contacts the free 3′ OH of the guanosine base in
390 2′3′-cGAMP (pink). dSTING N159 (blue) extends across the ligand binding pocket to contact the
391 the free 3′ OH of the adenosine base in 3′2′-cGAMP (orange).
392
393 **Figure 5 | 3′2′-cGAMP activates STING-dependent antiviral immunity in *Drosophila*.**

a, Synthetic 3′2′-cGAMP or 3′3′-c-di-GMP was injected into the body cavity of wildtype or mutant flies and gene expression measured after 24 h. *STING-regulated gene 1* (*Srg1*) RNA levels shown as fold induction compared to buffer control in wildtype. dSTING$_{Mut}$ = RXN mutant, as previously characterized[31,34]. Data represent RNA levels measured relative to the house-keeping gene *RpL32* and are from 3 independent experiments *P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001; and n.s., P > 0.05. P value n.s. unless otherwise noted. **b,** Viral RNA loads 3 days after infection with the RNA pathogen Drosophila C virus (DCV) demonstrate significantly diminished viral replication in wildtype flies injected with 3′2′-cGAMP. **c,** Survival analysis of animals injected with 3′2′-cGAMP or buffer control and infected with DCV. 3′2′-cGAMP injection results in a dSTING-dependent response that significantly delays mortality. **d,** Survival analysis directly comparing the effects of cGAMP isomers 7 days after DCV infection. cGAMP injection increases animal survival in a dose-dependent manner compared to buffer control and injection of 3′2′-cGAMP confers greater protection than 2′3′-cGAMP at each dose tested (see also **Extended Data Fig. 10 c,d**). **e**, Proposed model for the cGLR-STING signaling. Diverse animal cGLRs recognize distinct molecular patterns, respond by synthesizing a nucleotide second messenger, and induce activation of STING-dependent antiviral immunity.

Methods

Bioinformatics and Dipteran cGLR sequence analysis

Building on previous analyses[6-12], animal cGAS homologs suitable for crystallography were identified using the amino acid sequences of human cGAS and *D. melanogaster* CG7194 to seed a position-specific iterative BLAST (PSI-BLAST) search of the NCBI non-redundant protein database. The PSI-BLAST search was performed with an E value cutoff 0.005 for inclusion into the next search round, BLOSUM62 scoring matrix, gap costs settings existence 11 and extension 1, and using conditional compositional score matrix adjustment. Candidate homologs identified from this search included the uncharacterized human protein MB21D2 and the *Tribolium castaneum* sequence XP_969398.1. Pairwise structural comparison between human MB21D2, *Tc*-cGLR, and protein structures in the Protein Data Bank was performed using DALI[37] and Z-scores for homologs less than 90% identical to one another (PDB90) were plotted in GraphPad Prism. A Z-score of 15 for *Tc*-cGLR and 13 for human MB21D2 was selected as a lower cut-off to emphasize direct relevant homologs in analysis.

Following structure determination of human MB21D2 and *T. castaneum* XP_969398.1, predicted cGLRs were further identified in *Diptera* using PSI-BLAST searches seeded with either *D. melanogaster* CG7194 or the *Tc*-cGLR sequence selecting in each round for proteins matching known cGLR domain organization and active-site residues. *Diptera* cGLR sequences were aligned using MAFFT (FFT-NS-i iterative refinement method)[38] and used to construct a phylogenetic tree in Geneious Prime v2020.12.23 using the neighbor-joining method and Jukes-Cantor genetic distance model with no outgroup. Further manual analysis and curation of candidate cGLR sequences was performed based on alignments and predictive structural homology using HHPred[39] and Phyre2[40]. Sequences were selected for predicted structural homology to cGAS, including the presence of a conserved nucleotidyltransferase domain with a G[S/G] activation loop and [E/D]h[E/D] $X_{50–90}$ [E/D] catalytic triad. Manual refinement was also used to exclude duplicate sequences, gene isoforms, and proteins less than 250 or greater than

700 residues. NCBI available genomes from 42 species in *Diptera* are represented in the final tree, including 31 species in the genus *Drosophila*. Clustering of sequences in the final unrooted tree was used to define clades, with no more than 30% sequence identity shared between members of different clades. Further manual analysis of the tree was used to determine the number and distribution of predicted cGLRs by species (see Extended Data Fig. 2a). PROMALS3D[41] used for structure guided alignment of apo human cGAS (PDB 4KM5)[14], hMB21D2, and *Tc*-cGLR in Extended Data Fig. 1a. MAFFT (FFT-NS-i iterative refinement method)[38] was used to align STING sequences in Extended Data Fig. 8a. Geneious Prime software was used to generate the sequence alignments in Figure 4 and Extended Data Figs. 1a, 3a and 8a.

Protein expression and purification

Recombinant cGLR and dSTING proteins were expressed and purified used methods previously optimized for human cGAS[17]. Animal cGLR and dSTING sequences were codon-optimized for expression in *E. coli* and cloned from synthetic constructs (GeneArt or Integrated DNA Technologies) into a custom pET16 expression vector with an N-terminal 6×His-MBP fusion tag or an N-terminal 6×His-SUMO2 fusion. The full-length coding sequence was used except for human cGAS 157–522, mouse cGAS 147–607, human MB21D2 S29–F491, *Ds*-cGLR1 19–393, and *D. eugracilis* STING 150–340 as specified. The N-terminus of *D. eugracilis* STING 150–340 was fused to the full-length coding sequence of T4 lysozyme connected by a Gly-Ser linker sequence. Briefly, transformed BL21-CodonPlus(DE3)-RIL *E. coli* (Agilent) were grown in MDG media overnight prior to inoculation of M9ZB media at an OD_{600} of 0.0475. M9ZB cultures were grown to OD_{600} of 2.5 (approximately 5 h at 37°C with shaking at 230 rpm) followed by cooling on ice for 20 min. Cultures were induced with 500 μM IPTG prior to incubation at 16°C overnight with shaking at 230 rpm. Cultures were pelleted the following day and either flash frozen in liquid nitrogen for storage at −80°C or directly lysed for purification. Selenomethionine-substituted

462 proteins for crystallography experiments were purified using a modified growth protocol as
463 previously described[30].

464 For large-scale protein purification, proteins were expressed with a 6×His-SUMO2 (Tc-
465 cGLR, *Ds*-cGLR1, *Deu*-cGLR, *Lc*-cGLR, dSTINGs) or 6×His-MBP (*Dm*-cGLR1, *Der*-cGLR1)
466 fusion tag and grown as ~4–8× 1 L cultures in M9ZB media. Pellets were lysed by sonication in
467 lysis buffer (20 mM HEPES pH 7.5, 400 mM NaCl, 30 mM imidazole, 10% glycerol and 1 mM
468 DTT) and clarified by centrifugation at ~47,850 × g for 30 min at 4°C and subsequent filtration
469 through glass wool. Recombinant protein was purified by gravity-flow over NiNTA resin (Qiagen).
470 Resin was washed with lysis buffer supplemented to 1 M NaCl and then eluted with 20 mL of lysis
471 buffer supplemented to 300 mM imidazole. SUMO2-fusion proteins were cleaved by
472 supplementing elution fractions with ~250 µg of human SENP2 protease (D364–L589 with M497A
473 mutation) during overnight dialysis at 4°C against dialysis buffer (20 mM HEPES pH 7.5, 250 mM
474 KCl, 1 mM DTT). MBP-tagged fusion proteins were buffer exchanged into lysis buffer with 4%
475 glycerol and no imidazole to optimize conditions for overnight cleavage by recombinant TEV
476 protease at ~10°C. cGLR proteins were next purified by ion exchange chromatography using
477 HiTrap Heparin HP columns (GE Healthcare) and eluted across a 150–1000 mM NaCl gradient.
478 Target protein fractions were pooled and further purified by size-exclusion chromatography using
479 a 16/600 Superdex 75 column or 16/600 Superdex 200 column and storage buffer (20 mM
480 HEPES-pH 7.5, 250 mM KCl, 1 mM TCEP). Final proteins were concentrated to ~20–30 mg mL^{-1}
481 and flash-frozen with liquid nitrogen and stored at −80°C for crystallography or supplemented with
482 10% glycerol prior to freezing for biochemistry experiments. *Tc*-cGLR and *Ds*-cGLR1 mutant
483 proteins were purified from 1 L M9ZB cultures using NiNTA affinity chromatography and overnight
484 dialysis directly into storage buffer (20 mM HEPES-pH 7.5, 250 mM KCl, 10% glycerol, 1 mM
485 TCEP) without SUMO2 tag cleavage.

486 For small-scale protein purification used in the *Diptera* cGLR screen, recombinant proteins
487 were expressed with a 6×His-MBP fusion tag with the exception of human cGAS, mouse cGAS,

Tc-cGLR, *Deu*-cGLR, *Lc*-cGLR, and *Ds*-cGLR1 which were expressed with a 6×His-SUMO2 fusion tag. Small-scale cultures were grown in 20 mL of M9ZB media, lysed with sonication, and recombinant protein was purified as previously described[11]. Briefly, protein was purified directly from lysates by centrifuge flow-through over NiNTA resin (Qiagen) in 2 mL Mini Spin columns (Epoch Life Sciences). Following elution with elution buffer (20 mM HEPES pH 7.5, 400 mM NaCl, 300 mM imidazole, 10% glycerol and 1 mM DTT) proteins were buffer exchanged into storage buffer (20 mM HEPES-pH 7.5, 250 mM KCl, 10% glycerol, 1 mM TCEP). Fresh protein preparations were immediately used for *in vitro* nucleotide synthesis reactions.

Protein crystallization and structure determination

Crystals of native and selenomethionine-substituted human MB21D2 S29–F491, *Tc*-cGLR, and T4 lysozyme-dSTING L150–I340, were grown at 18°C using hanging-drop vapor diffusion. Optimized crystals were grown in EasyXtal 15-well trays (NeXtal Biotechnologies) with 350 µL of reservoir solution and 2 µL drops set with a ratio of 1 µL protein solution and 1 µL of reservoir solution. Human MB21D2 crystals were grown using a reservoir solution (1.2 M ammonium sulfate, 5 mM $MgCl_2$, 100 mM MES pH 6.2) previously identified by Pei Wang and Raven Huang (University of Illinois at Urbana-Champaign)[42] for 1 day prior to cryoprotection with reservoir solution supplemented with 30% glycerol and freezing in liquid nitrogen. *Tc*-cGLR crystals were grown using the reservoir solution (0.3 M potassium thiocyanate, 10–16% PEG-3350) for 5–16 days prior to cryoprotection with reservoir solution supplemented with 15% ethylene glycol and freezing in liquid nitrogen. Apo T4 lysozyme-dSTING crystals were grown using the reservoir solution (0.2 M sodium citrate, 0.1 M Tris-HCl, 22% PEG-3350) 7 days prior to cryoprotection with reservoir solution supplemented with 15% ethylene and freezing in liquid nitrogen. T4 lysozyme-dSTING–3′2′-cGAMP crystals were grown using the reservoir solution (0.1–0.2 M sodium acetate pH 4.8, 0.2 M ammonium formate, 20–22% PEG-3350) supplemented with 250 µM 3′2′-cGAMP (Biolog) for 10 days prior to cryoprotection with reservoir solution supplemented to 35% PEG-

3350 and freezing in liquid nitrogen. Growth of single MB21D2 and *Tc*-cGLR crystals was further optimized with streak seeding. X-ray diffraction data were collected at the Advanced Photon Source beamlines 24-ID-C and 24-ID-E and at the Advanced Light Source beamlines 5.0.1 and 8.2.2. Data were processed with XDS and Aimless[43] using the SSRL autoxds script (*A. Gonzales*). Experimental phase information for all proteins was determined using data collected from selenomethionine-substituted crystals. Anomalous sites were identified, and an initial map generated with AutoSol within PHENIX[44]. Structural modeling was completed in Coot[45] and refined with PHENIX. Refinement statistics are described in Supplementary Table 1, and all structure figures were generated with PyMOL 2.3.0.

Nucleotide product synthesis analysis

cGLR nucleotide synthesis activity was analyzed by thin-layer chromatography as previously described[11]. For the *Diptera* cGLR screen, recombinant protein preparations were incubated in 10 µL reactions containing 0.5 µL α-^{32}P labeled NTPs (~0.4 µCi each of ATP, CTP, GTP, and UTP), 200 µM unlabeled NTPs, 10 mM $MgCl_2$, and 1 mM $MnCl_2$ in a final reaction buffer of 50 mM Tris-HCl pH 7.5, ~50 mM KCl, 1 mM TCEP. Reactions were additionally supplemented with ~1 µg poly I:C or 5 µM ISD45 dsDNA as indicated. Reactions were incubated at 37°C overnight and subsequently treated with 1 µL Quick CIP phosphatase (New England Biolabs) for 20 min at 37°C to remove unreacted phosphate signal. Each reaction was diluted 1:10 in 100 mM sodium acetate pH 5.2 and 0.5 µL was spotted on a 20-cm × 20-cm PEI-cellulose thin-layer chromatography plate. Plates were run with 1.5 M KH_2PO_4 solvent until ~2.5 cm from top of the plate, dried at room-temperature, and exposed to a phosphor-screen prior to signal detection with a Typhoon Trio Variable Mode Imager System (GE Healthcare). For all other nucleotide synthesis reactions visualized by thin-layer chromatography, enzymes were tested at 5 µM with 5 µM nucleic acid ligands and either 1 mM $MnCl_2$ or 10 mM $MgCl_2$ for insect cGLRs or cGAS respectively. hMB21D2 activity was tested with 1 mM $MnCl_2$ and 10 mM $MgCl_2$ using the following

synthetic innate immune agonists: lipopeptide Pam3CSK4 (Invivogen), *S. aureus* lipoteichoic acid (LTA-SA) (Invivogen), *S. cerevisiae* cell wall preparation (Zymosan) (Invivogen), *B. subtilis* peptidoglycan (PGN-BS) (Invivogen), synthetic lipid A mimic (CRX-527) (Invivogen), *B. subtilis* flagellin (FLA-BS) (Invivogen), imidazoquinoline (Imiquimod) (Invivogen), CpG oligonucleotide (ODN 2006) (Invivogen), *S. aureus* 23S rRNA oligonucleotide (ORN Sa19) (Invivogen). Besides *Diptera* screen reactions, samples were not diluted in sodium acetate prior to spotting on PEI-cellulose TLC plates. TLC images were adjusted for contrast using FIJI[46] and quantified using ImageQuant (8.2.0). Nucleotide product formation was measured according to the ratio of product to total signal for each reaction. For Figs. 1c, 2d, 2g and Extended Data Figs. 3e, 4b, 5d relative activity was calculated as percent conversion for each reaction relative to maximal conversion observed by wildtype enzyme or in the presence of 40 bp dsRNA for insect cGLRs and 45 bp dsRNA for cGAS.

Electrophoretic mobility shift assay (EMSA)

Analysis of *in vitro* protein–nucleic acid complex formation was conducted as previously described[17]. Briefly, 1 μM 40-bp dsRNA or 45-bp dsDNA was incubated with *Ds*-cGLR1 or hcGAS NTase-domain (D157–522) respectively at a gradient of protein concentrations as indicated in each figure. Complex formation was performed with the final reaction buffer 20 mM HEPES-NaOH pH 7.8, 75 mM KCl, 1 mM DTT. 20 μL reactions were incubated at 4°C for 20 min before separation on a 2% agarose gel using 0.5 × TB buffer (45 mM Tris, 45 mM boric acid) as a running buffer. The agarose gel was post-stained in 0.5 × TB buffer supplemented with 10 μg mL^{-1} ethidium bromide with gentle shaking at 25°C for 45 min. Complex formation was visualized using a ChemiDoc MP Imaging System (Bio-Rad).

***In vitro* phase separation assays**

In vitro phase separation was performed as previously described with minor modifications[23,25]. Briefly, *Ds*-cGLR1 was labeled with AlexaFluor-488 (AF488) carboxylic acid (succinimidyl ester) (Thermo Fisher Scientific) according to manufacturer's manuals using a molar ratio of 1:10 at 4°C for 4 h. Excess free dye was removed by dialysis against buffer (20 mM HEPES-KOH pH 7.5, 250 mM KCl, 1 mM DTT) at 4°C overnight and AF488-labeled *Ds*-cGLR1 was then further purified on a PD-10 desalting column (GE Healthcare) eluted with storage buffer (20 mM HEPES-KOH pH 7.5, 250 mM KCl, 1 mM TCEP). Final AF488-labeled *Ds*-cGLR1 was concentrated to ~5 mg mL^{-1}, flash-frozen in liquid nitrogen, and stored as aliquots at −80°C. hcGAS and hcGAS NTase-domain (D157–F522) proteins were prepared as previously described[25].

To induce phase separation, *Ds*-cGLR1 (10 μM, containing 1 μM AF488-labeled *Ds*-cGLR1) was mixed with various lengths of RNA (10 μM each) in buffer (20 mM Tris-HCl pH 7.5, 1 mg mL^{-1} BSA, 1 mM TCEP) in the presence of various salt concentrations at 25°C in a total reaction volume of 20 μL. The details of proteins, nucleic acids, and salt concentrations are indicated in figures. *Ds*-cGLR1–RNA reactions were placed in 384-well non-binding microplates (Greiner Bio-One) and incubated at 25°C for 30 min prior to imaging to allow condensates to settle. Fluorescence microscopy images were acquired at 25°C using a Leica TCS SP5 X (Leica Microsystems) mounted on an inverted microscope (DMI6000; Leica Microsystems) with an oil immersion 63×/numerical aperture 1.4 objective lens (HCX PL APO; Leica Microsystems). AF488-labeled *Ds*-cGLR1, hcGAS and hcGAS NTase-domain proteins were detected with excitation at 488 nm (emission at 500–530 nm). Microscopy images were processed with FIJI[46], and contrast adjusted with a uniform threshold setup for each enzyme.

Cellular STING signaling assays

Human HEK293T cells were maintained in complete media (DMEM supplemented with penicillin, streptomycin, and 10% fetal bovine serum) at 37°C. For all assays 4.5×10^4 cells were plated in 96-well plates. STING and cGLR activity assays were performed using Dual-Luciferase Reporter

Assay System (Promega) as previously described[14], with modifications. Lipofectamine-2000 was used to transfect IFNβ-firefly luciferase and TK-Renilla luciferase reporters and 5 ng of pcDNA4-mouse STING or 15 ng pcDNA4-dSTING. For cGLR signaling assays 150 ng of *Drosophila* cGLR1, 30 ng human cGAS balanced with empty vector, or 150 ng empty vector were additionally transfected. Native cGLR and STING coding sequences were expressed from a pcDNA4 vector. 24–30 h after transfection luciferase was measured using a GloMax microplate reader (Promega) and relative IFNβ expression calculated by normalizing firefly to Renilla readings. For poly I:C stimulation of cGLR activity, cells were transfected with 100 ng poly I:C (6.125–200 ng for titration experiment) 5 h after plasmid transfection. For dSTING signaling assays a final concentration of 500 pM to 50 μM 2′3-cGAMP or 3′2′-cGAMP was delivered to cells using a digitonin permeabilization buffer[47] 10 h prior to luciferase measurement.

Nucleotide purification and HPLC analysis

Enzymatic synthesis of cGLR nucleotide products for HPLC analysis was performed using 100 μL reactions containing 10 μM cGLR enzyme, 200 μM ATP, 200 μM GTP, 10 μg poly I:C, 1 mM $MnCl_2$ and 50 mM Tris-HCl pH 7.5. Protein storage buffer (20 mM HEPES pH 7.5, 250 mM KCl, 1 mM TCEP) was used as necessary to adjust KCl concentration to ~100 mM. Reactions were incubated at 37°C for 1 h and then nucleotide product was recovered by filtering reactions through a 30-kDa cutoff concentrator (Amicon) to remove protein. Nucleotide products were separated on an Agilent 1200 Infinity Series LC system using a C18 column (Zorbax Bonus-RP 4.6 × 150 mm, 3.5 μm) at 40°C. Products were eluted at a flow rate of 1 mL min^{-1} with a buffer of 50 mM NaH_2PO_4 pH 6.8 supplemented with 3% acetonitrile.

To purify the *Deu*-cGLR product for mass-spectrometry analysis, nucleotide synthesis reaction conditions were scaled as previously described for bacterial cGAS/DncV-like Nucleotidyltransferase reactions[11,48]. Briefly, a 10 mL reaction containing 528 nM *Deu*-cGLR enzyme, 125 μM ATP, 125 μM GTP, ~250 μg poly I:C, 1 mM $MnCl_2$, 50 mM Tris-HCl 7.5, and ~25

mM KCl was incubated with gentle rotation for 36 h at 37°C follow by Quick CIP (NEB) treatment for 6 h. The reaction was monitored using a 20 µL aliquot supplemented with α-^{32}P labeled NTPs and to visualize product formation by thin-layer chromatography. Following incubation, the large-scale reaction was filtered through a 10-kDa concentrator (Amicon) and purified by anion exchange chromatography using a 1 mL Q-sepharose column washed with water and eluting with a 0–2 M ammonium acetate gradient. Fractions corresponding to main product 3′2′-cGAMP were differentiated from fractions corresponding to 2′3′-c-di-AMP by HPLC analysis. Product fractions were further purified by size-exclusion chromatography using a Superdex 30 Increase 10/300 GL with dH_2O as a running buffer. Peak fractions were eluted in 1 mL volumes, pooled, and evaporated for storage prior to mass-spectrometry analysis.

Nucleotide mass spectrometry analysis and 3′2′-cGAMP identification

Purified nucleotide product samples were evaporated at 40°C under a gentle nitrogen stream. The residual pellet was resuspended in 200 µL HPLC grade water (J.T. Baker), and 40 µL was then mixed with 40 µL of water containing 50 ng mL^{-1} tenofovir as internal standard and transferred to measuring vials.

Experiments for 3′2′-cGAMP identification were performed on an ACQUITY UPLC I-Class/Vion IMS-QTOF high-resolution LC-MS system (Waters Corporation). Reverse phase chromatographic separation was carried out at 30°C on a C18 column (Nucleodur Pyramid C18 50 x 3 mm; 3 µm Macherey Nagel, Düren, Germany) connected to a C18 security guard (Phenomenex, Aschaffenburg, Germany) and a 2 µm column saver. Separation was achieved using a binary gradient of water containing 10 mM ammonium acetate and 0.1% acetic acid (solvent A) and methanol (solvent B). The analytes were eluted at a flow rate of 0.6 mL min^{-1}. The eluting program was as follows: 0 to 4 min: 0% B, 4 to 7.3 min: 0 to 10% B. This composition of 10% B was held for 1 min, then the organic content was increased to 30% within 2.7 min. The column was then re-equilibrated to 0% B for 2 min. Total analysis run time was 13 min. High

643 resolution MS-Data were collected on a Vion IMS-QTOF mass spectrometer equipped with an
644 electrospray ionization source, operating in positive ionization mode. The capillary voltage was
645 set at 2.5 kV and the cone voltage at 40 V. The source temperature and desolvation gas
646 temperature was 150°C and 600°C, respectively. Analyte fragmentation was achieved using
647 argon as the collision gas. Collision energy of 10 V was used to obtain a low collision energy
648 spectrum. For high collision energy spectrum, the collision energy was ramped from 15 to 30 V.
649 Data acquisition was controlled by the UNIFI 1.9.4.0 software (Waters). For 3′2′-cGAMP
650 identification the retention times, drift times and fragment spectra of a synthetic 3′2′-cGAMP
651 standard (Biolog) were collected as a reference and compared to those of the suspected 3′2′-
652 cGAMP in the samples.
653
654 **3′2′-cGAMP quantification**
655 For quantification of 3′2′-cGAMP, chromatographic conditions were transferred to a API4000
656 mass spectrometer (Sciex) coupled to a Shimadzu HPLC-system (Shimadzu, Duisburg,
657 Germany). The analytes were ionized by means of electro spray ionization in positive mode
658 applying an ion spray voltage of 3000 V. Further ESI parameters were: curtain gas (CUR): 30 psi,
659 collision gas (CAD): 9, source temperature: 650°C, gas 1: 60 psi and gas 2: 45 psi, respectively.
660 Detection was performed in SRM mode, selecting first for the double-protonated parent ion of
661 3′2′-cGAMP and 3′3′-cGAMP (used in calibrator series). This resulted in the following mass
662 transitions: 3′2′-cGAMP and 3′3′-cGAMP: m/z 338.2 → 152 (quantifier), m/z 338.2 → 136
663 (identifier). Tenefovir served as the internal standard (m/z 288 → 176).
664 For 3′2′-cGAMP semi-quantitative quantification from lysate samples in the *Diptera* cGLR
665 screen, calibration curves were created by plotting peak area ratios of 3′3′-cGAMP as an internal
666 standard versus the nominal concentration of the calibrators. The calibration curve was calculated
667 using quadratic regression and 1/x weighing.
668

Synthetic cyclic dinucleotide standards

Synthetic nucleotide standards used for HPLC analysis and mass-spectrometry analysis were purchased from Biolog Life Science Institute: 3′3′-cGAMP (cat no. C 117), 2′3′-cGAMP (cat no. C 161), 3′2′-cGAMP (cat no. C 238), 2′3′-c-di-AMP (cat no. C 187) and 2′3′-c-di-GMP (cat no. C 182).

Nuclease P1 and poxin cleavage analysis

Nuclease P1 cleavage analysis was performed using *Dm*-cGLR1 reactions labeled with either α-^{32}P-ATP or α-^{32}P-GTP as previously described[11,26]. Briefly, radiolabeled nucleotide products were incubated with Nuclease P1 (80 mU, Sigma N8630) in 1× P1 buffer (30 mM NaOAc pH 5.3, 5 mM $ZnSO_4$, 50 mM NaCl) for 30 min in the presence of Quick CIP (NEB).

Poxin cleavage reactions were carried out using purified insect viral AcNPV enzyme as previously described[29,30]. For HPLC analysis of poxin cleavage, 100 µL reactions were performed using 100 µM synthetic 2′3′-cGAMP or 3′2′-cGAMP, 50 nM AcNPV poxin, 50 mM HEPES pH 7.5, 10 mM KCl, and 1 mM TCEP. Reactions were incubated at 37°C and at each specified time reactions were terminated by heat-inactivation at 95°C for 2 min prior to HPLC analysis as described above. For thin-layer chromatography analysis of poxin cleavage, reactions were performed using α-^{32}P-GTP-labeled 2′3′-cGAMP synthesized by mcGAS or 3′2′-cGAMP synthesized by *Deu*-cGLR in 5 µL reactions containing 2.5 µM nucleotide product and 1 µM AcNPV poxin, 50 mM HEPES pH 7.5, 10 mM KCl, and 1 mM TCEP. Reactions were incubated at 37°C and at each specified time reactions were terminated by heat-inactivation at 80°C for 5 min prior to PEI-cellulose thin-layer chromatography analysis as described above.

STING CDN thermal shift assay

A final concentration of 15 µM dSTING was mixed with 3× SYPRO orange dye and 100 µM synthetic CDN (Biolog) (or as described in figure) in 20 mM HEPES-KOH pH 7.5 and 100 mM

KCl. Samples were heated from 20–95°C in a BioRad CFX thermocycler with HEX channel fluorescence measurements every 0.5°C. The derivative of each curve over time was calculated using GraphPad Prism and graphed as a percent maximum change in fluorescence or used to calculate the melting temperature.

D. melanogaster cyclic dinucleotide injection and signaling analysis

Fly stocks were raised on standard cornmeal agar medium at 25°C. All fly lines used in this study were *Wolbachia* free. w^{1118}, $dSTING^{Control}$, and $dSTING^{Rxn}$ stocks have been described previously[31,34]. $Relish^{E20}$ flies isogenized to the DrosDel w^{1118} isogenic background were a kind gift from Luis Teixeira (Instituto Gulbenkian de Ciência)[49]. Cyclic dinucleotides including 3′2′-cGAMP (Biolog), 2′3′-cGAMP (Invivogen) and 3′3′-c-di-GMP (Invivogen) were dissolved in 10 mM Tris-HCl pH 7.5 and diluted to the indicated concentrations. 3–5 days old adult flies were injected with 69 nL of cyclic dinucleotide solution or 10 mM Tris-HCl pH 7.5 (negative control) by intrathoracic injection using a Nanoject II apparatus (Drummond Scientific). Flies were collected 24 h later in pools of 6 individuals (3 males and 3 females) and homogenized for RNA extraction and RT-qPCR analysis, as described[34].

D. melanogaster viral challenge assays

For 3′2′-cGAMP and virus co-injection, flies were injected with 69 nL of virus (DCV: 5 PFU, VSV: 2000 PFU) in 10 mM Tris-HCl pH 7.5 or in a 0.9 mg mL^{-1} 3′2′-cGAMP solution. For titration experiments comparing cGAMP isomers, 69 nL of DCV (5 pfu) in serial diluted concentrations of 2′3′-cGAMP or 3′2′-cGAMP were injected in the body cavity of the flies. Survival was monitored daily and flies were collected in pools of 6 individuals (3 males and 3 females) at the indicated time points to monitor the viral RNA load by RT-qPCR.

Extended References

3 Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B. L., Zillinger, T., Serganov, A. A., Liu, Y., Jones, R. A., Hartmann, G., Tuschl, T. & Patel, D. J. Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* (2013) **153**, 1094-1107. PMC4382009.

6 Kranzusch, P. J. cGAS and CD-NTase enzymes: structure, mechanism, and evolution. *Curr Opin Struct Biol* (2019) **59**, 178-187. PMC7127440.

7 Wu, J. & Chen, Z. J. Innate immune sensing and signaling of cytosolic nucleic acids. *Annu Rev Immunol* (2014) **32**, 461-488

8 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* (2015) **43**, 10633-10654. PMC4678834.

9 Wu, X., Wu, F. H., Wang, X., Wang, L., Siedow, J. N., Zhang, W. & Pei, Z. M. Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic Acids Res* (2014) **42**, 8243-8257. PMC4117786.

10 Kranzusch, P. J., Wilson, S. C., Lee, A. S., Berger, J. M., Doudna, J. A. & Vance, R. E. Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2',3' cGAMP Signaling. *Mol Cell* (2015) **59**, 891-903. PMC4575873.

11 Whiteley, A. T., Eaglesham, J. B., de Oliveira Mann, C. C., Morehouse, B. R., Lowey, B., Nieminen, E. A., Danilchanka, O., King, D. S., Lee, A. S. Y., Mekalanos, J. J. & Kranzusch, P. J. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* (2019) **567**, 194-199. PMC6544370.

12 de Oliveira Mann, C. C., Kiefersauer, R., Witte, G. & Hopfner, K. P. Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein MAB21L1. *Sci Rep* (2016) **6**, 27498. PMC4897736.

13 Civril, F., Deimling, T., de Oliveira Mann, C. C., Ablasser, A., Moldt, M., Witte, G., Hornung, V. & Hopfner, K. P. Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* (2013) **498**, 332-337. PMC3768140.

14 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* (2013) **3**, 1362-1368. PMC3800681.

17 Zhou, W., Whiteley, A. T., de Oliveira Mann, C. C., Morehouse, B. R., Nowak, R. P., Fischer, E. S., Gray, N. S., Mekalanos, J. J. & Kranzusch, P. J. Structure of the Human cGAS-DNA Complex Reveals Enhanced Control of Immune Surveillance. *Cell* (2018) **174**, 300-311 e311. PMC6084792.

23 Du, M. & Chen, Z. J. DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science* (2018) **361**, 704-709

25 Zhou, W., Mohr, L., Maciejowski, J. & Kranzusch, P. J. cGAS phase separation inhibits TREX1-mediated DNA degradation and enhances cytosolic DNA sensing. *Mol Cell* (2021) **81**, DOI 10.1016/j.molcel.2021.1001.1024.

26 Kranzusch, P. J., Lee, A. S. Y., Wilson, S. C., Solovykh, M. S., Vance, R. E., Berger, J. M. & Doudna, J. A. Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. *Cell* (2014) **158**, 1011-1021. PMC4157622.

28 Morehouse, B. R., Govande, A. A., Millman, A., Keszei, A. F. A., Lowey, B., Ofir, G., Shao, S., Sorek, R. & Kranzusch, P. J. STING cyclic dinucleotide sensing originated in bacteria. *Nature* (2020) **586**, 429-433. PMC7572726.

29 Eaglesham, J. B., Pan, Y., Kupper, T. S. & Kranzusch, P. J. Viral and metazoan poxins are cGAMP-specific nucleases that restrict cGAS-STING signalling. *Nature* (2019) **566**, 259-263. PMC6640140.

30 Eaglesham, J. B., McCarty, K. L. & Kranzusch, P. J. Structures of diverse poxin cGAMP nucleases reveal a widespread role for cGAS-STING evasion in host-pathogen conflict. *Elife* (2020) **9**. PMC7688311.

31 Goto, A., Okado, K., Martins, N., Cai, H., Barbier, V., Lamiable, O., Troxler, L., Santiago, E., Kuhn, L., Paik, D., Silverman, N., Holleufer, A., Hartmann, R., Liu, J., Peng, T., Hoffmann, J. A., Meignin, C., Daeffler, L. & Imler, J. L. The Kinase IKKbeta Regulates a STING- and NF-kappaB-Dependent Antiviral Response Pathway in Drosophila. *Immunity* (2018) **49**, 225-234 e224. PMC6267954.

34 Cai, H., Holleufer, A., Simonsen, B., Schneider, J., Lemoine, A., Gad, H. H., Huang, J., Huang, J., Chen, D., Peng, T., Marques, J. T., Hartmann, R., Martins, N. E. & Imler, J. L. 2'3'-cGAMP triggers a STING- and NF-kappaB-dependent broad antiviral response in Drosophila. *Sci Signal* (2020) **13**

37 Holm, L. DALI and the persistence of protein shape. *Protein Sci* (2020) **29**, 128-140. PMC6933842.

38 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* (2013) **30**, 772-780. PMC3603318.

39 Gabler, F., Nam, S. Z., Till, S., Mirdita, M., Steinegger, M., Soding, J., Lupas, A. N. & Alva, V. Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr Protoc Bioinformatics* (2020) **72**, e108

40 Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* (2015) **10**, 845-858. PMC5298202.

41 Pei, J. & Grishin, N. V. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol Biol* (2014) **1079**, 263-271. PMC4506754.

42 Wang, P. Discovery and characterization of novel RNA repair systems. (2015), http://hdl.handle.net/2142/78582

43 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* (2010) **66**, 125-132. PMC2815665.

44 Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* (2019) **75**, 861-877. PMC6778852.
45 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* (2004) **60**, 2126-2132
46 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P. & Cardona, A. Fiji: an open-source platform for biological-image analysis. *Nat Methods* (2012) **9**, 676-682. PMC3855844.
47 Woodward, J. J., Iavarone, A. T. & Portnoy, D. A. c-di-AMP secreted by intracellular Listeria monocytogenes activates a host type I interferon response. *Science* (2010) **328**, 1703-1705. PMC3156580.
48 Lowey, B., Whiteley, A. T., Keszei, A. F. A., Morehouse, B. R., Mathews, I. T., Antine, S. P., Cabrera, V. J., Kashin, D., Niemann, P., Jain, M., Schwede, F., Mekalanos, J. J., Shao, S., Lee, A. S. Y. & Kranzusch, P. J. CBASS Immunity Uses CARF-Related Effectors to Sense 3'-5'- and 2'-5'-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. *Cell* (2020) **182**, 38-49 e17. PMC7728545.
49 Ferreira, A. G., Naylor, H., Esteves, S. S., Pais, I. S., Martins, N. E. & Teixeira, L. The Toll-Dorsal Pathway Is Required for Resistance to Viral Oral Infection in Drosophila. *PLoS pathogens* (2014) **10**
50 Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G. & Sorek, R. Cyclic GMP-AMP signalling protects bacteria against viral infection. *Nature* (2019) **574**, 691-695
51 Zhao, Z., Ma, Z., Wang, B., Guan, Y., Su, X. D. & Jiang, Z. Mn(2+) Directly Activates cGAS and Structural Analysis Suggests Mn(2+) Induces a Noncanonical Catalytic Synthesis of 2'3'-cGAMP. *Cell Rep* (2020) **32**, 108053

Data Availability Statement: Coordinates and structure factors of human MB21D2, *T. castaneum* cGLR, *Drosophila* STING, and the *Drosophila* STING–3′2′-cGAMP complex have been deposited in PDB under the accession codes 7LT1, 7LT2, 7MWY, and 7MWZ. All other data are available in the manuscript or the supplementary materials.

Extended Data Figure Legends

Extended Data Figure 1 | Sequence and structural analysis of hMB21D2 and *Tc*-cGLR.

a, Structure guided sequence alignment of the catalytic domain of human cGAS (PDB 4KM5), human MB21D2, and *Tc*-cGLR. Strict secondary structure conservation further supports conserved structural homology despite primary sequence divergence. The [D/E]hD[$X_{50–90}$]D catalytic triad is highlighted with red outline and the human Zn-ribbon insertion that is absent in other cGLRs is denoted with magenta line. hMB21D2 contains an additional 61 residues which are not resolved in the crystal structure and are absent from the alignment. **b,c,** Zoom-in cutaways of the human MB21D2 and *Tc*-cGLR crystal structures highlighting positioning of conserved catalytic residues in the nucleotidyltransferase active site. In human cGAS the analogous residues coordinate two Mg^{2+} metal ions to control synthesis of 2′3′-cGAMP (inset, middle; PDB 6CTA).

The hMB21D2 structure is in an inactive state distinguished by misaligned catalytic residues and occlusion by an extended Gly-Gly-activation loop, indicating that catalytic activation is likely controlled by a conformational rearrangement. **d, e,** TLC analysis of *in vitro* tests for potential activating ligands of hMB21D2. No nucleotide products were identified upon stimulation with 40 nt or bp nucleic acid ligands (**d**) or ligands known to activate mammalian Toll-like receptors (**e**) **f,** Z-score structural similarity plot showing homology between human MB21D2 and *Tc*-cGLR with representative structures in the Protein Data Bank (PDB90). Increasing Z-score indicates greater homology confirming the close relationship between animal cGLR enzymes and more distantly related similarity with cGAS/DncV-like Nucleotidyltransferases (CD-NTases) in bacterial anti-phage defense systems[11,50]. Z-scores cutoffs are 13 and 15 for hMB21D2 and *Tc*-cGLR respectively.

Extended Data Figure 2 | Forward biochemical screen of predicted cGLRs in Diptera. **a,** Violin plot showing the number of predicted cGLRs in *Diptera* genomes. *Drosophila* genomes (n = 31 species) have a median of four predicted cGLRs in contrast to a median of two predicted cGLRs in other Dipteran insects (n = 11 species). **b,** Schematic of *in vitro* screen of predicted cGLRs in the order *Diptera*. 53 sequences were selected representing each clade in the phylogeny in **Fig. 2a**. Following recombinant protein expression in *E. coli*, lysates were split into two samples for parallel TLC analysis of *in vitro* enzymatic activity and HPLC-MS analysis of lysate nucleotide metabolites. **c, d,** Purified cGLR proteins were incubated overnight at 37°C with α^{32}P-radiolabeled nucleotides, a mixture of Mn^{2+} and Mg^{2+}, and the 45 bp immunostimulatory DNA ISD45 or the synthetic dsRNA analog poly (I:C) as potential nucleic acid ligands, and reactions were visualized by PEI-cellulose TLC. Wild-type and catalytically inactive mouse cGAS enzymes were used as controls for each sample set. Note that mouse cGAS exhibits dsDNA-independent activity in the presence of Mn^{2+}[51]. Predicted *Diptera* cGLRs grouped by clade (DC01–05) and numbered within each clade. Ligand-dependent activity was identified for DC02_01, 05_03,

05_19, and 05_21; species listed below. We observed ligand-independent activity for two enzymes in Clade 3. Data represent n = 2 independent experiments. **e,** SDS-PAGE and Coomassie stain analysis of NiNTA purified cGLR protein fractions used for the biochemical screen. **f,** SDS-PAGE and Coomassie stain analysis of final NiNTA, ion-exchange, and size-exclusion purified cGLR proteins used for biochemical studies.

Extended Data Figure 3 | Sequence analysis and mutagenesis of insect cGLRs.

a, Alignment of the catalytic domain of human cGAS and active cGLRs identified in *T. castaneum*, *D. eugracilis*, *L. cuprina*, *D. erecta*, *D. simulans*, and *D. melanogaster*. The EhD[$X_{50–90}$]D catalytic triad is highlighted with red outline and the human Zn-ribbon insertion that is absent in insect cGLRs is denoted with a dashed red outline. cGLRs from *D. erecta* and *D. simulans* are close homologs of *Dm*-cGLR1 (76% and 91% sequence identity, respectively) and thus are also referred to as "cGLR1". All biochemical experiments with *Ds*-cGLR1 were performed with a construct beginning at M19. **b–c,** *In vitro* nucleotide synthesis reactions demonstrating effect of mutations to catalytic residues (**b**) or putative ligand binding groove (**c**) on insect cGLR enzymatic activity. Catalytic active-site mutations ablate nucleotide product synthesis and ligand-binding groove mutations that disrupt predicted RNA contacting residues significantly impair product synthesis. **d,** SDS-PAGE and Coomassie stain analysis of purified wild-type and mutant proteins, as labeled in above TLC images. **e,** *Above*, structure of *Tc*-cGLR1 modeled with dsRNA as shown in **Fig. 2f**, indicating putative ligand binding residues in *Tc*-cGLR selected analysis. *Below*, quantification of *in vitro* activity of *Tc*-cGLR in **c**, as displayed in **Fig. 2g** for *Ds*-cGLR1. Data are mean ± SEM, quantified relative to wild-type activity and represent n = 3 independent experiments. **f,** IFN-β luciferase assay in which cGLRs are expressed in human cells and CDN-synthesis is detected by mammalian STING activation. IFN-β quantified relative to empty vector control. In comparison to human cGAS control which is activated by expression vector-plasmid DNA, *Dm*-cGLR1 (left) and *Ds*-cGLR1 (right) require poly I:C stimulation to activate a downstream

STING response. Mutation to catalytic residues or putative ligand binding residues (as indicated in **Fig. 2f**) ablates cGLR activity. See also **Fig. 2h**. Data are mean ± SEM of n = 3 technical replicates and representative of n = 3 independent experiments. *P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001; and n.s., P > 0.05. P value n.s. unless otherwise noted.

Extended Data Figure 4 | Analysis of RNA-recognition by insect cGLRs.

a–c, *In vitro* activity assays for each active insect cGLR demonstrating dsRNA recognition is required for enzyme activation. Reactions were performed with synthetic 40 nt or bp ligands. Weak *Deu*-cGLR ssRNA-stimulated activity may be explained by transient short duplex formation similar to observations that some ssDNA oligos can stimulate mouse cGAS dsDNA-dependent activity[3]. **b,** TLC and quantification for enzyme activation in the presence of a panel of 10–40 bp synthetic dsRNA ligands. 30 bp dsRNA is sufficient to stimulate maximal activity for *Tc*-, *Dm*-, and *Lc*- cGLRs, while *Ds*-cGLR1 requires 35 bp and *Deu*-cGLR can be activated by dsRNAs as short as 15 bp. **c,** Reactions with 146 bp *in vitro* transcribed dsRNAs either retaining a 5′ triphosphate or 5′ OH termini demonstrate that dsRNA-recognition by insect cGLRs does not involve 5′ end discrimination. Data are mean ± SEM, quantified relative to maximum observed activity and represent n = 3 independent experiments. **d,** Deconvolution of catalytic metal requirements for enzymatic activity by insect cGLRs. Insect cGLRs require Mn^{2+} for maximal catalytic activity with weak product formation observed in the presence of Mg^{2+}. Data represent n = 3 independent experiments. **e,** Poly I:C titration demonstrates dsRNA-stimulation of *Drosophila* cGLR1 activity in cells is dependent on RNA concentration. IFN-β luciferase assay in which cGLRs are expressed in human cells and CDN-synthesis is measured by mammalian STING activation, as in **Fig. 2h** and **Extended Data Fig. 3f**. IFN-β quantified relative to empty vector control. Data are mean ± SEM of n = 3 technical replicates and representative of n = 3 independent experiments.

Extended Data Figure 5 | Characterization of *Ds*-cGLR1–dsRNA condensate formation.

a, Electrophoretic mobility shift assay (EMSA) showing binding between *Ds*-cGLR1 or the C-terminal nucleotidyltransferase domain of human cGAS (hcGAS-NTase) and a 40 bp dsRNA or 45 bp dsDNA. *Ds*-cGLR1 preferentially binds dsRNA and more weakly interacts with dsDNA, consistent with observed binding between human cGAS and dsRNA[13]. **b–c,** Analysis of *Ds*-cGLR1 and human cGAS (hcGAS) phase separation with AF488-labeled protein. Mammalian cGAS contains a highly disordered N-terminal extension of ~150 residues, but this unstructured extension is absent in insect cGLR sequences. In the presence of dsDNA, full-length hcGAS forms highly dynamic liquid droplets, whereas the minimal hcGAS NTase-domain forms rigid protein–DNA condensates similar to those formed by *Ds*-cGLR1–RNA complexes. Human cGAS exhibits a preference for condensate formation in the presence of dsDNA while *Ds*-cGLR1 exhibits a preference for dsRNA (scale bar = 10 μm) (see also **Fig. 2e**). **c,** Analysis of *Ds*-cGLR1 dsRNA length-specificity for condensate formation demonstrates clear length-dependency and supports that long dsRNA and condensate formation are required for maximal *Ds*-cGLR1 activation. **d,** Analysis of the impact of AF488-labeling on *Ds*-cGLR1 enzymatic activity. Similar to previous observations with hcGAS[25], AF488-labeling negatively impacts enzymatic activity but has minimal effect at the ratio of 90% unlabeled and 10% labeled protein used for all imaging experiments. Data represent n = 3 independent experiments, and are quantified in **c** as the mean ± SEM.

Extended Data Figure 6 | Synthesis of 3′2′-cGAMP by *Diptera* cGLRs.

a, HPLC analysis of the nucleotide products of *Tc*-cGLR, *Dm*-cGLR1, *Ds*-cGLR1, *Lc*-cGLR, and *Deu*-cGLR reactions compared to relevant synthetic controls. Integration of major and minor product peaks in n = 3 independent experiments were used to calculate relative product ratios shown in **Fig. 3d**. **b,** The *Drosophila* cGLR major reaction product was purified from *Deu*-cGLR reactions and compared to synthetic 3′2′-cGAMP with tandem mass-spectrometry analysis. Parent mass extracted ion trace (left) and tandem mass spectra comparison (right) validate the

chemical identity of the *Drosophila* cGLR product as 3′2′-cGAMP. **c,** Heat map showing the relative concentrations of cGAMP isomers detected by HPLC-MS in bacterial lysates expressing *Diptera* cGLRs (as described in **Extended Data Fig. 2b**). In all cases 3′2′-cGAMP was present as the dominant product with trace amounts of 3′3′- and 2′3′-cGAMP detected in some samples as minor species.

Extended Data Figure 7 | Mechanism of 3′2′-cGAMP bond formation and resistance to degradation by viral poxin enzymes.

a, Analysis of *Dm*-cGLR1 reactions with pairwise combinations of α-^{32}P-labeled nucleotides and non-hydrolyzable nucleotides reveals reaction intermediates and identifies the order of bond formation during 3′2′-cGAMP synthesis. *Left*, TLC analysis demonstrates *Dm*-cGLR1 forms a linear intermediate in the presence of GTP and non-hydrolyzable ATP (Apcpp) indicating the 2′–5′ phosphodiester bond is synthesized first. Exposed γ-phosphates removed by phosphatase treatment prior to analysis indicated by parentheses. Note that while a linear intermediate cannot be formed in the presence of non-hydrolyzable GTP (Gpcpp), *Dm*-cGLR1 will synthesize the off-product 2′3′-c-di-AMP. Mouse cGAS, which synthesizes 2′3′-cGAMP through the linear intermediate pppG[2′–5′]pA, is shown here for comparison[26]. *Right*, Schematic of the reaction mechanism for each enzyme. Data are representative of n = 3 independent experiments. **b,** Poxins are 2′3′-cGAMP-specific viral nucleases that disrupt cGAS-STING signaling. HPLC analysis of synthetic 2′3′-cGAMP or 3′2′-cGAMP treated with poxin from the insect baculovirus *Autographa californica* nucleopolyhedrovirus (AcNPV)[29,30]. In 1 min, AcNPV poxin cleaves 2′3′-cGAMP into a mixture of intermediate and full cleavage product; and after 1 h turnover is complete. No cleavage of 3′2′-cGAMP is observed by AcNPV poxin under these reaction conditions. **c,** Using TLC as a more sensitive assay, we observed minimal cleavage of 3′2′-cGAMP following overnight incubation with AcNPV poxin. **d,** Schematic highlighting how an isomeric switch in phosphodiester linkage specificity makes 3′2′-cGAMP remarkably resistant to

poxin-mediated cleavage.

Extended Data Figure 8 | Structural and biochemical analysis of *Drosophila* STING

a, Alignment of the C-terminal cyclic dinucleotide (CDN)-binding domains of human STING, mouse STING, *D. eugracilis* STING and *D. melanogaster* STING. Architecture of the core CDN-binding domain is conserved across metazoans; the disordered C-terminal tail which controls IRF3-IFNβ signaling is specific to vertebrates[10,28]. Ligand-interacting residues selected for mutational analysis annotated with a navy circle; *Diptera*-specific adaptations highlighted with red outline. All structural and biochemical experiments were performed with a *D. eugracilis* STING construct ending at I340. **b,** *In vitro* thermal denaturation assay analyzing dSTING interactions with a panel of CDNs. Only 3′2′-cGAMP forms a thermo-stable complex with dSTING (see also **Fig. 4a**). Data are mean ± SEM of the average T_m calculated from technical duplicates in n = 3 independent experiments. **c,** *In vitro* thermal denaturation assay demonstrating concentration-dependent thermal shift induced by 3′2′-cGAMP. **d,** Dose titration of 2′3′-cGAMP and 3′2′-cGAMP in human cells demonstrating selective response by dSTING to 3′2′-cGAMP. *D. eugracilis* CDN binding domain (CBD) was adapted for downstream signaling in human cells by addition of N-terminal human transmembrane (hTM) domains and human C-terminal tail (hCTT). **e**, Comparison of the human STING–2′3′-cGAMP and dSTING–3′2′-cGAMP crystal structures reveals a conserved closed homodimer architecture in which apical “wings” are spread 32–36 Å, demonstrating high-affinity engagement with an endogenous ligand. **f**, Enlarged cutaway of 3′2′-cGAMP in the dSTING crystal structure, shown as simulated annealing F_O-F_C omit map. **g**, Full crystal structure used to determine structure of *D. eugracilis* STING in complex with 3′2′-cGAMP. T4-lysozyme is fused to the N-terminus of the *D. eugracilis* STING CDN binding domain. **h**, Thermal denaturation assay as in **Fig. 4a** demonstrating that N-terminal fusion of T4 lysozyme does not impair dSTING recognition of 3′2′-cGAMP. **i**, Mutational analysis of key ligand-interacting residues in dSTING; thermal denaturation assay used to analyze 3′2′-cGAMP recognition.

Mutations which conserve functional contacts with 3′2′-cGAMP (Y164F) maintain ligand recognition; mutations which ablate contacts abrogate ligand binding. N159S exhibits diminished ability to recognize 3′2′-cGAMP. Data in **b** and **i** are mean ± SEM of the average T_m calculated from n=2 technical replicates in n = 3 independent experiments. Data in **c** are representative of n = 3 independent experiments. Data in **d** are mean ± SEM of n = 3 technical replicates and representative of n = 3 independent experiments. **j,** SDS-PAGE and Coomassie stain analysis of purified wild-type and mutant proteins.

Extended Data Figure 9 | 3′2′-cGAMP induces the expression of dSTING-regulated genes.

a–d, Injection of 3′2′-cGAMP into *D. melanogaster* has a dose-dependent effect on the expression of dSTING-regulated genes. 2′3′-cGAMP was used as positive control as previously characterized[31,34]. Synthetic nucleotide was injected into the body cavity of wildtype (w^{1118}) flies and gene expression was measured after 24 h. RNA levels were measured relative to the house-keeping gene *RpL32*, and nucleotide concentrations are displayed in μg μL^{-1}. Note that for *srg2* measurement after injection of 9E-7 μg μL^{-1} 3′2′-cGAMP there was one outlier replicate with a value of 0.5977 (data not shown, included in mean analysis). **e–k,** As in **Fig. 5a**, RNA expression analysis of STING-regulated genes 24 h after injection with synthetic 3′2′-cGAMP or 3′3′-c-di-GMP. RNA levels are shown as fold induction compared to buffer control in wildtype, *dSTING*, or *Relish* mutant flies respectively. $dSTING_{Mut}$ = RXN mutant; $Relish_{Mut}$ = $Relish^{E20}$ mutant, as previously characterized[31,34]. All data represent the mean ± SEM of n = 3 independent experiments, and each point represents a pool of 6 flies *P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001; and n.s., P > 0.05. P value n.s. unless otherwise noted.

Extended Data Figure 10 | 3′2′-cGAMP functions as a potent antiviral ligand

a, Analysis of Drosophila C virus (DCV) viral RNA load in flies injected with 3′2′-cGAMP or buffer control. *dSTING* wildtype and mutant flies were injected with 3′2′-cGAMP or buffer control and

then infected with DCV. Viral RNA levels were measured at each time as indicated relative to the house-keeping gene *RpL32.* DCV is a picornavirus-like (+)ssRNA virus in the family *Dicistroviridae*. **b,** Analysis of vesicular stomatitis virus (VSV) viral RNA load in flies injected with 3′2′-cGAMP or buffer control. *dSTING* wildtype and mutant flies were injected with 3′2′-cGAMP or buffer control as in **a** and then infected with VSV. Viral RNA levels were measured 4 days post infection relative to the house-keeping gene *RpL32.* VSV is a (-)ssRNA virus in the *Rhabdoviridae* family. **c,** (As in **a**) Analysis of Drosophila C virus (DCV) viral RNA load in flies injected with 3′2′-cGAMP, 2′3′-cGAMP, or buffer control. Viral RNA levels were measured one, two, or three days post-infection (dpi) relative to house-keeping gene *RpL32.* **d,** Survival curves after DCV infection showing effect of injection with dose titration of 3′2′-cGAMP or 2′3′-cGAMP compared to buffer control. Both cGAMP isomers significantly delay mortality in a dose-dependent manner; 3′2′-cGAMP provides greater protection in comparison to 2′3′-cGAMP. All data represent the mean ± SEM of n = 3 independent experiments, and each point represents a pool of 6 flies (**a**, **b**) or 10 flies (**c, d**). *P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001; and n.s., P > 0.05. P value n.s. unless otherwise noted.

Supplementary Table 1 | Summary of X-ray crystallography data collection, phasing and refinement statistics.

Supplementary Table 2 | cGLR sequence information.

Fig. 1 | Structural remodeling in animal cGLRs enables divergent pattern recognition

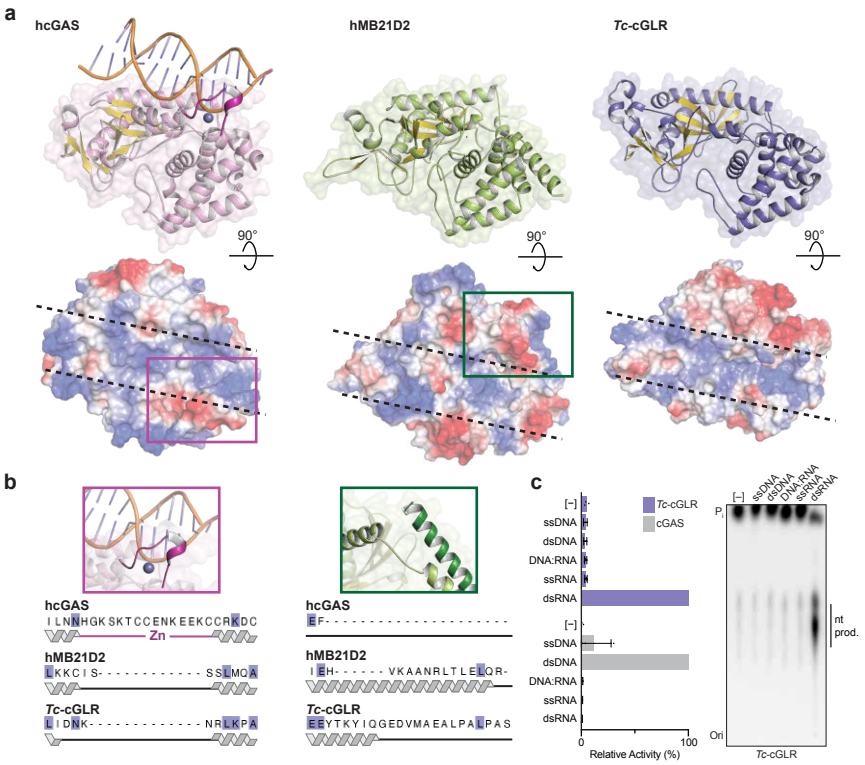


Fig. 2 | *Drosophila* cGLR1 senses long double-stranded RNA

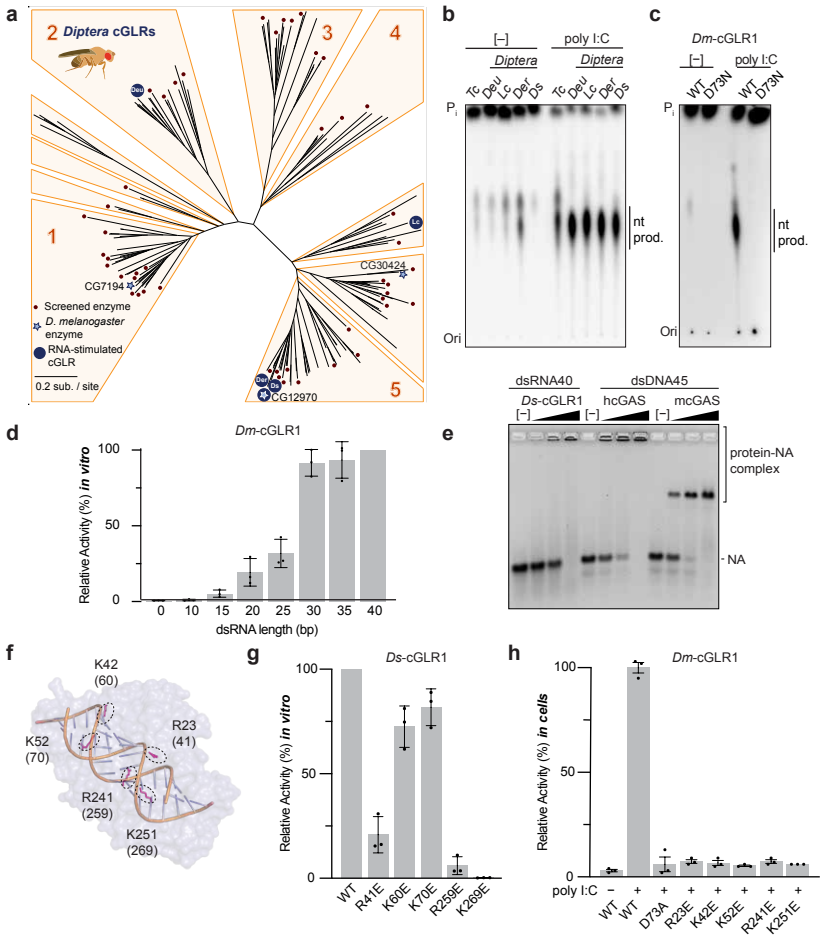


Fig. 3 | Discovery of 3′2′-cGAMP as a metazoan nucleotide second messenger

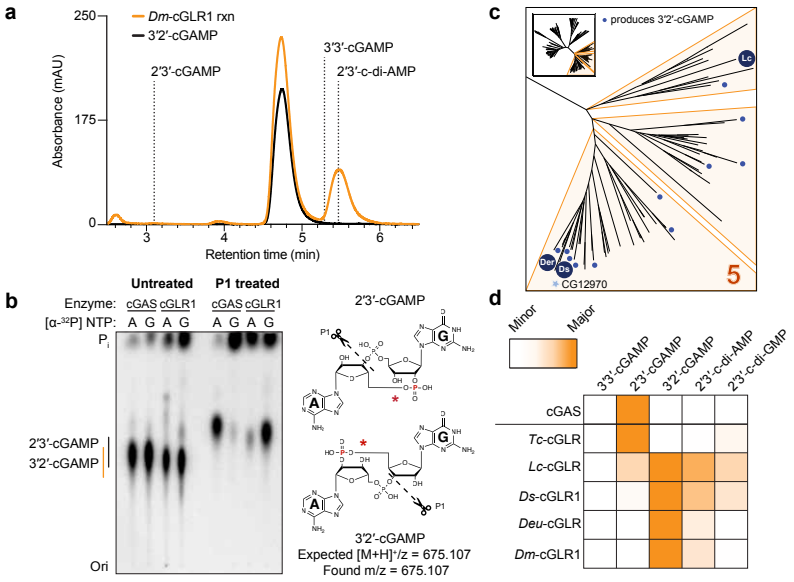


Fig. 4 | Structural basis for 3'2'-cGAMP recognition by *Drosophila* STING

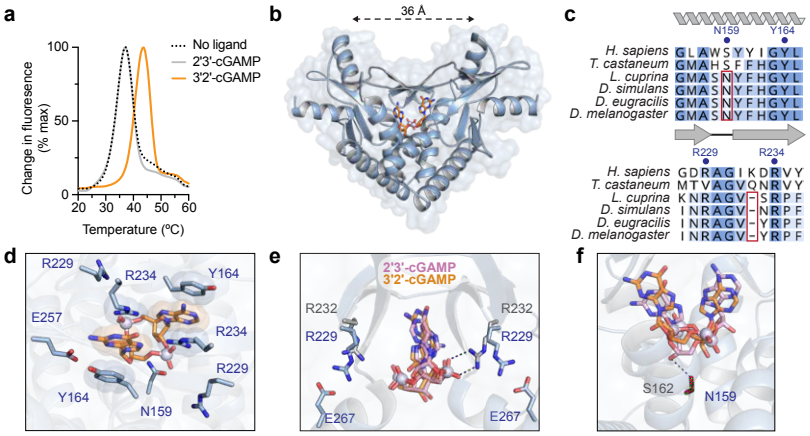
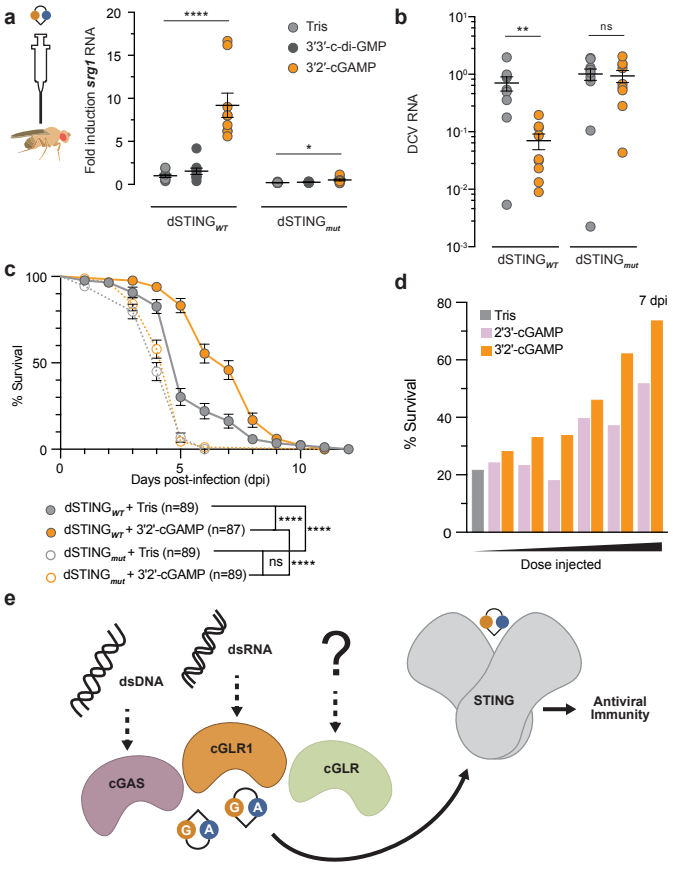
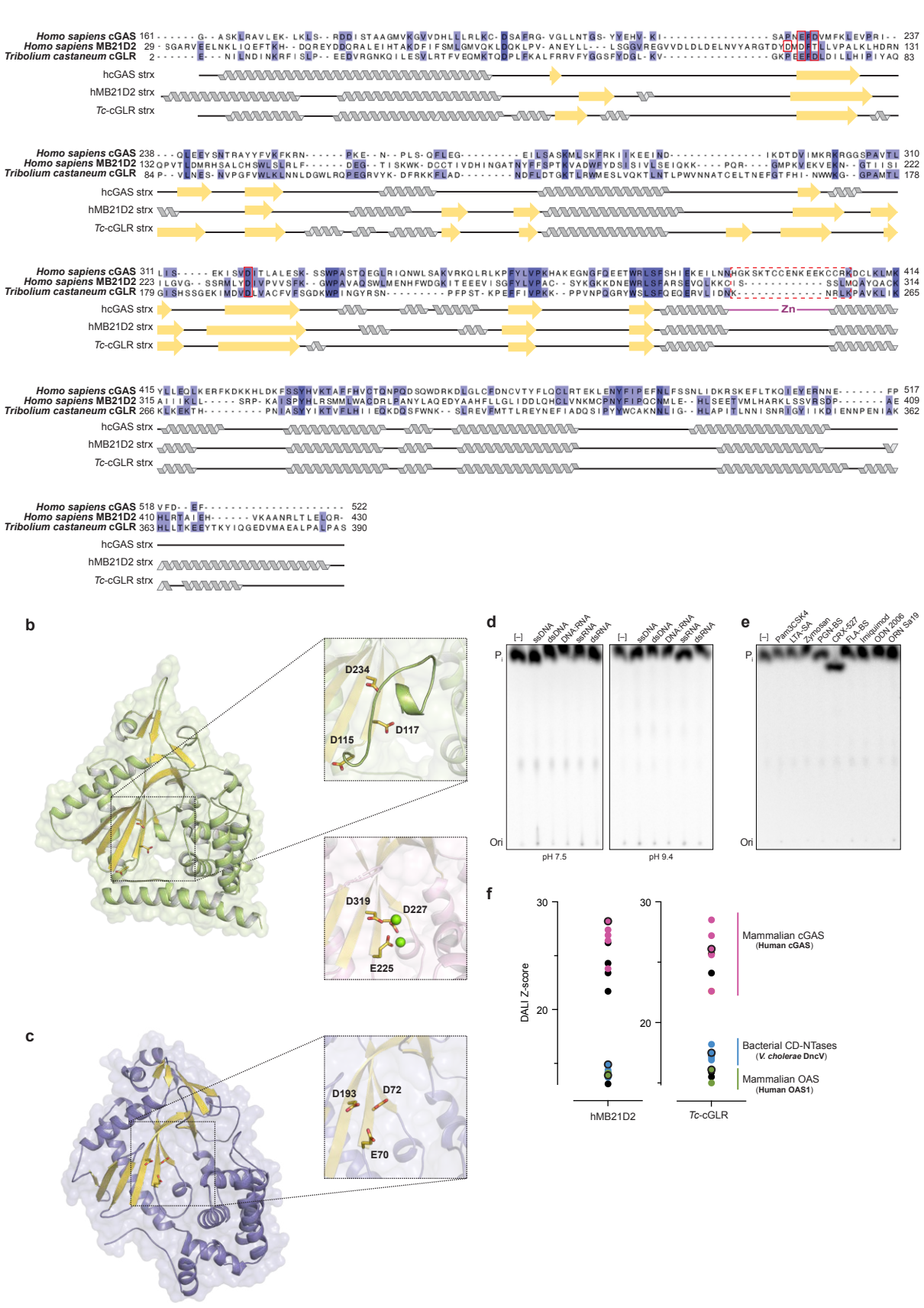


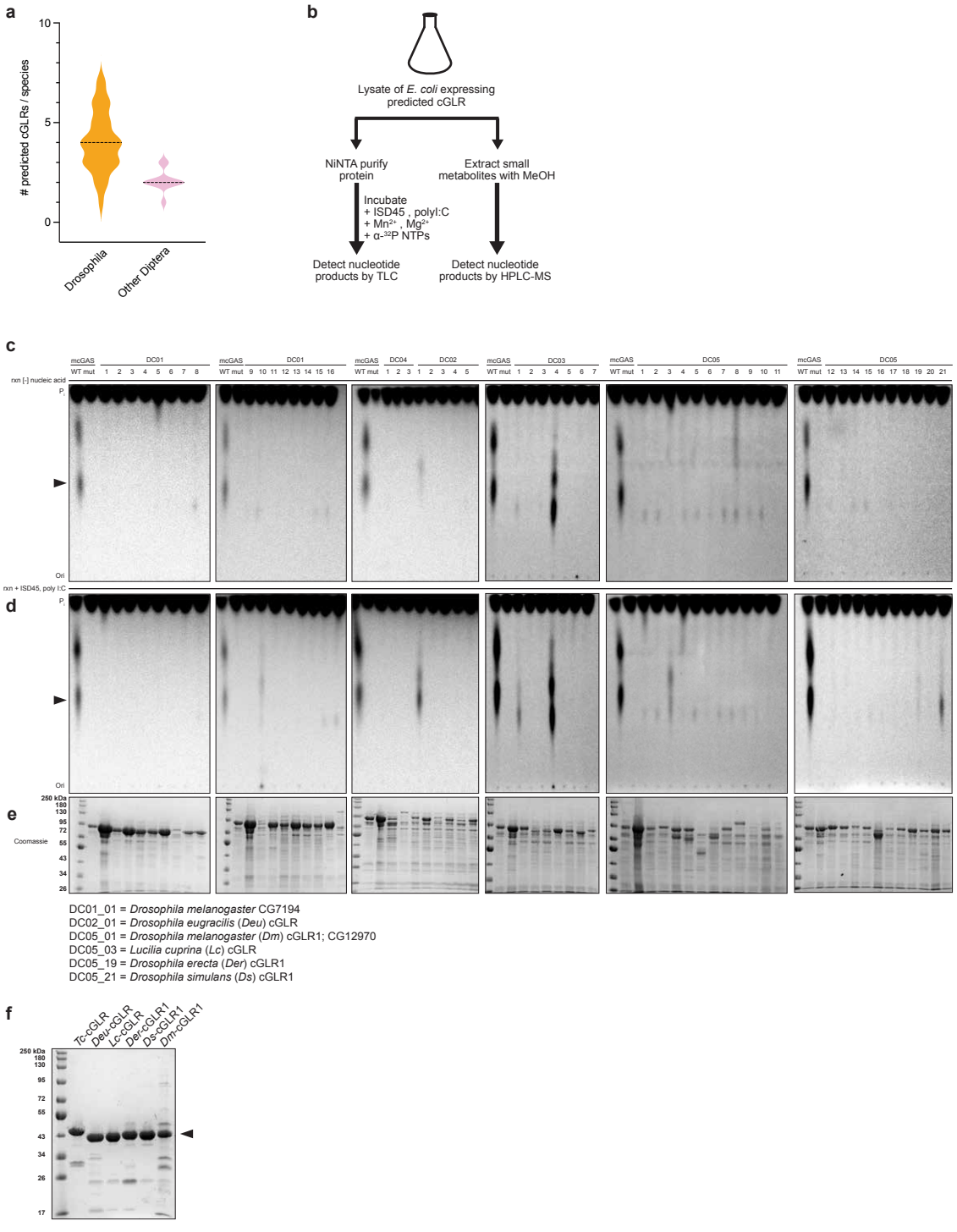
Fig. 5 | 3′2′-cGAMP activates STING-dependent antiviral immunity in *Drosophila*



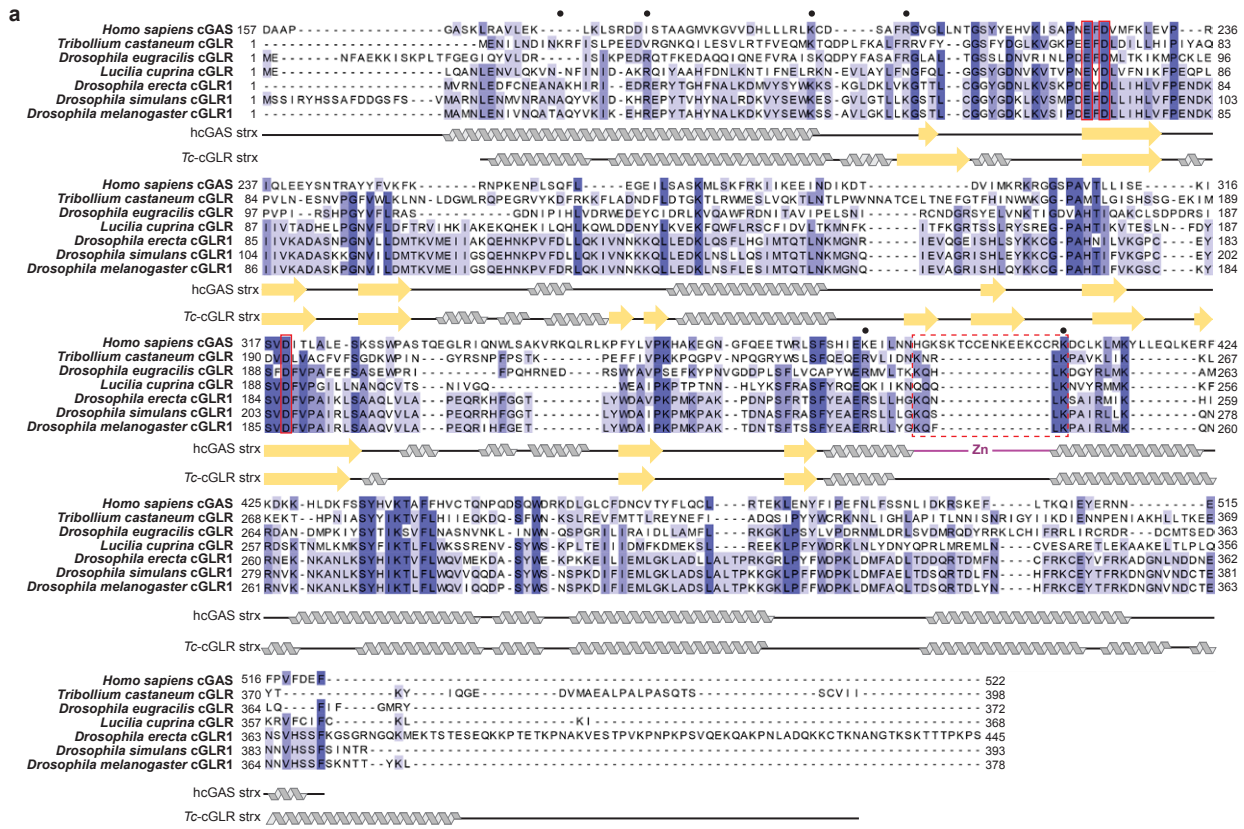
Extended Data Fig. 1 | Sequence and structural analysis of hMB21D2 and *Tc*-cGLR

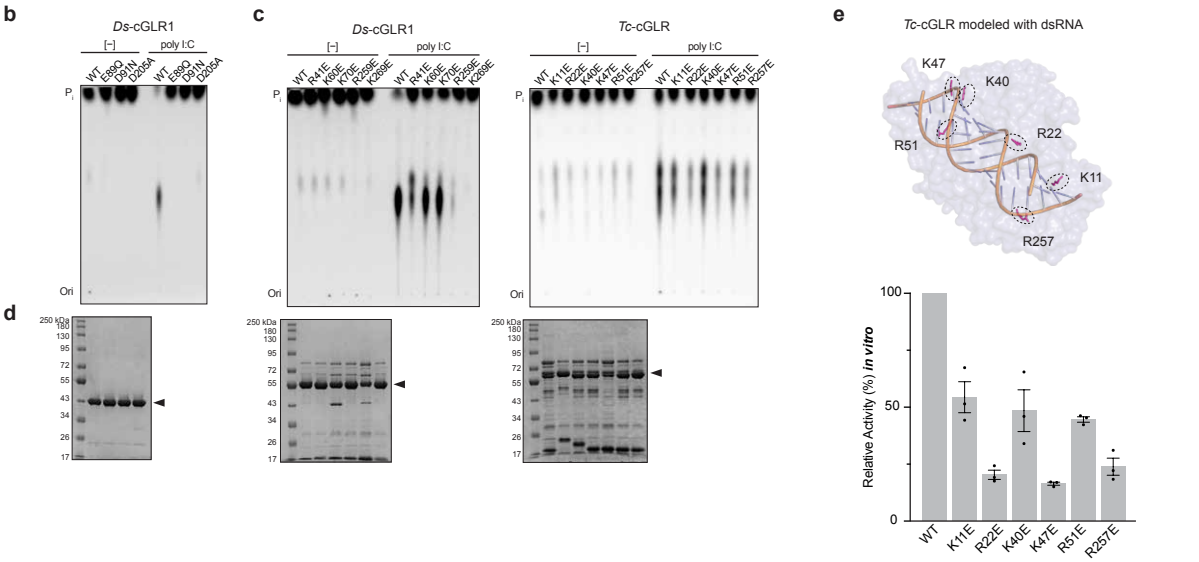


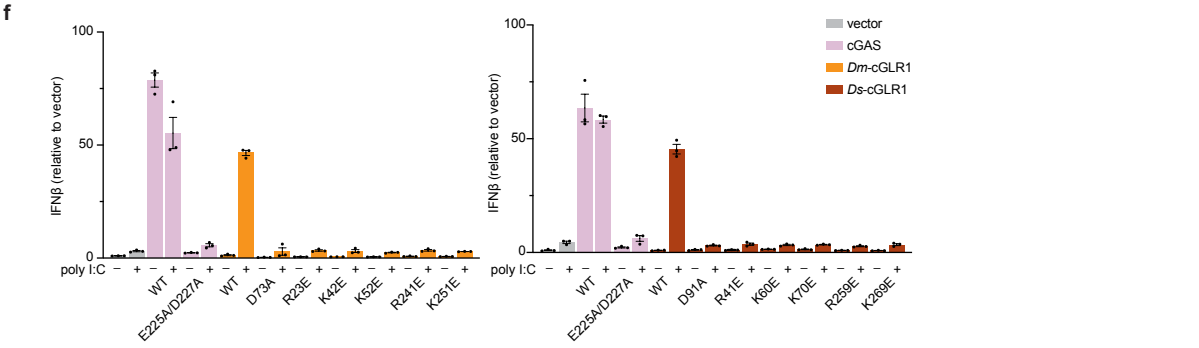
Extended Data Fig. 2 | Forward biochemical screen of predicted cGLRs in *Diptera*



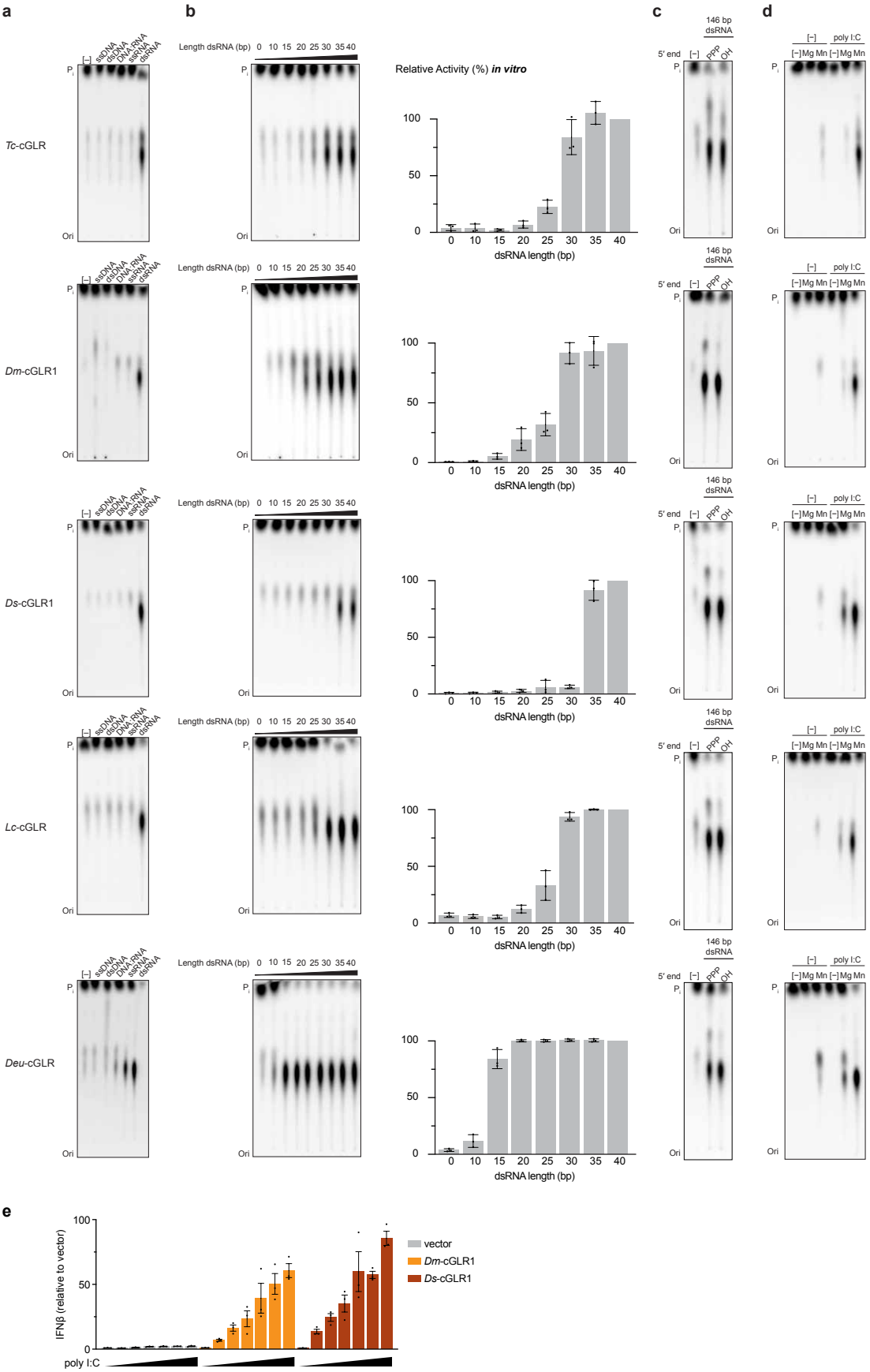
Extended Data Fig. 3 | Sequence analysis and mutagenesis of insect cGLRs



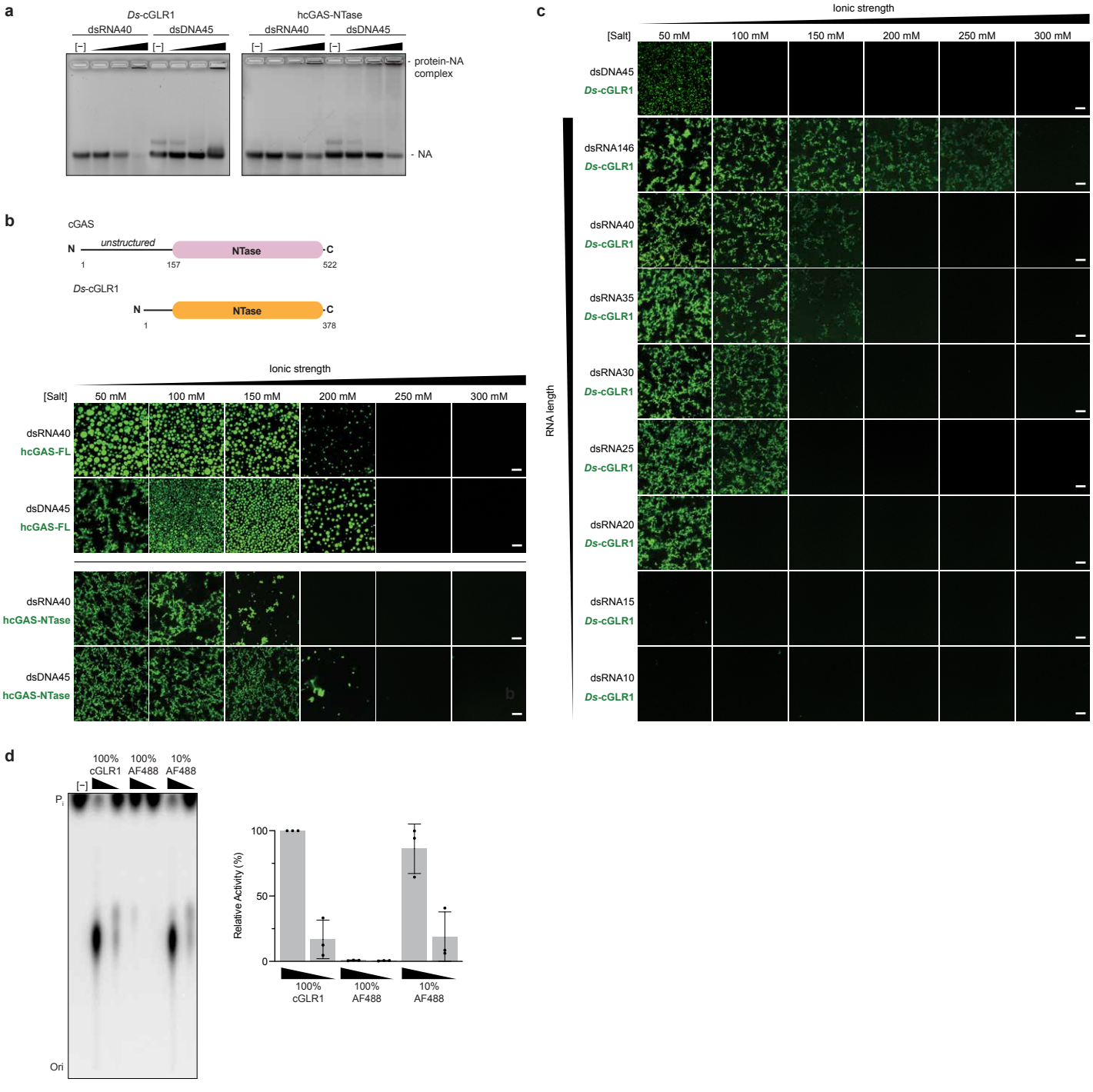




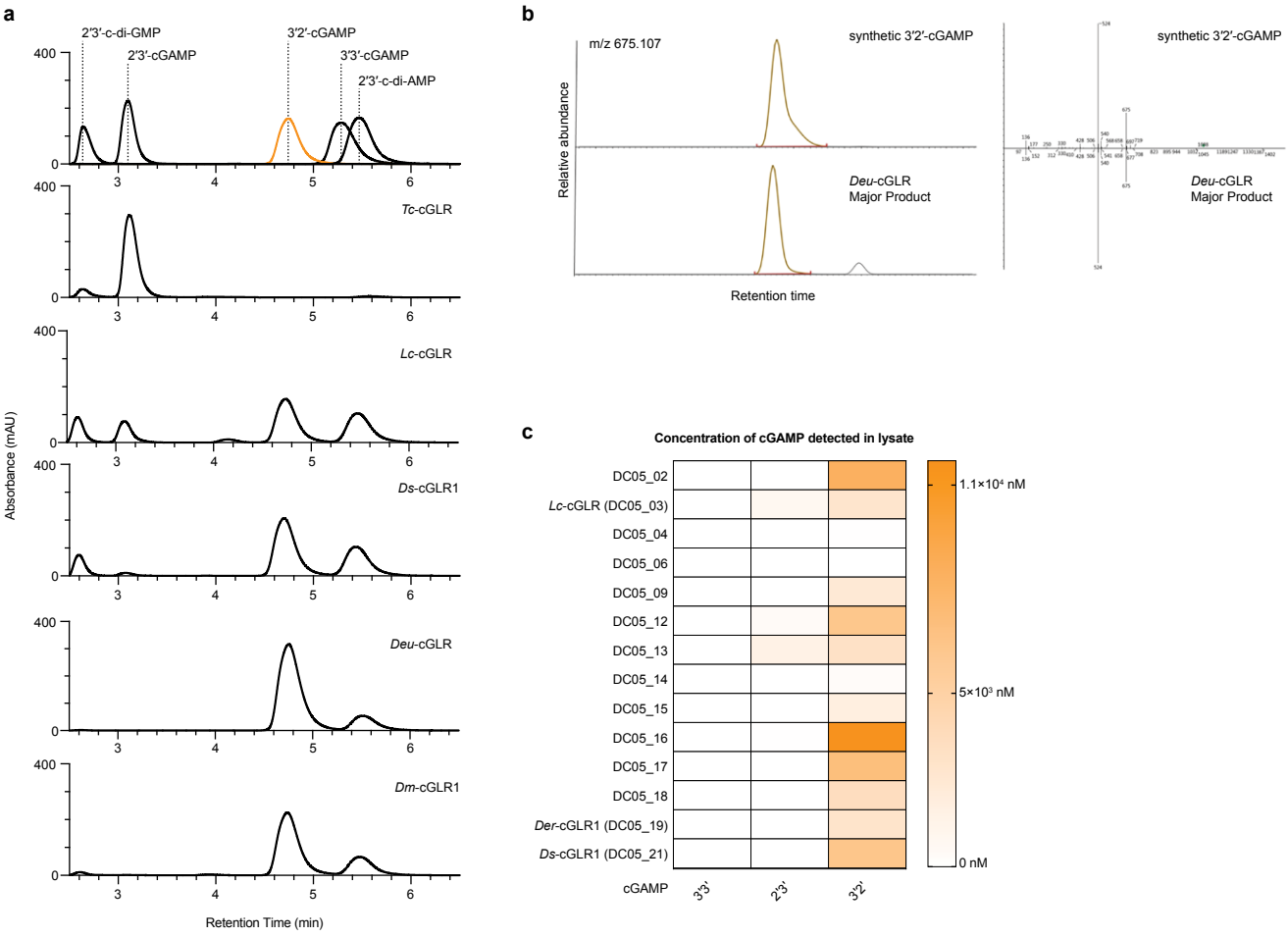
Extended Data Fig. 4 | Analysis of RNA-recognition by insect cGLRs



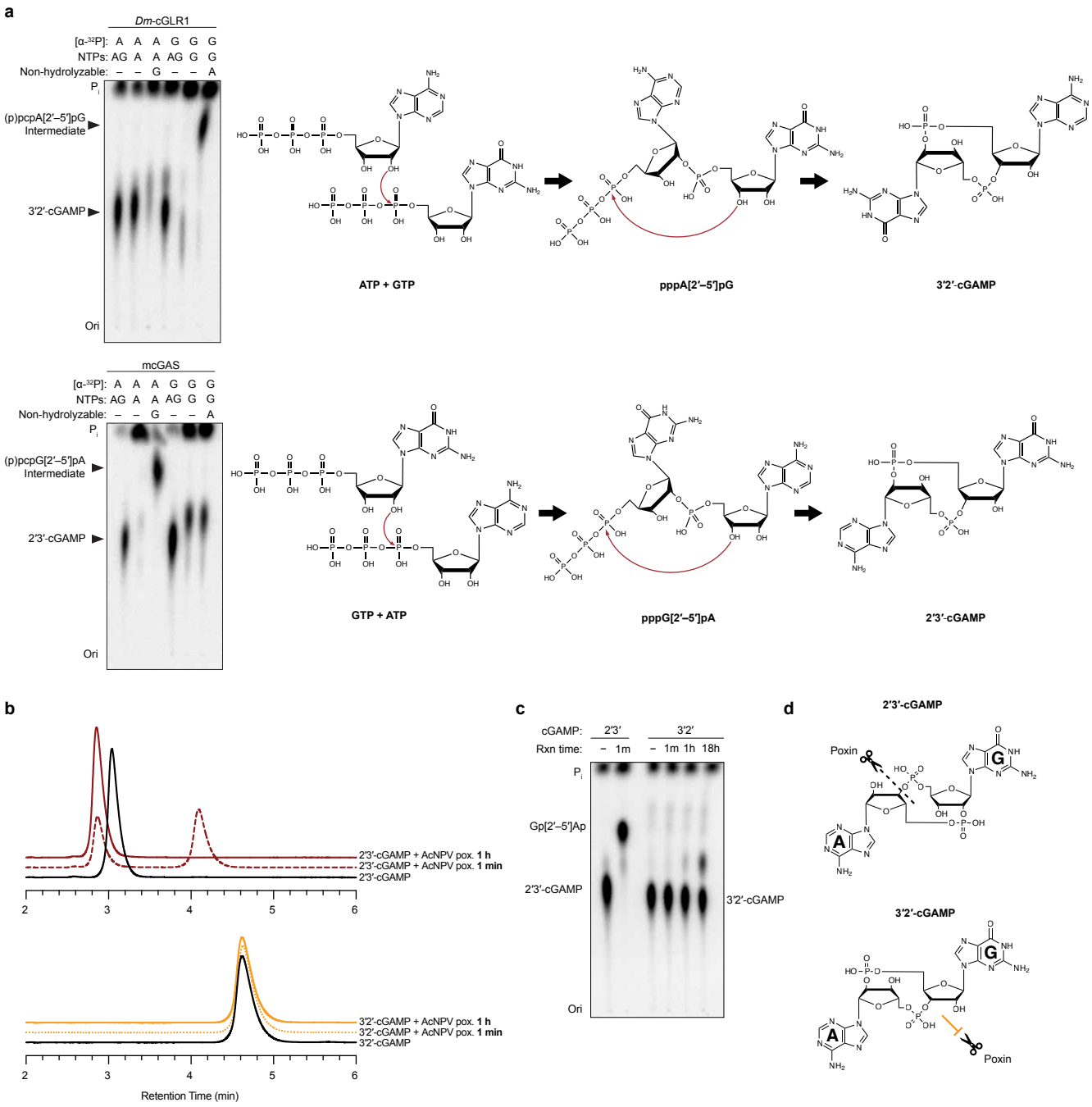
Extended Data Fig. 5 | Characterization of *Ds*-cGLR1–dsRNA condensate formation



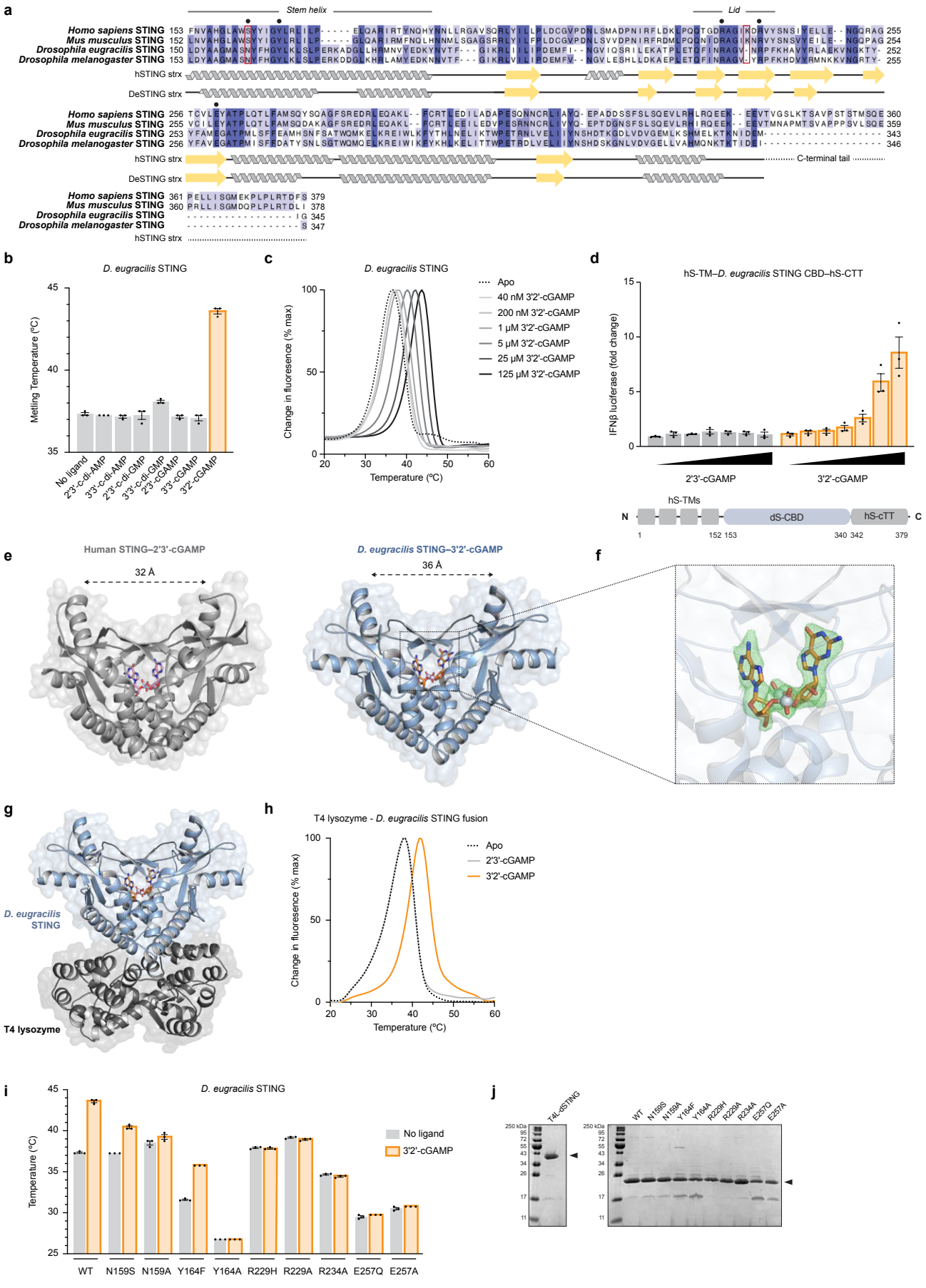
Extended Data Fig. 6 | Synthesis of 3′2′-cGAMP by *Diptera* cGLRs



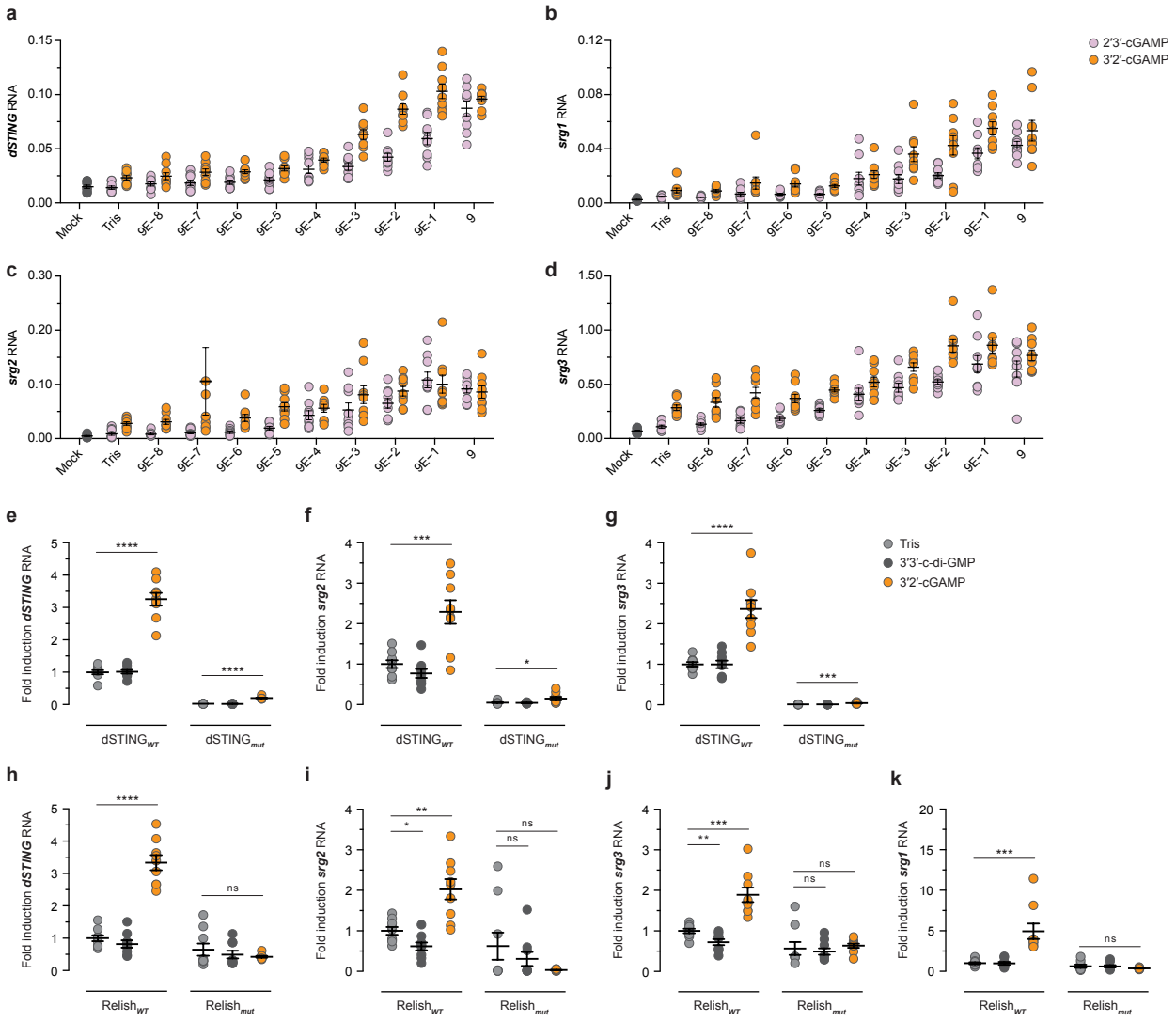
Extended Data Fig. 7 | Mechanism of 3′2′-cGAMP bond formation and resistance to degradation by viral poxin enzymes



Extended Data Fig. 8 | Structural and biochemical analysis of *Drosophila* STING



Extended Data Fig. 9 | 3′2′-cGAMP induces the expression of dSTING-regulated genes



Extended Data Fig. 10 | 3′2′-cGAMP functions as a potent antiviral ligand