SUPPLEMENTAL INFORMATION

The Cytosolic DNA Sensor cGAS Forms An Oligomeric Complex with DNA and Undergoes Switch-like Conformational Changes in the Activation Loop

Xu Zhang, Jiaxi Wu, Fenghe Du, Hui Xu, Lijun Sun, Zhe Chen, Chad A. Brautigam, Xuewu Zhang and Zhijian J. Chen

Supplemental Information contains a Table, 4 supplemental figures and legends, 4 supplemental movies and legends and Supplemental Experimental Procedures.

Table S1. Statistics of data collection and refinement of cGAS

Data	apo human cGAS	apo human cGAS (Se-Met)	human cGAS-cGAMP	human cGAS-sulfate	mouse cGAS-DNA
Space Group	P 2_12_12	P 2_12_12	P $2_12_12_1$	P6_122	C2
Unit Cell (Å,°)	162.194, 47.051, 58.404, 90, 90, 90	160.803 47.272 62.256 90.00 90.00 90.00	47.822, 118.542, 124.025, 90, 90, 90	101.027, 101.020, 241.331, 90, 90, 120	183.711, 96.330, 75.989, 90, 98.28, 90
Number of molecules in ASU	1	1	2	1	4
Wavelength (Å)	0.97899	0.97899	1.5418	0.97921	0.97918
Resolution (Å)	50-2.45 (2.49-2.45)	50-2.95 (3.00-2.95)	50-2.44 (2.48-2.44)	50-2.25 (2.29-2.25)	50-1.86 (1.89-1.86)
R_{merge} (%)	8.9 (79.3)	4.4 (86.8)	5.7 (33.8)	9.3 (70.6)	7.6 (40.9)
I/σ	17.6 (2.1)	40.5 (1.3)	24.7 (3.8)	29.6 (1.5)	23.0 (3.3)
Completeness (%)	97.4 (97.6)	99.2 (97.2)	99.1 (92.2)	99.0 (98.6)	99.5 (93.4)
Number of measured reflections	81463	80586	142767	591606	533161
Number of unique reflections	16956	10506	26688	35060	109036
Redundancy	4.8 (3.5)	7.7 (6.3)	5.3 (3.3)	16.9 (13.8)	4.9 (3.9)
Wilson B factor ($Å^2$)	31.7	59.8	44.1	48.3	27.5
R-factor (%)	18.0 (20.0)		20.0 (29.5)	19.0 (26.4)	18.7 (21.6)
R_{free} (%)	24.4 (34.5)		28.1 (37.5)	22.4 (29.7)	21.8 (23.8)
Number of atoms					
Macromolecule	2903		5703	2913	7228
Ligand	1		92	16	2
Water	90		89	169	1049
All atoms	2994		5884	3098	8279
Average B value ($Å^2$)					
Macromolecule	45.9		61.8	59.6	36.7
Ligand	27.1		52.6	108.3	23.1
solvent	40.2		49.8	55.9	40.0
All atoms	45.7		61.4	59.6	37.1
Rms deviations from ideal values					
Bonds (Å)	0.008		0.012	0.008	0.007
Angle (°)	1.251		1.48	1.064	1.189
Ramachandran plot statistics (%)					
Favored	95.7		96.2	97.4	97.0
Allowed	4.3		3.8	2.6	3.0
Outliers	0		0	0	0
Molprobity score	1.94		1.79	1.83	1.31

Values in parentheses are for the highest resolution shell. $R=\Sigma|F_{obs}-F_{calc}|/\Sigma F_{obs}$, where F_{calc} is the calculated protein structure factor from the atomic model.

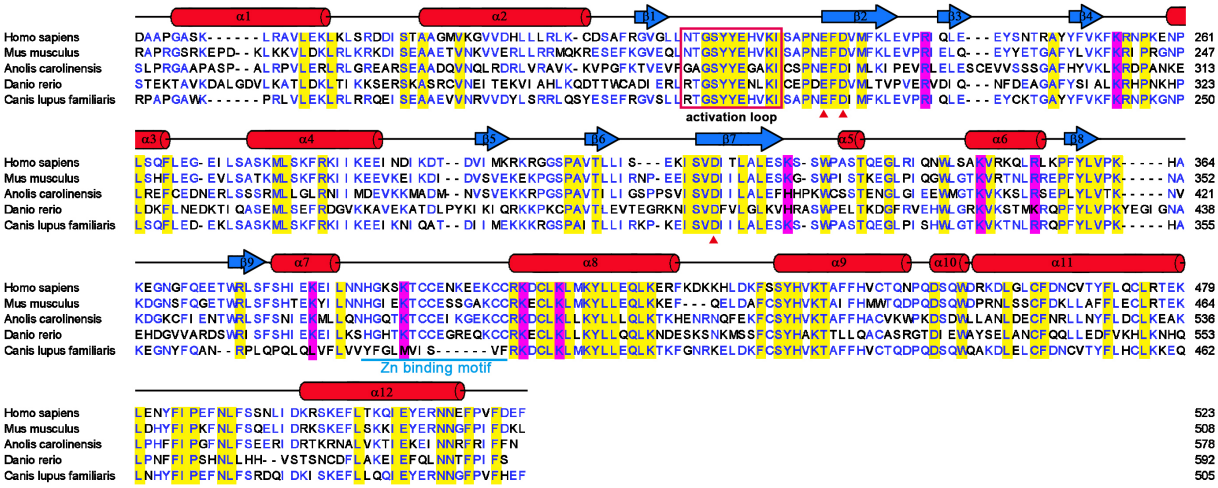

Figure S1. Sequence alignment of cGAS from different species. Secondary structural elements of cGAS are indicated above the sequence alignment. Residues of the activation loop are highlighted by red rectangle. The catalytic residues are indicated by solid red triangle. Residues that are strictly conserved among the species are shaded in yellow. The accession numbers for cGAS are: *Homo sapiens*, GI: 31581598; *Mus musculus*, GI: 27734190; *Anolis carolinensis*, GI: 327282750; *Danio rerio* GI: 68371332; *Canis lupus familiaris*, GI: 359320925. The sequences were aligned with ClustalW. Related to Figure 1.

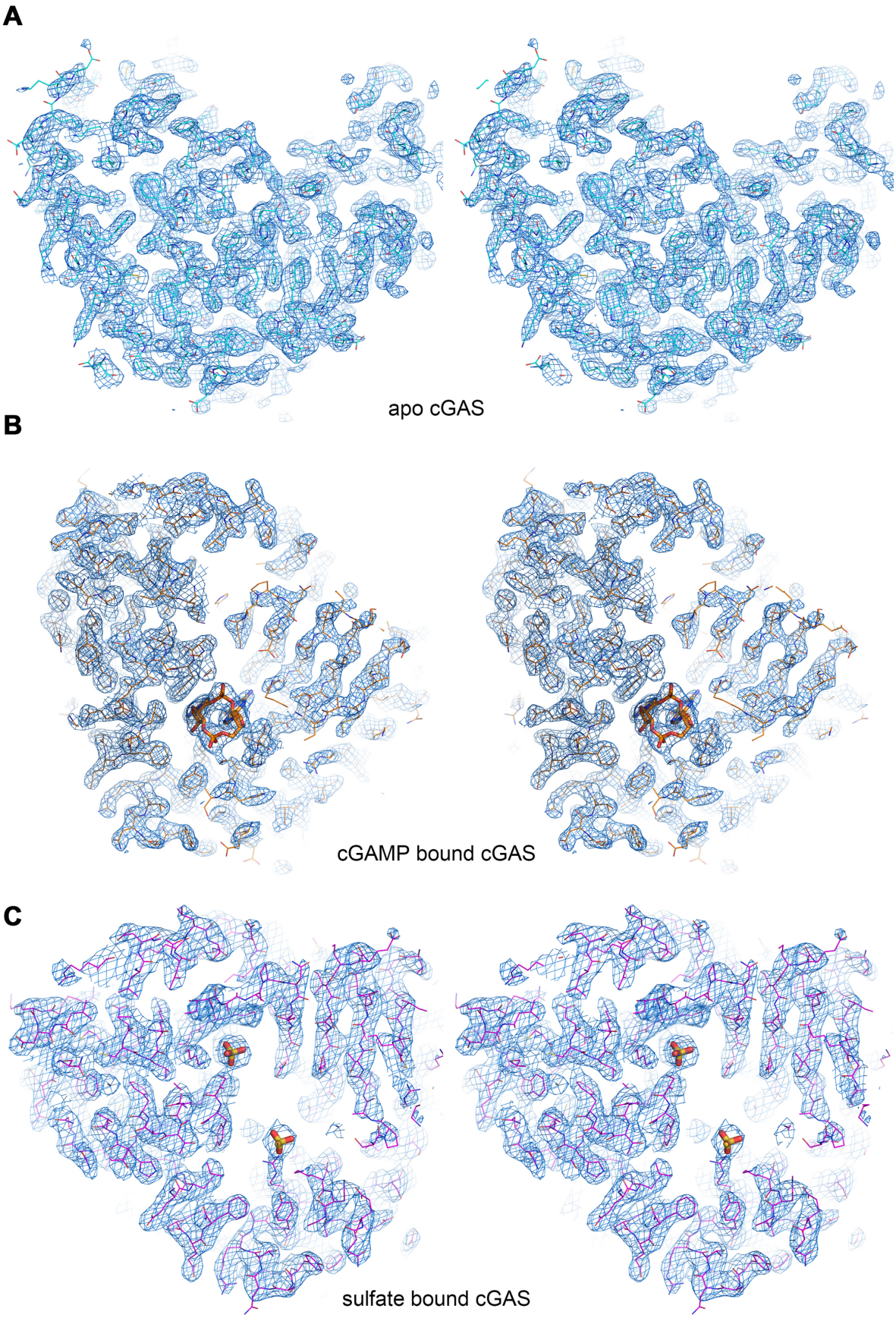

Figure S2. Representative electron density maps of the cGAS structures contoured at 1.0 σ. The stereo view of representative slabs of the $2F_o$-F_c electron density from apo cGAS **(A)**, cGAMP bound cGAS **(B)**, and sulfate bound cGAS **(C)**. Related to Figure 1.

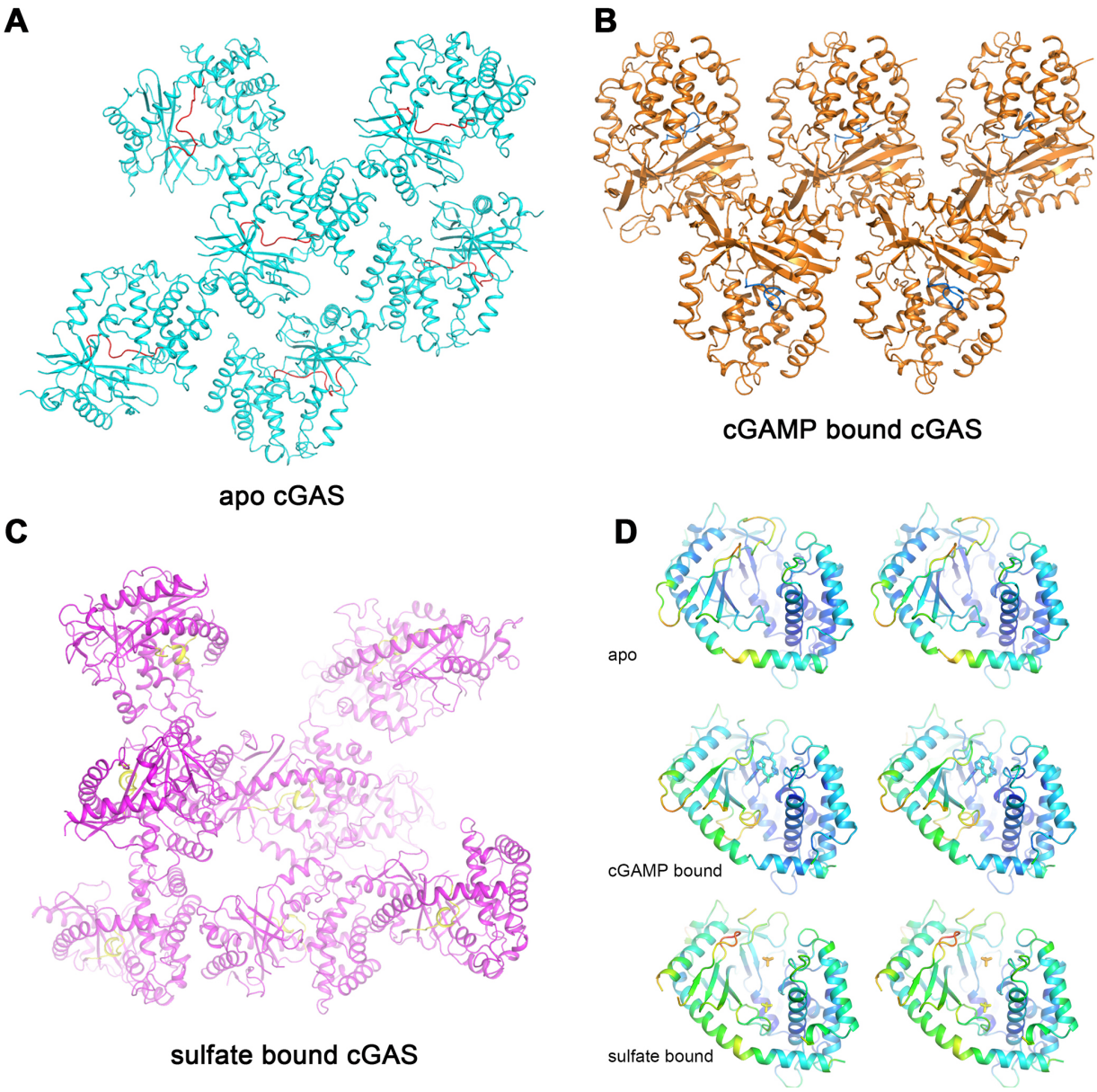

Figure S3. Crystallographic properties of the cGAS structures. (A-C) Crystal packing manners of the cGAS structures. The activation loop is colored in red in apo cGAS, blue in cGAMP bound cGAS, and yellow in sulfate bound cGAS, respectively. **(D)** The structures, generated in PyMol, are colored according to crystallographic temperature factors (blue<20.64 Å^2 to red>182.50 Å^2 for apo cGAS, blue<22.47 Å^2 to red>181.43 Å^2 for cGAMP bound cGAS, and blue<27.49 Å^2 to red>164.05 Å^2 for sulfate bound cGAS). Related to Figure 2.

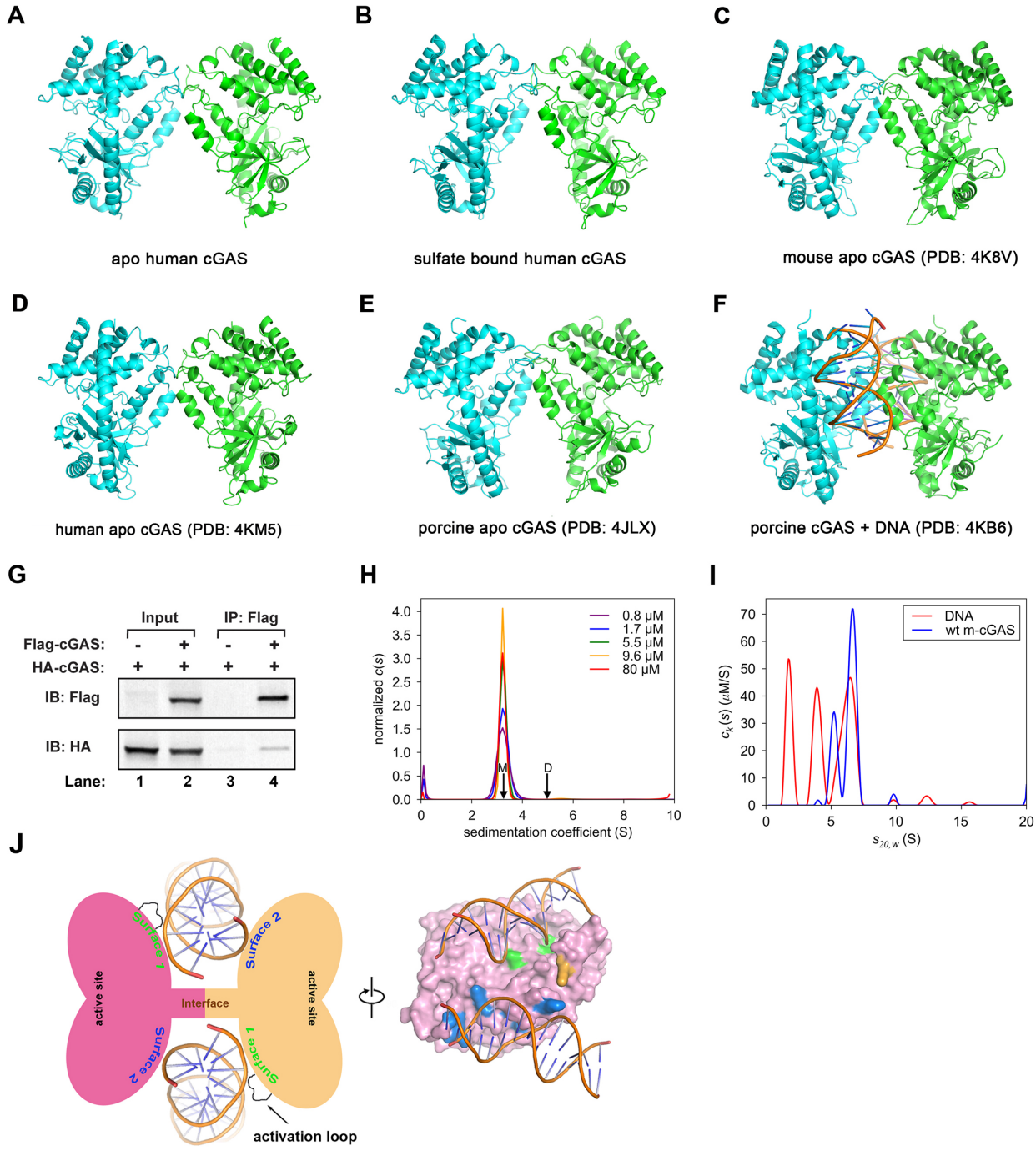

Figure S4. Characterization of cGAS oligomerization. (A and B) Both apo and sulfate bound human cGAS form dimers in the crystal through 447 Å^2 and 678 Å^2 contacts, respectively. **(C-F)** cGAS forms an identical dimer in the published structures as indicated. **(G)** Flag- and HA-tagged human cGAS expression plasmids were transfected into HEK293T cells. 24 hours after transfection, a Flag antibody was used for immunoprecipitation followed by immunoblotting with the indicated antibodies. (**H**) Analytical ultracentrifugation of m-cGAS at indicated concentrations. The theoretical sedimentation coefficients for m-cGAS monomer ('M") and dimer ("D") are shown by arrows. All distributions were normalized by the total amount of absorbance present in the respective experiment. (**I**) Multisignal sedimentation velocity analysis of m-cGAS and DNA**.** The area under respective peaks

represents the concentration of sedimenting material in μM units. In the ca. 6.5-S peak, integration shows that there are about 50 μM DNA and 56 μM m-cGAS. (**J**) Schematic representations of the cGAS-DNA complex in two perpendicular views. Key residues on the two DNA binding surfaces and protein-protein interface are mapped on the cGAS surface. The residues on the first and second DNA binding surfaces are colored in green and blue, respectively, and the residue involved in the protein-protein interface is colored in light orange. One cGAS molecule is omitted for clarity. Related to Figure 4.

SUPPLEMENTARY MOVIE LEGENDS

Movie S1. The model of DNA-dependent activation of cGAS. The apo cGAS structure is used as the start state, and the sulfate bound cGAS as the end state. The two structures are superimposed against the C lobe. The B-form DNA is docked based on the mutagenesis data shown in Figure 3C. The animation method is described in Experimental Procedures. Related to Figure 2.

Movie S2. The conformational shifts between the apo cGAS and the 2'3'-cGAMP bound cGAS structures. The two structures are superimposed against the C lobe. Related to Figure 2.

Movie S3. The conformational shifts between the apo cGAS and the sulfate bound cGAS structures. The two structures are superimposed against the C lobe. Related to Figure 2.

Movie S4. A model of cGAS oligomerization and activation induced by DNA. cGAS is colored in cyan and DNA in light orange. To illustrate the conformational changes in the formation of 2:2 complex, the activation loop and the active site residues are colored in magenta and yellow, respectively. Residues involved in cGAS DNA binding surface 1, surface 2 and protein-protein interface are colored in red, blue and purple, respectively. Related to Figure 4.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Protein Expression and Purification

The complementary DNA of truncated human cGAS (147-522 in apo form, 161-522 in 2'3'-cGAMP bound form and sulfate bound form) was cloned into pET-SUMO vector (Invitrogen^{TM}). Overexpression of cGAS was induced in *E. coli* BL21 (DE3) pLysS with 0.8 mM isopropyl-β-D-thiogalactoside when the cell density reached OD_{600nm} of 1.2. After growth at 18°C for 15 h, the cells were collected, resuspended in a buffer containing 25mM Tris-HCl, pH 8.0, 1 M NaCl and 10% glycerol, and disrupted by French Press with 2 passes at 15,000 p.s.i. Cell debris was removed by centrifugation at 27,000g for 50 min. The supernatant was loaded onto a Ni^{2+}-nitrilotriacetate affinity resin (Qiagen). Subsequently, the resin was rinsed three times with 50 ml buffer containing 25 mM Tris-HCl, pH 8.0, 1 M NaCl, 40 mM imidazole-HCl, 10% glycerol. The SUMO tag was then removed by digesting the proteins using the SUMO protease at 4°C overnight. The eluate was desalted and concentrated to about 15 mg/ml before applying to Heparin chromatography. The peak fractions of the protein were collected and concentrated to 6 mg/ml for crystallization trials. The selenomethionine derivatized protein and the mouse cGAS (147-507) was purified similarly as described above.

Enzymatic Synthesis and Purification of cGAMP

2'3'-cGAMP is synthesized in the buffer containing 20mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 10 mM $CoCl_2$, 0.01mg/ml herring testis DNA, 1 mM ATP, 1 mM GTP, and

0.1 μM recombinant SUMO-tagged human cGAS (147-522) was incubated at 37°C for 1 hr. The mixture was fractionated on a Hitrap Q column using a linear 0-0.5 M NaCl gradient; a UV peak corresponding to cGAMP was collected and loaded onto a C18 column (201TP510, 1 cm X 25 cm, Phenomenex, Hesperia, CA), and eluted with a linear 0-100% methanol gradient.

Crystallization

All the crystals were grown at 20°C by the hanging-drop vapor diffusion method. The apo cGAS yielded crystals in the buffer containing 6% PEG 3,350, 0.1 M Hepes, pH 7.8, 0.05 M NaCl, 0.01 M $MgCl_2$. The sulfate bound cGAS crystals were generated in the buffer containing 2 M $(NH4)_2SO_4$, 0.1 M Bis-Tris, pH 5.5. For the cGAMP bound cGAS, cGAMP was added into cGAS (161-522) in 6:1 molar ratio followed by incubation at room temperature for 30 min. The crystals of the complex were generated in the buffer containing 2% Isopropanol, 0.01 M $MgSO_4$, 0.1 M Tris-HCl, pH 8.5. The mouse cGAS and a 16bp dsDNA were mixed as previously described (Gao et al., 2013). It yielded crystals in a buffer containing 25% Methanol, 0.1M Tris-HCl, pH 8.0, 0.01M $MgCl_2$.

Data Collection, Processing and Structural Determination

For the apo form cGAS, the data were collected with Se high remote wavelength of 0.97899 Å at 19-ID of Argonne National Laboratory, Structural Biology Center at the Advanced Photon Source (APS). The data were integrated and scaled with HKL2000 package (Otwinowski and Minor, 1997). Further calculations were performed using

programs from PHENIX (Adams et al., 2010), unless otherwise specified. Initial phases were generated using anomalous differences of a 3.0 Å Se derivative dataset by SAD. The phase information was improved by the MR-SAD using another 2.45 Å Se derivative dataset. The initial model was built using the Autobuild suite aided by manual correction of the coordinates in the COOT (Emsley et al., 2010). The structure refinement was performed against the 2.45 Å dataset. The data set of the cGAMP bound cGAS was collected with a wavelength of 1.5418 Å using Rigaku FR-E copper rotating-anode generator and an R-AXIS IV^{++} imaging-plate area detector (Rigaku Americas, Houston, Texas, USA). The data set of the sulfate bound cGAS was collected with a wavelength of 0.97921 Å at APS. Both structures were determined by molecular replacement using apo cGAS as the search model, and then optimized by manual building with COOT. The data set of mouse cGAS DNA complex was collected with a wavelength of 0.97918 Å at APS. The structure was solved by molecular replacement using 4K98 as the search model. Data collection and refinement statistics are summarized in Table 1.

Analytical Ultracentrifugation

Prior to centrifugation, all protein and DNA samples were prepared and incubated overnight at 4°C. For m-cGAS experiments, the buffer was 25 mM Tris pH 8.0 and 25 mM NaCl. The h-cGAS experiments were performed in a buffer containing 25 mM Tris pH 8.0 and 300 mM NaCl. All analytical ultracentrifugation experiments were carried out in a Beckman Coulter Optima XL-I ultracentrifuge. Centerpieces (either 1.2 cm or 0.3 cm, depending on the expected absorbance) were sandwiched

between sapphire windows in centrifugation cells, then filled with solution (400 μL for the 1.2 cm centerpieces, 100 μL for the 0.3 cm centerpieces). After filling, the centrifugation cells were placed in an An50-Ti rotor that was subsequently incubated in the centrifuge under vacuum for at least 2.5 h at the experimental temperature, 20° C. Next, the centrifuge was accelerated to 50,000 rpm, and data acquisition was commenced. Data were collected using absorbance optics tuned to 260 nm for DNA-containing samples, 280 nm for protein-only samples, or 300 nm for high-concentration samples. Data were also acquired using interference optics. The data were analyzed using the *c*(*s*) methodology with algebraic noise decomposition (Schuck, 2000; Schuck and Demeler, 1999; Schuck et al., 2002) in the programs SEDFIT and SEDPHAT (www.analyticalultracentrifugation.com). For the multisignal analysis of the DNA, the signal increment for interference optics was obtained using the calculated extinction coefficient at 260 nm as a reference. This interference signal increment was then used as a standard to determine the extinction coefficient at 300 nm. For the protein, the estimated signal increment of the protein(Padrick et al., 2010) was used as a standard to determine the extinction coefficient at 300 nm. Theoretical values of sedimentation coefficients were also calculated in SEDFIT, assuming frictional ratios of 1.3 for both protein-DNA complexes and the protein alone. Partial-specific volumes were based on the weight averages of the partial-specific volumes of the cGAS protein (0.7343 mL/g) and DNA (0.54 mL/g). Solution density and viscosity and the partial-specific volumes of the proteins were calculated using SEDNTERP[7]. All figures featuring *c*(*s*) distributions were generated in GUSSI (biophysics.swmed.edu/MBR/software.html).

Cell Lines and Luciferase Reporter Assay

HEK293T-STING-IFNβ luciferase reporter cell line was constructed by lentiviral-mediated stable expression of human STING into previously described HEK293-IFNβ-luciferase cells (Chiu et al., 2009). Cells were cultured in DMEM supplemented with 10% calf serum and antibiotics. To test the IFNβ inducing ability of different cGAS mutants, expression plasmids were transfected into HEK293T-STING-IFNβ luciferase reporter cells. 24 hours after transfection, cells were lysed in the passive lysis buffer (Promega). An aliquot of the lysate was used to measure firefly luciferase activity, whereas another aliquot was measured renilla luciferase activity (Promega). IFNβ induction was calculated by normalizing firefly luciferase activity to renilla luciferase activity. All measurements were performed on a FLUOstar Optima Fluorometer (BMG Labtech).

Structural Animation

In order to generate the morph to visualize the conformational change between apo cGAS and sulfate bound cGAS, the two structures were superimposed. The shifted coordinates of the structures were used as the initial and end states, respectively, for morph generation. The intermediate morphs were obtained with the multiple-chain morphing script (Echols et al., 2003; Krebs and Gerstein, 2000) for Crystallography & NMR System (CNS) (Brunger, 2007; Brunger et al., 1998). The animations were finally produced using PyMol (DeLano, 2002).

SUPPLEMENTAL REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. Nat Protoc *2*, 2728-2733.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54*, 905-921.

Chiu, Y.-H., MacMillan, J.B., and Chen, Z.J. (2009). RNA Polymerase III Detects Cytosolic DNA and Induces Type I Interferons through the RIG-I Pathway. Cell *138*, 576-591.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System. on World Wide Web http://wwwpymolorg.

Echols, N., Milburn, D., and Gerstein, M. (2003). MolMovDB: analysis and visualization of conformational change and structural flexibility. Nucleic Acids Res *31*, 478-482.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B.L., Zillinger, T., Serganov, A.A., Liu, Y., Jones, R.A., Hartmann, G., *et al.* (2013). Cyclic [G(2',5')pA(3',5')p] Is the Metazoan Second Messenger Produced by DNA-Activated Cyclic GMP-AMP Synthase. Cell *153*, 1094-1107.

Krebs, W.G., and Gerstein, M. (2000). The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. Nucleic Acids Res *28*, 1665-1675.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Macromolecular Crystallography, Pt A *276*, 307-326.

Padrick, S.B., Deka, R.K., Chuang, J.L., Wynn, R.M., Chuang, D.T., Norgard, M.V., Rosen, M.K., and Brautigam, C.A. (2010). Determination of protein complex stoichiometry through multisignal sedimentation velocity experiments. Anal Biochem *407*, 89-103.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J *78*, 1606-1619.

Schuck, P., and Demeler, B. (1999). Direct sedimentation analysis of interference optical data in analytical ultracentrifugation. Biophys J *76*, 2288-2296.

Schuck, P., Perugini, M.A., Gonzales, N.R., Howlett, G.J., and Schubert, D. (2002). Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. Biophys J *82*, 1096-1111.