

## **Structure-guided reprogramming of human cGAS dinucleotide linkage specificity**

Philip J. Kranzusch<sup>1,4</sup>, Amy S.Y. Lee<sup>1,2</sup>, Stephen C. Wilson<sup>5</sup>, Mikhail S. Solovych<sup>1</sup>, Russell E. Vance<sup>1,4</sup>, James M. Berger<sup>3\*</sup> and Jennifer A. Doudna<sup>1,2,4-6\*</sup>

<sup>1</sup>Department of Molecular & Cell Biology, University of California, Berkeley, CA 94720, USA

<sup>2</sup>Center for RNA Systems Biology, University of California, Berkeley, CA 94720, USA

<sup>3</sup>Department of Biophysics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>4</sup>Howard Hughes Medical Institute (HHMI), University of California, Berkeley, CA 94720, USA

<sup>5</sup>Department of Chemistry, University of California, Berkeley, CA 94720, USA

<sup>6</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

\*Correspondence to: [jmberger@jhmi.edu](mailto:jmberger@jhmi.edu) (J.M.B.); [doudna@berkeley.edu](mailto:doudna@berkeley.edu) (J.A.D.)

## **Supplemental Information**

Supplemental information includes Extended Methods, 5 figures and 2 tables.

## SI Figure Legends

### SI Figure 1. *Vibrio* DncV amino-acid sequence conservation, Related to Figures 1 and 4

(A) Alignment of DncV amino-acid sequences generated with MAAFT and colored by BLOSUM62 conservation score (see methods). Cartoon schematic of the experimentally determined *Vibrio* DncV secondary structure is depicted below (green =  $\alpha$ -helix, blue =  $\beta$ -strand) and red dots denote the conserved active site residues involved in metal coordination. Regions with no observable electron density in the DncV crystal structure (F144–K149 and F218–T238) are faded. (B) Structural alignment of *Vibrio* DncV and human cGAS sequences based on superposition of pppA[3' –5' ]pG and cGAS pppG[2' –5' ]pG (PDB 4KM5 and 4K98) structures (see methods). Alignment is colored as in A, and experimentally determined secondary-structures are depicted for DncV above the sequence alignment (green =  $\alpha$ -helix, blue =  $\beta$ -strand) and for cGAS below the sequence alignment (purple =  $\alpha$ -helix, blue =  $\beta$ -strand).

### SI Figure 2. *Vibrio* DncV crystal contacts, Related to Figure 1

(A) Packing of individual DncV monomers in the crystallographic asymmetric unit. A red circle denotes the packing region between chain A (green/blue) and chain B (grey/blue), and the calculated buried surface area is indicated. Bound  $Mg^{2+}$  ions (pink) and pppA[3' –5' ]pG ligand (yellow/orange) are depicted as spheres. (B) Packing of individual DncV monomers in adjacent symmetry-related molecules. DncV chain A monomer is displayed as in A, and the symmetry-related molecule is depicted as a faded out monomer. Packing regions and calculated buried surface area are indicated in red as in A. (C) *Vibrio* DncV size-exclusion chromatography purification data demonstrating DncV migrates as a monomer during purification according to its predicted molecular weight. Migration of standards are indicated in grey (IgG ~158 kDa, Albumin ~67 kDa, Ovalbumin ~43 kDa, Myoglobin ~17 kDa).

### SI Figure 3. DncV activity is insensitive to nucleic acid, Related to Figures 2 and 5

(A) Analysis of DncV and human cGAS cyclic dinucleotide product formation as in *Figure 3*. Prior to substrate nucleotide addition, DncV reactions were supplemented with 20 nt ssRNA, 45 nt ssDNA or 45 bp dsDNA as indicated (see methods); these substrates have no apparent impact on product A[3'–5']pG[3'–5']p (cAG) formation. (B) DncV reactions prepared as in A, but with 5' radiolabeled nucleic acid substrates and separated by non-denaturing polyacrylamide gel-electrophoresis. Radiolabeled substrates include 20 nt ssRNA, 20 bp dsRNA, ~56 bp structured human 5' Jun UTR mRNA, 45 nt ssDNA, 45 bp dsDNA or a branched ss/dsDNA substrate as previously described to stimulate *Bacillus subtilis* DisA (Witte et al., 2008). Control reactions include corresponding radiolabeled substrate in the absence of purified DncV (left lanes).

**SI Figure 4. Reprogrammed cGAS enzymes have reduced guanine selectivity, Related to Figure 5**

(A) Analysis of DncV, human cGAS and reprogrammed cGAS cyclic dinucleotide formation as in *Figure 3D*. Reprogrammed cGAS enzymes have a reduced selectivity for order of guanine nucleotide selection but preferentially proceed in the same direction as wildtype cGAS. Additionally, consistent with relaxed guanine-specific base interactions, reprogrammed cGAS mutants produce cyclic di-A (A[3'–5']pA[3'–5']p) under low GTP conditions, as described in the main text. Reactions include labeled/unlabeled NTPs and nonhydrolyzable methylene-substituted ATP and GTP nucleotides as indicated. A[3'–5']pG (green), G[3'–5']pA (green) and G[2'–5']pA (purple) labels denote migration of trapped linear products from DncV and cGAS respectively, and images are representative of multiple independent experiments.

**SI Figure 5. DncV reprogramming mutations, Related to Figures 5 and 6**

(A) In-cell reconstitution of cyclic dinucleotide immune activation by *Vibrio* DncV as in *Figure 6*. Cells were co-transfected with indicated empty vector, human cGAS or *Vibrio* DncV plasmids in

the presence of wildtype STING or the R232H allele of STING incapable of responding to 3' – 5' cGAMP. Immune activation was monitored with a firefly luciferase reporter under control of the IFN-beta promoter (pIFN $\beta$ -FLuc). Data are normalized to wildtype human cGAS signal as in Figure 6 and error bars represent the SD from the mean of at least three independent experiments (\* denotes  $p < 0.002$ , and n.s. denotes not significant).

## **Extended Methods**

### **Protein structure and secondary-structure alignments**

The structures of DncV and cGAS (PDB 4KM5 and 4K98) trapped in linear-intermediate states were superposed in Coot. Potential sites for human cGAS reprogramming mutations were identified by manual inspection of the aligned active sites and relative ligand positions. Additionally, all non-protein atoms were removed and a sequence homology of the aligned structures was determined and extended to the human cGAS sequence using PROMALS3D (Pei et al., 2008). DncV and cGAS homologs were identified by BLAST-P using conservative cutoff scores of >30% similarity over >50% sequence coverage, and sequences were only included if the catalytic triad E/D residues were conserved. Amino-acid conservation alignments were generated using MAAFT (Kato and Standley, 2013), and colored according to BLOSUM62 conservation in Jalview (Waterhouse et al., 2009) (see SI Figure 1). Plots of sites 1, 2 and 3 local conservation were generated using WebLogo (Crooks et al., 2004), and yielded similar results using either all identified homolog sequences or only unique sequences (see SI Table 2).

## SI Literature Cited

- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica Section D, Biological crystallography* 66, 12-21.
- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome research* 14, 1188-1190.
- Karplus, P.A., and Diederichs, K. (2012). Linking crystallographic model and data quality. *Science* 336, 1030-1033.
- Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular biology and evolution* 30, 772-780.
- Pei, J., Kim, B.H., and Grishin, N.V. (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic acids research* 36, 2295-2300.
- Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189-1191.
- Weiss, M.S. (2001). Global indicators of X-ray data quality. *J Appl Crystallogr* 34, 130-135.
- Witte, G., Hartung, S., Buttner, K., and Hopfner, K.P. (2008). Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Molecular cell* 30, 167-178.

**A**

*V. cholerae*  
*V. narveyi*  
*E. coli*  
*K. pneumoniae*  
*P. syringae*  
*A. baumannii*  
*P. aeruginosa*

1 MA SMTWN FHQYYTNR --NDGLMGK LVLTDEEKNLKA LRKI IR LRTRDVFEEAKGI AKAVKKSALT FEI IQEKVSTTQIKHLSDS EQR EVAKL IY EMDDDARDEF LGLT PR FWTQGS FQYD 119  
 1 ---MAWMFHQYYTNR --MDGLMGK LVLTDEEKNLKA LRKI IR LRTRDVFEEAKGI AKAVKKSALT FEI IQEKVSKTKI HLS ENDLVAKL IY EMD EDARDEF LGLT PR FWTQGS FQYD 116  
 1 ---MHWDLNYYTNR --MDGLI SK LK LSKT ESTK ELRQI VR ERTRDVFEEAKGVKHT LT LEGVR LK LQTVNRY LSTADQAEVAR LI FEMDDARDF INLQPR FWTQGS FQYD 116  
 1 ---MTWNFHRYYSDS --TDGLI SK LR LAKER IEMLKA LRKVRVTKVEFAEAK ELAKT SNAS --LSLESTARASTRLK HLS PAAQVEAQL I FEMEDARDF LK FQPR FWTQGS FQYD 104  
 1 ---MNWVFHYYTNR --TDGLMGQLLSDKEDAKLAKLRDQVRETRDI FVEAKK LVNQAKKD --IGLEFLRVMSLTFKYLSDQEQNF AE LLIQLDSNAKAEFLKLT PR FWTQGS FQYD 115  
 1 M --PVFN LHSLLDSTVYADT FLAGETL ASDER EHMQSART EIRDR ---LRTLR PALLQRA LGSQQVR ---KPR FTFQGS SWAYK 76

120 T LNR PFQ --PGQEMDI DDGTY MPMP I FESE -PK IGHSLLI LLVDAS LK SLVA ENHGWF F -EAKQT CGR I KI EA EKTHI DV PMY A I PKDEFMKQKQI A LEAN -R S FVKGAI FESYVADSI --TD 234  
 117 T LNR PFQ --PGQEMDI DDGTY MPMP I FESE -PK IGHSLLI LLVDAS LK SLVA ENHGWF F -EAKQT CGR I KI EA EKTHI DV PMY A I PKDEFMKQKQI A LEAN -R S FVKGAI FESYVADSI --D 229  
 117 T LNR PFQ --PGQEMDI DDGTY MPMP I FESE -PK IGHSLLI LLVDAS LK SLVA ENHGWF F -EAKQT CGR I KI EA EKTHI DV PMY A I PKDEFMKQKQI A LEAN -R S FVKGAI FESYVADSI --D 229  
 105 T LNR PFH --PGQEMDI DDGTY MPMP I FESE -PS IGHSLLI LLVDAS LK SLVA ENHGWF F -EAKQT CGR I KI EA EKTHI DV PMY A I PKDEFMKQKQI A LEAN -R S FVKGAI FESYVADSI --D 217  
 116 T LNR PYAT --PGQEMDI DDGTY L PMA I FEDR -PV IGHSLLI LLVDAS LK SLVA ENHGWF F -EAKQT CGR I KI EA EKTHI DV PMY A I PKDEFMKQKQI A LEAN -R S FVKGAI FESYVADSI --D 230  
 116 T LNR PYAT --PGQEMDI DDGTY L PMA I FEDR -PV IGHSLLI LLVDAS LK SLVA ENHGWF F -EAKQT CGR I KI EA EKTHI DV PMY A I PKDEFMKQKQI A LEAN -R S FVKGAI FESYVADSI --D 229  
 77 T LNA PCK --DPOQAD LDDGTY L PFSY LEAT PP -VMSNV LFTCV EELQDLADEK -GWK LI DDNPCTR LEI ASDK -HI DV PAV SI PDEE FE ---LR ES -RA LAI CNA FDSV LAKAQ --LE 187

235 DS ETY ELI --DSENVN LALR EGDKWI NSDPK I V EDWFNS CGR I GKHLRKYCR FMKAWRDAQ -W -DVGG P SSI S LMAATVNI L DSVAHDA SD LG ET MK I I AKHL P S EFARGV ES PDST DEK P 352  
 230 DT D S EV --DSENVN LALR EGDKWI NSDPK I V EDWFNS CGR I GKHLRKYCR FMKAWRDAQ -W -DVGG P SSI S LMAATVNI L DSVAHDA SD LG ET MK I I AKHL P S EFARGV ES PDST DEK P 347  
 218 SR EAY LV -ESDKVN LALR EGDKWI NSDPK I V EDWFNS CGR I GKHLRKYCR FMKAWRDAQ -W -DVGG P SSI S LMAATVNI L DSVAHDA SD LG ET MK I I AKHL P S EFARGV ES PDST DEK P 347  
 230 GR EAY LV -ESDKVN LALR EGDKWI NSDPK I V EDWFNS CGR I GKHLRKYCR FMKAWRDAQ -W -DVGG P SSI S LMAATVNI L DSVAHDA SD LG ET MK I I AKHL P S EFARGV ES PDST DEK P 335  
 231 NR ADY EL --DENSVN LALR EGDKWI NSDPK I V EDWFNS CGR I GKHLRKYCR FMKAWRDAQ -W -DVGG P SSI S LMAATVNI L DSVAHDA SD LG ET MK I I AKHL P S EFARGV ES PDST DEK P 349  
 230 DQKRY KL --DSDCVN LALR EGDKWI NSDPK I V EDWFNS CGR I GKHLRKYCR FMKAWRDAQ -W -DVGG P SSI S LMAATVNI L DSVAHDA SD LG ET MK I I AKHL P S EFARGV ES PDST DEK P 342  
 188 EDDNWD LMPHTGV L MATK - -DRGWRDNDPR I KDWI E SEVA LKT EQLR LMR Y I KGRWDY QW -E SEDPK S I LLMVA YAKA LDVAV PRRDDIA --LLK VVYKA LPK I LQGRV FNB -A I AMK P 302

353 L ----FPPSYKHGPR EMDI MSKLERL PEI LLSA -ESADSK SEALKI NMA FGNRV - --TNS ELI V LAKA I PAFAQEP S SASK PEK I SSTMW SG 437  
 348 L ----FPPSYKHGPR EMDI MSKLERL PEI LLSA -ESADSK SEALKI NMA FGNRV - --TNS DL I V LAKA I PAFAQEP S SASK PEK I SSTMW SG 432  
 348 L ----FPPSYKHGPR EMDI MSKLERL PEI LLSA -ESADSK SEALKI NMA FGNRV - --TNS DL I V LAKA I PAFAQEP S SASK PEK I SSTMW SG 432  
 336 L ----FPAEWDNQVHQKTI V ETMKTLY ELVDA -ENANTR EDA LHKMNEA FGRV - --TNAQLIT STAAA PA FHV S P RE PRK I NKT MW SG 420  
 350 L ----FPPHYEHQVREI MEKLSLEGL LDOA -EQSETR EEA LRK I NFAEGRV - --TNAQLIT STAAA PA FHV S P RE PRK I NKT MW SG 420  
 349 L ----FPPY S EHGVER E I I EKMQT L L I NLENA -EVAOTKHEA L N L N FGRV - --KDY S L I V SQT A PA FEDEA SQA STT FOI SSTMI SG 433  
 303 V EEQEDLAAR LDKDC I RADVVAR LTR LGQOOLEA I YR S S PEOACR LLRDA FGNRV PY DGRV I DT VQST PA -- --DK SKAV S P VGI -- --MTAG 390

**B**

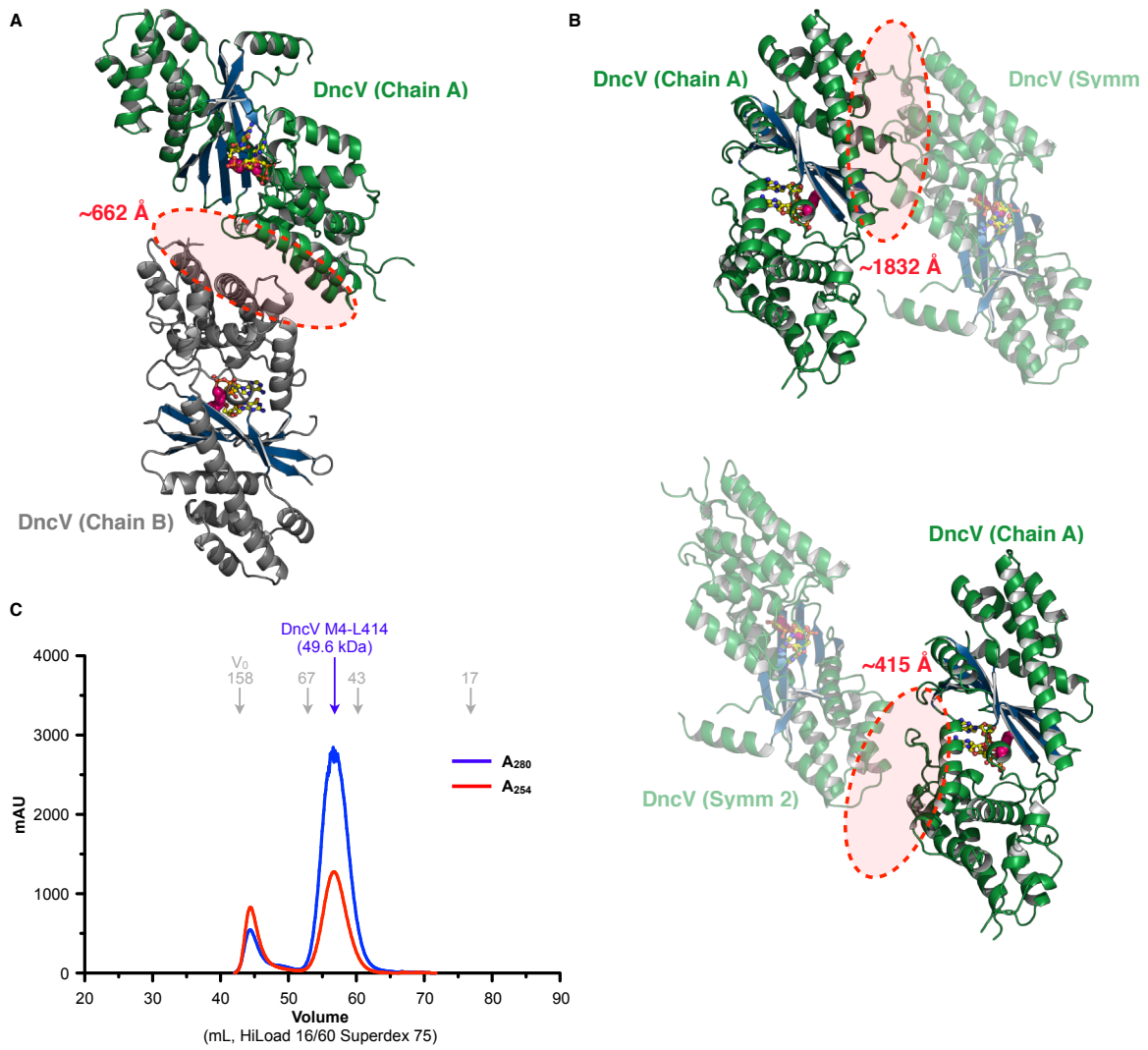
3 SMTWN FHQYYTNR NDGLMGK LVLTDEEKNLKA LRKI IR LRTRDVFEEAKGI AKAVKKSALT FEI IQEKVSTTQIKHLSDS EQR EVAKL IY EMDDDARDEF LGLT PR FWTQGS FQYD LNR 123  
 161 ---GA S -K LR ---AVLEK LK SRDDI STAAGMY KGVVDH LLL R -K ---AVLEK LK SRDDI STAAGMY KGVVDH LLL R -K ---CD S ---AF -RGVGLL NTGSY EHV K I S 221

124 P FOPGQEMDI DDGTY MPMP I ---FES EPK I GHSLLI LLVDAS LK SLVA EN --HGWK F -EAK --OTCCR I KI EA EKTHI DV PMY A I PKDEFMKQKQI 210  
 222 A ---PNE F DV MFK LEV P R I QLE EY S NTRAY FVKFRNPKENLSQF LEGEI L SASK M LSK FRKI I KEE I NDI KDTDV I MKR RGG SPATV L L L --SEKI I SVDI T L A L ESK S S --WPA S 333

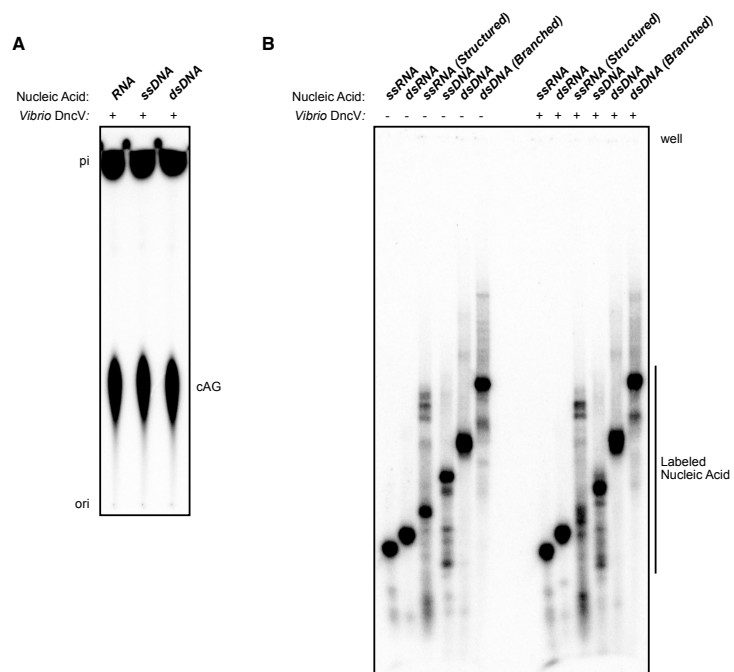
211 A LEANR S FVKGAI FESYVADSI TD SETY ELDS ENVN LALR E ---GDRKWIN SDPK I V EDWFNS CGR I --I GKHLRKYCR FMKAWRDAQ WMD ---VGG P SSI S LMAATVNI 310  
 334 TQEGLR I QNWL SAKVRK ---QLR LK PFY LV PKHAK EGV FQEEWR L S F SHI KE I L NNHGK SKT CC ENK EKCRCR DCLK LMKY L L EQLK ER KDK KHL DSK S YHV K T A FF 443

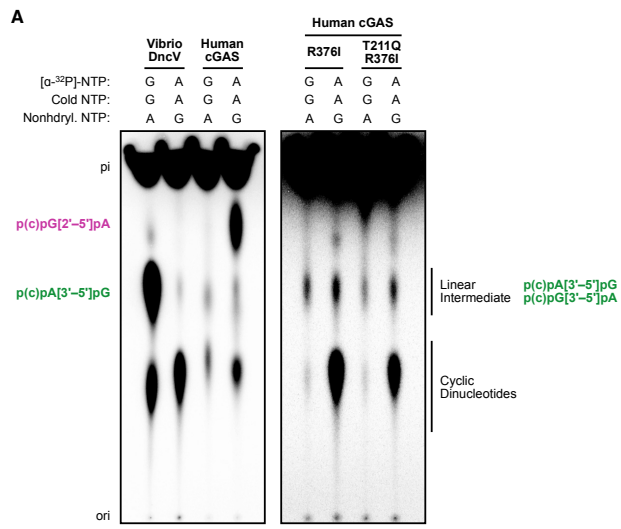
311 NI L DSV ---AHDASDLG ET MK I I AKHL P S EFAR -CV ES PDST DEK P L F PPSYKHGPR EMDI MSKLERL PEI LLSA E SADS K SEALKI NMA FGNRV TNS ELI V LAKA 413  
 444 HVCTQN PQD SQWRK RDLG LCFDNCVYI F LQCLRT E K L E N Y F I -P E N L I S S N L I D R K S E F L T K Q I E T E R N N E ---F -P V F E D F --- 522

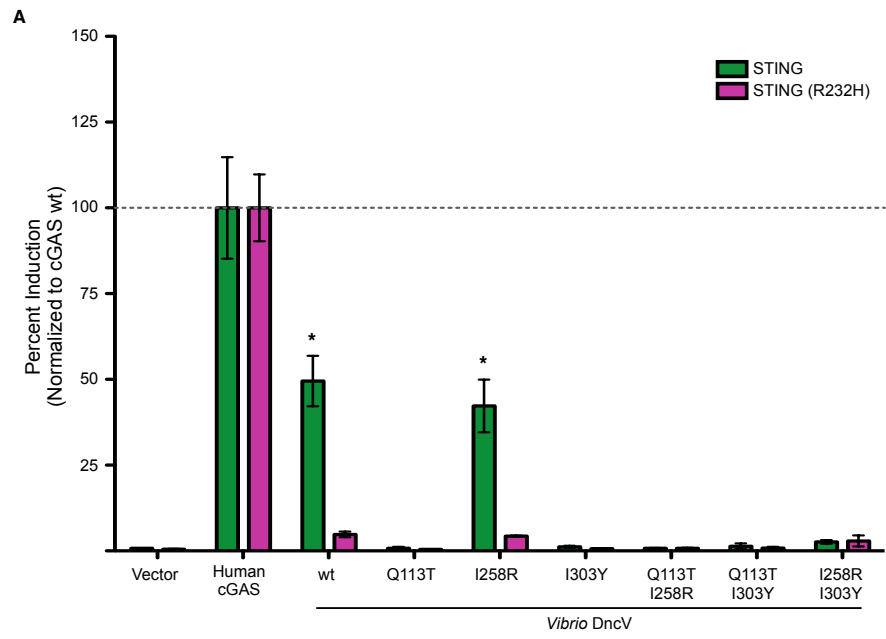
Zn Ribbon











SI Table 1. Crystallographic Statistics, Related to Figures 1, 2, 3 and 4

	DncV • GTP • Apcpp [5' pp(c)pA[3' -5' ]pG]	DncV • Gpcpp	Apo DncV	Apo DncV (Se Anomalous)
<b>Data Collection</b>				
Resolution (Å)	45.34–1.80 (1.89–1.80)	45.19–2.80 (2.95–2.80)	45.12–3.00 (3.16–3.00)	45.37–3.50 (3.83–3.50)
Wavelength (Å)	1.28180	0.97946	0.97946	0.97930
Space group	P 1 2 <sub>1</sub> 1	P 1 2 <sub>1</sub> 1	P 1 2 <sub>1</sub> 1	P 1 2 <sub>1</sub> 1
Unit cell dimensions: a, b, c (Å)	70.33, 59.59, 102.16	70.05, 59.37, 103.53	70.17, 59.11, 102.54	70.44, 59.51, 104.07
Unit cell dimensions: $\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 96.5, 90.0	90.0, 96.0, 90.0	90.0, 95.5, 90.0	90.0, 95.6, 90.0
Molecules per ASU	2	2	2	2
No. reflections: total	565968	72396	58583	139943
No. reflections: unique	76362	20746	16701	11031
Completeness (%)	97.2 (90.9)	97.9 (95.0)	98.0 (98.6)	99.7 (98.8)
Multiplicity	7.4 (7.1)	3.5 (3.5)	3.5 (3.6)	12.7 (11.7)
$I/\sigma$	9.7 (0.6)	5.8 (0.6)	6.4 (1.3)	15.6 (8.4)
CC(1/2) <sup>1</sup> (%)	99.8 (11.9)	98.8 (33.3)	98.5 (40.9)	99.5 (95.7)
Rpim <sup>2</sup> (%)	6.0 (124.2)	10.7 (73.0)	12.4 (69.3)	4.9 (11.1)
No. sites				26
<b>Refinement</b>				
Resolution (Å)	45.34–1.80	45.19–2.80	45.12–3.00	
Free reflections (%)	10	10	10	
R-factor/R-free	17.3/20.6	22.8/24.2	22.5/25.0	
R.M.S. deviation: bond distances (Å)	0.009	0.003	0.004	
R.M.S. deviation: bond angles (°)	1.201	0.800	0.810	
<b>Structure/Stereochemistry</b>				
No. atoms: nonhydrogen, protein	5967	5884	5910	
No. atoms: ligand	124	68		
No. atoms: water	530	5		
Average B-factor: nonhydrogen, protein	40.4	51.0	56.10	
Average B-factor: ligand	33.3	83.1		
Average B-factor: water	43.2	34.0		
Ramachandran plot: most favored regions	99.2%	98%	96%	
Ramachandran plot: additionally allowed	0.8%	2%	4%	
MolProbity <sup>3</sup> score	1.18	1.66	2.27	
Protein Data Bank ID	4TYO	4TXZ	4TXY	

<sup>1</sup>(Karpus and Diederichs, 2012)<sup>2</sup>(Weiss, 2001)<sup>3</sup>(Chen et al., 2010)