Structure-guided reprogramming of human cGAS dinucleotide linkage specificity

Philip J. Kranzusch[1,4], Amy S.Y. Lee[1,2], Stephen C. Wilson[5], Mikhail S. Solovykh[1], Russell E. Vance[1,4], James M. Berger[3]* and Jennifer A. Doudna[1,2,4-6]*

[1]Department of Molecular & Cell Biology, University of California, Berkeley, CA 94720, USA

[2]Center for RNA Systems Biology, University of California, Berkeley, CA 94720, USA

[3]Department of Biophysics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

[4]Howard Hughes Medical Institute (HHMI), University of California, Berkeley, CA 94720, USA

[5]Department of Chemistry, University of California, Berkeley, CA 94720, USA

[6]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

*Correspondence to: jmberger@jhmi.edu (J.M.B.); 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 = α-helix, blue = β-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 = α-helix, blue = β-strand) and for cGAS below the sequence alignment (purple = α-helix, blue = β-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 it's 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 (pIFNβ-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 (Katoh 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.

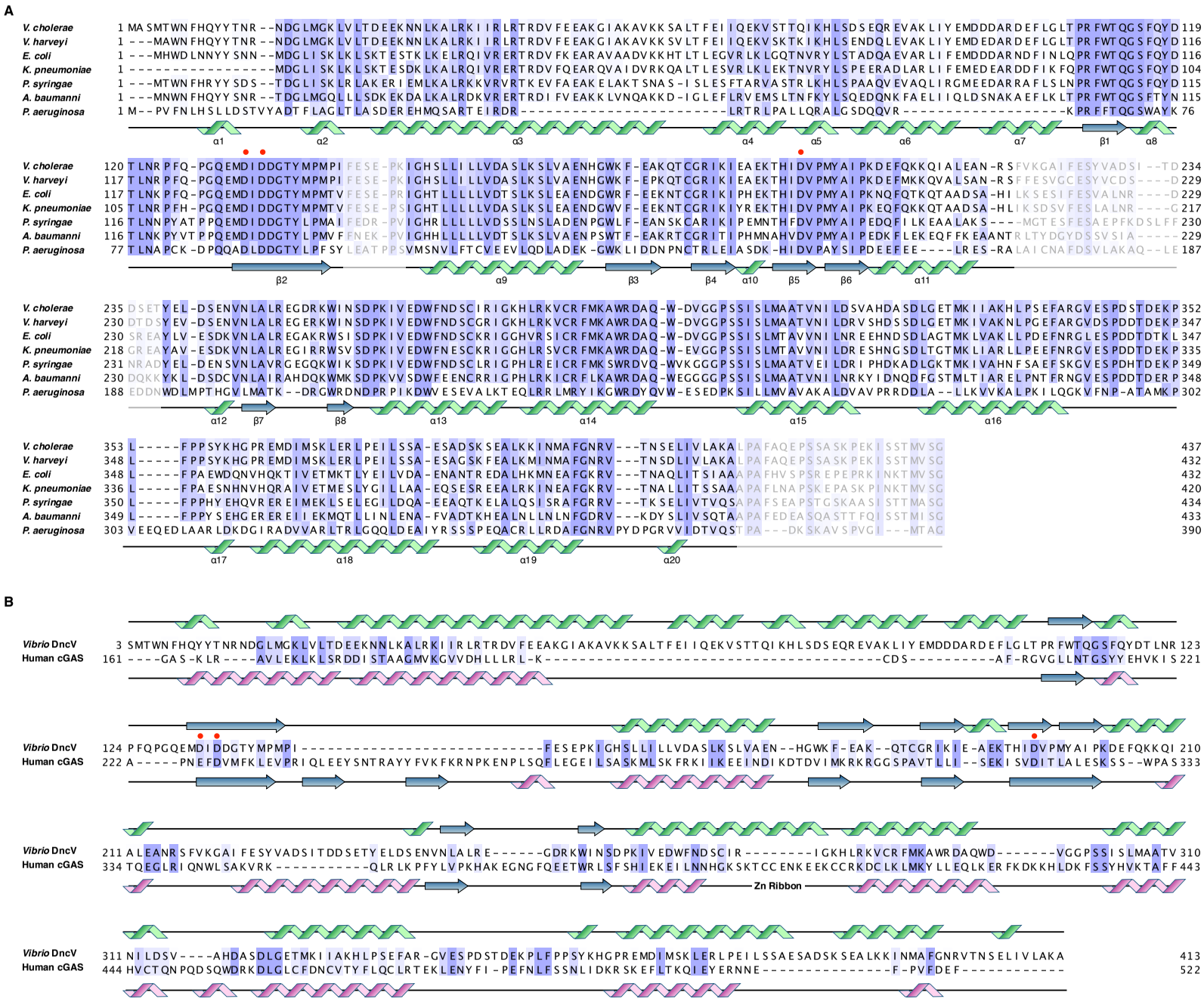
Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular biology and evolution *30*, 772-780.

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.

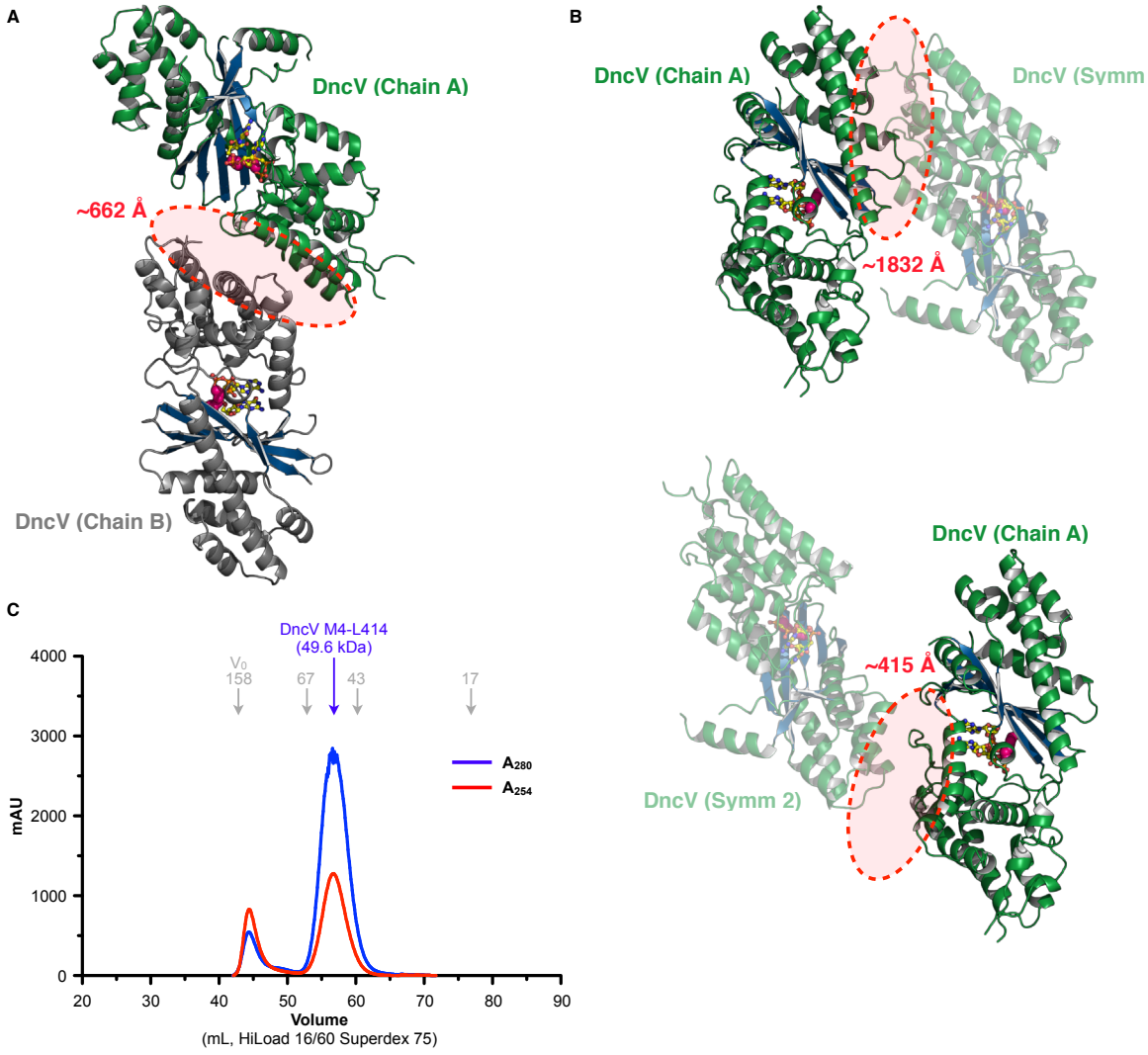
Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.

Weiss, M.S. (2001). Global indicators of X-ray data quality. J Appl Crystallogr *34*, 130-135.

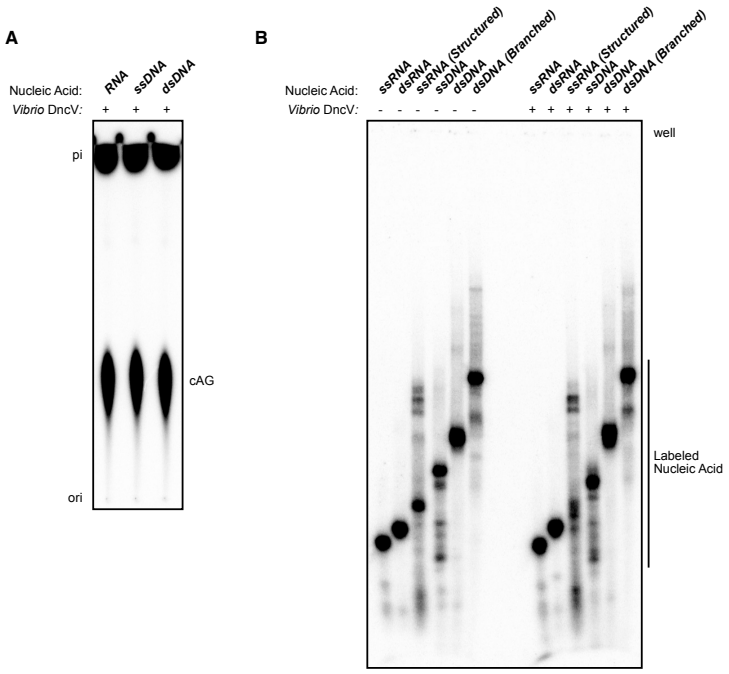
Witte, G., Hartung, S., Buttner, K., and Hopfner, K.P. (2008). Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. Molecular cell *30*, 167-178.


A
V. cholerae
V. harveyi
E. coli
K. pneumoniae
P. syringae
A. baumanni
P. aeruginosa
α1 α2 α3 α4 α5 α6 α7 β1 α8
β2 α9 β3 β4 α10 β5 β6 α11
α12 β7 β8 α13 α14 α15 α16
α17 α18 α19 α20
B
Vibrio DncV
Human cGAS
Zn Ribbon

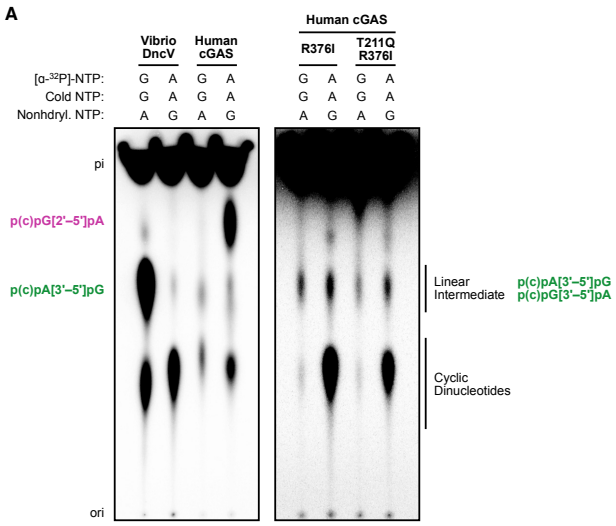
Supplemental Figure 2



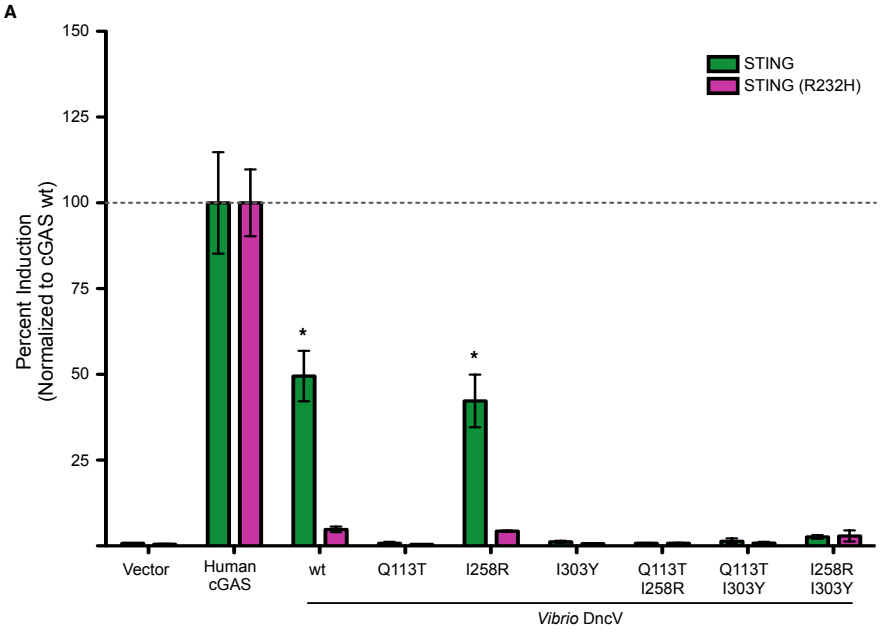
Supplemental Figure 3



Supplemental Figure 4



Supplemental Figure 5



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)**
Data Collection				
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_1 1	P 1 2_1 1	P 1 2_1 1	P 1 2_1 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: α, β, γ (°)	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)
$I/\sigma I$	9.7 (0.6)	5.8 (0.6)	6.4 (1.3)	15.6 (8.4)
CC(1/2)[1] (%)	99.8 (11.9)	98.8 (33.3)	98.5 (40.9)	99.5 (95.7)
Rpim[2] (%)	6.0 (124.2)	10.7 (73.0)	12.4 (69.3)	4.9 (11.1)
No. sites				26
Refinement				
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	
Structure/Stereochemistry				
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[3] score	1.18	1.66	2.27	
Protein Data Bank ID	4TY0	4TXZ	4TXY	

[1](Karplus and Diederichs, 2012)
[2](Weiss, 2001)
[3](Chen et al., 2010)