Supporting materials

DNA-binding Mechanisms of Human and Mouse cGAS: A Comparative MD and MM/GBSA Study

Xiaowen Wang[1,2], Honghui Zhang[1] and Wenjin Li[1,*]

[1] Institute for Advanced Study, Shenzhen University, Shenzhen 518060, China

[2] College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China

[*]**Corresponding author:**

Wenjin Li, Room 341, Administration Building, Institute for Advanced Study, Shenzhen University, Shenzhen 518060, China; E-mail: liwenjin@szu.edu.cn; Tel: +86-755-26942336

Contents

Build complete DNA structures

Initial DNA structures in hGhD (PDB ID: 6CT9) and mGmD (PDB ID: 4O6A) directly downloaded from PDB are structurally incomplete (see Figure S6). The dsDNA structures are then modelled and optimized via several approaches given below to obtain the full-length structure. We here take the hGhD as an example to explain the procedure in detail. The missing DNA segments in mGmD is generated based on the same protocol.

We first rebuilt a standard B-form dsDNA structure with an 18-bp DNA sequence of TTTCGTCTTCGGCAATTT using Avogadro software.[1] Then the new B-form DNA was aligned to the reference structure of crystal DNA by minimizing the RMSD of their DNA segments in the black box in Figure S6.[2] The initial guess of the coordinates of DNA nucleotides highlighted in red and blue in Figure S6 were then taken from the modeled B-form DNA, while the coordinates from crystal structures were used for other nucleotides. This gives the full structure of the cGAS–DNA complex.

This structure is then optimized. First, the structure is solvated using TIP3P water model in a cubic box with the solute being at least 10 Å away from the box. Sodium and chloride ions are added to achieve a salt concentration of 100 mM. The solvated system is then energy minimized with position restraints on the heavy atoms of the DNA with a spring constant of 500 kcal(mol·Å^2). In the resulted DNA structure, the hydrogen bonds between T18 and A19 nucleotides are partly lost. Thus, position restraints on these two nucleotides are then removed, and further energy minimization is performed. After that, energy minimization is again performed with no restraint on the DNA. Finally, 100 ps NVT simulation at 300 K is followed without position restraint to obtain a stable cGAS–DNA structure. The resulted structure is then used as the starting model for hGhD after removing one terminal base pair of the DNA. The subsequent simulation is described in the method “Molecular dynamics simulation” in the main text.

Supplementary Figures

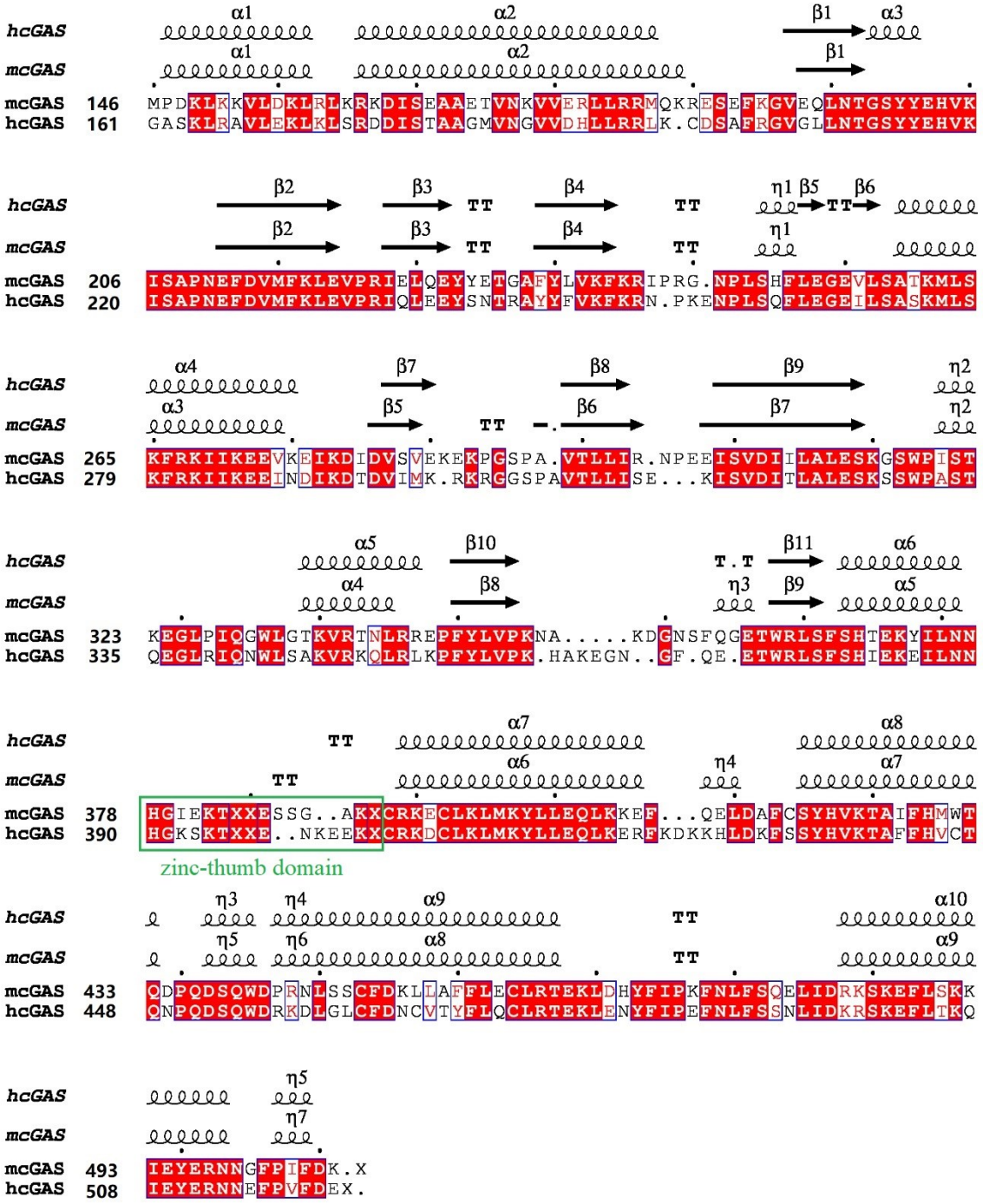

Figure S1. The structure-based sequence alignment for hcGAS (PBD ID: 6CT9) and mcGAS (PDB ID: 4O6A). The secondary structure analysis from the aligned sequence was completed by ESPript 3.0 tool.[3] The hcGAS from hGhD and hGmD possess the residues 161–521 and mcGAS from mGmD and mGhD possess the residues 146–506. The red boxes represent the identically overlapped regions. The elements above the sequences show the secondary structures of hcGAS and mcGAS. The medium (α) and small squiggles (η) describe the α-helices and 3_{10}-helices, respectively. Arrows represent β-strands and TT letters show the β-

turn. The “X” letters represent the CYM residue connected to one zinc ion. The green box is the zinc-thumb domain.

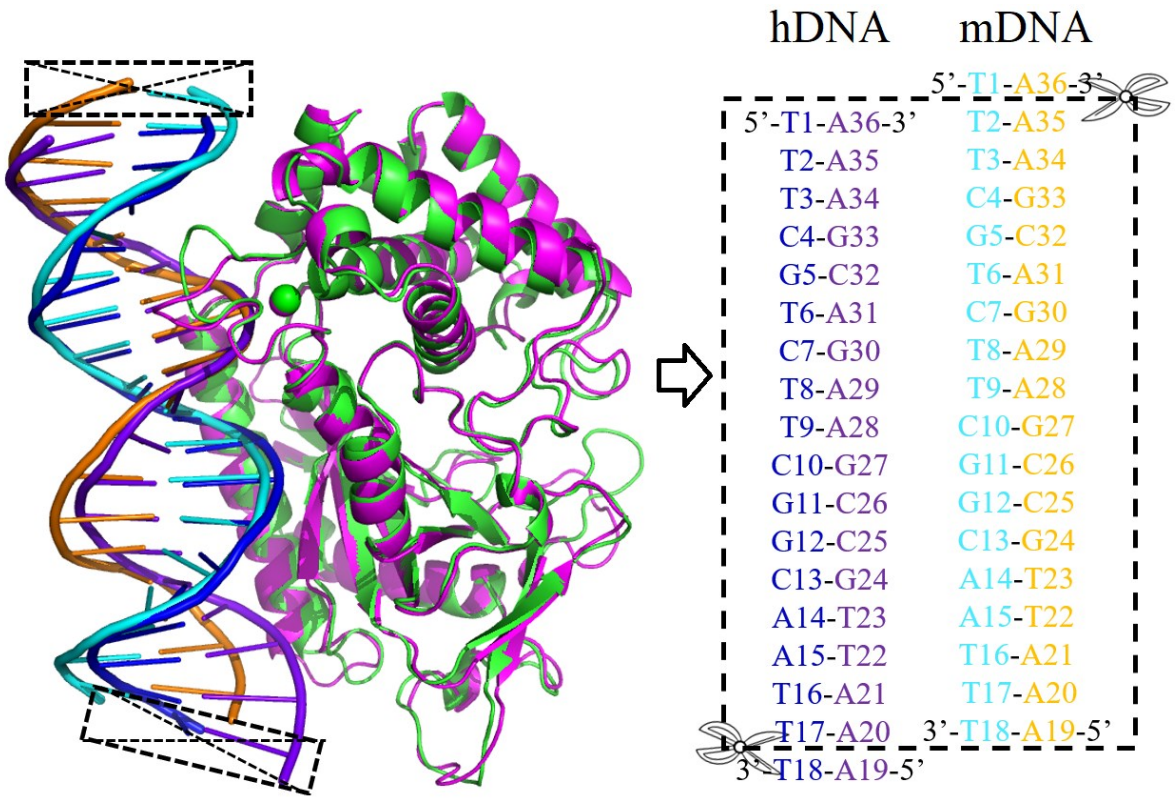

Figure S2. The treatment of the hDNA and mDNA sequences. Left side is the overlapped structure between the hcGAS–DNA complex and the mcGAS–DNA complex, with the structure of the full DNA length being modelled from the crystal structure of PDB IDs 6CT9 and 4O6A, respectively. Right side is the alignment of the two DNA sequences according to their positions in the overlapped structure in the left. The colors of cGAS and DNA are the same as the description of Figure 1 in the main text. After structural alignment between hGhD and mGmD, we found that two identical DNA structures did not completely overlap, but shifted by one base relative to each other. To achieve a one-to-one correspondence between the residue–nucleotide interactions in hcGAS–DNA and mcGAS–DNA complexes using cGAS structure as the reference, one terminal base pair of the original DNA sequence in the two structures (T18–A19 in the hcGAS–DNA complex and T1–A36 in the mcGAS–DNA complex) was removed, which gave the hDNA of the sequence d(TTTCGTCTTCGGCAATT) and the mDNA of the sequence d(TTCGTCTTCGGCAATTT).

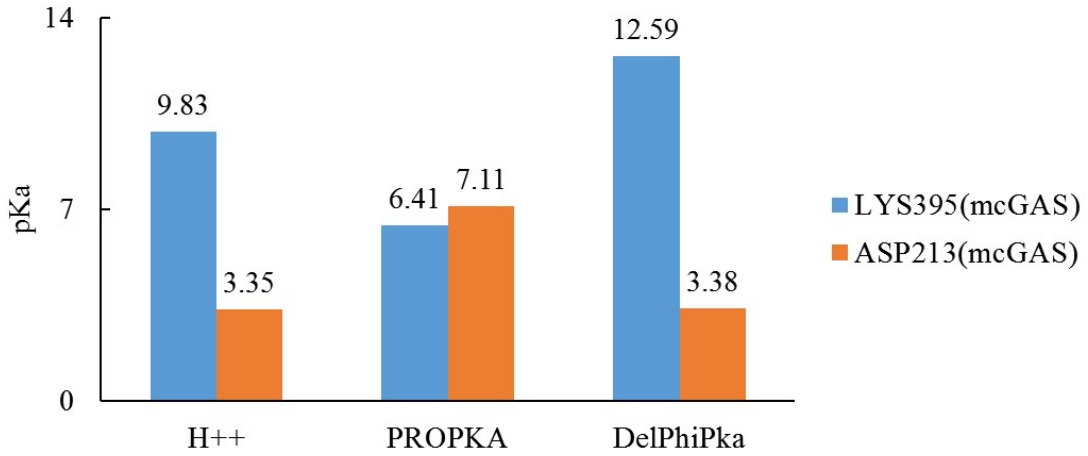

Figure S3. The pKa values for the residues LYS395 and ASP213 in mcGAS as predicted by H++, PROPKA, and DelPhiPKa tools. Most residues are predicted to possess the default protonation states by all the three methods above, while inconsistent protonation states are assigned for two residues: mcGAS LYS395 and ASP213. As shown, the pKa values of mcGAS LYS395 and ASP213 predicted by PROPKA are 6.41 and 7.11, respectively, while they are considered as the default protonation states by H++ and DelPhipKa predictions. Therefore, we choose to treat the hcGAS and mcGAS residues ARG and LYS in a protonation state and HIS, GLU, ASP in a deprotonation state.

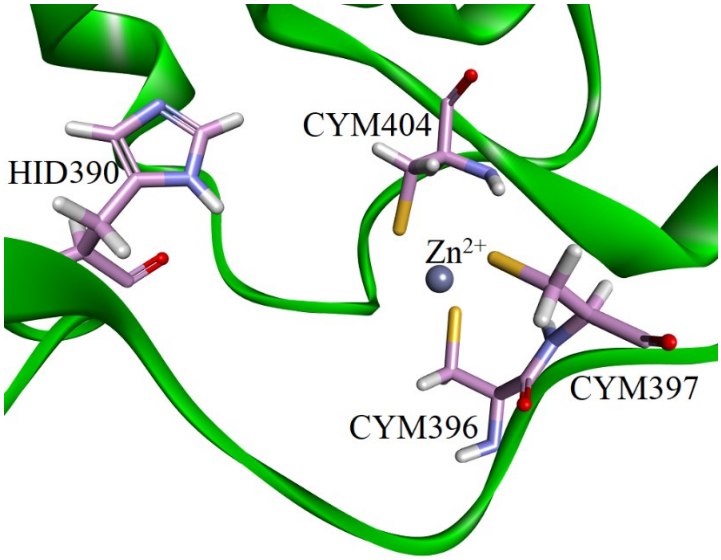

Figure S4. A snapshot of the zinc-thumb domain in hcGAS during 50 ns simulation without restraints. The four residues of HID390, CYM396, CYM397, and CYM404 are shown as sticks, where C-, N-, O-, H-, and S-atoms are displayed as pink, blue, red, white, and yellow, respectively. hcGAS protein is shown as green cartoon and zinc ion is shown as grey ball.

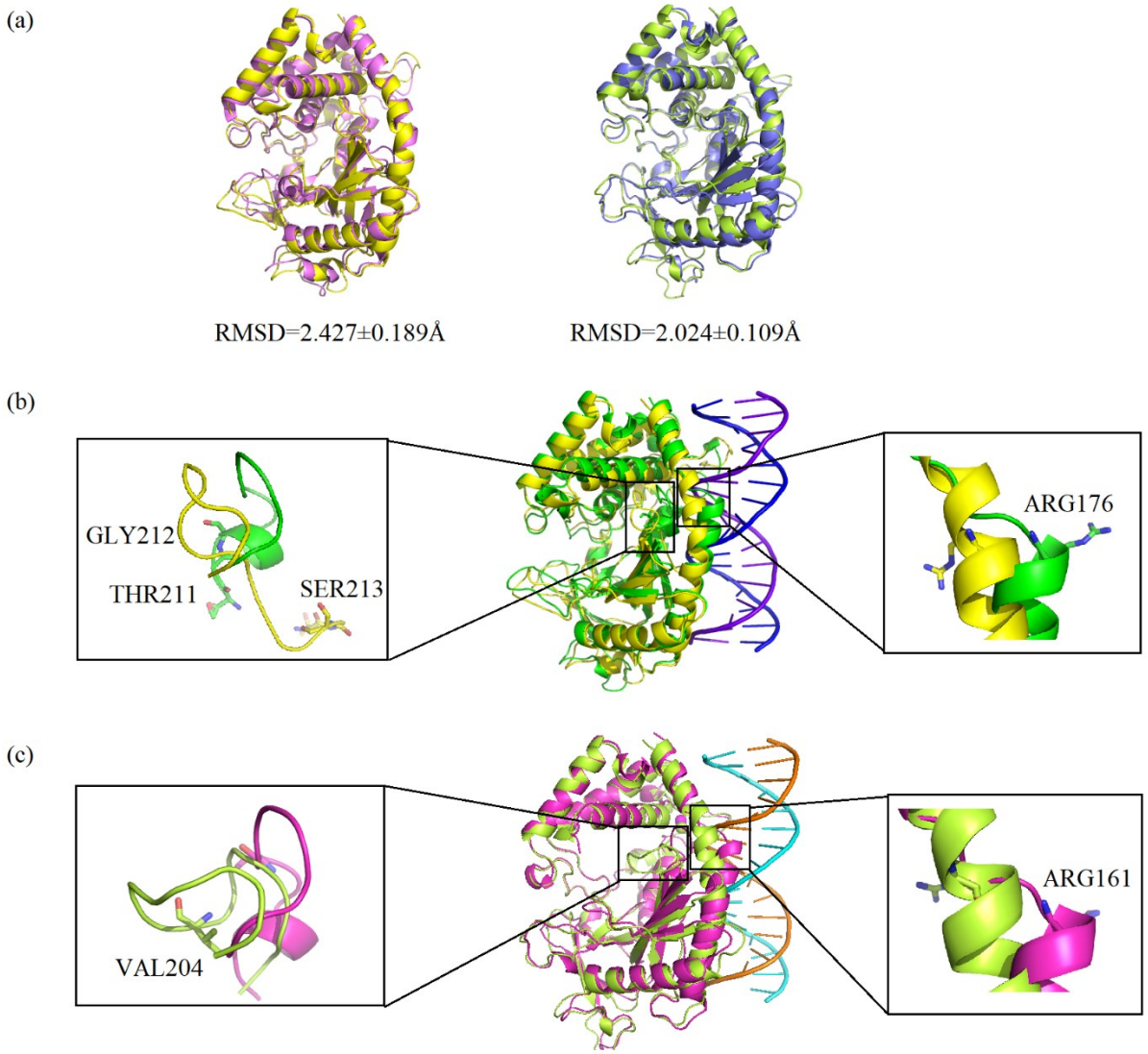

Figure S5. (a) Left: Superposition of apo hcGAS between MD structure (yellow) and crystal one (PDB ID: 4KM5, pink). Right: Superposition of apo mcGAS between MD structure (limon) and crystal one (PDB ID: 4K8V, blue). Note that wild-type LYS185 and LEU195 in the hcGAS crystal structure (PDB ID: 4KM5) were mutated to ASN185 and ARG195, respectively, in order to compare with mutant hcGAS–DNA complex (PDB ID: 6CT9). (b) Superposition of apo hcGAS (yellow) and the complex hGhD (PDB ID: 6CT9, green protein and blue&purple strands) obtained from MD trajectory. (c) Superposition of apo mcGAS (limon) and the complex mGmD (PDB ID: 4O6A, magenta protein and cyan&orange strands) obtained from MD trajectory. The close-up pictures of the activation loop close to its catalytic center and the kink region are shown in (b) and (c) for hcGAS and mcGAS, respectively. The predicted important residues were shown as sticks.

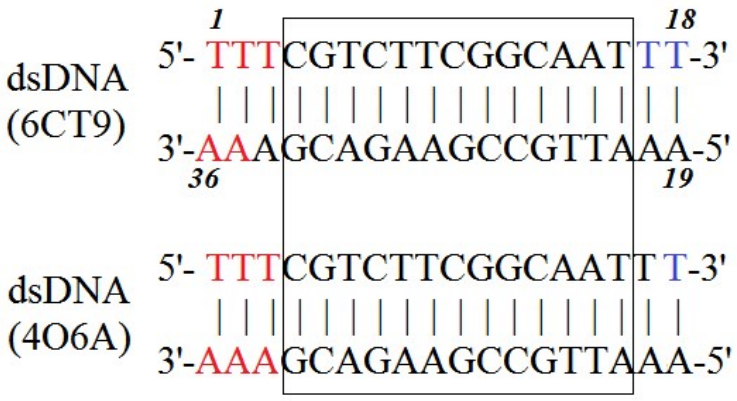

Figure S6. Double stranded DNA sequences in hcGAS–DNA (PDB code: 6CT9, top) and mcGAS–DNA (PDB code: 4O6A, bottom) complexes. The nucleotides resolved in the PDB structures are in black, while the missing nucleotides are in either red or blue. The segment in the black box is used to align the modelled B-form DNA using Avogadro software to the one in the crystal structure.

Supplementary Tables

Table S1. The averaged RMSD (Å) values on three parallel MD trajectories of whole systems, cGAS alone removing zinc ion, DNA alone, and zinc-thumb for hGhD, mGmD, hGmD, and mGhD.

models[a]	whole	protein	DNA	zinc-thumb
hGhD	2.44(0.17)	2.28(0.18)	1.92(0.11)	0.63(0.11)
mGmD	2.36(0.06)	2.18(0.04)	2.04(0.09)	0.51(0.10)
hGmD	2.43(0.13)	2.13(0.13)	2.10(0.12)	0.54(0.11)
mGhD	2.39(0.08)	2.15(0.05)	1.92(0.09)	0.48(0.08)

[a] The standard deviations (Å) of RMSDs are shown in parentheses. All-atoms, backbone-atoms, heavy-atoms, and heavy-atoms RMSDs for the whole system, protein alone, DNA alone, and zinc-thumb.

Table S2. Various components for the averaged binding free energies of LYS173 in hcGAS and ARG158 in mcGAS.

residues	ΔE_{vdw}	ΔE_{ele}	$\Delta G_{gb/solv}$	ΔG_{gb}[a]	$\Delta G_{np/solv}$	ΔG_{bind}
LYS173(hcGAS)	–2.3	–308.7	306.8	–1.8	–0.4	–4.5
ARG158(mcGAS)	–3.6	–331.0	324.4	–6.6	–0.5	–10.8

[a] $\Delta G_{gb} = \Delta G_{gb/solv} + \Delta E_{ele}$; All the units are in kcal/mol.

References

1. M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminformatics*, 2012, **4**, 17.
2. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605-1612.
3. X. Robert and P. Gouet, *Nucleic Acids Res.*, 2014, **42**, W320-W324.