PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2020, **22**, 26390

DNA-binding mechanisms of human and mouse cGAS: a comparative MD and MM/GBSA study†

Xiaowen Wang,[ab] Honghui Zhang[a] and Wenjin Li *[a]

Received 6th August 2020,
Accepted 23rd October 2020

DOI: 10.1039/d0cp04162a

rsc.li/pccp

Cyclic GMP–AMP synthase (cGAS) can detect the presence of cytoplasmic DNA and activate the innate immune system *via* the cGAS–STING pathway. Although several structures of cGAS–DNA complexes were resolved recently, the molecular mechanism of cGAS in its recognition of DNA has not yet been fully understood. In order to reveal the subtle differences between human and mouse cGAS in terms of their DNA-binding mechanisms, four systems, both human and mouse cGAS in complex with two different DNA sequences of equal length, were studied by molecular dynamics simulations and molecular mechanics/generalized Born surface area analysis. Several residues, including ARG176/ARG161, ARG195/ARG180, ASN210/ASN196, LYS384/LYS372, CYM397/CYM385, LYS403/LYS391, LYS407/LYS395, and LYS411/LYS399, were identified to be the common key residues in the recognition of DNA for cGAS in both humans and mice. In addition, four residue pairs LYS173/ARG158, ASP177/LYS162, CYS199/LYS184, and GLU398/SER387 were suggested to be the major residues that make human cGAS and mouse cGAS different in terms of their binding to DNA. Besides the well-known zinc-thumb domain, two residues at the kink of the spine helix were also proposed for the first time to be the major binding motifs in cGAS–DNA interaction.

Introduction

Mitochondrial stress, genomic instability or invasion of foreigners can lead to the existence of cytoplasmic double stranded DNA (dsDNA), which is then detected by cyclic GMP–AMP synthase (cGAS), a kind of nucleotidyltransferase (NTase).[1] Upon activation by DNA, cGAS produces second messenger cyclic GMP–AMP (cGAMP) to activate the stimulator of interferon genes (STING) and the downstream innate immune response.[2–4] Although the stimulation of cGAS by dsDNA from pathogens helps the host to defend against the infection of pathogens, unconstrained activation of cGAS due to endogenous DNA can also activate the cGAS–STING pathway, which in turn causes autoimmune diseases such as systemic lupus erythematosus and Aicardi–Goutières syndrome.[5–7] Therefore, cGAS is a potentially vital drug target for such autoimmune diseases.

Human cGAS (hcGAS) contains 522 residues and shares ~60% sequence homology with mouse cGAS (mcGAS), which is composed of 507 residues. Recently, the structures of several cGAS–DNA complexes were resolved.[8–11] For example, Gao *et al.* solved the crystal structures of apo mcGAS and its complexation with DNA.[10] The structures of complexes composed of dimeric cGAS bound to two molecules of DNA (2 : 2 complex) for both mcGAS and hcGAS were also reported.[8,9,12,13] In most of the known structures of cGAS–DNA complexes, the structural information of a highly conserved C terminal domain (specifically, residues 161–522 in hcGAS and residues 147–507 in mcGAS) instead of the full-length cGAS protein was provided as the N terminal domain is largely unstructured.[14] The C terminal domain, also known as the NTase domain, contains an NTase core domain, a male abnormal 21 homology domain, and one unique zinc-thumb (or zinc-ribbon) motif.[3,11] There are two distinct DNA binding sites (A-site and B-site) in the NTase domain as identified in the 2 : 2 complex.[8,13,14] The DNA A-site is the cGAS–DNA binding interface in the 1 : 1 complex (the complex of one cGAS protein and one DNA molecule), and it contains one zinc-thumb domain and a so-called 'spine' helix, which is composed of α1 and α2 helices and a kink or turn in between. The DNA B-site is the interface formed in the dimer of the 1 : 1 complex. Very recently, an additional binding site in hcGAS was also identified.[12]

In this study, we focus on the DNA A-site, as it provides more than 60% of the buried surface between cGAS and DNA and is the most studied one with a handful of both structural and kinetic data available.[8,9,13,15] Many valuable residues at the DNA A-site were identified by mutation experiments. For example, single-residue mutations of R158E, K160E, R161E, S165E, K372E, and K395E in the last half of the spine helix or the zinc-thumb

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

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

† Electronic supplementary information (ESI) available: The details on building of complete DNA structures, Fig. S1–S6, and Tables S1, S2. See DOI: 10.1039/d0cp04162a

domain (significantly) and K151E, K162E, and R180E in the spine helix (moderately) reduced the mcGAS activity.[13] It was reported in a mutagenesis study that several positively charged residues in hcGAS contributed strongly to its binding to DNA, and they were K173 and R176 in the spine helix, K384 near the activation loop, K400 and K403 at zinc-thumb domain, and K407 and K411 near the zinc-thumb domain.[9] Moreover, the combination of K400E and K403E mutants abolished the hcGAS activity.[16] Zhou and coworkers suggested the importance of N187 and R195 in hcGAS as K187N and L195R mutations can enhance its binding affinity to the short DNA.[8] It was reported that hcGAS and mcGAS showed differences in both binding affinity and length-dependence in their interaction with DNA in various assays, such as isothermal titration calorimetry (ITC) and fluorescence anisotropy (FA).[11,15] Specifically, the dissociation constant (K_d) of wild type (WT) hcGAS and mcGAS in its binding with 20-base pair (bp) DNA were determined by ITC to be 16.3 μM and 9.4 μM, respectively, while the K_d of WT hcGAS and mcGAS were 0.59 μM and 1.72 μM, respectively, as measured by FA.[15] In another study with ITC, the K_d of WT mcGAS to a 20-bp DNA was reported to be 19.1 μM.[13] Zhou and coworkers reported that WT mcGAS can be activated in a length-dependent manner by both short (17-bp) and long (45-bp) DNA, while WT hcGAS can be activated by 45-bp DNA and show marginal activity in the presence of 17-bp DNA.[8] However, human-specific K187N/L195R mutation resulted in strikingly high activity in response to 17-bp DNA of low concentration.[8]

Although extensive experimental studies on the cGAS–DNA interaction have been made, we are still far from a thorough understanding of its molecular mechanism, especially at the atomistic level. As an important complementary to experimental results, molecular dynamics simulations can provide real-time dynamic behavior of biomolecules and have been widely used in elucidating the binding mechanism of protein–DNA complexes.[17–19] Binding free energy estimations can provide a direct comparison to the binding affinity measured in experiments, and thus is a common way to validate the computational model. In addition, functionally important residues can be easily identified from their energetic contributions to the protein–DNA interaction. Among the various free energy calculation approaches, the molecular mechanics generalized-Born surface area (MM/GBSA) was a popular choice.[20–22] In MM/GBSA, only the free energy difference between two end-points was considered and the solvation effects were approximated by a continuum solvation model, and thus the binding free energy can be evaluated with a moderate accuracy and a very low computational cost.[23] To date, a combination of molecular dynamics (MD) and MM/GBSA approaches has been successfully applied in mechanistic studies of, for example, protein–protein interactions, protein–DNA interactions, and protein–ligand interactions.[20,22,24–27]

In order to unveil at the single-residue level the differences between hcGAS and mcGAS in their recognition of dsDNA from a dynamic and energetic viewpoint, we thus paid our attention to the cGAS–DNA complexes in humans and mice resolved by experiments. In addition, to explore the sequence specificity of cGAS, two different DNA sequences, d(TTTCGTCTTCGGCAATT) or hDNA and d(TTCGTCTTCGGCAATTT) or mDNA, with one molecule (1 : 1 complex) configuration of cGAS were selected as well. Thus, the complexation of both hcGAS and mcGAS with hDNA and with mDNA were investigated in this study. In total, there were four cGAS–DNA models: the hcGAS–hDNA complex (hGhD), the hcGAS–mDNA complex (hGmD), the mcGAS–hDNA complex (mGhD), and the mcGAS–mDNA complex (mGmD).

In this study, MD and MM/GBSA methods were applied to the four cGAS–DNA models. The computational models were validated by root mean square displacement (RMSD) analysis and systematic comparisons to experimental results, such as binding affinity and mutation studies. The dynamic details in the cGAS–DNA interaction were revealed *via* hydrogen bonding and salt-bridge analysis. From an energetic viewpoint, we provided a qualitative comparison between the importances of residues at the protein–DNA interface in their recognition of DNA. Most importantly, the similarities and differences between hcGAS and mcGAS were identified based on the results of per-residue energy decomposition analysis. Furthermore, residue–nucleotide pairwise interaction analysis revealed the 'fingerprint' of cGAS–DNA interaction, and the sequence preference was also touched.

Materials and computational methods

Starting models

The hDNA is derived from the DNA sequence in the crystal structure (PDB ID: 6CT9 with the resolution of 2.26 Å) of a hcGAS–DNA complex in the Protein Data Bank, in which a hcGAS mutant (K187N, L195R) is used.[8] The crystal structure (PDB ID: 6CT9) was thus used as the starting model for the hGhD model. The hcGAS mutant instead of the WT hcGAS was studied, because most of the experimental results available are for the mutant and thus provide a better comparison to the computational work here. In the following, we used hcGAS to denote the mutant hcGAS, while the WT hcGAS was named as $hcGAS^{WT}$. For hcGAS, only the residues 161–521 were included in the model. The structure of several residues (loop structures: 255–258, 292–294, 300–302, and 366–368) was unresolved in the crystal structure, and was modelled using the MODELLER software.[28] In the process of modelling protein structure, another hcGAS–DNA complex (PDB ID: 6EDB with the resolution of 3.21 Å) was chosen as the template, sharing 100% identity in sequence 161 to 521 with the entry 6CT9. The missing segments were further refined by the standard loop modelling protocol.[29] The mDNA is derived from the DNA sequence in the crystal structure (PDB ID: 4O6A with the resolution of 1.86 Å) of a mcGAS–DNA complex,[9] and thus the crystal structure (PDB ID: 4O6A) was the starting model for the mGmD model. For mcGAS, only the residues 146–506 were included in the model. For a direct comparison between hcGAS and mcGAS, the two cGAS were prepared to be of the same length and the positions of their residues were largely

overlapped in a structural comparison (Fig. S1, ESI†). Note that the two DNA sequences used to obtain the structures of PDB IDs 6CT9 and 4O6A in the X-ray experiments are identical;[8,9] mcGAS and hcGAS however bind to the dsDNA in different configurations. Actually, when overlapping the two structures with the structure of cGAS as a reference, the two identical DNA structures do not completely overlap, and instead their positions are shifted by one base relative to each other (Fig. S2, ESI†). To achieve a one-to-one correspondence between the residue–nucleotide interactions in hcGAS–DNA and mcGAS–DNA complexes using the 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 hDNA of a sequence d(TTTCGTCTTCGGCAATT) and mDNA of a sequence d(TTCGTCTTCGGCAATTT). The residues in hcGAS were named according to their index in the crystal structure (PDB ID: 6CT9), and the index in the crystal structure (PDB ID: 4O6A) was used to name the residues in mcGAS.

Missing DNA nucleotides in the two crystal structures were constructed from the structures of their standard B-DNA form with the Avogadro software.[30] More details about the predictions of complete DNA structures are provided in the section ''Build complete DNA structures'' in the ESI.† All crystal water molecules were retained. The hGmD model was thus generated by docking the mDNA in the mGmD model to the hcGAS in the hGhD model, while the mGhD model was constructed by docking the hDNA in the hGhD model to the mcGAS in the mGmD model. All the homology modeling and molecular docking were performed with the UCSF Chimera software.[31] All four models constructed are shown in Fig. 1. For the zinc-thumb, HIS390 in hcGAS and HIS378 in mcGAS were protonated in the N_δ position, residues CYS396, CYS397, and CYS404 in hcGAS and CYS384, CYS385, and CYS392 in mcGAS were deprotonated (Fig. 2). The protonation states for other titratable residues in both hcGAS and mcGAS were determined to possess their default protonation states by combining the results from the H++, DelPhiPKa, and PROPKA website tools (see Fig. S3 for more details, ESI†).[32] All the 3D structural graphics were rendered *via* PyMOL 1.6.0.[33]

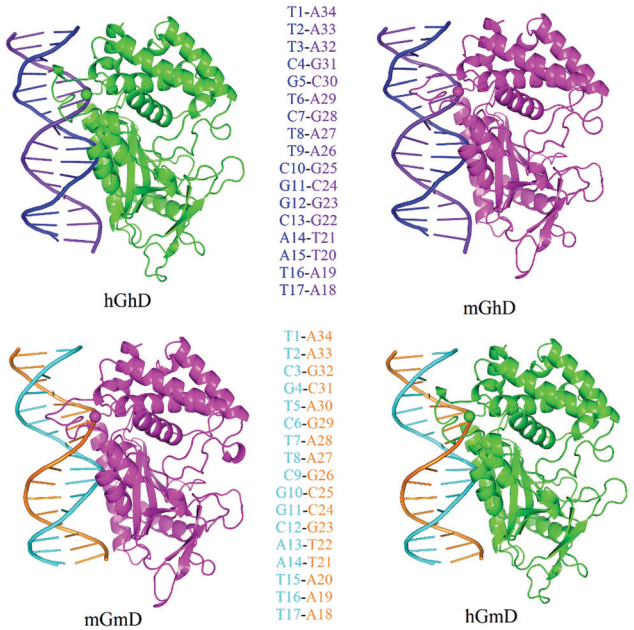

Fig. 1 The starting models of hGhD, mGmD, mGhD, and hGmD are displayed as cartoons. hcGAS and mcGAS are shown in green and magenta, respectively. The zinc ions are shown in spheres. The two strands of hDNA are colored in blue and purple, respectively, while the ones of mcGAS are in cyan and orange, respectively. The sequences of the two DNA are shown in the middle in the same color as their structures.

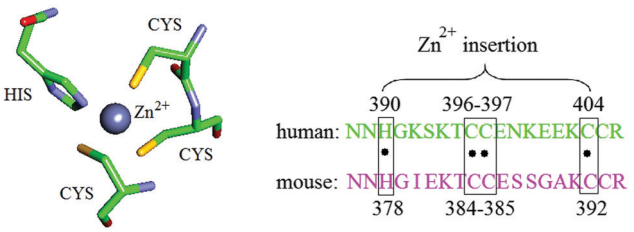

Fig. 2 Left: The structure of the zinc-thumb Zn[HIS][CYS]$_3$ in cGAS. The grey ball is the zinc ion. The four zinc-ion-coordinated residues are given as sticks, where C, O, N, and S atoms are shown in green, red, blue and yellow, respectively. The hydrogen atoms are hidden for clarity and only heavy atoms are displayed. Right: The full sequence of the zinc-thumb domain. Four zinc-ion-coordinated residues HID390, CYM396, CYM397, and CYM404 in hcGAS and HID378, CYM384, CYM385, and CYM392 in mcGAS are highlighted in black boxes.

Molecular dynamics simulation

MD simulations were performed using the 2019 version of the GROMACS program.[34] We chose the AMBER14SB force field for the proteins (hcGAS and mcGAS) and the parmbsc1 parameters for DNA.[35,36] All the initial models were located inside the cubic box (78 Å × 78 Å × 78 Å, the dimension of the cubic box) using the TIP3P water model.[37] Approximately 31 000 water molecules were added in the cubic box of each system to solvate the complex. Water molecules were then randomly replaced by 76 sodium and 56 chloride ions to reach electroneutrality and to mimic the experimental ionic strength of 100 mM in the crystallization of both hcGAS–DNA and mcGAS–DNA complexes.[8,9] Sodium salt was often used in the simulation of protein–DNA complexes,[38–41] although intracellular concentrations of sodium are generally much lower than potassium in eukaryotic cells. For each model, energy minimization with a maximum of 5000 steps was carried out as follows: first, four energy minimization steps were performed with position restraints on the heavy atoms, in which the spring constants of 500, 250, 100, and 50 kcal (mol Å^2)$^{-1}$ were used, respectively. Then, the resulted system was minimized without position restraints. After energy minimization, 1 ns MD simulation in the *NVT* ensemble and 1 ns simulation in the *NPT* ensemble were successively performed with the positions of all heavy atoms being restrained with a force constant of 500 kcal (mol Å^2)$^{-1}$. The position restraints were gradually released *via* five steps of 500 ps *NPT* simulations with the force constants of 250, 100, 50, 10,

and 5 kcal (mol Å^2)$^{-1}$ forces, respectively, for the heavy atoms. Then, 500 ps simulation in the *NPT* ensemble with position restraints (force constant of 5 kcal (mol Å^2)$^{-1}$) on backbone atoms was followed.

Finally, with the starting structures taken from the final 500 ps simulation, three 50 ns production runs were carried out at the *NPT* ensemble. Since HID390 in hcGAS and HID378 in mcGAS were observed to lose their interactions with Zn^{2+} ions in 50 ns simulations without restraints to the zinc-thumb domain (Fig. S4, ESI†), a distance restraint with a force constant of 5 kcal (mol Å^2)$^{-1}$ was thus applied in the production runs for the coordinations between Zn^{2+} ion and four special residues (HID390, CYM396, CYM397, and CYM404 in hcGAS and HID378, CYM384, CYM385, and CYM392 in mcGAS) (Fig. 2). The time step was 2 fs. Temperature and pressure were kept constant at 300 K and 1 bar, respectively. Two temperature coupling groups were used, one for cGAS, Zn^{2+} and dsDNA, and the other for water, Na^+ and Cl^-. In the production runs, the velocity-rescaling thermostat was applied for temperature coupling,[42,43] while the Parrinello–Rahman approach was applied for constant pressure control.[44,45] The SHAKE algorithm was used to constrain covalent bonds involving hydrogen atoms.[46,47] The long-range electrostatic interactions were treated by particle mesh Ewald (PME).[48] The cutoff values of van der Waals and electrostatic were set to 12 Å. The simulation trajectories were saved every 100 ps.

MM/GBSA binding estimation

In this work, we applied an MM/GBSA method to evaluate the binding free energy between cGAS and DNA.[49] Generally, the free energy change when the cGAS protein binds to DNA in solvents could be written as:

$$\Delta G_{\text{bind}} = G_{\text{cGAS–DNA}} - (G_{\text{cGAS}} + G_{\text{DNA}}), \quad (1)$$

where, G_{cGAS}, G_{DNA} and $G_{\text{cGAS–DNA}}$ represent the free energies of isolated cGAS, DNA and cGAS–DNA complex, respectively. The free energy can be expressed as:

$$G = G_{\text{sol}} - TS + E_{\text{MM}}, \quad (2)$$

where G_{sol} is the free energy of solvation, the free energy required to move the solute from the vacuum into the solvent. T and S are the absolute temperature and entropy, respectively. E_{MM} is the free energy defined by the potential energy in vacuum, which was the summation of bonded and nonbonded interactions. The bonded part of E_{MM} was assumed to be zero in a single-trajectory setup, which was adapted in this study due to its simplicity and similar accuracy compared to a multi-trajectory or three-average setup.[25,50] Hence eqn (1) can be rewritten as:

$$\Delta G_{\text{bind}} = \Delta G_{\text{solv}} - T\Delta S + \Delta E_{\text{vdW}} + \Delta E_{\text{ele}}, \quad (3)$$

where E_{vdW} and E_{ele} are the nonbonded part of E_{MM} due to van der Waals (vdW) and electrostatic interactions, respectively. The solvation free energy ΔG_{sol} can be further decomposed into polar ($\Delta G_{\text{gb/solv}}$) and nonpolar ($\Delta G_{\text{np/solv}}$) contributions.

To explore the impact of each residue in the cGAS–DNA interaction, the per-residue binding free energy decomposition was usually performed. By neglecting the entropic part ($T\Delta S$), the free energy of a single residue (ΔG_{res}) can be divided into three terms:

$$\Delta G_{\text{res}} = \Delta G_{\text{gb/solv}} + \Delta G_{\text{ng/solv}} + \Delta E_{\text{MM}}, \quad (4)$$

where, ΔE_{MM} contains the electrostatic and vdW interactions of the residue in vacuum, and $\Delta G_{\text{gb/solv}}$ and $\Delta G_{\text{np/solv}}$ are the polar and nonpolar parts of the solvation free energy of the residue, respectively. The interaction free energy between residue and nucleotide was also evaluated as follows:

$$\Delta G_{\text{RD}}^{\text{inter}} = \sum_{i\in\text{R}}\sum_{j\in\text{D}} E_{ij}^{\text{vdW}} + \sum_{i\in\text{R}}\sum_{j\in\text{D}} G_{ij}^{\text{np/solv}} + \sum_{i\in\text{R}}\sum_{j\in\text{D}} E_{ij}^{\text{ele}} + \sum_{i\in\text{R}}\sum_{j\in\text{D}} G_{ij}^{\text{gb/solv}}. \quad (5)$$

In eqn (5), $\Delta G_{\text{RD}}^{\text{inter}}$ is the intermolecular free energy between a residue (R) in cGAS and a nucleotide (D) in DNA.[51,52] i and j represent the ith atom and jth atom of each residue or nucleotide, respectively.

Binding free energy was generally calculated by using MM/GBSA and molecular mechanics–Poisson Boltzmann-solvent accessible surface area (MM/PBSA) methods, especially for molecular docking and modelling.[20,24,27,53,54] The performance of MM/GBSA and MM/PBSA was usually assessed for the large biomolecules, such as protein–ligand, protein–protein and protein–DNA interactions,[20,25,27,53] which were developed using fast and reliable tools to estimate the binding free energy based on several molecular dynamics software and APBS programs.[55–57] The g_mmpbsa was a useful software to calculate the binding free energy, combined with the GROMACS program. However, it held limitations to predict the protein–DNA interactions due to the overestimation of electrostatics without non-bonded interaction cutoff settings.[57] The MMPBSA.py was an efficient post-processing program to estimate the binding free energy included in the AmberTools package,[55] which treated the MM calculation with cutoff settings more accurately for the highly charged systems over the g_mmpbsa. MM/GBSA was assessed as a reliable method for the study of nucleic acids.[49] All the MM/GBSA calculations were performed using the AMBER18 program and the post-processing program MMPBSA.py in this work.[58] The solvation free energy was evaluated using the generalized Born (GB) and molecular surface (MS) models. Specifically, the GB method was applied to calculate the $G_{\text{gb/solv}}$.[59] The temperature was set to 300 K and the grid spacing was set to 0.5 Å. The concentration of charged ions was 100 mM with the radius of chloride and sodium ions being 1.81 and 0.95 Å, respectively. The total non-polar solvation $G_{\text{np/solv}}$ free energy was estimated from two terms: the dispersion term and cavity term. The dispersion term was computed with a surface-based interaction method closely related to the PCM (polarized continuum model) solvent for the quantum mechanics program.[60] The probe radius and water density were set to 0.557 Å and 1.129 g cm^{-3}, respectively. The cavity term was obtained from the function of γMS + b.

The solvent-accessible volume was used for the MS model.[61] The surface tension of 0.0072 kcal (mol Å^2)$^{-1}$ was used. The default constant γ and fitting parameter b were 0.0090 kcal (mol Å^2)$^{-1}$ and −0.136 kcal mol^{-1}, respectively. The probe radius was 1.4 Å.

Results and discussion

Stability of cGAS–DNA complexes

We evaluated the stability of all four cGAS–DNA complexes by monitoring the change of RMSD along with the simulation time (Fig. 3), which quantifies the fluctuation of simulated structures to the corresponding starting models. The RMSDs of each cGAS–DNA complex for all four systems reached equilibrium within 5 ns and thereafter fluctuated at around 2.4 Å. By averaging the RMSDs of the three 50 ns production runs, the averaged RMSD values of the cGAS protein, the DNA, and zinc-thumb were about 2.2, 2.0, and 0.5 Å (Table S1, ESI†), respectively. Thus, all the complexes were stable within the simulation time, indicating the reliability of our models in predicting the structural dynamics of all the complexes.

Hydrogen bonding and salt-bridge analysis

Due to the importance of hydrogen bonding (HB) in protein–DNA interactions, we calculated the total number of HBs at the cGAS–DNA interface. The number of HBs fluctuated around 15, 14, 13, and 13 for the hGhD, hGmD, mGhD, and mGmD models, respectively. The magnitude of the fluctuation for all the systems were around 2–3. To identify the important HBs involved in the cGAS–DNA interaction, the lifetime and the percentage of HB occupancy were analyzed using the hGhD and mGmD models as the representatives of the hcGAS–DNA and the mcGAS–DNA interactions, which are listed in Table 1.

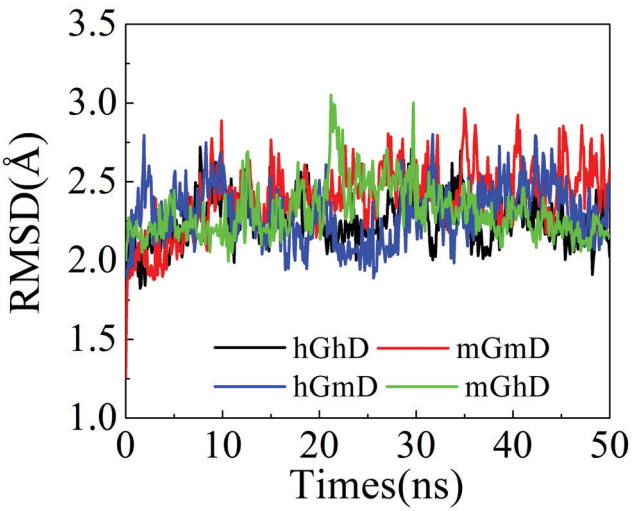

Fig. 3 The RMSD distribution as a function of time in the process of multiple MD simulations with respect to the initial models for the whole systems hGhD (black line), mGmD (red line), hGmD (blue line), and mGhD (green line). For clarity, only one of the three trajectories for each model is shown here. A summary of the averaged RMSD for all the systems can be found in Table S1 (ESI†).

Table 1 The averaged lifetime (ns) and percentage of occupancy (%) of the key HB interactions between proteins (hcGAS/mcGAS) and DNA. The HBs reported in both this work and existing works are in red, the HBs in close vicinity of previously reported ones are in blue, the ones occurring only in this work are in black, and the HBs reported only in existing works are in green. The values are the averages of the results from three parallel MD simulations, with the standard deviations in parentheses

Protein type	Residue:atom	DNA:atom	Lifetime (ns)	Percentage of occupancy (%)
hcGAS	ARG166:NE	T2:O2P	2.7(1.4)	37.7(8.7)
	LYS173:NZ	A29:O1P	—	—
	LEU174:N	A29:O1P	11.5(1.9)	66.7(16.5)
	ARG176:N	A29:O3'	0.7(0.1)	12.0(6.1)
	ARG176:N	C30:O1P	1.3(1.0)	12.0(2.6)
	ARG176:NH1	C7:O2	0.2(0.2)	14.0(8.5)
	ARG176:NH1	C7:O4'	0.1(0.0)	5.0(0.0)
	ARG176:NH2	C7:O4'	—	—
	SER180:OG	T9:O1P	17.8(7.4)	42.0(21.9)
	ASN187:ND2	C10:O2P	4.1(1.3)	29.3(20.5)
	ASN210:ND2	G11:O2P	13.6(5.7)	65.7(2.3)
	TYR214:OH	C10:O1P	9.1(2.1)	34.3(0.6)
mcGAS	ARG158:NE	G29:O2P	2.6(1.6)	14.3(2.5)
	ARG161:NH1	T8:O4'	—	—
	ARG161:NH2	T8:O2	7.2(2.4)	26.5(7.8)
	ARG161:N	A30:O1P	2.8(0.7)	15.3(1.5)
	SER165:OG	C9:O1P	7.9(2.4)	41.3(22.4)
	ASN196:ND2	G11:O2P	6.6(1.7)	36.7(11.9)
	TYR200:OH	G10:O1P	12.0(1.2)	68.3(4.6)
	ASN376:ND2	A27:O1P	2.3(2.8)	14.0(15.1)

For the hcGAS–DNA interaction, among the proposed 10 HBs, 5 of them (ARG176:NH1–C7:O4′, SER180:OG–T9:O1P, ASN187:ND2–C10:O2P, ASN210:ND2–G11:O2P, and TYR214:OH–C10:O1P) were identified in a previous work.[8] LYS173:NZ–A29:O1P and ARG176:NH2–C7:O4′ were also suggested to be key players in the cGAS–DNA interaction in existing works,[8] instead of the exact HBs predicted by MD simulation, and several HBs in their close vicinity (LEU174:N–A29:O1P, ARG176:N–A29:O3′, ARG176:N–C30:O1P, and ARG176:NH1–C7:O2) were observed to possess significant lifetimes and occupancies in our prediction. Specifically, A29:O1P formed a HB with LEU174:N instead of the previously reported LYS173:NZ, and ARG176:NH1 formed a HB with C7:O2 instead of C7:O4′. Therefore, almost all the HBs identified in our work are well consistent with the ones reported in existing works. SER180:OG–T9:O1P and ASN210:ND2–G11:O2P were the most stable HBs. Specifically, SER180:OG–T9:O1P possessed the lifetime of 17.8 ns and the percentage of occupancy of 42.0%, while ASN210:ND2–G11:O2P possessed the lifetime of 13.6 ns and the percentage of occupancy of 65.7%. In addition, the lifetime and the percentage of occupancy of ARG166:NE–T2:O2P was 2.7 ns and 37.7%, respectively. LYS151 in mcGAS (equivalent to ARG166 in hcGAS) was previously reported to play an important role in DNA binding and cGAS activity,[13] while ARG166 in hcGAS was suggested for the first time to be an additional key HB in the hcGAS–DNA interaction. For the mcGAS–DNA interaction, 6 out of the 7 proposed key HBs were

Table 2 The averaged lifetime (ns) and percentage of occupancy (%) of the key SB interactions between proteins (hcGAS/mcGAS) and DNA. The color code is analogue to the one in Table 1. The values are the averages of the results from three parallel MD simulations, with the standard deviations in parentheses

Protein type	Residue:atom	DNA:atom	Lifetime (ns)	Percentage of occupancy (%)
hcGAS	ARG195:NH1	T20:O1P	0.5(0.4)	8.3(5.8)
	LYS198:NZ	A19:O2P	0.5(0.1)	21.7(5.8)
	LYS384:NZ	G11:O1P	5.2(2.2)	41.3(15.6)
	LYS400:NZ	C4:O2P	0.4(0.1)	31.7(14.2)
	LYS407:NZ	G28:O1P	0.4(0.1)	16.7(7.2)
mcGAS	LYS151:NZ	T2:O1P	0.8(0.6)	34.0(30.0)
	ARG158:NH2	G29:O2P	5.7(5.5)	29.0(12.8)
	ARG180:NH1	A20:O2P	4.5(2.6)	32.3(26.0)
	LYS372:NZ	G11:O1P	3.2(3.2)	26.3(12.2)
	LYS395:NZ	A28:O1P	2.0(1.2)	15.3(8.7)

previously reported to be important for the DNA-binding of mcGAS in experimental works.[8,9] Among them, the most stable HB was TYR200:OH–G10:O1P, whose lifetime and percentage of occupancy were 12.0 ns and 68.3% respectively. In addition, ASN362:ND2–A27:O1P was also suggested to play an important role and possessed a lifetime and a percentage of occupancy of 2.3 ns and 14.0%, respectively.

Also, Table 2 lists the important salt-bridges (SBs) proposed by MD simulations. In the hcGAS–DNA interaction, besides the SBs of ARG195:NH1–T20:O1P, LYS384:NZ–G11:O1P, LYS400:NZ–C4:O2P, and LYS407:NZ–G28:O1P that were identified in previous experimental works,[8,9] LYS198:NZ–A19:O2P was suggested to be a vital one as well, with the lifetime of 0.5 ns and the percentage of occupancy of 21.7%, respectively. In the mcGAS–DNA interaction, 4 out of 5 SBs were also consistent with existing experimental results.[9] Consistent with a mutation study in which K151E mutation in the spine helix resulted in a moderate decrease of the mcGAS activity,[13] LYS151:NZ–T2:O1P was also identified to be a key HB in our prediction with a lifetime of 0.8 ns and a percentage of occupancy of 34.0%.

In general, the lifetime of HB or SB was observed to be positively correlated with the percentage of occupancy. However, there were some exceptions. For instance, the lifetimes of HBs SER180:OG–T9:O1P, ASN210:ND2–G11:O2P, and LEU174:N–A29:O1P in the hcGAS–DNA interaction were 17.8, 13.6, and 11.5 ns, respectively, while their percentages of occupancy were 42.0, 65.7, and 66.7%, respectively, which was in reverse order. Here, the lifetimes of salt bridges are generally shorter than the ones of hydrogen bonds. Similar results were observed in the MD simulations of some protein–DNA complexes,[17] although the opposite was reported in the MD simulations of other systems.[17,62]

Binding free energy analysis

In order to understand the mechanism of the cGAS–DNA interaction from the energetic viewpoint, we adapted the MM/GBSA method to estimate the binding free energies and their components for all four models, which are listed in Table 3. The cGAS–DNA interactions are favored, as all the binding free energies (ΔG_{bind}) are negative. The cGAS–DNA interaction is mainly driven by the electrostatic interactions (ΔE_{ele}) between the negatively charged DNA molecule and the highly charged DNA-binding interface of cGAS, which is counterbalanced by the polar part of the solvation free energy ($\Delta G_{gb/solv}$), as the values of ΔE_{ele} and $\Delta G_{gb/solv}$ are one order of magnitude higher than the ones of other components. The binding free energies in both the hGhD and hGmD models are significantly lower than the ones in the mGhD and mGmD models, indicating that the binding affinity of the mutant hcGAS to DNA is higher than the one of mcGAS. This is in agreement with a recent experimental result,[8] in which mutant hcGAS showed a higher percentage of conversion in cGAMP synthesis compared to WT mcGAS in the presence of a low concentration of 17-bp DNA. The results are also consistent with the previous HB analysis, in which the numbers of HBs in the hGhD and hGmD models are slightly more than the ones in the mGhD and mGmD models.

As shown in Table 3, the binding free energies of cGAS–DNA complexes are significantly overestimated by the MM/GBSA method, similar to many results on protein–DNA interactions.[40,41] The uncertainties of MM/GBSA are largely due to the approximation in the continuum solvation model used in estimating the polar contribution, ignoring the conformational changes upon DNA binding, and neglecting entropy.[63] Despite the existence of limitations, MM/GBSA is usually considered as a suitable tool to predict the relative binding free energy,[63] especially for comparative free energy analysis on closely related systems, such as the hcGAS–DNA and mcGAS–DNA systems in this work.

Common key residues in the cGAS–DNA interaction

To further reveal the key residues of cGAS in its binding to dsDNA, the contributions of each residue to the binding free energy were calculated by the per-residue binding free energy analysis. Here, we selected the residues whose binding energies were greater than ±5 kcal mol^{-1} in one of four models and listed them in Table 4. For a better comparison between the results of hcGAS and the ones of mcGAS, a one-to-one correspondence was then constructed for the residues at the cGAS–DNA binding interface by aligning the structure of hcGAS to the one of mcGAS (Fig. S1, ESI†), and the residue pairs between hcGAS and mcGAS are also listed in Table 4. The one-to-one

Table 3 Averaged binding free energies (kcal mol^{-1}) and their components obtained *via* the MM/GBSA approach for the four models. Here, $\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv}$; $\Delta E_{MM} = \Delta E_{vdW} + \Delta E_{ele}$; $\Delta G_{solv} = \Delta G_{gb/solv} + \Delta G_{np/solv}$. The standard errors of the mean are listed in parentheses

Energy	hGhD	hGmD	mGhD	mGmD
ΔE_{vdW}	−152.5(1.0)	−148.3(1.2)	−131.2(1.1)	−114.1(1.0)
ΔE_{ele}	−4807.6(14.6)	−5043.7(10.3)	−4044.6(17.8)	−3939.0(19.0)
$\Delta G_{gb/solv}$	4830.9(13.6)	5049.2(9.9)	4049.8(16.8)	3950.9(18.5)
$\Delta G_{np/solv}$	172.9(0.6)	−169.1(0.6)	−149.4(0.6)	−130.9(0.5)
ΔE_{MM}	−4960.1(9.6)	−5192.0(6.4)	−4175.8(11.8)	−4053.1(12.7)
ΔG_{solv}	4810.5(9.2)	5028.4(6.6)	4031.6(11.4)	3934.0(12.7)
ΔG_{bind}	−149.6(1.9)	−163.5(1.7)	−144.2(1.8)	−119.0(1.9)

correspondence for hcGAS and mcGAS is mentioned in the form of residue_A/residue_B, in which residue_A is from hcGAS and residue_B is the corresponding one from mcGAS.

As expected, most of the key residues in Table 4 are located on the zinc-thumb and the spine helix. Consistent with a previous experimental report that the binding affinity of hcGAS–DNA interaction was largely enhanced through mutations K187N and L195R,[8] ASN187/ASN172 and ARG195/ARG180 were predicted to contribute a significantly favored binding free energy in cGAS–DNA interaction. Specifically, the averaged binding free energies of ARG195/ARG180 in hcGAS and mcGAS are −5.7 and −8.9 kcal mol^{-1}, respectively; the averaged binding free energy of ASN187 in hcGAS is −6.2 kcal mol^{-1}, and the corresponding one in mcGAS is −2.6 kcal mol^{-1}, slightly lower than 5 kcal mol^{-1} in magnitude. In addition, residues LYS173, ARG176, LYS384, LYS400, LYS403, LYS407, and LYS411 in hGhD and LYS151, ARG158, LYS160, ARG161, LYS162, LYS372, and LYS395 in mGmD were previously proposed to play vital roles in the recognition of DNA according to experimental mutagenesis studies.[9,13] The residues ASN210, HIS217, and CYM397 in hGhD and ASN196, HIS203, CYM385, and LYS399 in mGmD were the contacts directly observed in crystal structures of hcGAS–DNA and mcGAS–DNA.[8] Furthermore, these residues were mainly positively charged ones, reconfirming the dominant role of electrostatic contributions in the cGAS–DNA interactions.

Here, we proposed that the residues with a binding free energy higher than 5 kcal mol^{-1} in magnitude in both hcGAS and mcGAS (columns hG and mG in Table 4) are the common ones that play vital roles in the DNA recognition by cGAS, and we identified the residue pairs ARG176/ARG161, ARG195/ARG180, ASN210/ASN196, LYS384/LYS372, CYM397/CYM385, LYS403/LYS391, LYS407/LYS395, and LYS411/LYS399 (bold residue names in Table 4). Interestingly, these residues are of the same type in both hcGAS and mcGAS. Importantly, residues (LYS403/LYS391 and LYS411/LYS399) near the zinc-thumb domain were also predicted to be key players in the cGAS–DNA interactions.

To describe how these common residues are involved in DNA-binding, we show the interaction modes of ARG176/ARG161, ASN210/ASN196, LYS384/LYS372, LYS403/LYS391, LYS407/LYS395, and LYS411/LYS399 in Fig. 4, except for the previously reported ARG195/ARG180 and CYM397/CYM385 at thte zinc-thumb domain with positive binding free energies. ASN210/ASN196, LYS384/LYS372, LYS403/LYS391, LYS407/LYS395, and LYS411/LYS399 mainly interact with the phosphate backbone of either the Crick or the Watson strand using their side-chains (Fig. 4b–f). ARG176/ARG161 protrudes into the minor groove its main-chain and side-chain N atoms, which interact with the Crick strand and Watson strand, respectively (Fig. 4a). Such an interaction mode of ARG176/ARG161 well explains their strong binding free energies of −14.1 and −10.9 kcal mol^{-1} in both hcGAS and mcGAS, respectively. It is worth mentioning that ARG176/ARG161 was identified in the above HB analysis as well (Table 1).

Key residues that make hcGAS and mcGAS different

One of our goals in this study is to answer the question of what makes hcGAS and mcGAS different. It is reported that hcGAS and mcGAS possess different binding affinities and length-dependence on dsDNA.[8,11,15] By comparing the average binding free energy of each residue pair in hcGAS and mcGAS listed in Table 4, we found that the binding free energies in the residue

Table 4 The per-residue binding free energy obtained from MM/PBSA analysis of 22 selected residues for hGhD, hGmD, mGhD, and mGmD models. The energy values over ±5 kcal mol^{-1} are shown in black bold letters

No.	Residue		Location[a]	Binding free energies (kcal mol^{-1})					
	hcGAS	mcGAS		hGhD	hGmD	mGhD	mGmD	hG[b]	mG[c]
1	ARG166	LYS151	Spine, α1/α1	−3.3	−2.4	**−6.1**	**−9.0**	−2.9	**−7.5**
2	LYS173	ARG158	Spine	−4.8	−4.2	**−9.0**	**−12.6**	−4.5	**−10.8**
3	SER175	LYS160	Spine	−4.9	−3.4	**−5.8**	**−6.1**	−4.1	**−6.0**
4	**ARG176**	**ARG161**	Spine	**−13.6**	**−14.6**	**−10.1**	**−11.6**	**−14.1**	**−10.9**
5	ASP177	LYS162	Spine, α2/α2	**5.2**	3.4	−0.9	−2.9	4.3	−1.9
6	ASN187	ASN172	Spine, α2/α2	**−7.8**	−4.6	**−5.5**	0.2	**−6.2**	−2.6
7	**ARG195**	**ARG180**	Spine, α2/α2	**−11.1**	−0.3	**−8.8**	**−9.1**	**−5.7**	**−8.9**
8	LYS198	GLN183	Spine, α2/α2	−2.8	**−6.8**	−2.6	0.3	−4.8	−1.2
9	CYS199	LYS184	Spine, α2/α2	0.2	0.2	−4.6	**−5.7**	0.2	**−5.1**
10	ARG204	LYS190	Before β1/β1	−0.4	−1.6	−1.4	**−9.4**	−1.0	**−5.4**
11	LEU209	LEU195	β1/β1	**−5.2**	−1.9	−1.6	−4.7	−3.6	−3.2
12	**ASN210**	**ASN196**	β1/β1	**−7.4**	**−6.0**	**−6.6**	**−10.0**	**−6.7**	**−8.3**
13	HIS217	HIS203	After α3/β1	**−5.6**	**−6.5**	−1.4	**−5.1**	**−6.0**	−3.2
14	**LYS384**	**LYS372**	α6/α5	**−11.9**	**−8.7**	**−8.1**	**−14.7**	**−10.3**	**−11.4**
15	LYS394	LYS382	Zinc-thumb	**−9.6**	**−7.1**	**−8.1**	−0.4	**−8.3**	−4.2
16	**CYM397**	**CYM385**	Zinc-thumb	**6.4**	**7.3**	**7.8**	**5.1**	**6.9**	**6.5**
17	GLU398	SER387	Zinc-thumb	4.6	**6.6**	−0.6	−0.3	**5.6**	−0.5
18	LYS400	GLY389	Zinc-thumb	**−5.6**	−1.3	−4.6	1.4	−3.4	−1.6
19	**LYS403**	**LYS391**	Zinc-thumb	**−7.0**	**−10.7**	**−10.7**	−2.0	**−8.8**	**−6.4**
20	**LYS407**	**LYS395**	α7/α6	**−6.1**	**−14.9**	**−7.9**	**−7.3**	**−10.5**	**−7.6**
21	**LYS411**	**LYS399**	α7/α6	**−5.0**	**−6.2**	**−9.8**	**−5.0**	**−5.6**	**−7.4**
22	ARG457	PRO442	η4/η6	**−7.1**	−0.5	−1.1	−0.5	−3.8	−0.8

[a] The location labels like "Spine, α1/α1" describe the corresponding residues in hcGAS and mcGAS distributing at the spine helix. [b] Averaged binding free energy for the hcGAS residues in hGhD and hGmD. [c] Averaged binding free energy for the mcGAS residues in mGhD and mGmD.

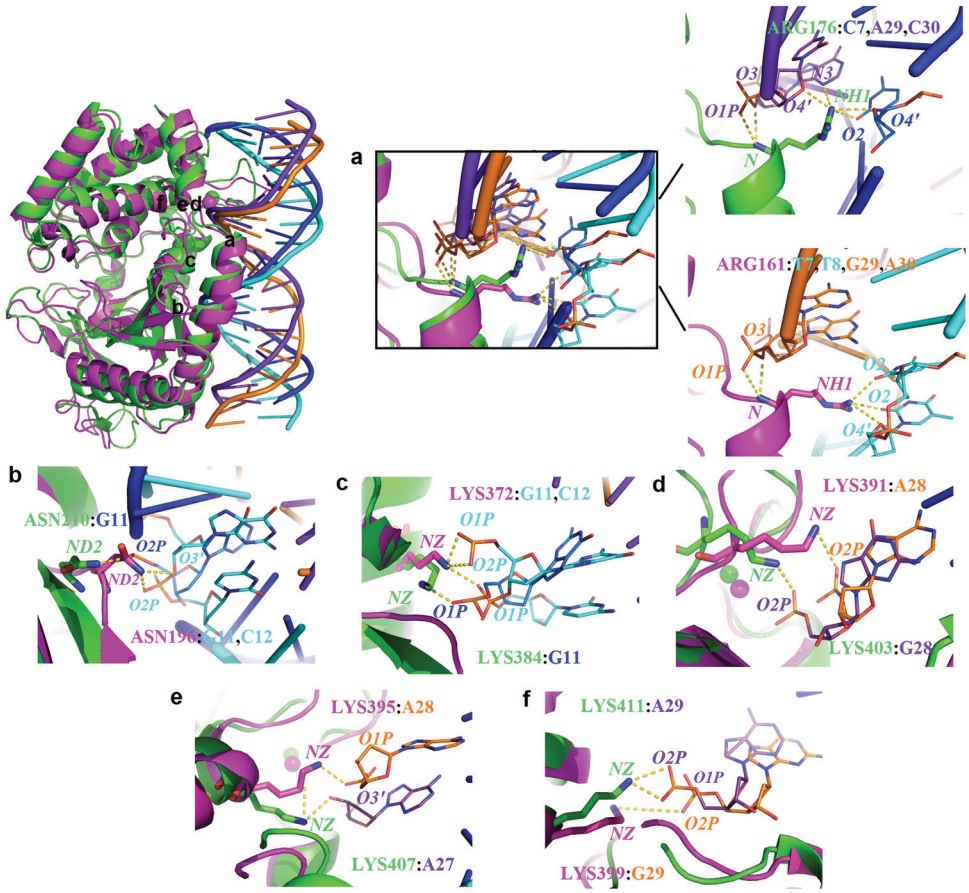

Fig. 4 Six common residue pairs in hcGAS and mcGAS predicted by MM/GBSA based on the structural overlaps of hcGAS–DNA and mcGAS–DNA interactions displayed in the top left side. The mcGAS (magenta), hcGAS (green), and dsDNA (cyan and blue: Watson-strands, orange and purple: Crick-strands) are shown as cartoon representations. Zinc ions are shown as spheres. The locations of six common residues are labelled as the corresponding positions (a to f). We zoom in on the interaction details of the predicted new active residues as stick representations, whereas the C-atom colors of the sticks in hcGAS and mcGAS are consistent with associated proteins and the N and O atoms are blue and red colors, respectively. The DNA nucleotides binding with new active residues are shown as lines, consistent with associated strand colors. The colors of residue and nucleotide names including atom names labelled in italics in hcGAS–DNA and mcGAS–DNA interactions are the same as the associated protein and DNA. The yellow dashed lines express the HB interactions.

pairs LYS173/ARG158, ASP177/LYS162, CYS199/LYS184, and GLU398/SER387 differed the most and their differences were greater than 5 kcal mol^{-1}. The average binding free energies of the residues ASP177, CYS199, and GLU398 in hcGAS are 4.3, 0.2, and 5.6 kcal mol^{-1}, respectively, while the average binding free energies of their corresponding residues in mcGAS are −1.9, −5.1, and −0.5 kcal mol^{-1}, respectively. The differences in their binding free energies are due to the charge differences in the corresponding residue pairs, as the charges in the residue pairs ASP177/LYS162, CYS199/LYS184, and GLU398/SER387 are −1 e/+1 e, 0 e/+1 e, and −1 e/0 e, respectively.

Interestingly, the energy difference between LYS173 in hcGAS and ARG158 in mcGAS is 6.3 kcal mol^{-1}, the greatest one among the proposed four residue pairs that make hcGAS and mcGAS different. More specifically, the averaged binding free energy of LYS173 in hcGAS is −4.5 kcal mol^{-1}, while the averaged binding free energy of its corresponding residue ARG158 in mcGAS is −10.8 kcal mol^{-1}. Since these two residues are of the same change of +1 e, it is thus necessary to check their interaction details in order to rationalize the difference in binding free energies. As shown in Fig. 5, both residues mainly interact with the phosphate backbone of the Crick strand. However, ARG158 in mcGAS extends its side-chain deeper into the DNA major groove than LYS173 in hcGAS, as the side-chain of ARG158 is longer than the one of LYS173. In the switch from LYS173 in hcGAS to ARG158 in mcGAS, the change in the binding free energy is largely due to the change in electrostatic interactions, while the vdW contributions remain unchanged (Table S2, ESI†).

Sequence-selectivity of cGAS

To check the potential sequence-selectivity of cGAS, we compared the per-residue binding free energies between the hGhD and hGmD models or the ones between the mGhD and mGmD models (Table 4). Indeed, the binding free energies of several residues change significantly when hDNA is replaced by mDNA. For example, the binding free energy of ARG195 in hGhD is −11.1 kcal mol^{-1}, while the one in hGmD is −0.3 kcal mol^{-1}.

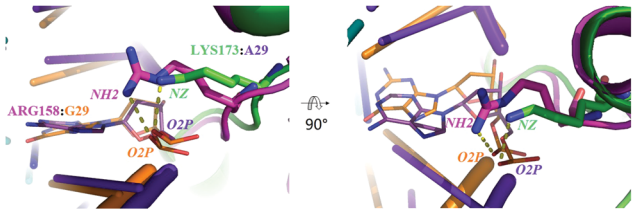

Fig. 5 The interaction details of LYS173 in hcGAS and ARG158 in mcGAS binding with DNA. The side and top views are displayed in the left and right sides, respectively. The colors of cGAS and dsDNA are the same as in the description of Fig. 4.

However, the binding free energies of the corresponding residue ARG180 in mcGAS in mGhD and mGmD are comparable. Since ARG195 in hGhD and ARG180 in mcGAS are of the exactly same type, it is thus illusive that ARG195 in hGhD is sensitive to the DNA sequence. It is thus not conclusive on the sequence-selectivity of cGAS based on just two sequences. Additional novel DNA sequences should be included to further explore the sequence-selectivity of cGAS. Without further MD simulations, the sequence specificity of protein–DNA interactions can also be investigated by constructing a position weight matrix.[17,64]

cGAS mainly interacting with three nucleotides of DNA

Which nucleotide does the cGAS predominantly interact with? To answer this question, we decomposed the total binding free energy into the components on each nucleotide of the DNA. The average binding free energies of each nucleotide over the four models are shown in Fig. 6. Interestingly, the binding free energies on most of the nucleotides except the 27th, 28th, and 29th nucleotides (AGA in hGhD and mGhD, AAG in mGmD and hGmD) are lower than 3.0 kcal mol^{-1} in magnitude. Specifically, the average binding free energy on these three nucleotides are −4.4, −10.6, and −15.2 kcal mol^{-1}, respectively.

Residue–nucleotide pairwise interaction analysis

In order to identify the fingerprint of the cGAS–DNA interaction, the per-residue binding free energy was further decomposed into its components due to the residue–nucleotide pairwise interaction. The averaged residue–nucleotide interaction matrixes between the selected 22 residues are shown in Table 4 and all the 34 nucleotides over the four models are shown in Fig. 7a. Obviously, the interaction matrixes are very sparse, indicating that only a minority of residue–nucleotide pairs are involved in the

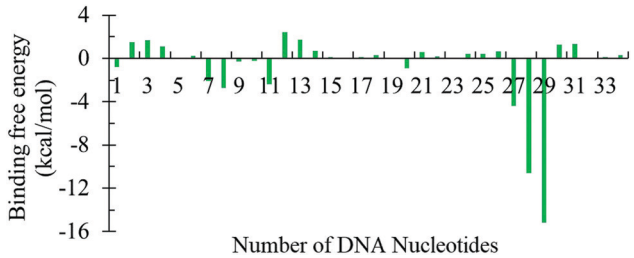

Fig. 6 The average binding free energy of each nucleotide over the hGhD, hGmD, mGhD, and mGmD models.

cGAS–DNA interactions. From such a 2D interaction map, the major interaction mode for the cGAS–DNA interaction was identified to be the interaction between the 27th–29th nucleotides of the DNA and the residue pairs of LYS403/LYS391, LYS407/LYS395, LYS411/LYS399, LYS173/ARG158, SER175/LYS160, and ARG176/ARG161. By taking a close look at the structural snapshot of these nucleotides and residues (Fig. 7b), these three nucleotides 27th–29th in the Crick strand were found to closely interact with residues LYS403/LYS391, LYS407/LYS395, and LYS411/LYS399 surrounding the zinc-thumb domain, reconfirming the importance of the zinc-thumb in the recognition of DNA.[11] In addition, these three nucleotides were also observed to interact with residues LYS173/ARG158, SER175/LYS160, and ARG176/ARG161, which were located in the kink between the α1 helix and α2 helix of the spine helix (Fig. 7c). Specifically, the residue LYS173/ARG158 inserted its long side chain into the major groove and interacted with the 29th nucleotide, while the side chain of residue ARG176/ARG161 protruded into the minor groove and interacted with the same nucleotide. Furthermore, ARG176/ARG161 interacted with the 7th nucleotide as well, with the binding free energy of −5.2 kcal mol^{-1}. Thus, we suggested that the kink region and the zinc-thumb together account for the most part of the cGAS–DNA interaction. This was the first time that the kink region was identified in cGAS–DNA interactions to the best of our knowledge. The kinks were reported to play vital structural and functional roles in many other proteins,[65–68] indicating that they may be generally adopted in protein function. Other than the interaction with the zinc-thumb and kink regions, the 27th–29th nucleotides are also observed to weakly interact with HIS217/HIS213, LYS394/LYS382, and CYM397/CYM385. Specifically, the binding free energy between 27th–28th nucleotides and HIS217/HIS213 is around −2.0 kcal mol^{-1}; LYS394/LYS382 is involved in a weak interaction with the 27th nucleotide. The interaction between the 28th nucleotide and CYM397/CYM385 at the zinc-thumb domain is marginally disfavored with the binding free energy of 1–2 kcal mol^{-1}.

There are two minor interaction modes. One is the minor interaction between residue LYS384/LYS372 and the 11th nucleotide, as identified in previous HB analysis (Fig. 4c). The other is the interaction between residue ARG195/ARG180 and the 20th nucleotide and the structural snapshot is shown in Fig. 7d, in which their side-chains interact with the DNA phosphate backbone of the Crick strand instead of the minor or major groove.

In addition, 50 ns MD simulations on hcGAS (PDB ID: 4KM5) and mcGAS (PDB ID: 4K8V) in apo states were also carried out. Structural comparison of the apo-cGAS and the DNA-bound cGAS revealed that the large-scale structural changes occur in the activation loop close to its catalytic center and the kink region (Fig. S5, ESI†). It seems that these structural changes share great similarity in both hcGAS and mcGAS, highlighting their importance in revealing further details of the molecular mechanism of cGAS. However, the timescale of these structural changes is significantly longer than the simulated timescale of this study, as no significant structural change of cGAS in 50 ns simulations was observed.

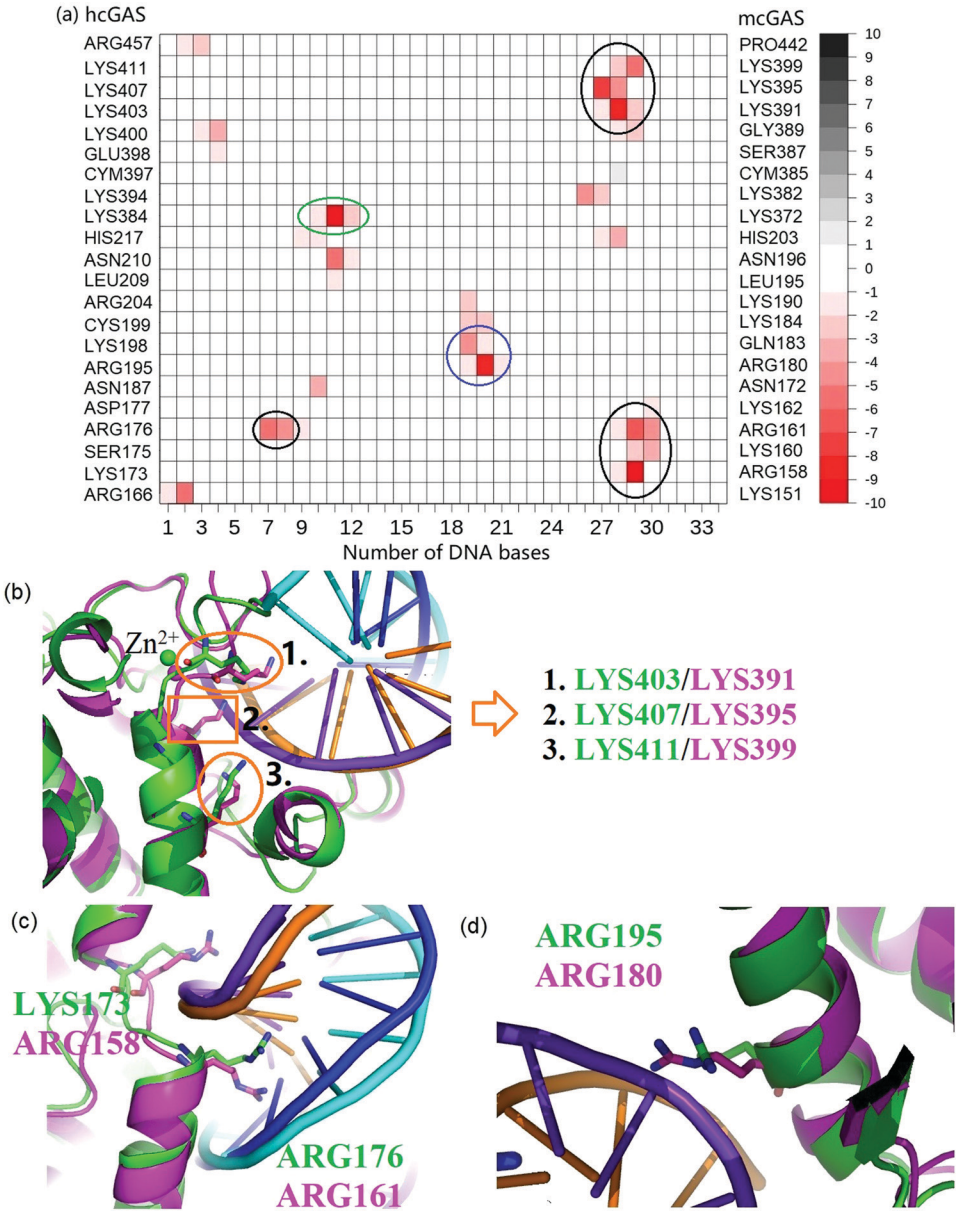

Fig. 7 (a) 2D maps of the averaged residue–nucleotide pairwise interaction free energy (kcal mol^{-1}) for 22 selected amino acids in hcGAS and mcGAS. Black, blue, and green circles describe three major motifs. The energy bar represents the interaction value range. (b) LYS403/LYS391, LYS407/LYS395, and LYS411/LYS399 at the zinc-thumb domain. (c) One kink region composed of LYS173/ARG158, SER175/LYS160, and ARG176/ARG161. (d) The interaction modes of ARG195 in hcGAS and ARG180 in mcGAS. The colors of cGAS and dsDNA are the same as in the description of Fig. 4.

Thus, the exploration of large-scale collective motions is left for future research.

Conclusion

For a better understanding of the cGAS–DNA interaction in both humans and mice, all-atom MD simulations were performed for four cGAS–DNA complexes, which are the complexation of the hcGAS and mcGAS with two 17-bp DNA of different sequences. The simulated structures were shown to be stable by RMSD analysis. Several key HB and SB were identified and are consistent with the related experimental results.

The MM/GBSA results suggested that the total binding affinity of hcGAS was higher than the one of mcGAS, which is in agreement with previous experimental reports.[8] The cGAS–DNA interaction was dominated by electrostatic interactions between positively charged residues in cGAS and the highly charged DNA. Importantly, the similarities and differences between hcGAS and mcGAS were revealed. The common key residues in both hcGAS and mcGAS are suggested to be residue pairs ARG176/ARG161, ARG195/ARG180, ASN210/ASN196, LYS384/LYS372, CYM397/CYM385, LYS403/LYS391, LYS407/LYS395, and LYS411/LYS399, while four residue pairs LYS173/ARG158, ASP177/LYS162, CYS199/LYS184, and GLU398/SER387 make the hcGAS and

mcGAS differ in their recognition of DNA, as they differ either in charge or the length of their side-chains.

In addition, cGAS was observed to mainly interact with three nucleotides of the DNA. From the 'fingerprint' of the cGAS–DNA interaction obtained using residue–nucleotide pairwise decomposition analysis, one major interaction mode was identified. Specifically, these three nucleotides were found to be bound by the zinc-thumb domain and the kink region, which connects the α1 and α2 helices of the spine helix. To the best of our knowledge, it is the first time that the kink region is suggested to play the vital role in cGAS–DNA interactions, together with the zinc-thumb domain. Nevertheless, these theoretical predictions should be validated by experimental studies such as site-specific mutagenesis assay.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

We would like to thank the National Supercomputer Center in Guangzhou for the computing and technical support services. This work was financially supported by the National Natural Science Foundation of China under grant number 31770777, Start-up Foundation for Peacock Talents (827-000365), and the Start-up Grant for Young Scientists (860-000002110384), Shenzhen University.

References

1 B. Zhang, M. M. Davidson and T. K. Hei, *Life Sci. Space Res.*, 2014, **1**, 80–88.
2 X. Cai, Y.-H. Chiu and Z. J. Chen, *Mol. Cell*, 2014, **54**, 289–296.
3 L. Sun, J. Wu, F. Du, X. Chen and Z. J. Chen, *Science*, 2013, **339**, 786–791.
4 Q. Chen, L. Sun and Z. J. Chen, *Nat. Immunol.*, 2016, **17**, 1142–1149.
5 D. S. Pisetsky, *Nat. Rev. Rheumatol.*, 2016, **12**, 102–110.
6 A. Ablasser and M. F. Gulen, *J. Mol. Med.*, 2016, **94**, 1085–1093.
7 M.-M. Hu and H.-B. Shu, *Annu. Rev. Immunol.*, 2020, **38**, 79–98.
8 W. Zhou, A. T. Whiteley, C. C. de Oliveira Mann, B. R. Morehouse, R. P. Nowak, E. S. Fischer, N. S. Gray, J. J. Mekalanos and P. J. Kranzusch, *Cell*, 2018, **174**(300–311), e311.
9 X. Zhang, J. Wu, F. Du, H. Xu, L. Sun, Z. Chen, C. A. Brautigam, X. Zhang and Z. J. Chen, *Cell Rep.*, 2014, **6**, 421–430.
10 P. Gao, M. Ascano, Y. Wu, W. Barchet, B. L. Gaffney, T. Zillinger, A. A. Serganov, Y. Liu, R. A. Jones and G. Hartmann, *Cell*, 2013, **153**, 1094–1107.
11 Philip J. Kranzusch, Amy S.-Y. Lee, James M. Berger and Jennifer A. Doudna, *Cell Rep.*, 2013, **3**, 1362–1368.
12 W. Xie, L. Lama, C. Adura, D. Tomita, J. F. Glickman, T. Tuschl and D. J. Patel, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 11946.
13 X. Li, C. Shu, G. Yi, C. T. Chaton, C. L. Shelton, J. Diao, X. Zuo, C. C. Kao, A. B. Herr and P. Li, *Immunity*, 2013, **39**, 1019–1031.
14 F. Civril, T. Deimling, C. C. de Oliveira Mann, A. Ablasser, M. Moldt, G. Witte, V. Hornung and K.-P. Hopfner, *Nature*, 2013, **498**, 332–337.
15 A. Lee, E.-B. Park, J. Lee, B.-S. Choi and S.-J. Kang, *FEBS Lett.*, 2017, **591**, 954–961.
16 K. Kato, R. Ishii, E. Goto, R. Ishitani, F. Tokunaga and O. Nureki, *PLoS One*, 2013, **8**, e76983.
17 L. Etheve, J. Martin and R. Lavery, *Nucleic Acids Res.*, 2016, **44**, 9990–10002.
18 M. Y. Hamed and G. Arya, *J. Biomol. Struct. Dyn.*, 2016, **34**, 919–934.
19 X. Wang, N. Singh and W. Li, in *Systems Medicine*, ed. O. Wolkenhauer, Academic Press, Oxford, 2021, pp. 182–189.
20 T. Hou, J. Wang, Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, **51**, 69–82.
21 E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J. Z. H. Zhang and T. Hou, *Chem. Rev.*, 2019, **119**, 9478–9508.
22 E. Wang, G. Weng, H. Sun, H. Du, F. Zhu, F. Chen, Z. Wang and T. Hou, *Phys. Chem. Chem. Phys.*, 2019, **21**, 18958–18969.
23 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Acc. Chem. Res.*, 2000, **33**, 889–897.
24 Y. Chen, Y. Zheng, P. Fong, S. Mao and Q. Wang, *Phys. Chem. Chem. Phys.*, 2020, **22**, 9656–9663.
25 S. Genheden and U. Ryde, *Expert Opin. Drug Discovery*, 2015, **10**, 449–461.
26 S. R. Peddi, S. K. Sivan and V. Manga, *J. Biomol. Struct. Dyn.*, 2018, **36**, 486–503.
27 W. Zhao, M. Xiong, X. Yuan, M. Li, H. Sun and Y. Xu, *J. Chem. Inf. Model.*, 2020, **60**, 3265–3276.
28 A. Sali and T. L. Blundell, *J. Mol. Biol.*, 1993, **234**, 779–815.
29 A. Fiser, R. K. G. Do and A. Šali, *Protein Sci.*, 2000, **9**, 1753–1773.
30 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminf.*, 2012, **4**, 17.
31 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.
32 S. J. Fisher, J. Wilkinson, R. H. Henchman and J. R. Helliwell, *Crystallogr. Rev.*, 2009, **15**, 231–259.
33 W. L. DeLano, *CCP4 Newsletter on protein crystallography*, 2002, vol. 40, pp. 82–92.
34 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.
35 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
36 I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella,

F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case and M. Orozco, *Nat. Methods*, 2016, **13**, 55–58.
37 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
38 V. Gapsys and B. L. de Groot, *J. Chem. Theory Comput.*, 2017, **13**, 6275–6289.
39 A. Esadze, C. Chen, L. Zandarashvili, S. Roy, B. M. Pettitt and J. Iwahara, *Nucleic Acids Res.*, 2016, **44**, 6961–6970.
40 J. Lee, J.-S. Kim and C. Seok, *J. Phys. Chem. B*, 2010, **114**, 7662–7671.
41 B. Yang, Y. Zhu, Y. Wang and G. Chen, *J. Comput. Chem.*, 2011, **32**, 416–428.
42 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.
43 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
44 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
45 S. Nosé and M. L. Klein, *Mol. Phys.*, 1983, **50**, 1055–1076.
46 S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962.
47 H. C. Andersen, *J. Comput. Phys.*, 1983, **52**, 24–34.
48 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
49 V. Tsui and D. A. Case, *J. Am. Chem. Soc.*, 2000, **122**, 2489–2498.
50 C. Wang, D. A. Greene, L. Xiao, R. Qi and R. Luo, *Front. Mol. Biosci.*, 2018, **4**, 1–18.
51 P. Tamamis, D. Morikis, C. A. Floudas and G. Archontis, *Proteins*, 2010, **78**, 2655–2667.
52 P. Tamamis, A. López de Victoria, R. D. Gorham Jr, M. L. Bellows-Peterson, P. Pierou, C. A. Floudas, D. Morikis and G. Archontis, *Chem. Biol. Drug Des.*, 2012, **79**, 703–718.
53 F. Chen, H. Liu, H. Sun, P. Pan, Y. Li, D. Li and T. Hou, *Phys. Chem. Chem. Phys.*, 2016, **18**, 22129–22139.
54 C. Wang, P. H. Nguyen, K. Pham, D. Huynh, T.-B. N. Le, H. Wang, P. Ren and R. Luo, *J. Comput. Chem.*, 2016, **37**, 2436–2446.
55 B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theory Comput.*, 2012, **8**, 3314–3321.
56 N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 10037–10041.
57 R. Kumari, R. Kumar and A. Lynn, *J. Chem. Inf. Model.*, 2014, **54**, 1951–1962.
58 S. R. B. D. A. Case, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. H. G. Ghoreishi, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. S. Lee, S. P. L. LeGrand, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, I. Omelyan, H. Nguyen, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. C. L. S. Shen, J. Smith, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER*, University of California, San Francisco, 2018.
59 G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem.*, 1996, **100**, 19824–19839.
60 V. Barone, M. Cossi and J. Tomasi, *J. Chem. Phys.*, 1997, **107**, 3210–3221.
61 S. Genheden, J. Kongsted, P. Söderhjelm and U. Ryde, *J. Chem. Theory Comput.*, 2010, **6**, 3558–3568.
62 S. Pylaeva, M. Brehm and D. Sebastiani, *Sci. Rep.*, 2018, **8**, 13626.
63 N. Homeyer and H. Gohlke, *Mol. Inf.*, 2012, **31**, 114–122.
64 G. Ambrosini, I. Vorontsov, D. Penzar, R. Groux, O. Fornes, D. D. Nikolaeva, B. Ballester, J. Grau, I. Grosse, V. Makeev, I. Kulakovskiy and P. Bucher, *Genome Biol.*, 2020, **21**, 114.
65 E. C. Law, H. R. Wilman, S. Kelm, J. Shi and C. M. Deane, *PLoS One*, 2016, **11**, e0157553.
66 A. Krokhotin, A. Liwo, G. G. Maisuradze, A. J. Niemi and H. A. Scheraga, *J. Chem. Phys.*, 2014, **140**, 025101.
67 H. R. Wilman, J. Shi and C. M. Deane, *Proteins*, 2014, **82**, 1960–1970.
68 D. J. Barlow and J. M. Thornton, *J. Mol. Biol.*, 1988, **201**, 601–619.