Computational and Structural Biotechnology Journal 19 (2021) 6108–6124

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Atomic-scale insights into allosteric inhibition and evolutional rescue mechanism of *Streptococcus thermophilus* Cas9 by the anti-CRISPR protein AcrIIA6

Xinyi Li [a,b,1], Chengxiang Wang [b,1], Ting Peng [b,1], Zongtao Chai [c,1], Duan Ni [d], Yaqin Liu [e], Jian Zhang [b,e], Ting Chen [a,*], Shaoyong Lu [b,e,*]

[a] *Department of Cardiology, Changzheng Hospital, Naval Medical University, Shanghai 200003, China*
[b] *Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai Jiao Tong University, School of Medicine, Shanghai 200025, China*
[c] *Department of Hepatic Surgery VI, Eastern Hepatobiliary Surgery Hospital, Naval Medical University, Shanghai 200438, China*
[d] *The Charles Perkins Centre, University of Sydney, Sydney, NSW 2006, Australia*
[e] *Medicinal Chemistry and Bioinformatics Centre, Shanghai Jiao Tong University, School of Medicine, Shanghai 200025, China*

ARTICLE INFO

Article history:
Received 16 September 2021
Received in revised form 10 November 2021
Accepted 11 November 2021
Available online 16 November 2021

Keywords:
CRISPR-Cas
Anti-CRISPR proteins
Molecular dynamics simulations
Conformational dynamics
Gene editing

ABSTRACT

CRISPR-Cas systems are prokaryotic adaptive immunity against invading phages and plasmids. Phages have evolved diverse protein inhibitors of CRISPR-Cas systems, called anti-CRISPR (Acr) proteins, to neutralize this CRISPR machinery. In response, bacteria have co-evolved Cas variants to escape phage's anti-CRISPR strategies, called anti-anti-CRISPR systems. Here we explore the anti-CRISPR allosteric inhibition and anti-anti-CRISPR rescue mechanisms between *Streptococcus thermophilus* Cas9 (St1Cas9) and the anti-CRISPR protein AcrIIA6 at the atomic level, by generating mutants of key residues in St1Cas9. Extensive unbiased molecular dynamics simulations show that the functional motions of St1Cas9 in the presence of AcrIIA6 differ substantially from those of St1Cas9 alone. AcrIIA6 binding triggers a shift of St1Cas9 conformational ensemble towards a less catalytically competent state; this state significantly compromises protospacer adjacent motif (PAM) recognition and nuclease activity by altering interdependently conformational dynamics and allosteric signals among nuclease domains, PAM-interacting (PI) regions, and AcrIIA6 binding motifs. Via *in vitro* DNA cleavage assays, we further elucidate the rescue mechanism of efficiently escaping AcrIIA6 inhibition harboring St1Cas9 triple mutations (G993K/K1008M/K1010E) in the PI domain and identify the evolutionary landscape of such mutational escape within species. Our results provide mechanistic insights into Acr proteins as natural brakes for the CRISPR-Cas systems and a promising potential for the design of allosteric Acr peptidomimetics.

1. Introduction

The evolutionary arms race between bacteria and viruses (phages) has been ongoing for millions of years [1–5]. Suffering from constant predation by viruses, bacteria have evolved a number of either innate or adaptive immune mechanisms to protect themselves against frequent assault [6–10]. Among the diverse anti-viral defences, the clustered regularly interspaced short palindromic repeats (CRISPR) system, which cleaves invading genetic elements in a sequence-specific manner, is a form of bacterial adaptive immunity against bacteriophages and represents a milestone in this coevolutionary arms race [11–14]. This specificity and easily programmable nature of the single-protein (type II) CRISPR-Cas9 endonuclease led to its exploitation by researchers to achieve facile gene manipulation [11,15–22].

CRISPR-Cas9 function begins with assembly of the surveillance complex, in which the Cas9 endonuclease associates with single guide RNA (sgRNA) carrying CRISPR RNA (crRNA) transcribed from previously encountered invading genetic elements to form a ribonucleoprotein complex (RNP) [13,23,24]. Subsequent double-strand DNA (dsDNA) recognition and interference by Cas9 requires identification of the protospacer adjacent motif (PAM) located

* Corresponding authors at: Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai Jiao Tong University, School of Medicine, Shanghai 200025, China (S. Lu).
E-mail addresses: pipidan@126.com (T. Chen), lushaoyong@sjtu.edu.cn (S. Lu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2021.11.010

immediately downstream of the target sequence before further substrate DNA targeting [24–31]. Previous structural characterisation of *Streptococcus thermophilus* Cas9 (St1Cas9), the model system that kicked off the CRISPR craze, revealed a bilobed architecture composed of a recognition (REC) and nuclease (NUC) lobes connected through an arginine-rich bridge helix (BH) [2,32–35] (Fig. 1A–D). The REC lobe primarily engages in nucleic acid binding, mediating sgRNA loading and RNA-DNA heteroduplex recognition [24,36–38]. The NUC lobe includes two important catalytic HNH and RuvC nuclease domains that perform synchronised site-specific cleavage of the corresponding target DNA (tDNA) and nontarget DNA (ntDNA) strands, and the two nuclease domains are interconnected by L1 and L2 linkers. It also includes an α/β wedge (WED) and PAM-interacting (PI) domains [collectively referred to as topoisomerase-homology (TOPO) and C-terminal domains (CTD)] connected to the catalytic cores via a phosphate lock loop (PLL). The PAM-containing DNA duplex is recognised and accommodated by WED and PI domains. Identification of a matching PAM sequence initiates dsDNA unwinding and heteroduplex formation, which ultimately triggers St1Cas9 conformational activation characterised by prominent RuvC and HNH relocation for efficient substrate DNA cleavage [2,39–42].

To counteract this robust anti-viral immunity, phages utilise anti-CRISPR proteins (Acrs) that bind and interfere with CRISPR-Cas function in a sequence-independent manner [5,43–48]. As natural antagonists of Cas endonuclease, Acrs have garnered intense research interest for their potential in restraining the powerful CRISPR-Cas machinery and enhancing the precision at which genetic perturbations can be made [49–51]. To date, ~ 45 diverse Acrs targeting different CRISPR-Cas systems have been identified [45,52–54], and bioinformatic analysis has suggested that Acrs are likely far more widespread than expected [55–58]. Acrs are found to interfere with CRISPR-Cas in a variety of manners [53,59], including inhibiting nuclease recruitment during RNP assembly (AcrIF3) [60], competing with sgRNA loading (AcrIIC2) [61,62], PAM binding (AcrIIA2, AcrIIA4) [63–65], DNA hybridisation (AcrIF1) [66], and blocking catalytic residues (AcrIIC1) [47]. Among the identified Acrs (AcrIIA1–AcrIIA6 and AcrIIC1–AcrIIC3) that function against the type II CRISPR-Cas system [43,46,67], only the newly reported AcrIIA6 possesses specific inhibition towards St1Cas9 [2,45,55]. A cryo-electron microscopy (cryo-EM) structure shows that AcrIIA6 binds to the back face of St1Cas9 RNP which is not primarily engaged in substrate DNA recognition and binding, nonoverlapping with the sgRNA and dsDNA binding sites (Fig. 1E). AcrIIA6 consists of 183 amino acids and predominantly exists as a symmetric homodimer (monomers referred to as AcrIIA6^{A} and AcrIIA ^{B}) in solution (Fig. 1F and Supplementary Fig. 1). Notably, static structural snapshots of St1Cas9 RNP in the AcrIIA6-bound and -unbound states are nearly identical (root-mean-square deviation of St1Cas9 is only 0.72 Å for 768 equivalent Cα atoms), rendering conformationally mechanistic insights into the allosteric inhibition of St1Cas9 RNP by AcrIIA6 and residue level-based allosteric signals propagated between St1Cas9 RNP and AcrIIA6 unrecognized.

The evolutionary arms race between bacteria and their predators has already surpassed the above-mentioned defence mechanisms. Under the strong selective pressure of bacteriophages capable of bypassing CRISPR-Cas immunity, naturally occurring strains possessing residue variations were found resistant to this anti-CRISPR strategy [2,68,69], giving rise to the so-called "anti-anti-CRISPR systems". Indeed, a St1Cas9 variant from *S. thermophilus* strain LMG18311 containing triple residue substitutions (G993K/K1008M/K1010E, *mut* St1Cas9) in the PI domain can escape allosteric inhibition by AcrIIA6. The molecular underpinning underlying the triple mutations-induced anti-anti-CRISPR rescue mechanism of St1Cas9 remains unexplored.

To uncover dynamic conformational changes of CRISPR-Cas systems, single-molecule Förster resonance energy transfer, solution nuclear magnetic resonance, and computational methods have been widely used. Among these approaches, molecular dynamics (MD) simulations are a powerful tool for exploring the conformational landscapes at an atomic level and directly uncovering biomolecular mechanisms [41,70–74]. Indeed, previous investigations utilising large-scale MD simulations have successfully elucidated the mechanism behind a series of key events including PAM recognition [27,75,76], HNH dynamics [42,77,78], and off-target effects [36,79] to obtain dynamic biophysical information that is otherwise inaccessible with the currently available experimental techniques [80].

Here, we performed extensive all-atom MD simulations of St1Cas9 in multiple different states (Supplementary Fig. 2, in total 120 μs) to elucidate the conformational reorganisation of St1Cas9, characterise the detailed allostery underlying AcrIIA6-mediated St1Cas9 inhibition, and explore the mechanism through which triple mutations of St1Cas9 escape inhibition by AcrIIA6. We found that AcrIIA6 binding induced a conformational transition of St1Cas9 endonuclease towards a 'departing' inactive state characterized by WED and PI contraction as well as HNH outward relocation, resulting in both steric occlusion for PAM recognition and increased spatial distance between nuclease and cleavage sites as well as compromising the propagation of allosteric signals in the PAM recognition and nuclease domains. Multiple mutagenesis experiments revealed that a St1Cas9 variant with triple residue substitutions (G993K/K1008M/K1010E) is efficiently resistant to AcrIIA6 inhibition. Our study not only provides a deep atomic-level understanding of St1Cas9 allosteric inhibition and rescue mechanism, but also offers a promising potential for the design of allosteric AcrIIA6 peptidomimetics.

2. Results

2.1. AcrIIA6 induces varied conformational dynamics of St1Cas9

To elucidate the detailed allosteric inhibitory mechanism of St1Cas9 by AcrIIA6 at the atomic level, multiple MD simulations of St1Cas9 in different states were performed for eight systems, including *apo* St1Cas9, St1Cas9·sgRNA, St1Cas9·sgRNA·dsDNA, St1Cas9·sgRNA·AcrIIA6, St1Cas9·sgRNA·dsDNA·AcrIIA6, *mut* St1Cas9·sgRNA, *mut* St1Cas9·sgRNA·AcrIIA6, and *Streptococcus pyogenes* Cas9 RNP (SpCas9·sgRNA) (Supplementary Fig. 2). Each system underwent 3 μs × 5 independent runs with random initial velocities, leading to a cumulative simulation timescale of 120 μs. After reaching relative equilibrium after 500-ns MD simulation, a ~1.3-fold increase in root-mean-square deviation (RMSD) value for St1Cas9 in the St1Cas9·sgRNA RNP was observed in the presence of AcrIIA6, indicating considerable conformational variation compared with that of the initial structural complex (Supplementary Results 2.1 and Supplementary Fig. 3). Moreover, comparing the average structures of St1Cas9·sgRNA·AcrIIA6 and SpCas9·sgRNA throughout the simulations revealed that, owing to the absence of the WED domain in SpCas9 that was involved in both PAM recognition and AcrIIA6 interaction in St1Cas9, AcrIIA6 was incapable of binding to SpCas9·sgRNA, as well as other bacterial species lacking the WED domain (Supplementary Fig. 4). This result was consistent with the biolayer interferometry experiments that revealed the insensitivity of SpCas9 to AcrIIA6 [2].

To uncover the domain-specific influence of AcrIIA6, we analysed the per-residue root mean square fluctuations (RMSF) of St1Cas9 (Supplementary Results S2.1 and Supplementary Fig. 5),

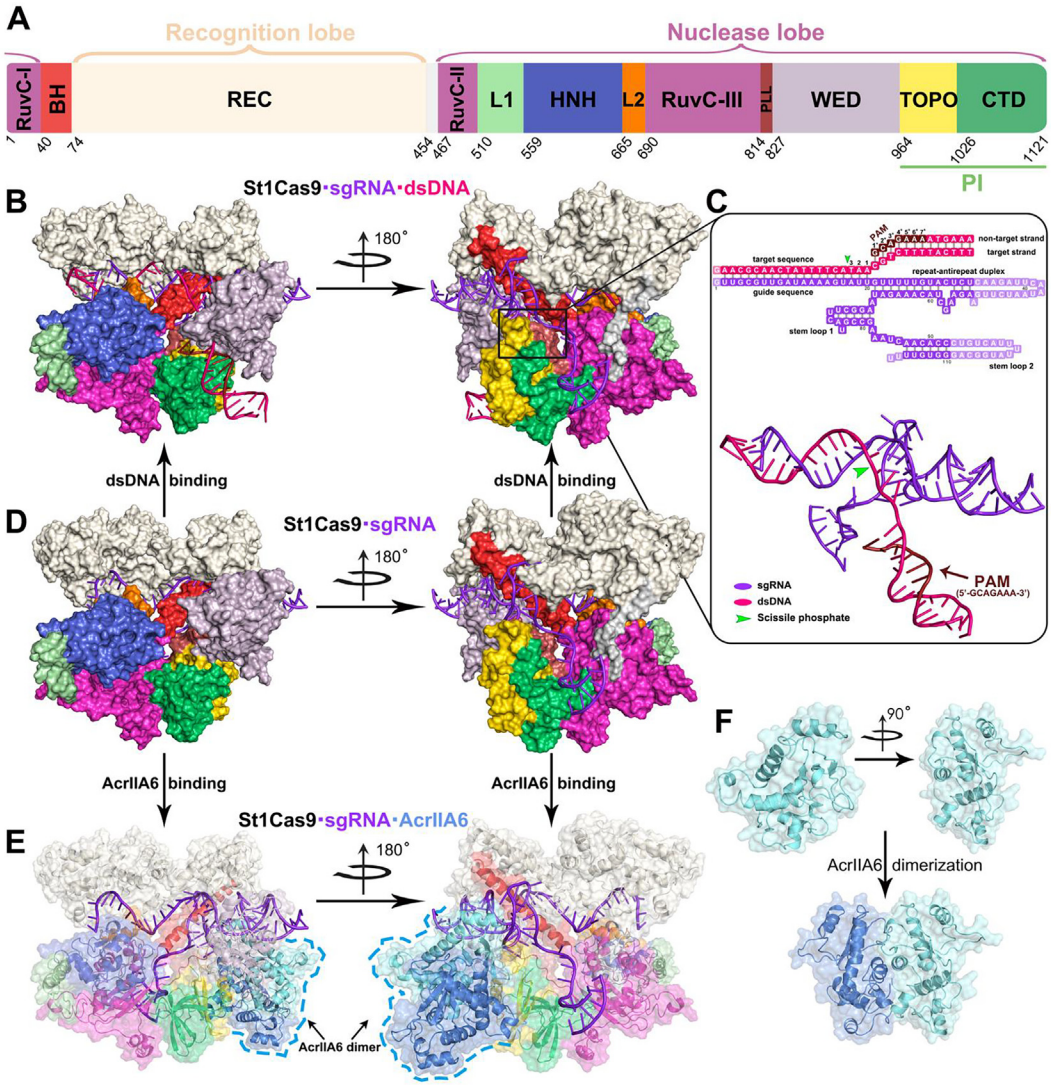

Fig. 1. Overall structure of St1Cas9·sgRNA in complex with dsDNA or AcrIIA6. (A) Domain organisation of St1Cas9. (B) Surface representation of St1Cas9·sgRNA·dsDNA, with St1Cas9 colour coded as in (A). (C) Nucleotide sequence (top) and a magnified cartoon diagram (bottom) of sgRNA (purple) in complex with dsDNA (hotpink) containing the PAM sequence (brown). Catalytic site on tDNA is indicated with a green arrow. (D) Surface representation of St1Cas9·sgRNA. (E) Surface representation of St1Cas9·sgRNA·AcrIIA6; St1Cas9 is made translucent to highlight the AcrIIA6 dimer. (F) Surface representation of AcrIIA6 monomer (top) and dimer (bottom). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and projected the RMSF difference between simulation systems with dissimilar compositions onto the initial St1Cas9 structure for more intuitive representations (Supplementary Fig. 6). Despite significantly increased residue fluctuations of HNH and WED domains in the presence of AcrIIA6, the most prominent RMSF was observed within the L1 that bridges the catalytic RuvC and HNH domains. In addition, L2 exhibited diverse fluctuations under different situations, in which AcrIIA6 binding was found to enhance L2 flexibility. Notably, the catalytic HNH domain resides topologically distant from the direct AcrIIA6 binding interface. Thus, the increased HNH flexibility in St1Cas9·sgRNA·AcrIIA6 is a good demonstration of the allosteric effects of AcrIIA6, in which the flexible L1 and L2 linkers may play a pivotal role as allosteric transducers [76]. Furthermore, although it is involved in direct interaction with AcrIIA6, the WED domain had a more primary function as a key component in mediating target DNA recognition and binding together with the PI domains (containing TOPO and CTD), whose fluctuations remained largely unchanged after AcrIIA6 binding. The prominently altered conformational flexibility of proximal WED and distal HNH domains support the notion that AcrIIA6 acts through modifying DNA-associated conformational dynamics of the WED domain, which also allosterically enhanced plasticity of HNH and positioned the nuclease domain in an unfavourable conformation for subsequent DNA cleavage (Supplementary Results S2.1).

2.2. AcrIIA6 induces St1Cas9 conformational transitions from 'approaching' to 'departing'

Principal component analysis (PCA) was performed to characterise the global conformational transition pattern and domain dynamics (Supplementary Methods). The conformational space sampled by St1Cas9 in each system was investigated using the two most representative collective principal components (PC1 and PC2). Distinct conformations were adopted by St1Cas9 dependent on AcrIIA6 binding, as evidenced by the dominant conformational clusters that were well distinguished along PC1, with PC1 < 0 corresponding to the AcrIIA6-inactivated St1Cas9 (Fig. 2A and Supplementary Fig. 7). Further insights into the domain dynamics along PC1 revealed a 'departing' conformation, characterised

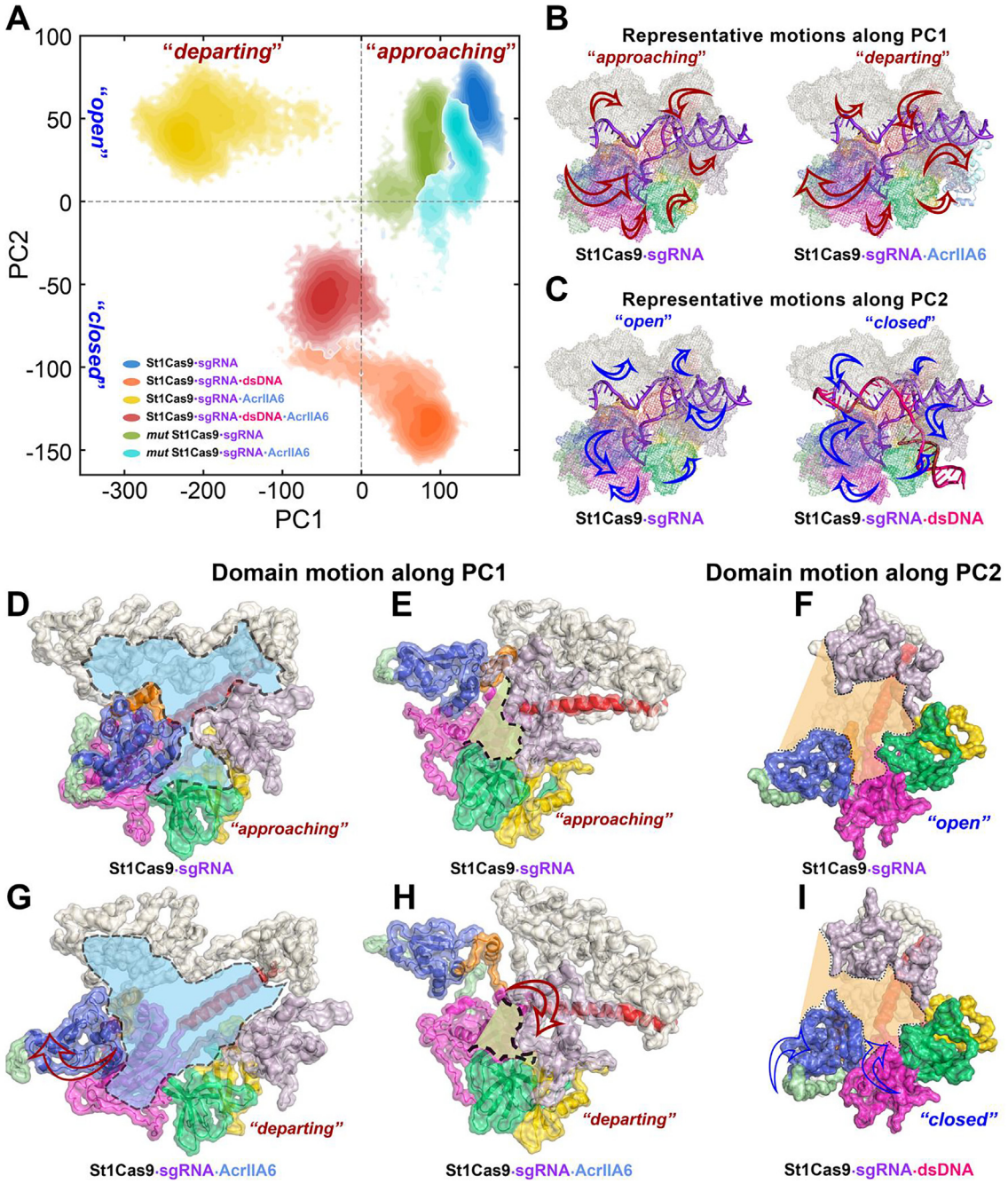

Fig. 2. Approaching to departing and open to closed conformational transition of St1Cas9. (A) Representative PCA projection results on the conformational subspace defined by the first two collective principal components (PC1 and PC2) of St1Cas9 in St1Cas9·sgRNA (blue), St1Cas9·sgRNA·dsDNA (orange), St1Cas9·sgRNA·AcrIIA6 (yellow), St1Cas9·sgRNA·dsDNA·AcrIIA6 (crimson), *mut* St1Cas9·sgRNA (yellow green), and *mut* St1Cas9·sgRNA·AcrIIA6 (turquoise). The panel is divided into regions of departing (PC1 < 0) and approaching (PC1 > 0) as well as open (PC2 > 0) and closed (PC2 < 0) along the PC1 and PC2 axes, respectively. (B) Comparison of principal mode of motion along PC1 in St1Cas9·sgRNA (left) and St1Cas9·sgRNA·AcrIIA6 (right), indicating an approaching to departing conformational transition (animation in Supplementary Movie S1). (C) Comparison of principal mode of motion along PC2 in St1Cas9·sgRNA (left) and St1Cas9·sgRNA·dsDNA (right), indicating an open to closed conformational transition (animation in Supplementary Movie S2). (D) Front view (HNH domain indicated) and (E) side view (WED and PI domain indicated) of the approaching conformation along PC1 in St1Cas9·sgRNA. (G) Front view (HNH domain indicated) and (H) side view (WED and PI domain indicated) of departing conformation along PC1 in St1Cas9·sgRNA·AcrIIA6. (F) Side view of the open conformation along PC2 in St1Cas9·sgRNA. (I) Side view of the closed conformation along PC2 in St1Cas9·sgRNA·dsDNA. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mainly by relocations of the HNH, WED, and PI domains in the St1Cas9·sgRNA·AcrIIA6 complex relative to the 'approaching' conformation observed without AcrIIA6 (Fig. 2B and Supplementary Movie 1). In the departing conformation, the WED and PI domains synchronously moved towards AcrIIA6 at the back of the RNP; this movement constricted the DNA recognition groove (Fig. 2E, H), which may be too narrow to accommodate the DNA duplex when searching throughout the genomic landscape for matching PAM sequences—the prerequisite for initiating DNA unwinding and melting. Meanwhile, the catalytic HNH domain shifted away from the central concave region responsible for heteroduplex accommodation (Fig. 2D, G), which can result in increased spatial distance between the catalytic and cleavage sites and is characteristic of Cas9 under an inactivated state. Additionally, comparing the prin-

cipal motions of St1Cas9 in St1Cas9·sgRNA and St1Cas9·sgRNA·dsDNA, whose conformational clusters were separated along PC2, indicated closure of the bilobed architecture with inward movement of the catalytic domains after dsDNA loading (Fig. 2A, C, F, I, and Supplementary Movie 2), which was in agreement with the classic open-to-closed conformational transition [27,42,76,81,82]. Based on these observations, PC1 can be viewed as a parameter measuring the ability of St1Cas9 RNP to recognise PAM sequences and associate with complementary dsDNA (DNA-relevant dynamics), as both WED and PI domains are primarily implicated in DNA recognition and accommodation-relevant events, while PC2 indicates the potential of the surveillance complex to transit into the 'closed' catalytically active state, in which prominent inward relocations of REC and NUC lobes occurred. Our simulation results indicate that AcrIIA6 showed compromised inhibition towards St1Cas9·sgRNA·dsDNA, as evidenced by the location of major conformational clusters for St1Cas9·sgRNA·dsDNA·AcrIIA6 with PC2 < 0, which represented the catalytically active closed conformation (Fig. 2A).

Collectively, the results indicate that association of St1Cas9 with the allosteric effector AcrIIA6 leads to altered St1Cas9 conformational dynamics with departing characteristics. In addition to the prominent conformational rearrangements of WED and PI domains, AcrIIA6 triggered relocation of the nuclease HNH domain away from the central concave for heteroduplex accommodation, resulting in unfavourable PAM recognition and catalytic activation and ultimately decreasing DNA binding affinity and catalytic activity of St1Cas9 RNP (Supplementary Results S2.2).

2.3. Contraction of WED and PI domains of St1Cas9 upon AcrIIA6 binding hinders efficient PAM interrogation

Based on the results from PCA analysis where distinct behaviour of WED and PI domains that might be unfavourable for the binding of incoming substrate DNA in the presence of AcrIIA6, we conducted further experiments to dissect the detailed dynamics and molecular basis underlying such conformational shift. Direct interactions of AcrIIA6 with St1Cas9 induced significant structural rearrangements of the WED and PI domains, whose back face is responsible for PAM recognition – the first of the sequential events leading to target cleavage. To investigate the dynamic features of WED and PI associated with both PAM recognition and AcrIIA6 binding, two order parameters were calculated based on structural information from the cumulative 15-μs MD sampling of each simulation system. One parameter was defined as the perimeter of the triangle formed by the WED residue K837 and two PI residues K1049 and N1079 ($\Delta_{K837-K1049-N1079}$), which reflects the relative degree of openness of regions associated with PAM recognition and subsequent dsDNA accommodation. The other parameter was defined by the volume of one of the major AcrIIA6 binding pockets for AcrIIA6^A β2-β3 hairpin insertion, which is sandwiched between WED and PI domains and constitutes a major interface for intramolecular recognition between AcrIIA6 and St1Cas9 RNP [2], and can monitor the binding affinity between St1Cas9 RNP and AcrIIA6 (graphic representation of the two order parameters is shown in Supplementary Fig. 8).

The potential of mean force (PMF) as a function of the selected order parameters indicated that PAM recognition and AcrIIA6 binding, both mediated by WED and PI domains, were mutually exclusive events to some extent, leading to activation and inhibition of St1Cas9 RNP, respectively (Fig. 3 and Supplementary Table S1). In St1Cas9·sgRNA (Fig. 3A), two distinct states were observed: one with a smaller $\Delta_{K837-K1049-N1079}$ perimeter but a larger AcrIIA6 binding pocket volume (C^L), and the other was the opposite (C^R). The coexistence of C^L and C^R coincided with the propensity of St1Cas9 to be bound by either AcrIIA6 or dsDNA, indicating an innate equilibrium between the two states with different WED and PI domain configurations (Supplementary Fig. 9A). Upon association with dsDNA (Fig. 3B), as a result of WED and PI conformational rearrangements, the pocket volume for AcrIIA6^A β2-β3 hairpin insertion decreased significantly compared with that of C^L before DNA loading. This reduction was accompanied by a change in pocket shape, hampering St1Cas9·sgRNA·dsDNA complex recruitment of AcrIIA6. Meanwhile, the smaller $\Delta_{K837-K1049-N1079}$ perimeter in the St1Cas9·sgRNA·dsDNA compared with that of C^R in the St1Cas9·sgRNA supported the regional contraction of WED and PI domains for tight association with dsDNA (Supplementary Fig. 9B–D), in which residues from St1Cas9 WED, TOPO, and CTD domains (K867, S961, V962, S963, K1025, and N1026) established stable polar contacts with the phosphoribosyl backbone of the PAM sequence, especially with nucleotides dC2*, dA3*, dG4*, and dA7* (Fig. 3G, H and Supplementary Fig. 10). Upon binding of AcrIIA6 (Fig. 3C), the pocket volume was retained at a high level; however, a drastic decrease in $\Delta_{K837-K1049-N1079}$ perimeter was observed. This indicated a shrinkage of the groove encompassed by WED and PI domains, which is insufficient for PAM recognition and dsDNA accommodation (Supplementary Fig. 9C, E, F). Even in the presence of dsDNA, contraction of the PAM recognition groove was only slightly relieved (Fig. 3C, D). The flat groove resulting from AcrIIA6 binding markedly weakened overall interactions between PAM and St1Cas9 by increasing conformational plasticity of dsDNA, which underwent an approximate 5.2 Å inward shift towards St1Cas9 to avoid potential structural conflicts with the deformed PAM recognition groove (Fig. 3G, I, Supplementary Figs. 10, 11, and Supplementary Table S2). Altogether, the results indicate that contraction of WED and PI domains in the St1Cas9·sgRNA induced by AcrIIA6 binding would compromise PAM recognition and DNA binding capability of St1Cas9 RNP, leading to inhibition of the surveillance complex (Supplementary Results S2.3).

2.4. Counter-effects of WED and PI conformational dynamics on AcrIIA6

Given the significant weakening of intermolecular recognition between dsDNA and St1Cas9 in the presence of AcrIIA6, we subsequently probed the AcrIIA6-St1Cas9 RNP binding interface of different systems. Our MD simulation results demonstrated a mixed protein-nucleotide interface recognised by AcrIIA6, which inserted structural motifs from both monomers into the clefts formed within St1Cas9 RNP to facilitate stable anchoring (Fig. 4A, B). The backbone of stem loop 1 (ST1) from sgRNA (G71, C72, U73, C75) strongly interacted with the AcrIIA6^A β2-β3 hairpin (T118) and L9 loop (G167, R168, N170), constituting the nucleotide interface for AcrIIA6 docking in St1Cas9 RNP. The protruding hairpin (R120, Y122, A124, N127, Y128, A130) and L8 loop (G146, S147, Q151) of AcrIIA6^A extensively interacted with St1Cas9 residues from WED (G943, D947) and PI (D987, Q989, K1010) domains. In addition, the AcrIIA6^B α5-α6 helix residue N83 established strong polar interactions with the PI (F1019, D1115) domain. However, the density and intensity of the interaction network between St1Cas9 RNP and AcrIIA6 markedly changed in response to dsDNA binding (Fig. 4C, D, Supplementary Fig. 12, 13, and Supplementary Table 3, 4). Despite a few weakened contacts (e.g. AcrIIA6^A R168 and U73; AcrIIA6^A A124 and K1010), the overall intermolecular interaction network was strengthened owing to a few newly established interactions. However, the molecular mechanics/Poisson Boltzmann surface area (MM/PBSA) binding free energy calculations between AcrIIA6 and St1Cas9 RNP with/without dsDNA suggested reduced affinity towards St1Cas9·sgRNA·dsDNA

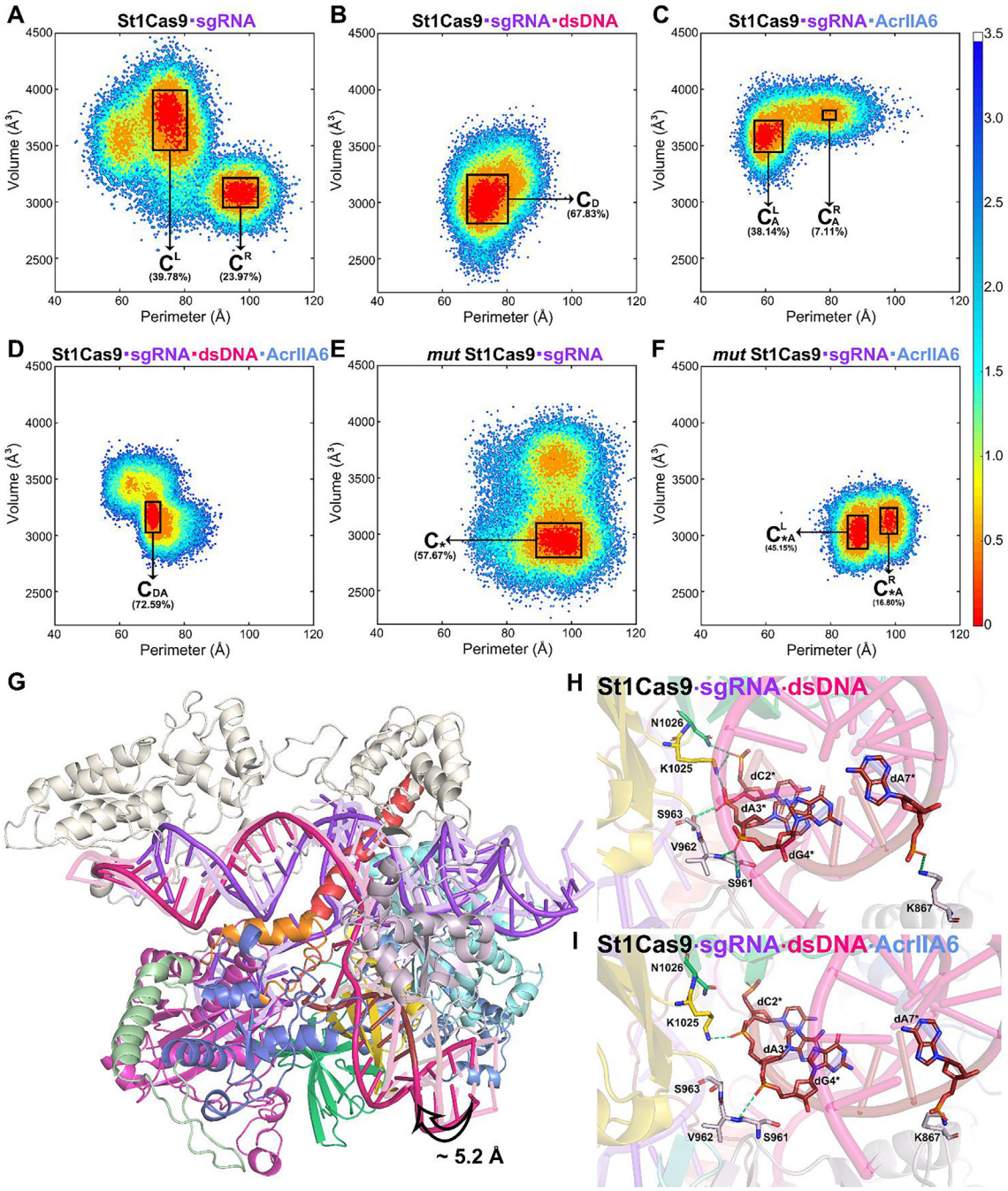

Fig. 3. Contraction of WED and PI domains hinders efficient PAM interrogation. Conformation FEL associated with PAM recognition and AcrIIA6 binding in (A) St1Cas9·sgRNA, (B) St1Cas9·sgRNA·dsDNA, (C) St1Cas9·sgRNA·AcrIIA6, (D) St1Cas9·sgRNA·dsDNA·AcrIIA6, (E) *mut* St1Cas9·sgRNA, and (F) *mut* St1Cas9·sgRNA·AcrIIA6. The landscape was generated with the perimeter of triangle $\Delta_{K837-K1049-N1079}$ and the pocket volume for AcrIIA6^{A} β2-β3 hairpin insertion. The fraction of each cluster is indicated in brackets. (G) Representative structure extracted from energy basin C_{DA} in the St1Cas9·sgRNA·dsDNA·AcrIIA6 system. Nucleotide components (translucent ones) from C_D in St1Cas9·sgRNA·dsDNA is superimposed onto the nucleic backbone in C_{DA} as contrast; the ~5.2 Å inward shift of the dsDNA duplex in the presence of AcrIIA6 is denoted by a curved arrow. St1Cas9 domains are colour coded as in Fig. 1A. Molecular interactions between PAM nucleotides and St1Cas9 residues in (H) St1Cas9·sgRNA·dsDNA and (I) St1Cas9·sgRNA·dsDNA·AcrIIA6.

(Supplementary Methods and Supplementary Table 5). Such unfavourable binding may be attributed to a prominent decrease in the interfacial area between St1Cas9 RNP and AcrIIA6 upon dsDNA loading, in which both sgRNA and St1Cas9 engaged a smaller area for interaction with AcrIIA6 (Supplementary Table 6). Interestingly, AcrIIA6 as a dimer was also destabilised upon binding to DNA-loaded St1Cas9 RNP, as revealed by the reduced binding free energy between AcrIIA6 monomers (Supplementary Table 5). Thus, together with the compromised binding affinity towards St1Cas9·sgRNA·dsDNA, the counter-effect on the AcrIIA6 dimer by conformational dynamics associated with dsDNA may contribute to the inefficient inhibition of St1Cas9·sgRNA·dsDNA by AcrIIA6.

2.5. AcrIIA6 alters conformational plasticity of the distal HNH domain

The outward displacement of the nuclease HNH domain is a unique feature of the AcrIIA6-induced global conformational tran-

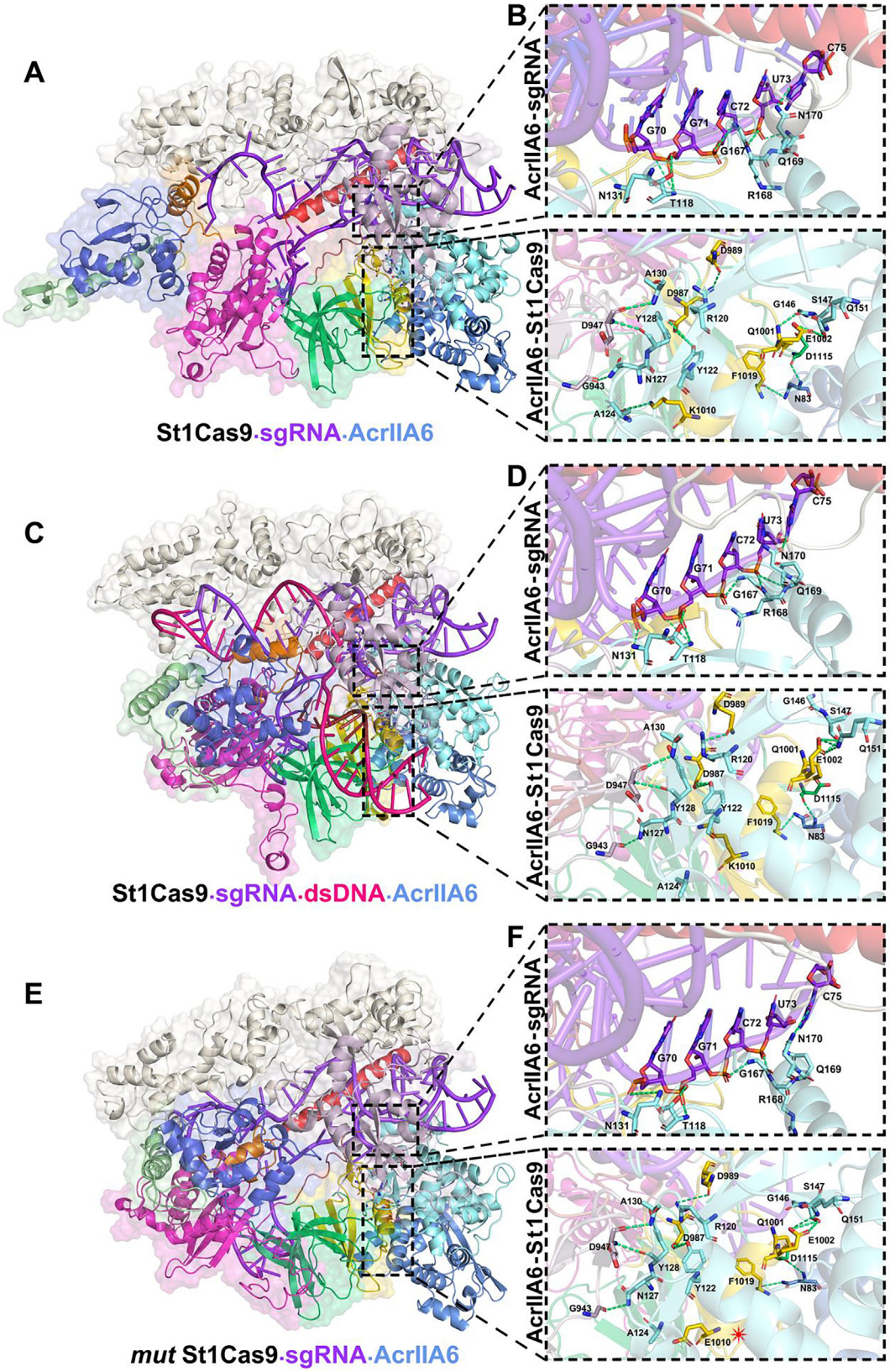

Fig. 4. AcrIIA6-St1Cas9 RNP interaction network in different systems. Structure of representative conformations and magnified view of the molecular interactions established between AcrIIA6–sgRNA (top) and AcrIIA6–St1Cas9 (bottom) in (A, B) St1Cas9·sgRNA·AcrIIA6, (C, D) St1Cas9·sgRNA·dsDNA·AcrIIA6, and (E, F) *mut* St1Cas9·sgRNA·AcrIIA6 (✶ K1010E mutation). The landscape was generated with the perimeter of triangle $\Delta_{K837-K1049-N1079}$ and the pocket volume for AcrIIA6^{A} β2-β3 hairpin insertion. The fraction of each cluster is indicated in brackets. (G) Representative structure extracted from energy basin C_{DA} in the St1Cas9·sgRNA·dsDNA·AcrIIA6 system. Nucleotide components (translucent ones) from C_D in St1Cas9·sgRNA·dsDNA is superimposed onto the nucleic backbone in C_{DA} as contrast; the ∼5.2 Å inward shift of the dsDNA duplex in the presence of AcrIIA6 is denoted by a curved arrow. St1Cas9 domains are colour coded as in Fig. 1A. Molecular interactions between PAM nucleotides and St1Cas9 residues in (H) St1Cas9·sgRNA·dsDNA and (I) St1Cas9·sgRNA·dsDNA·AcrIIA6.

sition to the departing conformation. We then investigated into St1Cas9 HNH domain conformational dynamics upon AcrIIA6 binding. As revealed by the PCA analysis, AcrIIA6 binding induced outward movement of the distal HNH domain. We suspected that such departure from the catalytic core would result in significant barrier for St1Cas9·sgRNA·AcrIIA6 to attain the catalytically active state even if dsDNA can be loaded onto the complex, thus, we employed the most prevalent conformation from St1Cas9·sgRNA·dsDNA (C_D) which well reflected the catalytically competent state to draw a comparison with representative St1Cas9 structure from St1Cas9·sgRNA·AcrIIA6 (C_A^L). Comparison between the two representatives showed that, compared with the catalytically-competent state, the HNH domain of St1Cas9 exhibited ~59° rotation around the L2 hinge that led to departure from the central concave upon AcrIIA6 binding (Fig. 5A, B, measured when superimposing the two structures with Pymol). This outward relocation further increased the distance by 12.57 Å between the catalytic residue H599 in the HNH domain and the presumed location for scissile phosphate on the target strand (Fig. 5C). Compared with St1Cas9·sgRNA, dsDNA loading did not affect HNH conformational plasticity but significantly enhanced structural flexibility of L1 and rendered L2 less stable (Fig. 5D and Supplementary Fig. 14A, B). This allowed the inward movement of the HNH nuclease domain during conformational transition toward the closed catalytically active state in the presence of substrate DNA (also detailed in above PCA section). In contrast, the association of St1Cas9·sgRNA with AcrIIA6 reduced the original structural stability of both HNH and L1 but stabilised L2 (Fig. 5E and Supplementary Fig. 14C), thereby permitting rotation-based outward movement of the HNH nuclease domain, which was also opposite to the moving direction induced by dsDNA. The opposite structural stability of L1 and L2 linker of St1Cas9·sgRNA would indeed result in distinct HNH motion: when L2 loosens, the HNH domain trends to move towards L2 which resides near the main body of St1Cas9 and thus enables activation; while L1 loosens, the HNH domain falls to the rim of St1Cas9 nuclease as L1 lies in the relatively outer part and ultimately contributed to inhibition (Fig. 5D, E). In the presence of either dsDNA or AcrIIA6 binding, interactions between HNH and the remaining domains in St1Cas9 decreased significantly—as revealed by the reduced number of hydrogen bonds established between them (Supplementary Fig. 15) – with greater HNH flexibility in St1Cas9·sgRNA·AcrIIA6.

2.6. Correlated motion between functional domains mediates AcrIIA6 allosteric inhibition

Generalised correlation (GC_{ij}) analysis, which captures both linear and nonlinear correlations based on Shannon entropy, was conducted to identify dynamic correlations between residues [83]. The results showed that binding of the allosteric inhibitor AcrIIA6 reshaped the overall correlation structure and strengthened the correlated motions within St1Cas9 (Supplementary Fig. 16A, C, G), as evidenced by the altered GC_{ij} distribution in St1Cas9·sgRNA (~0.67) to a more correlated state in St1Cas9·sgRNA·AcrIIA6 (~0.72). This increase in coupled motion was consistent with the allosteric characteristic of AcrIIA6, which relied on concerted motions between distal sites for global conformational reassembly. Notably, compared with substrate DNA binding, an association with AcrIIA6 induced more significant enhancement of correlated movements between residues, especially between REC components and the remaining part of St1Cas9. Previous studies have suggested that REC motional dynamics play a role in the open to closed conformational transitions of Cas9 [38]. The altered correlation in our study suggests that AcrIIA6 binding may disrupt REC-mediated motional correlation and hinder further conformational transitions underlying endonuclease activation.

We further investigated interdependent motion between functional domains based on the obtained residue GC_{ij} matrix (Fig. 6). The St1Cas9·sgRNA complex possessed the weakest inter-domain correlation network, in which only the REC lobe displayed a relatively strong correlation with domains from the NUC lobe (Fig. 6A). Loading of dsDNA increased overall domain connectivity (Fig. 6B), with enhanced correlation between HNH and PI domains, confirming the ability of direct AcrIIA6-binding sites to allosterically regulate the catalytic activity of remote nuclease domains. Meanwhile, HNH and RuvC nuclease domains established stronger motion correlation upon dsDNA loading, which is the basis of synchronised DNA double-strand cleavage. AcrIIA6 binding also strengthened the correlation map within St1Cas9 and particularly enhanced the interdependent conformational dynamics between the REC and NUC lobes (Fig. 6C), indicating the role of REC in coordinating nuclease activity in AcrIIA6-induced allostery. In addition to the strong correlation between the nuclease RuvC and HNH domains, WED and PI domains also displayed stronger inter-correlation, which was supported on the structural level by their synchronous movement towards a contracted conformation with a shrunken DNA binding groove during AcrIIA6-induced conformational structural transition to the departing state.

Collectively, the results indicate that changes in domain motion correlations induced by either substrate DNA or AcrIIA6 binding underlie the structural transition to the catalytically active closed or inactive departing state as well as the coupled motion mediated by WED and PI along the direct AcrIIA6 binding interface. Furthermore, the RuvC and HNH nuclease domains are recognised as key elements in inhibitor-induced allostery.

2.7. AcrIIA6 induces community reorganisation and allosteric signalling

Dynamic network analysis was performed to uncover the correlation structure in different simulation systems, and closely related residues were clustered into 'communities' inter-connected with each other. The obtained community networks were visualised as spheres connected by sticks, whose widths were proportional to the connectivity strength of the edges bridging them (Fig. 7A–F). In St1Cas9 RNP, community 9 (Fig. 7A), composed predominantly of BH residues, served as a major hub for information transfer between the REC and NUC lobes, while community 10, which represented PI domains, only loosely interacted with the remaining contact network (domain composition for each community is listed in Supplementary Table 7). The recognition and loading of dsDNA reduced the community number and enhanced allosteric information transfer (Fig. 7B). Under such circumstances, the PI domain-containing community 10 established stronger correlation with both WED (community 7) and RuvC (community 16) domains, while the latter exhibited enhanced connection with community 14, which largely consisted of the HNH nuclease domain and L2 linker through community 13. This strong signal transduction was in consistence with the experimentally identified tight correlations between the catalytic activity of the two nuclease cores mediated by the linker bridging them. Binding of AcrIIA6, as shown in Fig. 7C, fragmented the community structure of St1Cas9 RNP, but strengthened global allosteric signalling. AcrIIA6 was mainly found deposited in two communities (community Acr^A and Acr^B), each of which consisted predominantly of residues from an AcrIIA6 monomer. $AcrIIA6^A$ was found to establish extensive contacts with St1Cas9 residues/nucleotides, reflected by its coexistence with WED, TOPO, CTD, and sgRNA components in the same community (community 4, 6, and 10), whereas $AcrIIA6^B$ exhibited

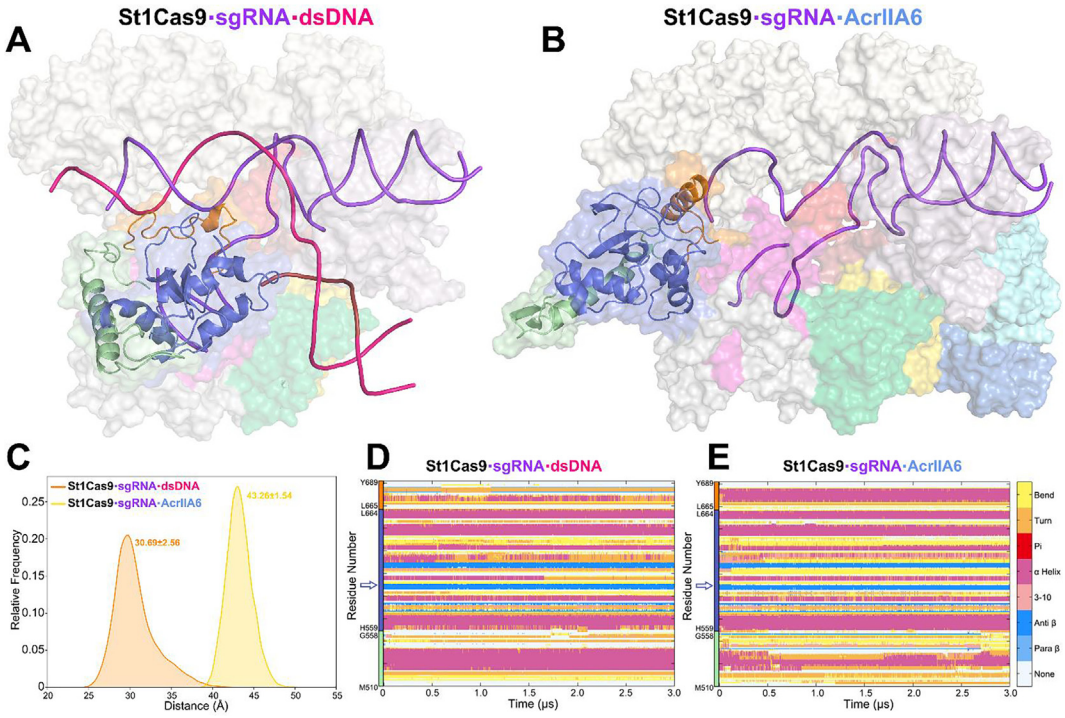

Fig. 5. AcrIIA6 alters conformational plasticity of the distal HNH domain. Molecular surface representation with L1, HNH, and L2 in representative conformations from (A) C_D in St1Cas9·sgRNA·dsDNA and (B) C_A^L in St1Cas9·sgRNA·AcrIIA6. (C) Probability distribution map of the distance between catalytic H599 and scissile phosphate (T3) in St1Cas9·sgRNA·dsDNA (orange) and St1Cas9·sgRNA·AcrIIA6 (yellow; presumed cleavage site inferred from structural alignment with St1Cas9·sgRNA·dsDNA). All five productive runs were calculated as DSSP plots for real-time secondary structure transitions within the L1, HNH, and L2 regions in (D) St1Cas9·sgRNA·dsDNA and (E) St1Cas9·sgRNA·AcrIIA6. Position of catalytic H599 is indicated by a blue outlined arrow along the vertical axis indicating residue number. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

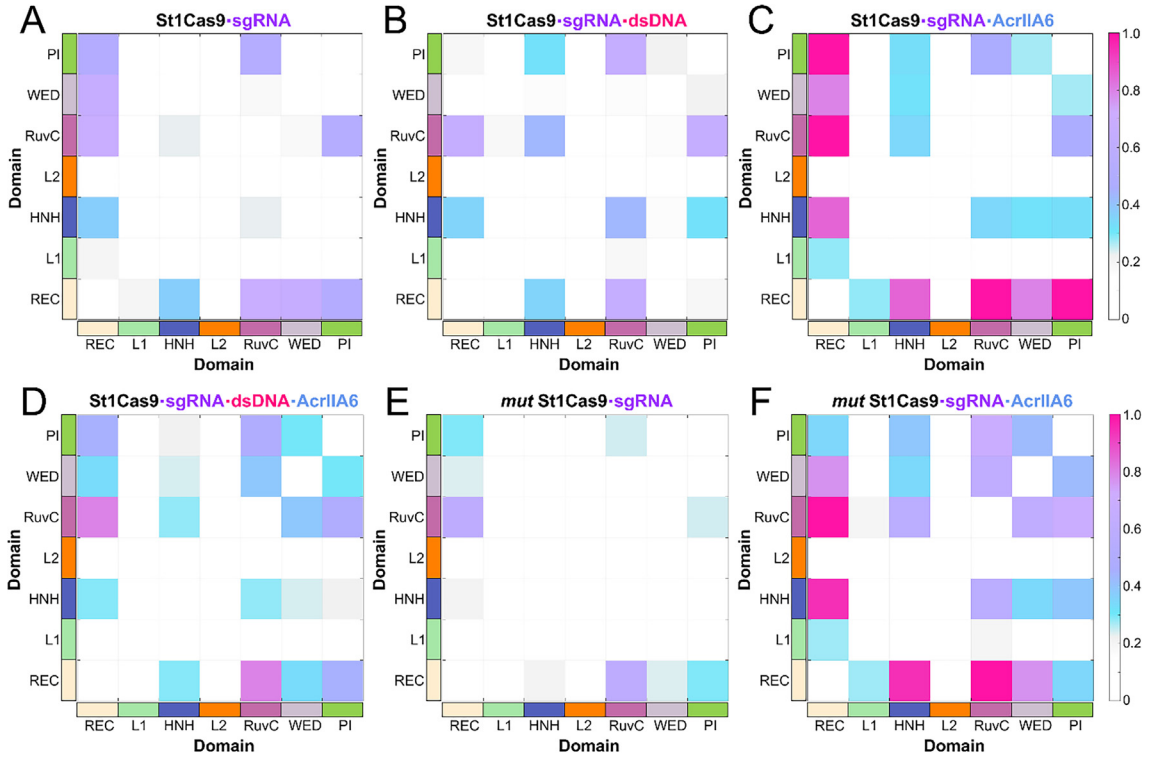

Fig. 6. Inter-domain correlation of St1Cas9 in different simulation systems. Interdomain correlations calculated by accumulating per-residue CS^{inter} within each domain in (A) St1Cas9·sgRNA, (B) St1Cas9·sgRNA·dsDNA, (C) St1Cas9·sgRNA·AcrIIA6, (D) St1Cas9·sgRNA·dsDNA·AcrIIA6, (E) *mut* St1Cas9·sgRNA, and (F) *mut* St1Cas9·sgRNA·AcrIIA6. The BH and PLL domains were left out for clarity. RuvC collectively refers to RuvC-I, RuvC-II, and RuvC-III, while PI contains both TOPO and CTD domains.

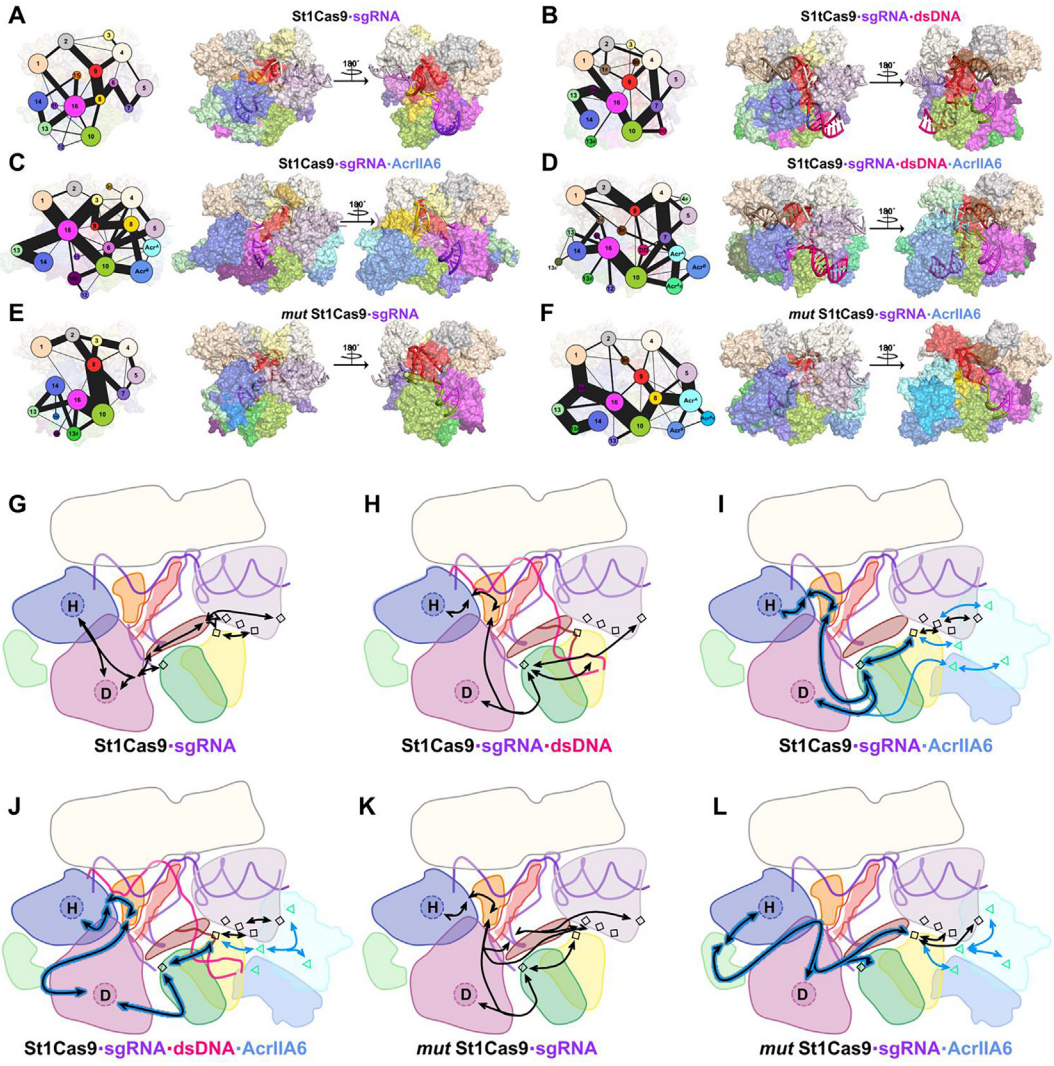

Fig. 7. Community network and the allosteric signalling pathway in each simulation system. Community network representation of (A) St1Cas9·sgRNA, (B) St1Cas9·sgRNA·dsDNA, (C) St1Cas9·sgRNA·AcrIIA6, (D) St1Cas9·sgRNA·dsDNA·AcrIIA6, (E) *mut* St1Cas9·sgRNA, and (F) *mut* St1Cas9·sgRNA·AcrIIA6 (with nucleotides, detailed in Supplementary section 1.7; community components are summarised in Supplementary Table 7). Each sphere represents an individual community whose residue number is indicated with a sphere area, and the thickness of sticks connecting communities is proportional to the corresponding edge connectivity. Schematic representation of the domain-level allosteric signalling pathway connecting catalytic H599 in HNH domain (denoted with (H)), D9 in the RuvC domain (denoted with (D)), PAM interacting residues (PIR) in WED and PI domains (WED: K867, S961, V962, S963; TOPO: K1025; CTD: N1026; denoted with □), and AcrIIA6 residues ($T118^A$, $A124^A$, $S147^A$, $N83^B$; denoted with △). The bidirectionally curved black arrow indicates the information transfer pathway; sky blue arrow indicates AcrIIA6-specific signalling pathways; signalling pathways shared between catalytic residues and PIRs and those from AcrIIA6 are highlighted with sky-blue contour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

more exclusive allosteric information transfer through the CTD domain, which was consistent with the stable interaction observed between $AcrIIA6^B$ N83 and CTD residues in our analysis. The presence of AcrIIA6 further wired community 16 as the signal transduction hub in the community network of St1Cas9 RNP, in which it received regulatory signals from community 10 and further allosterically promoted them toward the distal REC lobe (community 1, 3) and catalytic HNH domain (community 14). Under both dsDNA-loaded and AcrIIA6-bound conditions, community 13, whose major components were L1 residues, mediated informational flow between the two catalytic domains (community 14 and community 16), which is in line with previous studies showing L1 as an allosteric transducer [42,81]. The more fragmented community architecture with considerable connectivity in St1Cas9·sgRNA·dsDNA·AcrIIA6 may result from the coexistent allosteric effects imposed by both PAM and AcrIIA6 (Fig. 7D) because previous studies also demonstrated the allosteric effects of the PAM sequence [27].

Similar effects on the community network of St1Cas9 RNP were observed by either AcrIIA6 binding or dsDNA loading—both exhibited enhanced signal transduction between communities composed of WED, PI, and more distal subgroups within HNH and REC via the RuvC domain (community 16). Such similarities in community reorganisation may contribute to the prominent HNH relocation induced by both dsDNA and AcrIIA6, although in opposite directions. Nonetheless, the inhibitor indeed strengthened the crosstalk between communities to a greater extent, as represented by thicker inter-community sticks, which may well underlie the more prominent outward displacement/rotation of HNH domain during AcrIIA6-induced transition from approaching to departing conformation.

We next aimed to elucidate the detailed information transfer pathway underlying AcrIIA6-induced WED and PI contraction and HNH outward reorientation as well as how the catalytic domains coordinate with each other in this process. To this end, we computed the optimal residue-by-residue signalling pathway

connecting AcrIIA6 (T118^{A}, A124^{A}, S147^{A}, N83^{B}), PAM-interacting residues (PIRs, including WED: K867, S961, V962, S963; TOPO: K1025; CTD: N1026), HNH (H599), and RuvC (D9) catalytic sites. The obtained signalling pathways are summarised on both the residue (Supplementary Fig. 17, 18) and simplified domain level (Fig. 7G–L). In St1Cas9 RNP alone (Fig. 7G), signals from PIRs were transmitted to HNH via RuvC through an unchanged pathway bypassing the PLL domain that structurally connects WED and nuclease domains. Meanwhile, catalytic D9 communicated with H599 only through RuvC and HNH residues. Generally, dsDNA loading and AcrIIA6 binding both wired the communication pathway between the above components, mainly by enabling direct signalling from PIRs to RuvC without transmitting through PLL, and introducing L2 residues along the signalling path between D9 and H599 catalytic sites to achieve a more intense and efficient transduction (Fig. 7H, I and Supplementary Fig. 17). Specifically, PAM nucleotides were employed to enable efficient signal transmission between WED and RuvC domains in St1Cas9·sgRNA·dsDNA. However, stronger signalling was achieved by AcrIIA6 binding, as evidenced by the shorter path length and smaller path degeneracy, indicating that it is a more potent allosteric effector than the PAM duplex (Supplementary Fig. 17 and Supplementary Table 8).

Common signalling routes were adopted by inhibitory signals from AcrIIA6 when propagating within St1Cas9 RNP to the catalytic sites and PIRs; however, the presence of dsDNA hindered such efficient signalling (Fig. 7J, Supplementary Fig. 18, and Supplementary Table S8). In St1Cas9·sgRNA·dsDNA·AcrIIA6, inhibitory signals from AcrIIA6 were transmitted via an elongated path consisting of more residues before reaching the AcrIIA6-St1Cas9 RNP interface. Such compromised inhibitory signalling may underlie the preservation of St1Cas9·sgRNA·dsDNA catalytic activity when bound by AcrIIA6. In general, PI residues K1010 and F1019 served as direct recipients of the AcrIIA6-derived inhibitory signal, whereas L601 (HNH), N683 (L2), and S961 (RuvC) maintained the mediation of allosteric information transfer within St1Cas9 RNP, suggesting that mutagenesis of these residues is a potential modulatory strategy for fine-tuning the activity and selectivity of St1Cas9.

2.8. Bacteria Cas9 employ residue variations to protect against AcrIIA6 assault

The presence of St1Cas9 strains carrying point mutations (G993K/K1008M/K1010E) that's capable of evading the phage invading system depicted the ongoing coevolution between species. Based on the previous findings [2], we further conducted multiple sequence alignment of PI domains of all 53 *S. thermophilus* Cas9 homologs from different *S. thermophilus* strains (Supplementary Fig. S19), and found that nine *S. thermophilus* Cas9 proteins (from strain MTCC 5460, S9, EPS, etc.) have the same residue substitutions at the implicated sites (G993K/K1008M/K1010E), indicating that this defence mechanism may transfer rapidly among the *S. thermophilus* genomes during the evolution of its Cas9 endonuclease to bypass AcrIIA6 inhibition.

Using large-scale MD simulations, we further clarified the molecular basis underlying such triple mutational escape (G993K/K1008M/K1010E, referred to as *mut* St1Cas9). Based on the dynamics exhibited by *mut* St1Cas9·sgRNA throughout 15 μs sampling (Supplementary Figs. 3, 5, and 6) and its similarity with St1Cas9 RNP in terms of conformational distribution in the principal subspace (Fig. 2A and Supplementary Fig. 7), the triple mutations were unable to significantly disturb the overall conformation; this was predictable because residue variations are naturally present in *S. thermophilus* with normal endonuclease activity. In addition, the original approaching (along PC1) and open (along PC2) essential dynamics were well preserved along each order (Supplementary Fig. 20), with relaxed PAM recognition grooves encompassed by WED and PI domains as well as HNH nuclease domain docking near the cleavage site. Nonetheless, triple mutations weakened the innate correlation within *mut* St1Cas9 (Fig. 6E and Supplementary Fig. 16E, G) and altered its community network by enhancing direct informational flow from mutation-harbouring community 10 to community 16 (representing the RuvC nuclease domain), ultimately resulting in weakened signalling between HNH and RuvC catalytic sites and regions relevant to PAM recognition in WED and PI. The original WED and PI architecture associated with PAM recognition was largely retained in the mutated systems, as evidenced by the similar $\Delta_{K837-K1049-N1079}$ perimeter in the major conformational clusters from *mut* St1Cas9 RNP under both AcrIIA6 unbound/bound states (Fig. 3E, F; C_{*} for *mut* St1Cas9·sgRNA, C^{L}_{*A} and C^{R}_{*A} for *mut* St1Cas9·sgRNA·AcrIIA6) as that in the unmutated systems (Fig. 3A, B and Supplementary Table 1). However, introduction of the triple residue substitutions reduced the AcrIIA6^{A} β2-β3 binding pocket volume by inducing greater counter-movement between WED and PI domains resembling the closure of a nutcracker, which resulted in disrupted *mut* St1Cas9 RNP shape complementarity as well as decreased interface area for AcrIIA6 recognition (Supplementary Fig. 21 and Supplementary Table 6). In contrast, the MM/PBSA binding free energy demonstrated that the affinity between St1Cas9 and AcrIIA6 was significantly compromised in the presence of triple mutations, further destabilising the AcrIIA6 dimer (Supplementary Table 5). Indeed, the original interaction network between St1Cas9 RNP and AcrIIA6 was markedly weakened upon triple residue substitutions, with complete loss of the strong polar interaction between K1010 and AcrIIA6^{A} A124 (Fig. 4E, F, Supplementary Fig. 12C, 13C, and Supplementary Tables 3, 4). K1010 was also directly implicated in AcrIIA6 allosteric signalling toward St1Cas9 as directly recipient of signals from AcrIIA6^{A} A124 (Fig. 7 and Supplementary Fig. 18); its charge-reversal substitution from lysine to glutamic acid resulted in a switch to a more roundabout and less effective lane (through TOPO residue D987 or K984) for inhibitory signals to be transduced from AcrIIA6 to the *mut* St1Cas9.

2.9. Determination of key mutations in St1Cas9 that affect AcrIIA6 inhibition

To explore the molecular determinants of coevolution of phage-host interactions, we monitored the role of naturally occurring mutations of St1Cas9 in conferring AcrIIA6-mediated inhibition. For this, we generated three single mutants, G993K, K1008M and K1010E, three double mutants, G993K/K1008M, G993K/1010E and K1008M/1010E, and one triple mutant, G993K/K1008M/K1010E, and subsequently stably expressed them in *E. coli* cells. To exclude the effect of mutations on the thermodynamic stability of St1Cas9, thermal shift assays were performed by the quantitative reverse transcription PCR (RT-qPCR). The results showed that the melting temperature (T_{m}) of all variants was comparable to that of wild-type St1Cas9 (Fig. 8A and Supplementary Fig. 22), suggesting that the mutations had a minor effect on the thermodynamic stability of St1Cas9.

An *in vitro* DNA cleavage assay was further performed to assess the inhibitory effect of AcrIIA6 on enzymatic activity of wild-type St1Cas9 RNP and variants. We concomitantly incubated St1Cas9 RNP with 150 nM AcrIIA6 and 5 nM 1664-bp long dsDNA substrates, harboring target DNA and PAM sequences. The dsDNA cleavage products were resolved using electrophoresis at different time points (Fig. 8B). When we normalized the *in vitro* DNA cleavage activities of the St1Cas9 variants to that of wild-type enzyme, we found that all the variants decreased the cleavage of target DNA

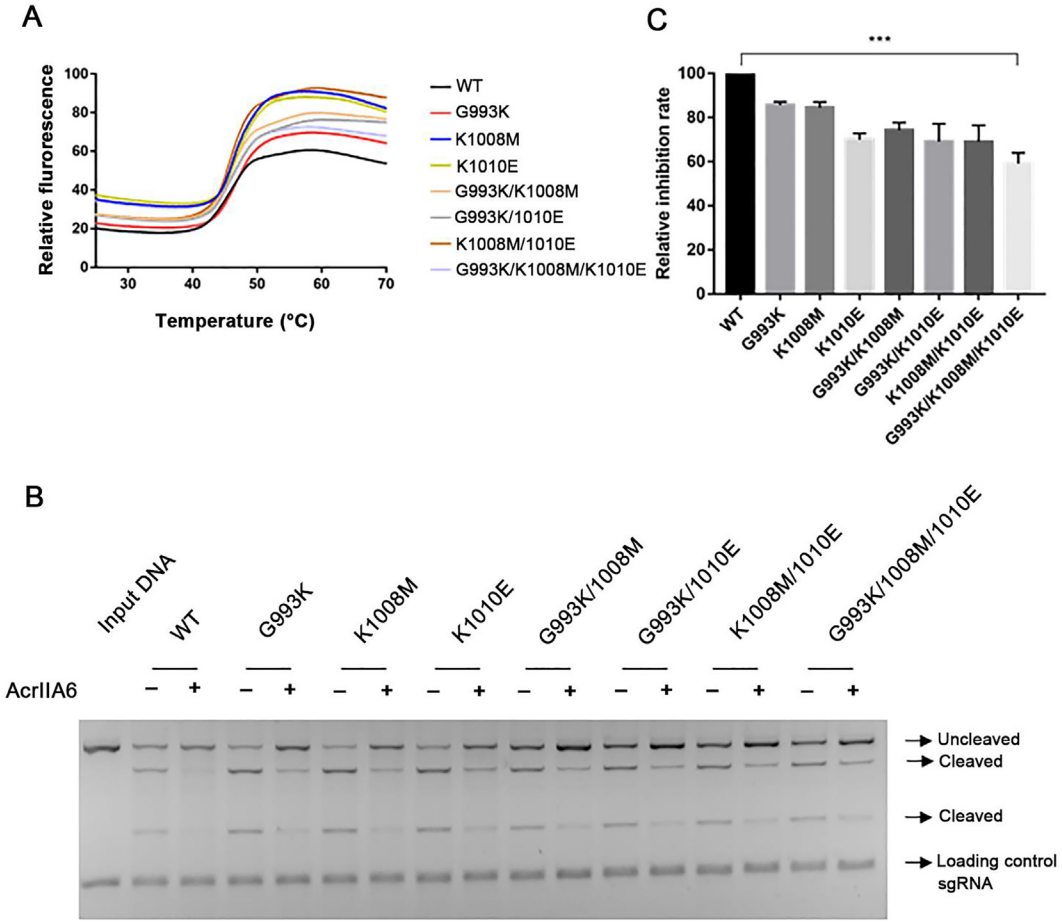

Fig. 8. St1Cas9 variants escape AcrIIA6 inhibition in cells. (A) Protein thermal shift assays of different St1Cas9 variants. The experiment was performed three times and each time had triplicates. (B) *In vitro* DNA cleavage assay of the wild-type and mutants of St1Cas9, using dsDNA fragments containing a target sequence and a PAM motif. Mobilities of input DNA (uncleaved) and cleavage products (cleaved) are indicated with arrows. The experiments were performed at least four times and yielded equivalent results. (C) Quantification of relative inhibition rate in (B), showing St1Cas9 variants having different sensitivities towards AcrIIA6. The experiments were performed in biological quadruplicates. Data are presented as the mean ± s.d. from four independent experiments. ***$p < 0.001$ by *t*-test.

in the presence of AcrIIA6 with varying extents. As shown in Fig. 8C and Supplementary Table S9, a moderate (∼15%) reduction in AcrIIA6-dependent inhibition of St1Cas9 RNP nuclease activity was observed in both G993K and K1008M single mutants, while the K1010E mutant exhibited an appreciable (∼30%) reduction in its inhibition rate in the presence of AcrIIA6. This result indicated that the K1010E mutation affects a key residue, consistent with the MD results that K1010E mutation resulted in loss of strong AcrIIA6–St1Cas9 interaction and hindered effective allosteric signal transduction. The two double mutants (G993K/1010E and K1008M/1010E), to a lesser extent, the G993K/K1008M double mutant, showed a similar reduction of nuclease activities upon treatment with AcrIIA6 to the K1010E single mutant. Strikingly, the triple mutant (G993K/K1008M/K1010E) exhibited a marked (∼40%) reduction in its nuclease activity upon treatment with AcrIIA6. This implied that the St1Cas9 triple mutant can become an efficient molecular evolutionary pattern of bacteria to bypass phages AcrIIA6 inhibition.

3. Discussion

The exploitation of CRISPR-Cas systems in various molecular, biotechnological, and medical applications is growing; however, fine-tuning the efficacy and safety of this toolbox is of vital importance in addressing potential problems such as off-target effects [20,84–86]. Previous attempts to prevent such unwanted activity include mutating the Cas endonuclease [87,88], fusing with regulatory domains [89,90], or employing pre-formed RNP with controllable fidelity and nuclease activity [91,92]. However, such strategies have been faced with fatal flaws in delivery modality and ortholog universality [45]. The natural *bona fide* CRISPR-Cas antagonist Acrs—evolved during the evolutionary arms race between bacteria and bacteriophages—represent a genetically encodable, post-translational regulation of CRISPR-derived technologies to keep this powerful machinery in check [45]. Indeed, a number of studies integrating sequence analysis, machine learning, etc. have attempted to exploit this natural CRISPR-Cas regulator repertoire [18,56–58]; moreover, computational-guided engineering of existing Acr-based structural information has been recognised as a promising approach for expanding Acr inhibition [93]. Currently, the broad spectrum of Cas endonucleases regulated by Acrs compensates for the current scarcity of regulation strategies for Cas endonucleases with different sizes, fidelity, and PAM promiscuity from that of the established CRISPR-SpCas9-based toolkit [4,7,45,46,94], which would enable controllable and customised CRISPR-based techniques for different purposes (i.e. manipulation of differently sized genes).

We explored the detailed atomic-level AcrIIA6 inhibitory mechanism through extensive MD simulations, which demonstrated the AcrIIA6-induced conformational transition of St1Cas9 from the approaching conformation—in which the nuclease HNH domain is positioned near the cleavage site while the WED and PI domains

adopt a stretched posture ready for PAM interrogation—to a departing ensemble—in which outward movement of the catalytic HNH core was accompanied by shrinkage of the regions responsible for PAM recognition. Furthermore, through the intensive interaction network established based on shape complementarity between AcrIIA6 and St1Cas9 RNP, inhibitory signals were propagated to the St1Cas9 endonuclease while the correlated motion between functional domains mediated further transduction of inhibitory signals; these signals allosterically resulted in increased conformational plasticity of the catalytic HNH domain residing far from the direct AcrIIA6 recognition interface, allowing for its significant relocation in the presence of the inhibitor. Moreover, by computing the residue-level communication pathway coordinating the functional synchroneity between HNH, RuvC catalytic sites, and PAM interaction regions as well as how inhibitory signals are involved, we identified the key role of K1010 in the PI domain as a direct recipient of allosteric signals from AcrIIA6, further propagating them towards the St1Cas9 RNP. Variations in K1010, along with residue substitutions in K993 and K1008, have been shown to bypass the anti-CRISPR strategy by reshaping the intermolecular recognition interface between AcrIIA6 and *mut* St1Cas9, rewiring the original signalling pathway. The widespread distribution of residue polymorphisms of the implicated residues, especially K1010, suggests that such variations represent a strategy of AcrIIA6 inhibition escape, reflecting the coevolution of bacteria and phages.

Notably, we observed the inhibition failure of dsDNA-loaded St1Cas9 RNP by AcrIIA6, in which the inhibitor induced a spatial shift of the substrate DNA duplex, resulting in weakened PAM-St1Cas9 interactions but retaining the nuclease HNH configuration. This observation indicates that although AcrIIA6 inhibits St1Cas9 RNP functionality by interfering with both DNA binding and nuclease activation, its more primary mode of action may be modifying the conformational dynamics of WED and PI domains associated with PAM recognition; the altered dynamics of WED and PI further resulted in HNH outward relocation, which would otherwise be located closer to the presumed cleavage site. While successfully establishing the mechanistic framework underlying AcrIIA6 allosteric inhibition, we also observed the conformational transition tendency of St1Cas9·sgRNA·dsDNA towards a catalytically active state along the 3 μs simulation; however, due to the limited simulation time scale, the fully activated St1Cas9·sgRNA·dsDNA structure for catalysis was not completely captured, posing a challenge for future studies to obtain continuous large-scale MD simulations of large macromolecular systems, such as CRISPR-Cas9.

Taken together, our findings help elucidate the driving force behind the emergence of AcrIIA6 during the coevolution of bacteria and phages as well as the detailed mechanism of allosteric inhibition of St1Cas9 RNP by the anti-CRISPR protein AcrIIA6 [95–98]. The results provide a dynamic picture of this mechanism. The identified AcrIIA6-meidated allosteric inhibitory mode can provide further guidance for designing regulatory molecules/peptidomimetics to fine-tune the bioactivity of the CRISPR-Cas machinery when performing *in vivo* gene manipulation, improving the safety of gene therapy. Meanwhile, the distinct allosteric mechanism displayed by AcrIIA6 further emphasises the amazing diversity of inhibitors that phages have evolved to counteract the existing challenges posed by CRISPR-Cas systems. Moreover, we provided insights into anti-CRISPR strategies exemplified by Acrs to the Acr-driven mutational escape of prokaryotes, uncovering the basis of mutational escape from AcrIIA6 inhibition through a few residue variations. The robustness of different *S. thermophilus* CRISPR-Cas9 systems with residue variations will be further explored to identify economically suitable strains that can resist the selective pressure imposed by rapidly reproducing phages. By providing an atomic-level understanding of AcrIIA6 allosteric inhibition of St1Cas9 and St1Cas9 anti-anti-CRISPR rescue mechanism, we can offer a guidance for devising 'off-switches' for the CRISPR-Cas9 machinery with improved regulatability and efficacy, enabling temporal, spatial, and conditional genome manipulation for both basic and clinical applications.

4. Materials and methods

4.1. Setup of simulation systems

All simulation systems in our study were constructed based on the solved cryo-EM crystal structure of St1Cas9· sgRNA·tDNA20·AcrIIA6 (PDB ID: 6RJA), St1Cas9·sgRNA·tDNA59-ntPAM (PDB ID: 6RJD), and St1Cas9·sgRNA·AcrIIA6·tDNA59-ntPAM (PDB ID: 6RJG) complexes at 3.0, 3.3, and 3.2 Å resolution [2]. Structural superposition of St1Cas9 from the three complexes was first performed to identify the missing residues in common. For the unsolved large segments around the HNH domain (MET510–TYR689) and RuvC-III domain (Leu751–Lys802), structural information was modelled based on sequence homology with *Staphylococcus aureus* Cas9 (SaCas9, PDB ID: 5AXW [99]) using SWISS-MODEL, and a structural model of reliable quality was obtained with QMEAN of −1.29 for HNH and −2.15 for RuvC domain, respectively. In addition, other sporadic missing residues in the original crystal profile were remodelled using Discovery Studio. After completion of the above-mentioned operations, the obtained systems were first subjected to 10000-step minimisation using the steepest descent algorithm typed with CHARMM forcefield.

The *apo* St1Cas9, St1Cas9·sgRNA, and St1Cas9·sgRNA·AcrIIA6 systems were built based on the remodelled St1Cas9·sgRNA·tDNA20·AcrIIA6 (PDB ID: 6RJA, without PAM sequence) by deleting constituent sgRNA·tDNA20·AcrIIA6, tDNA20·AcrIIA6, and tDNA20, respectively; meanwhile, the remodelled St1Cas9·sgRNA·tDNA59·ntPAM (PDB ID: 6RJD, with PAM sequence) and St1Cas9·sgRNA·AcrIIA6·tDNA59·ntPAM (PDB ID: 6RJG, with PAM sequence) well represented the conformation of dsDNA-loaded Cas9 RNP under AcrIIA6 unbound-(St1Cas9·sgRNA·dsDNA) and bound-(St1Cas9·sgRNA·dsDNA·AcrIIA6) states. Triple mutations along the AcrIIA6 binding interface (G993K, K1008M, K1010E) were introduced using Discovery Studio to construct *mut* St1Cas9·sgRNA and *mut* St1Cas9·sgRNA·AcrIIA6 systems. To further investigate the species-specific inhibition of St1Cas9 by AcrIIA6, an additional *Streptococcus pyogenes* Cas9 RNP model was constructed (PDB ID: 6O0Z [39]), whose missing residues were remodelled based on coordinate information of spCas9 under other activation states (PDB ID: 4OO8); the obtained complex was also subjected to 10000-step steepest descent minimisation (Supplementary Fig. 2).

4.2. MD simulation details

The above-mentioned systems were prepared with the AMBER18 program using the *ff14SB* force field [100] to describe the ribonucleoprotein complex. All systems were first solvated in an orthorhombic transferable intermolecular potential three point (TIP3P) water box [101], followed by the addition of Na^+ and Cl^- counterions to neutralise the system while mimicking the *in vivo* physiological cleavage condition (Supplementary Table S10). The generated topology and coordinate files served as inputs for subsequent MD simulations. Two rounds of energy minimisation were first carried out as previously reported [102–106], first with the whole protein scaffold fixed, followed by removing all constraints for 5000- and 10000-step maximum minimisation cycles, respectively. Subsequently, all systems were equilibrated in a canonical

ensemble (NVT) for 700 ps after heating from 0 K to 300 K within 300 ps. Finally, five independent 3-μs MD simulation replicas with random initial velocities were performed for all systems embedded in an isothermal and isobaric ensemble (NPT) with periodic boundaries, generating 40 independent trajectories and accumulating ~ 120-μs conformational sampling in total. Langevin dynamics using 1 ps^{-1} collision frequency was used to control the temperature during the simulation. Long-range electrostatic interactions were analysed using the particle mesh Ewald (PME) method, and a 10 Å non-bonded cut-off was introduced for short-range electrostatics and van der Waals interactions. Covalent bond interactions involving hydrogen atoms were constrained by the SHAKE algorithm. Subsequent analysis of simulation trajectories was based on all five replicas (consisting of ~15 μs of sampling) for each simulation system to strengthen our statistical analysis. Based on the time evolution trend of the root mean square deviation of St1Cas9 (Supplementary Fig. S3), all productive runs reached relative equilibrium after the first 500-ns MD simulations.

4.3. Generalised correlation analysis

Assessment of the correlated motions within our model systems under investigation was achieved with the generalised correlation (GC_{ij}) analysis approach proposed by Oliver and Helmut [83], which enabled further insights into the allosteric signal transduction process. The GC_{ij} approach surpasses the established Pearson correlation analysis in that it enables the description of nonlinear correlations while being independent of atomic fluctuation orientations. Rest on the fundamental definition of independence of random variables, in GC_{ij} analysis the variables x_i, x_j are considered correlated when the product of their marginal distribution $p(x_i) \cdot p(x_j)$ is larger than their joint distribution $p(x_i, x_j)$. To measure the degree of correlation between selected variables, mutual information (MI) between x_i and x_j was introduced and defined as:

$$MI[x_i, x_j] = \iint p(x_i, x_j) ln \frac{p(x_i, x_j)}{p(x_i) \cdot p(x_j)} dx_i dx_j \# \quad (1)$$

where the right side of equation [1] defines MI as closely related to the well-known Shannon entropy $H[x]$, which is calculated as:

$$H[x] = \int p(x) lnp(x) dx \# \quad (2)$$

thus, the correlation between pairs of atoms x_i and x_j is described by MI and calculated using the marginal Shannon entropy $H[x_i]$, $H[x_j]$, and the joint entropy term $H[x_i, x_j]$ as follows:

$$MI[x_i, x_j] = H[x_i] + H[x_j] - H[x_i, x_j] \# \quad (3)$$

The g_correlation tool in Gromacs 3.3 [107] calculates the entropy terms $H[x_i]$, $H[x_j]$, and $H[x_i, x_j]$ with the k-nearest neighbour distance algorithm using atomic fluctuation information. The $MI[x_i, x_j]$ values were further normalised to obtain the normalised generalised correlation coefficients (GC_{ij}):

$$GC_{ij} = \left\{ 1 - e^{-\frac{2MI[x_i, x_j]}{d}} \right\}^{\frac{1}{2}} \# \quad (4)$$

where d represents the dimensionality of x_i and x_j, which equals 3 in our analysis.

Inspired by Palermo et al. [27,76] to further represent the extent to which a generalised correlation is represented by each residue with the remaining part, we introduced the residue correlation score (CS) in our analysis. For residue i, its corresponding CS_i is defined as:

$$CS_i = \sum_{i \neq j}^{N} GC_{ij} \# \quad (5)$$

where N represents the total number of residues within the protein (in our case N is 1121, total number of residues in St1Cas9). Depending on the domain to which residue j belongs, CS_i^{intra} (residue i and j belong to the same domain) and CS_i^{inter} (residue j belongs to domains excluding that of residue i) were calculated to describe intra- and inter-domain correlations mediated by each residue. Moreover, to gain insights into the more macroscopic inter-domain correlation, CS_i^{inter} over residues j from a specific domain were accumulated, and the obtained matrix was plotted to denote inter-domain correlations. Notably, all GC_{ij} analyses were based on the filtered GC_{ij} matrix with a minimum threshold value of 0.65.

4.4. Dynamic network analysis

To describe and investigate residue-residue interactions within biomolecular systems, we employed concepts from network theories [108–111]. The whole ribonucleotide protein complex was defined as a set of nodes, which were assigned to the Cα atom within each amino acid residue or P atom in the nucleotide backbone for each nucleotide. Edges were drawn between nodes that stayed within a cut-off distance of 4.5 Å for at least 75% of the conformational sampling consisting of all equilibrated simulations (the last 2.5 μs in 5 replicas, 12.5 μs in total) for each system. The edges between node i and node j were further weighed by the corresponding GC_{ij} coefficients as follows:

$$d_{ij} = -logGC_{ij} \# \quad (6)$$

to indicate the relative distance between nodes i j during information transfer in the connection network. The obtained weighed dynamical network was used as a basis for further 'community' [112] and 'optimal path' analysis [113–116]. To obtain the community substructure, the Girvan-Newman divisive algorithm was applied, in which 'edge betweenness' that defines the number of shortest pathways across a given edge serves as an important partition criterion. By iteratively removing edges with the highest betweenness from the network and recalculating the betweenness of remaining edges until each node represents an isolated community, the optimal substructure of the network—in which connections are only dense within a community but sparse in between—can be obtained. A 'modularity' parameter Q indicating the quality of a given community partition was defined as:

$$Q = \sum_m (e_{mn} - (\sum_n e_{mn})^2) \# \quad (7)$$

where e_{mn} indicates the fraction of edges that connect nodes in community m to nodes in community n, while $\sum_n e_{mn}$ represents the fraction of edges that links to nodes in community n. Falling within the range of 0 to 1, a larger modularity value indicates a higher quality of the obtained community structure. The package *gncommunities* was employed to gain insights into the community structure of the St1Cas9·sgRNA (Supplementary Table 7-1), St1Cas9·sgRNA·dsDNA (Supplementary Table 7-2), St1Cas9·sgRNA·AcrIIA6 (Supplementary Table 7-3), St1Cas9·sgRNA·dsDNA·AcrIIA6 (Supplementary Table 7-4), *mut* St1Cas9·sgRNA (Supplementary Table 7-5), and *mut* St1Cas9·sgRNA·AcrIIA6 (Supplementary Table 7-6) systems, with their corresponding modularity values reported.

To track the detailed information transfer pathway between pairs of residues (defined as source and sink), we used the Floyd-Warshall algorithm to identify the 'shortest pathway' by comparing the sum of d_{ij} (path length, PL) of all edges involving in the path. The source and sink were laid on catalytic D9 in RuvC, H599 in HNH, PAM-interacting residues (PIR, including WED: K867, S961, V962, S963; TOPO: K1025; CTD: N1026), and AcrIIA6 residues ($T118^A$, $A124^A$, $S147^A$, $N83^B$) to disclose their residue-by-residue inter-signalling path. The number of suboptimal pathways

with lengths that are not larger than the PL of the shortest pathway by 2% is defined as pathway degeneracy (PD) to indicate specificity of the optimal communication path.

4.5. Plasmids preparation for protein expression in E. coli

The full-length St1Cas9 gene from *Streptococcus thermophilus (strain ATCC BAA-491/LMD-9)* was synthesized by Tsingke Biological Technology (Beijing, China) and cloned into pET28a vector using the NdeI and BamH I restriction sites adding an N-terminal 6x histidine tag to ease protein purification. The full-length AcrIIA6 of phage D1811 was also synthesized and cloned into pET28a vector with N-terminal 6x histidine.

4.6. Site-directed mutagenesis

Point mutations were introduced into the plasmid encoding the wild-type St1Cas9 by Mut Express II Fast Mutagenesis Kit V2 (Vazyme Biotech co.,ltd, Nanjing, Jiangsu, China). The successful introduction of the mutations was verified by DNA sequencing (Personalbio, Shanghai, China). The detailed nucleotide sequences for constructing all the macromolecule models (including both wildtype and mutated complexes) were supplemented in Supporting Information Table S11.

4.7. Protein expression and purification

Plasmid was transformed into *E. coli* BL21-Star(DE3) pLysS cells (Weidi, Shanghai, China) for protein purification. The cells were induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 18 °C. The St1Cas9 protein was firstly purified with nickel column (GE healthcare, Little Chalfont, Buckinghamshire, UK) in the lysis buffer (50 mM HEPES pH 7.4, 500 mM NaCl, and 10 mM imidazole), 0.1 mg/mL Dnase I, 0.1 mg/mL lysozyme and 1x protein inhibitor cocktail were added before French Press. Next, the St1Cas9-containing fractions were buffer exchanged in low salt buffer (20 mM HEPES pH 7, 100 mM NaCl) and applied onto a 5 mL HiTrap SP HP column (GE Healthcare, Little Chalfont, Buckinghamshire, UK) for cation exchange chromatography [2]. St1Cas9 containing fractions were buffer exchanged into storage buffer (20 mM Hepes, pH 7.5, 150 mM NaCl), flash-frozen in liquid nitrogen, and stored at −80 °C. Purification of mutation forms of St1cas9 protein was carried out analogously. For AcrIIA6 protein, gel filtration chromatography was used instead of cation exchange chromatography and stored in the same storage buffer [55].

4.8. Preparation of sgRNA and dsDNA

The transcription templates of sgRNAs were synthesized by GENEWIZ (Suzhou, Jiangsu, China) and amplified by a PCR. sgRNAs were prepared by *in vitro* transcription using the T7 High Yield RNA Transcription Kit (Vazyme Biotech co., ltd, Nanjing, Jiangsu, China) and further purified using phenol/chloroform extraction followed by ethanol precipitation [117]. dsDNA used as cleavage substrate was synthesized by GENEWIZ (Suzhou, Jiangsu, China) and amplified by a PCR.

4.9. In vitro DNA cleavage assay

1664-bp long dsDNA substrates containing the target (underlined) and PAM (bold) sequences were generated by PCR (target and PAM sequences inserted into the pUC19 plasmid: **TTTCTGC**AATACTTTTATCAACGCAAG) as previously described [2]. *In vitro* cleavage reactions were performed in 20 μL reaction buffer (20 mM HEPES pH 7.5, 100 mM KCl, 1 mM DTT, 10 mM $MgCl_2$) containing 5 nM linearized dsDNA substrates, 50 ng Loading control, 150 nM St1Cas9 and 200 nM sgRNA, with or without 150 nM AcrIIA6. The reaction was incubated in 37 °C for 15 min and was quenched by adding 50 mM EDTA, 20 μg Proteinase K and 1× Loading buffer for 30 min in room temperature [31]. Products were analysed by electrophoresis in a 2% agarose, 0.5× TBE gel stained with 4S red plus dye (Sangon, Shanghai, China). Gels were imaged using a Tanon-3500 gel imaging system (Tanon, Shanghai, China) and quantification were performed using Image J. loading control dsDNA wasn't cleaved by St1Cas9, as a correction for quantification. The experiment was independently repeated four times. Data are presented as the mean ± s.d. of three independent experiments.

4.10. Protein thermal shift assays

A total of 2 μM St1Cas9 wild-type or mutant proteins was mixed with 5x SYPRO orange dye (Sigma-Aldrich, St. Louis, MO, USA) in 1x PBS. Samples were analysed in LightCycler 480 Real-time PCR instrument System II (Roche, Basel, Switzerland). The temperature was increased at a rate of 0.05 °C/s over a range of 25–95 °C, and the fluorescence was monitored with the SYPRO orange channel. T_m values were obtained from the minimum of the first derivative $-(dF/dT)$ plots of the unfolding protein curves through analysis using LightCycler 480 software provided by the manufacture (Roche, Basel, Switzerland). GraphPad Prism Version 7.00 (GraphPad Software, La Jolla, CA, USA) was used to normalize the dF/dT curves. At least three trials were independently performed for each experiment.

5. Ethics approval and consent to participate

Not applicable.

6. Consent for publication

All the authors have approved and agreed to publish this manuscript.

7. Availability of data and material

All data generated and analysed to support the conclusion of this study is included in this published manuscript and its additional supporting files.

8. Authors' contributions

S.L. conceived and supervised the project, designed the experiments, and revised the manuscript; X.L. and T.C. performed MD and analysis; C.W. and Z.C. performed biochemical assays; T.P. performed bioinformatic analysis; D.N., Y.L., and J.Z. acquired the data and revised the manuscript; S.L. was responsible for study conception and project oversight. All authors discussed the results and reviewed the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (22077082).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.11.010.

References

[1] Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 2007;315:1709–12.
[2] Fuchsbauer O, Swuec P, Zimberger C, Amigues B, Levesque S, Agudelo D, et al. Cas9 Allosteric Inhibition by the Anti-CRISPR Protein AcrIIA6. Mol Cell 2019;76:922–937.. e7.
[3] Garneau JE, Dupuis M-È, Villion M, Romero DA, Barrangou R, Boyaval P, et al. The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 2010;468:67–71.
[4] Maxwell KL. The anti-CRISPR Story: A battle for survival. Mol Cell 2017;68:8–14.
[5] Luo G, Gao Q, Zhang S, Yan B. Probing infections disease by single-cell RNA sequencing: Progresses and perspectives. Comput Struct Biotechnol J 2020;18:2962–71.
[6] Kronheim S, Daniel-Ivad M, Duan Z, Hwang S, Wong AI, Mantel I, et al. A chemical defence against phage infection. Nature 2018;564:283–6.
[7] Maxwell KL. Bacterial twist to an antiviral defence. Nature 2019;574:638–9.
[8] Swarts DC, Jore MM, Westra ER, Zhu Y, Janssen JH, Snijders AP, et al. DNA-guided DNA interference by a prokaryotic Argonaute. Nature 2014;507:258–61.
[9] Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic discovery of antiphage defense systems in the microbial pangenome. Science 2018;359:eaar4120.
[10] Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, et al. BREX is a novel phage resistance system widespread in microbial genomes. EMBO J 2015;34:169–83.
[11] Li Q, Qin Z, Wang Q, Xu T, Yang Y, He Z. Applications of genome editing technology in animal disease modeling and gene therapy. Comput Struct Biotechnol J 2019;17:689–98.
[12] Nishimasu H, Ran FA, Hsu P, Konermann S, Shehata S, Dohmae N, et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 2014;156:935–49.
[13] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012;337:816–21.
[14] Pausch P, Al-Shayeb B, Bisom-Rapp E, Tsuchida CA, Li Z, Cress BF, et al. Crispr-casf from huge phages is a hypercompact genome editor. Science 2020;369:333–7.
[15] Adli M. The CRISPR tool kit for genome editing and beyond. Nat Commun 2018;9:1911.
[16] Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. Science 2014;346:1258096.
[17] Jiang F, Doudna JA. CRISPR-Cas9 structures and mechanisms. Annu Rev Biophys 2017;46:505–29.
[18] Gasiunas G, Young JK, Karvelis T, Kazlauskas D, Urbaitis T, Jasnauskaite M, et al. A catalogue of biochemically diverse CRISPR-Cas9 orthologs. Nat Commun 2020;11:5512.
[19] Wang H, La Russa M, Qi LS. CRISPR/Cas9 in genome editing and beyond. Annu Rev Biochem 2016;85:227–64.
[20] Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nat Biotechnol 2020;38:824–44.
[21] Xu Y, Li Z. CRISPR-Cas systems: overview, innovations and applications in human disease research and gene therapy. Comput Struct Biotechnol J 2020;18:2401–15.
[22] Chen LQ, Zhao WS, Luo GZ. Mapping and editing of nucleic acid modifications. Comput Struct Biotechnol J 2020;18:661–7.
[23] Liu G, Zhang Y, Zhang T. Computational approaches for effective CRISPR guide RNA design and evaluation. Comput Struct Biotechnol J 2020;18:35–44.
[24] Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science 2014;343:1247997.
[25] Huai C, Li G, Yao R, Zhang Y, Cao M, Kong L, et al. Structural insights into DNA cleavage activation of CRISPR-Cas9 system. Nat Commun 2017;8:1375.
[26] Shibata T, Iwasaki W, Hirota K. The intrinsic ability of double-stranded DNA to carry out D-loop and R-loop formation. Comput Struct Biotechnol J 2020;18:3350–60.
[27] Palermo G, Ricci CG, Fernando A, Basak R, Jinek M, Rivalta I, et al. Protospacer adjacent motif-induced allostery activates CRISPR-Cas9. J Am Chem Soc 2017;139:16028–31.
[28] Hirano S, Abudayyeh OO, Gootenberg JS, Horii T, Ishitani R, Hatada I, et al. Structural basis for the promiscuous PAM recognition by Corynebacterium diphtheriae Cas9. Nat Commun 2019;10:1968.
[29] Menon AV, Sohn J, Nam J-W. CDG: Comprehensive guide designer for CRISPR-Cas system. Comput Struct Biotechnol J 2020;18:814–20.
[30] Sternberg SH, Redding Sy, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 2014;507:62–7.
[31] Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 2014;513:569–73.
[32] Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, et al. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol 2008;190:1390–400.
[33] Karvelis T, Gasiunas G, Young J, Bigelyte G, Silanskas A, Cigan M, et al. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome Biol 2015;16:253.
[34] Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res 2011;39:9275–9282.
[35] Horvath P, Romero DA, Coûté-Monvoisin AC, Richards M, Deveau H, Moineau S, et al. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J Bacteriol 2008;190:1401–12.
[36] Ricci CG, Chen JS, Miao Y, Jinek M, Doudna JA, McCammon JA, et al. Deciphering off-target effects in CRISPR-Cas9 through accelerated molecular dynamics. ACS Cent Sci 2019;5:651–62.
[37] Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. A Cas9-guide RNA complex preorganized for target DNA recognition. Science 2015;348:1477–81.
[38] Palermo G, Chen JS, Ricci CG, Rivalta I, Jinek M, Batista VS, et al. Key role of the REC lobe during CRISPR-Cas9 activation by "sensing", "regulating", and "locking" the catalytic HNH domain. Q Rev Biophys 2018;51:e91.
[39] Zhu X, Clarke R, Puppala AK, Chittori S, Merk A, Merrill BJ, et al. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. Nat Struct Mol Biol 2019;26:679–85.
[40] Sternberg SH, LaFrance B, Kaplan M, Doudna JA. Conformational control of DNA target cleavage by CRISPR-Cas9. Nature 2015;527:110–3.
[41] Palermo G, Ricci CG, McCammon JA. The invisible dance of CRISPR-Cas9. Phys Today 2019;72:30–6.
[42] Palermo G, Miao Y, Walker RC, Jinek M, McCammon JA. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. Proc Natl Acad Sci U S A 2017;114:7260–5.
[43] Jiang F, Liu J-J, Osuna BA, Xu M, Berry JD, Rauch BJ, et al. Temperature-Responsive Competitive Inhibition of CRISPR-Cas9. Mol Cell 2019;73:601–10.
[44] Rauch BJ, Silvis MR, Hultquist JF, Waters CS, McGregor MJ, Krogan NJ, et al. Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. Cell 2017;168:150–8.
[45] Marino ND, Pinilla-Redondo R, Csörgő B, Bondy-Denomy J. Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. Nat Methods 2020;17:471–9.
[46] Pawluk A, Davidson AR, Maxwell KL. Anti-CRISPR: Discovery, mechanism and function. Nat Rev Microbiol 2018;16:12–7.
[47] Harrington LB, Doxzen KW, Ma E, Liu J-J, Knott GJ, Edraki A, et al. A broad-spectrum inhibitor of CRISPR-Cas9. Cell 2017;170:1224–33.
[48] Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature 2013;493:429–32.
[49] Cox DBT, Platt RJ, Zhang F. Therapeutic genome editing: Prospects and challenges. Nat Med 2015;21:121–31.
[50] Dai W-J, Zhu L-Y, Yan Z-Y, Xu Y, Wang Q-L, Lu X-J. CRISPR-Cas9 for in vivo gene therapy: promise and hurdles. Mol Ther - Nucleic Acids 2016;5:e349.
[51] Hoffmann MD, Aschenbrenner S, Grosse S, Rapti K, Domenger C, Fakhiri J, et al. Cell-specific CRISPR-Cas9 activation by microRNA-dependent expression of anti-CRISPR proteins. Nucleic Acids Res 2019;47:e75.
[52] Trasanidou D, Gerós AS, Mohanraju P, Nieuwenweg AC, Nobrega FL, Staals RHJ. Keeping crispr in check: Diverse mechanisms of phage-encoded anti-crisprs. FEMS Microbiol Lett 2019;366:fnz098.
[53] Bondy-Denomy J, Garcia B, Strum S, Du M, Rollins MF, Hidalgo-Reyes Y, et al. Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. Nature 2015;526:136–9.
[54] Hwang S, Maxwell KL. Meet the anti-CRISPRs: widespread protein inhibitors of CRISPR-Cas systems. Cris J 2019;2:23–30.
[55] Hynes AP, Rousseau GM, Agudelo D, Goulet A, Amigues B, Loehr J, et al. Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range of Cas9 proteins. Nat Commun 2018;9:2919.
[56] Zhang G, Dai Z, Dai X. C-RNNCrispr: Prediction of CRISPR/Cas9 sgRNA activity using convolutional and recurrent neural networks. Comput Struct Biotechnol J 2020;18:344–54.
[57] Gussow AB, Park AE, Borges AL, Shmakov SA, Makarova KS, Wolf YI, et al. Machine-learning approach expands the repertoire of anti-CRISPR protein families. Nat Commun 2020;11:3784.
[58] Hwang G-H, Yu J, Yang S, Son WJ, Lim K, Kim HS, et al. CRISPR-sub: Analysis of DNA substitution mutations caaused by CRISPR-Cas9 in human cells. Comput Struct Biotechnol J 2020;18:1686–94.
[59] Wang J, Ma J, Cheng Z, Meng Xu, You L, Wang M, et al. A CRISPR evolutionary arms race: Structural insights into viral anti-CRISPR/Cas responses. Cell Res 2016;26:1165–8.
[60] Rollins MF, Chowdhury S, Carter J, Golden SM, Miettinen HM, Santiago-Frangos A, et al. Structure reveals a mechanism of CRISPR-RNA-guided nuclease recruitment and anti-CRISPR viral mimicry. Mol Cell 2019;74:132–42.
[61] Zhu Y, Gao A, Zhan Qi, Wang Y, Feng H, Liu S, et al. Diverse mechanisms of CRISPR-Cas9 inhibition by type IIC anti-CRISPR proteins. Mol Cell 2019;74:296–309.

[62] Thavalingam A, Cheng Z, Garcia B, Huang X, Shah M, Sun W, et al. Inhibition of CRISPR-Cas9 ribonucleoprotein complex assembly by anti-CRISPR AcrIIC2. Nat Commun 2019;10:2806.
[63] Liu L, Yin M, Wang M, Wang Y. Phage AcrIIA2 DNA mimicry: structural basis of the CRISPR and anti-CRISPR arms race. Mol Cell 2019;73:611–20.
[64] Dong D, Guo M, Wang S, Zhu Y, Wang S, Xiong Z, et al. Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. Nature 2017;546:436–9.
[65] Shin J, Jiang F, Liu J-J, Bray NL, Rauch BJ, Baik SH, et al. Disabling Cas9 by an anti-CRISPR DNA mimic. Sci Adv 2017;3:e1701620.
[66] Chowdhury S, Carter J, Rollins MF, Golden SM, Jackson RN, Hoffmann C, et al. Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex. Cell 2017;169:47–57.
[67] Ka D, An SY, Suh JY, Bae E. Crystal structure of an anti-CRISPR protein, AcrIIA1. Nucleic Acids Res 2018;46:485–492.
[68] Bolotin A, Quinquis B, Renault P, Sorokin A, Ehrlich SD, Kulakauskas S, et al. Complete sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus. Nat Biotechnol 2004;22:1554–8.
[69] Hynes AP, Rousseau GM, Lemay M-L, Horvath P, Romero DA, Fremaux C, et al. An anti-CRISPR from a virulent streptococcal phage inhibits Streptococcus pyogenes Cas9. Nat Microbiol 2017;2:1374–80.
[70] Palermo G, Gravina Ricci C, Rivalta I, Batista VS, McCammon JA. An allosteric signaling governs the CRISPR-Cas9 function. Biophys J 2019;116:485a.
[71] Palermo G, Casalino L, Magistrato A, Andrew McCammon J. Understanding the mechanistic basis of non-coding RNA through molecular dynamics simulations. J Struct Biol 2019;206:267–79.
[72] Palermo G. Structure and dynamics of the CRISPR-Cas9 catalytic complex. J Chem Inf Model 2019;59:2394–406.
[73] Pérez-Mejías G, Velázquez-Cruz A, Guerra-Castellano A, Baños-Jaime B, Díaz-Quintana A, González-Arzola K, et al. Exploring protein phosphorylation by combining computational approaches and biochemical methods. Comput Struct Biotechnol J 2020;18:1852–63.
[74] Pantsar T. The current understanding of KRAS protein structure and dynamics. Comput Struct Biotechnol J 2020;18:189–98.
[75] Zhang G, Zeng T, Dai Z, Dai X. Prediction of CRISPR/Cas9 single guide RNA cleavage efficiency and specificity by attention-based convolutional neural networks. Comput Struct Biotechnol J 2021;19:1445–57.
[76] Palermo G, Miao Y, Walker RC, Jinek M, McCammon JA. Striking plasticity of CRISPR-Cas9 and key role of non-target DNA, as revealed by molecular simulations. ACS Cent Sci 2016;2:756–63.
[77] East KW, Newton JC, Morzan UN, Narkhede YB, Acharya A, Skeens E, et al. Allosteric motions of the CRISPR-Cas9 HNH nuclease probed by NMR and molecular dynamics. J Am Chem Soc 2020;142:1348–58.
[78] Palermo G, Casalino L, Jinek M. Two-metal ion mechanism of DNA cleavage in CRISPR-Cas9. Biophys J 2020;118:64a.
[79] Niezwicki Ł, Arantes PR, Saha A, Palermo G. Establishing the allosteric mechanism in CRISPR-Cas9. WIREs Comput Mol Sci 2021;11:e1503.
[80] Aguirre de Cárcer D. Experimental and computational approaches to unravel microbial community assembly. Comput Struct Biotechnol J 2020;18:4071–81.
[81] Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science 2016;351:867–71.
[82] Byrne LC. Check twice, cut once-Improved CRISPR-Cas9 genome editing accuracy. Sci Transl Med 2017;9:e8165.
[83] Lange OF, Grubmüller H. Generalized correlation for biomolecular dynamics. Proteins Struct Funct Genet 2006;62:1053–61.
[84] Dabrowska M, Czubak K, Juzwa W, Krzyzosiak WJ, Olejniczak M, Kozlowski P. QEva-CRISPR: A method for quantitative evaluation of CRISPR/Cas-mediated genome editing in target and off-target sites. Nucleic Acids Res 2018;46:e101.
[85] Luo X, He Y, Zhang C, He X, Yan L, Li M, et al. Trio deep-sequencing does not reveal unexpected off-target and on-target mutations in Cas9-edited rhesus monkeys. Nat Commun 2019;10:5525.
[86] Shen CC, Hsu MN, Chang CW, Lin MW, Hwu JR, Tu Y, et al. Synthetic switch to minimize CRISPR off-target effects by self-restricting Cas9 transcription and translation. Nucleic Acids Res 2019;47:e13.
[87] Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature 2016;529:490–5.
[88] Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. Science 2016;351:84–8.
[89] Nihongaki Y, Kawano F, Nakajima T, Sato M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. Nat Biotechnol 2015;33:755–60.
[90] Senturk S, Shirole NH, Nowak DG, Corbo V, Pal D, Vaughan A, et al. Rapid and tunable method to temporally control gene editing based on conditional Cas9 stabilization. Nat Commun 2017;8:14370.
[91] Kim S, Kim D, Cho SW, Kim J, Kim J-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res 2014;24:1012–9.
[92] Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife 2014;3:e04766.
[93] Mathony J, Harteveld Z, Schmelas C, Upmeier zu Belzen J, Aschenbrenner S, Sun W, et al. Computational design of anti-CRISPR proteins with improved inhibition potency. Nat Chem Biol 2020;16:725–30.
[94] Byun JA, VanSchouwen B, Akimoto M, Melacini G. Allosteric inhibition explained through conformational ensembles sampling distinct "mixed" states. Comput Struct Biotechnol J 2020;18:3803–18.
[95] Lu S, Shen Q, Zhang J. Allosteric methods and their applications: facilitating the discovery of allosteric drugs and the investigation of allosteric mechanisms. Acc Chem Res 2019;52:492–500.
[96] Lu S, Zhang J. Small molecule allosteric modulators of G-protein-coupled receptors: Drug-target interactions. J Med Chem 2019;62:24–45.
[97] Lu S, He X, Ni D, Zhang J. Allosteric modulators discovery: From serendipity to structure-based design. J Med Chem 2019;62:6405–21.
[98] Ni D, Chai Z, Wang Y, Li M, Yu Z, Liu Y, et al. Along the allostery stream: Recent advances in computational methods for allosteric drug discovery. WIREs Comput Mol Sci 2021. https://doi.org/10.1002/wcms.1585.
[99] Nishimasu H, Cong Le, Yan W, Ran FA, Zetsche B, Li Y, et al. Crystal structure of Staphylococcus aureus Cas9. Cell 2015;162:1113–26.
[100] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput 2015;11:3696–713.
[101] Price DJ, Brooks CL. A modified TIP3P water potential for simulation with Ewald summation. J Chem Phys 2004;121:10096–103.
[102] Lu S, He X, Yang Z, Chai Z, Zhou S, Wang J. et al. Activation patwhay of a G protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design. Nat Commun 2021;12:4721.
[103] Ni D, Wei J, He X, Rehman AU, Li X, Qiu Y, et al. Discovery of cryptic allosteric sites using reversed allosteric communication by a combined computational and experimental strategy. Chem Sci 2021;12:464–76.
[104] Lu S, Chen Y, Wei J, Zhao M, Ni D, He X, et al. Mechanism of allosteric activation of SIRT6 revealed by the action of rationally designed activators. Acta Pharm Sin B 2021;11:1355–61.
[105] Jang H, Zhang M, Nussinov R. The quaternary assembly of KRas4B with Raf-1 at the membrane. Comput Struct Biotechnol J 2020;18:737–48.
[106] Lu S, Ni D, Wang C, He X, Lin H, Wang Z, et al. Deactivation pathway of Ras GTPase underlies conformational substates as targets for drug design. ACS Catal 2019;9:7188–96.
[107] Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. J Mol Model 2001;7:306–17.
[108] Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA: Protein complexes. Proc Natl Acad Sci U S A 2009;106:6620–5.
[109] Wang Y, Ji D, Lei C, Chen Y, Qiu Y, Li X, et al. Mechanistic insights into the effect of phosphorylation on Ras conformational dynamics and its interactions with cell signaling proteins. Comput Struct Biotechnol J 2021;19:1184–99.
[110] Li X, Dai J, Ni D, He X, Zhang H, Zhang J, et al. Insight into the mechanism of allosteric activation of PI3Kα by oncoprotein K-Ras4B. Int J Biol Macromol 2020;144:643–55.
[111] Qiu Y, Yin X, Li X, Wang Y, Fu Q, Huang R, et al. Untangling dual-targeting therapeutic mechanism of epidermal growth factor receptor (EGFR) based on reversed allosteric communication. Pharmaceutics 2021;13:747.
[112] Girvan M, Newman MEJ. Community structure in social and biological networks. Proc Natl Acad Sci U S A 2002;99:7821–6.
[113] Del Sol A, Fujihashi H, Amoros D, Nussinov R. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. Mol Syst Biol 2006;2006:0019.
[114] Liang Z, Zhu Y, Long J, Ye F, Hu G. Both intra and inter-domain interactions define the intrinsic dynamics and allosteric mechanism in DNMT1s. Comput. Struct. Biotechnol. J. 2020;18:749–64.
[115] Foutch D, Pham B, Shen T. Protein conformational switch discerned via network centrality properties. Comput. Struct. Biotechnol. J. 2021;19:3599–608.
[116] Marasco M, Kirkpatrick J, Nanna V, Sikorska J, Carlomagno T. Phosphotyrosine couples peptide binding and SHP2 activation via a dynamic allosteric network. Comput. Struct. Biotechnol. J. 2021;19:2398–415.
[117] Zhang Y, Zhang H, Xu X, Wang Y, Chen W, Wang Y, et al. Catalytic-state structure and engineering of Streptococcus thermophilus Cas9. Nat Catal 2020;3:813–23.