

# UC Irvine

## UC Irvine Previously Published Works

### Title

Allosteric regulation in CRISPR/Cas1-Cas2 protospacer acquisition mediated by DNA and Cas2

### Permalink

<https://escholarship.org/uc/item/2651h9tt>

### Journal

Biophysical Journal, 120(15)

### ISSN

0006-3495

### Authors

Long, Chunhong

Dai, Liqiang

E, Chao

et al.

### Publication Date

2021-08-01

### DOI

10.1016/j.bpj.2021.06.007

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# Allosteric regulation in CRISPR/Cas1-Cas2 protospacer acquisition mediated by DNA and Cas2

Chunhong Long,<sup>1</sup> Liqiang Dai,<sup>2,3</sup> Chao E,<sup>3</sup> Lin-Tai Da,<sup>4</sup> and Jin Yu<sup>5,\*</sup>

<sup>1</sup>School of Science, Chongqing University of Posts and Telecommunications, Chongqing, China; <sup>2</sup>Shenzhen JL Computational Science and Applied Research Institute, Shenzhen, China; <sup>3</sup>Beijing Computational Science Research Center, Beijing, China; <sup>4</sup>Shanghai Center for Systems Biomedicine, Shanghai JiaoTong University, Shanghai, China; and <sup>5</sup>Departments of Physics and Astronomy and Chemistry, NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, California

**ABSTRACT** Cas1 and Cas2 are highly conserved proteins across clustered-regularly-interspaced-short-palindromic-repeat-Cas systems and play a significant role in protospacer acquisition. Based on crystal structure of twofold symmetric Cas1-Cas2 in complex with dual-forked protospacer DNA (psDNA), we conducted all-atom molecular dynamics simulations to study the psDNA binding, recognition, and response to cleavage on the protospacer-adjacent-motif complementary sequence, or PAMc, of Cas1-Cas2. In the simulation, we noticed that two active sites of Cas1 and Cas1' bind asymmetrically to two identical PAMc on the psDNA captured from the crystal structure. For the modified psDNA containing only one PAMc, as that to be recognized by Cas1-Cas2 in general, our simulations show that the non-PAMc association site of Cas1-Cas2 remains destabilized until after the stably bound PAMc being cleaved at the corresponding association site. Thus, long-range correlation appears to exist upon the PAMc cleavage between the two active sites (~10 nm apart) on Cas1-Cas2, which can be allosterically mediated by psDNA and Cas2 and Cas2' in bridging. To substantiate such findings, we conducted repeated runs and further simulated Cas1-Cas2 in complex with synthesized psDNA sequences psL and psH, which have been measured with low and high frequency in acquisition, respectively. Notably, such intersite correlation becomes even more pronounced for the Cas1-Cas2 in complex with psH but remains low for the Cas1-Cas2 in complex with psL. Hence, our studies demonstrate that PAMc recognition and cleavage at one active site of Cas1-Cas2 may allosterically regulate non-PAMc association or even cleavage at the other site, and such regulation can be mediated by noncatalytic Cas2 and DNA protospacer to possibly support the ensued psDNA acquisition.

**SIGNIFICANCE** Cas1-Cas2 protein responsible for initial acquisition of protospacer or psDNA are highly conserved among all types of CRISPR-Cas systems. We employed all-atom molecular dynamics simulations to probe how a two-fold symmetric Cas1-Cas2 possibly conducts two cleavages quickly on the psDNA, containing only one PAMc sequence for recognition. We hypothesized that communication exists between two remotely separated active sites of Cas1-Cas2, in association with one PAMc and one non-PAMc DNA, respectively. Consistently, we found long-range correlation or allosteric communication to allow the non-PAMc-binding site to be stabilized or cleaved upon the PAMc cleavage at the other site, coordinated by the psDNA and associated Cas2. The study here brings a working scenario of Cas1-Cas2 acquisition in the absence of additional protein factors.

## INTRODUCTION

Bacteria and archaea use clustered regularly interspaced short palindromic repeats (CRISPRs) along with associated protein (CRISPR-Cas) adaptive immune systems to protect against invading foreign nucleic acids from phages and plasmids (1–5). The CRISPR-Cas system captures the foreign

DNA segments or protospacers and integrate them into the CRISPR array that consists of identical short repeats and variable spacers of similar sizes (1,2,6). The CRISPR-Cas system works in three steps: first, in the protospacer acquisition or adaptation stage, a new spacer is captured and stored in the host genomic CRISPR array (7–9), which is achieved by the highly conserved Cas1-Cas2 protein complex and possibly with assistances from additional protein factors (10). Second, the CRISPR locus is transcribed and processed into short mature CRISPR RNA (crRNA), which then binds to additional Cas proteins and forms a protein-

Submitted January 6, 2021, and accepted for publication June 4, 2021.

\*Correspondence: [jin.yu@uci.edu](mailto:jin.yu@uci.edu)

Editor: Anatoly Kolomeisky.

<https://doi.org/10.1016/j.bpj.2021.06.007>

© 2021 Biophysical Society.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

crRNA complex (2,11). Last, at the targeted interference step, the secondary invading nucleic acid complementary to crRNA is recognized and degraded precisely by the protein-crRNA complex (12,13). Although the overall mechanism of CRISPR-Cas immune system has been well received, some of key steps remain to be elucidated. In particular, the acquisition conducted by Cas1-Cas2 is a least-understood step, even though this step is fundamental and ubiquitous to all types of CRISPR systems (9,14–19).

According to the evolutionary classification of the CRISPR-Cas system and Cas genes, Cas1 and Cas2 are the only two Cas proteins conserved across all CRISPR-Cas systems (20–24). Previous biochemical studies identify Cas1 and Cas2 as metal-dependent nucleases (25,26). Cas1 is capable of cleaving DNA of various forms in a sequence-independent manner and demonstrates catalytic activity in the protospacer acquisition. Cas2 is found to be able to cleave single-stranded DNA and double-stranded DNA (dsDNA), yet no catalytic activity of Cas2 has been shown in the protospacer acquisition (17,27). Correspondingly, the functional role of Cas2 in the acquisition remains to be elucidated.

The conserved Cas1 and Cas2 proteins can assemble into twofold symmetric or dimeric Cas1-Cas2 complex (28,29), with two active sites formed by four Cas1 proteins (Cas1a and Cas1b and Cas1a' and Cas1b'), respectively, whereas two Cas2 proteins (Cas2 and Cas2') are sandwiched in between the dimeric Cas1. Such a Cas1-Cas2 complex captures the protospacer DNA (psDNA) before integrating it into the host CRISPR locus. Such an acquisition or adaptation process can possibly include four substeps: psDNA binding and selection, the 3' overhang cleavages, integration, and DNA repair (5,18,19,28,30–32) (see Fig. 1 A). In this study, we focus on the first two substeps, i.e., the psDNA binding and selection and cleavages at two active sites, assuming that the dimeric Cas1-Cas2 plays a dominant role in the absence of additional protein factors. During the psDNA binding, the protospacer-adjacent motif (PAM) recognition is key for the selection (28,29,33). The protospacer-adjacent-motif complementary sequence (PAMc) recognition via hydrogen bonding (HB) and amino-acid side-chain stacking in the binding pocket are shown (see Fig. 1 B, bottom).

It has been shown that PAM is important for recognition and selection of the protospacer during the acquisition (15,16,34). In particular, psDNAs flanked by the correct PAM (GAA) can be cleaved and integrated efficiently into the CRISPR array, which ensure that the foreign psDNA containing the right PAM is incorporated (33,35,36). Meanwhile, in the crystal structures captured for the psDNA acquisition complex of Cas1-Cas2 (28), symmetrical PAM complementary sequence, or PAMc (CTT) were constructed at the nucleotides (nts) 28–30 in the two 3' overhangs of the DNA protospacer, and both active sites (labeled as site1 and site2) formed by Cas1a and Cas1a' are indeed bound with PAMc. Correspondingly, the 3' PAMc sequence in this sys-

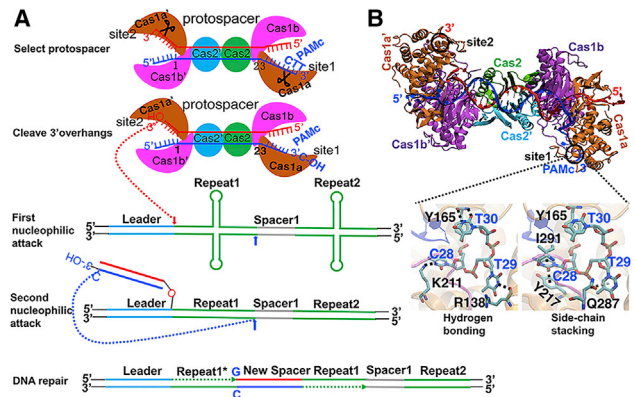


FIGURE 1 (A) Schematics of psDNA acquisition or adaptation into the host CRISPR array for the CRISPR-Cas adaptive immunity. The acquisition process includes four substeps (from top to bottom): protospacer binding and selection, 3' overhang cleavage, integration, and DNA synthesis and repair. During the integration step, PAMc sequence C'-OH (cleavage at site1 of Cas1-Cas2) is integrated into the spacer side of Repeat1 in the CRISPR array (likely via the second nucleophilic attack), whereas non-PAMc (cleavage at site2) is integrated into the leader side of Repeat1 (likely via the first nucleophilic attack) (19,28). (B) Crystal structure of *E. coli* Cas1-Cas2 bound to dual-forked psDNA (Protein Data Bank: 5DQZ) (28). Cas1a and Cas1a' are colored in orange. The Cas1b and Cas1b' are colored in magenta. Cas2 and Cas2' are shown in cyan and green, respectively. The two active site1 and site2 are shown with black circles. In the crystal structure, both sites are bound with PAMc (CTT). The PAMc recognition via hydrogen bonding (HB; shown between heavy atoms in participation) and amino-acid side-chain stacking in the binding pocket are indicated in black and gray dotted lines (bottom left and right), respectively. To see this figure in color, go online.

tem is supposed to be cleaved between C and T, and C remains with the spacer to be integrated to the CRISPR array (see the schematics in Fig. 1 A, bottom).

Upon the psDNA association and the 3' overhang PAMc recognition, the CTT can be cleaved by the Cas1-Cas2 complex at one active site because of the high specificity of such reaction (18,19,28,29). Meanwhile, the other active site, which generally binds to a non-PAMc at the other 3' overhang, is also supposed to conduct the catalytic cleavage, though it is not clear whether the cleavage can proceed sufficiently fast on the non-PAMc. Notably, high-resolution structural studies have caught a half-site intermediate integration complex of Cas1-Cas2 (30), which has its non-PAMc side 3'-OH linking to the CRISPR between the leader and Repeat1, as if it happens as the first nucleophilic attack on the CRISPR array (see Fig. 1 A, schematics). Assuming only Cas1-Cas2 works on the psDNA during this process, if the non-PAMc cleavage conducted by Cas1-Cas2 happens much slower than the first PAMc cleavage, then it is likely that an alternative half-site intermediate integration complex dominates, which has its PAMc 3'-OH linked to the CRISPR as the first nucleophilic attack (i.e., between the Repeat1 and the previous spacer or spacer1). The existence of the first half-site integration complex of Cas1-Cas2, with the non-PAMc 3' overhang linked to CRISPR, thus suggests

that the non-PAMc can be cleaved (at site2, labeled for convenience) almost as fast or immediately after the PAMc cleavage (at site1) when Cas1-Cas2 is the dominant player in charge of this acquisition process. We therefore want to explore how the non-PAMc site can be cleaved sufficiently fast in the Cas1-Cas2 acquisition complex. We hypothesized that certain communication exists between the two active sites in the Cas1-Cas2 complex so that the efficient non-PAMc cleavage is enabled.

We accordingly conducted atomistic molecular dynamics (MD) simulations on the high-resolution crystal structure of the twofold Cas1-Cas2 complex in association with a dual-forked psDNA (28,29), which contains two PAMc on the two fork regions, respectively. In this dual-PAMc complex, the psDNA lies on the surface of Cas1-Cas2 in a head-to-head orientation (see Fig. 1 B). Two identical PAMc are bound by the two active sites (site1 and site2), respectively. We simulated the original Cas1-Cas2 binding complex with the dual PAMc and identified instead an asymmetrical equilibrium binding pattern between the two active sites. Then, we modified one PAMc to a non-PAMc (at site2) while keeping the other PAMc intact (at the site1) to mimic PAMc binding and recognition in general. We subsequently examined the Cas1-Cas2 systems from the psDNA binding to precatalytic and to a half-site postcatalytic state (right after the site1-PAMc cleavage) and monitored structural and correlation dynamics between the two active sites in respective systems. To substantiate our findings revealing certain correlations between the active sites upon the site1-PAMc cleavage modeled in the original Cas1-Cas2 complex, two synthetic psDNA sequences, which were experimentally identified with low and high acquisition frequencies or efficiencies (37), were further tested in the Cas1-Cas2-psDNA simulation systems at the half-site postcatalytic state. Via comparative studies, we demonstrate negative cooperativity between the two active sites, which seems to rely on allosteric propagation  $\sim 10$  nm from the site1-PAMc to the site2-non-PAMc upon the site1-PAMc cleavage, as being mediated by the psDNA in acquisition and noncatalytic Cas2 and Cas2' in close association with the psDNA, which are harbored in between Cas1a and Cas1b (containing site1) and Cas1a' and Cas1b' (containing site2).

## MATERIALS AND METHODS

By employing all-atom MD simulations, we analyzed the Cas1-Cas2 proto-spacer binding and recognition via examining formation of potential HBs and base-stacking characteristics at the active site (see Fig. 1). We first examined the original psDNA binding complex captured from the crystal structure with two identical PAMc bound at both sites (site1 and site2); then, we focused on the modified system, with one PAMc (CTT) bound at site1 and one non-PAMc (TTT) in association with site2, as supposed to be recognized by Cas1-Cas2 in general. For such a one-PAMc system, we examined not only the psDNA binding state (with PAMc but not non-PAMc stably bound) but also a precatalytic state (with catalytic magnesium ions incorporated to the active site1) and a half-site postcatalytic state (with

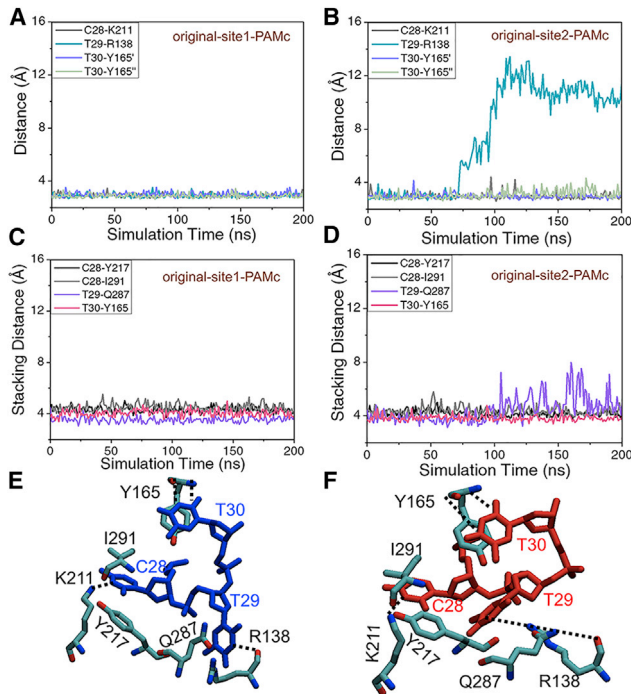
PAMc modeled cleaved at site1, as if being catalyzed via an endonuclease reaction). See [Supporting materials and methods](#) and [Figs. S1–S4](#) for modeling at active site, simulation equilibration, water box impacts, and histidine protonation status. For each state of the one-PAMc system, the protein-DNA internal correlations with the active site1 are measured for individual residues and compared among different simulation systems to reveal allosteric signal from the PAMc-binding site to the non-PAMc-binding site and the rest of the Cas1-Cas2 psDNA complex. Repeating equilibrium simulations have been conducted for above systems (each system for up to three repetitive simulation runs of 200 ns individually) to ensure robust results. Because above analyses indicated that Cas2-psDNA in bridging the two active sites play particular roles in the allosteric communication, Cas1-Cas2 in complex with the modified synthetic pSL and pSH DNA sequences are further examined (with extended simulations from 200 to 500 ns for computational convergence) to verify the allosteric effects and the role of psDNA and Cas2.

## RESULTS

### Cas1-Cas2 bound with two identical PAMcs are asymmetrically stabilized at one site and nonstabilized at the other site

We first performed an equilibrium MD simulation to the original psDNA binding complex of Cas1-Cas2, which was crystalized with twofold symmetry, and with two identical PAMcs bound symmetrically at both active site1 and -2 (Protein Data Bank: 5DQZ) (28) (see Fig. S4). According to the HBs and stacking interactions formed between the protein and PAMc at site1 and site2 in the crystal structure (also in Fig. S5), we examined whether these interaction characteristics maintain in the equilibrium simulation. To keep simulation model close to the crystal structure, positional restraints were implemented for up to 10 ns at the beginning of the simulation (see [Supporting materials and methods](#)). When simulated within a large water box ( $\sim 180$ -Å size or with a minimal distance between the protein-DNA complex and boundary reaching  $\sim 25$  Å, and the simulation system  $\sim 0.5$  million atoms), the HBs and stacking interactions at site1 were maintained well during the simulation (see Fig. 2, A–C; see Fig. S6 for results from two repetitive simulation runs). Note that using a small simulation box ( $\sim 150$ -Å size or a minimal distance  $\sim 11$  Å), the PAMc-site1 could not be stabilized indeed (see Fig. S3); testing with an even larger simulation box ( $\sim 200$  Å, a minimal distance  $\sim 35$  Å), site1 can become stabilized as in the 180-Å case (see Figs. S6 and S7). Note that the water-box-size-dependent behaviors have been demonstrated systematically in a previous work simulating tetrameric complex of human hemoglobin via all-atom MD (38), which highlights the importance of simulating bulk water environment and hydrophobicity of the protein system in conformational sampling and energetics. It is also possible that the multimeric or the protein quaternary structure matters essentially in the comparatively large assembly of the systems.

In contrast to site1, at the site2, we found that the expected HB between residues R138 and T29 of PAMc (CTT) was broken at  $\sim 70$  ns of the simulation, and the



**FIGURE 2** The equilibrium MD simulation of the Cas1-Cas2 complex with dual-PAMc psDNA and respective association characteristics at both site1 and site2. (A and B) The expected HB interactions with distances measured between residues R138, Y165, K211, and PAMc (CTT) at binding site1 and site2, respectively. (C and D) The expected stacking interactions with distances measured between the centers of mass of residues Y165, Y217, Q287, I291, and PAMc at binding site1 and site2, respectively. (E and F) The structural views at site1 and site2, at the end of the equilibrium simulation. The PAMc nts are colored blue and red at site1 (E) and site2 (F), respectively. The amino acids are colored by atom types (cyan, red, and blue for carbon, oxygen, and nitrogen, respectively). The expected HBs (according to the crystal structure, may not form in the simulation, e.g., at site2) are highlighted by dotted lines. To see this figure in color, go online.

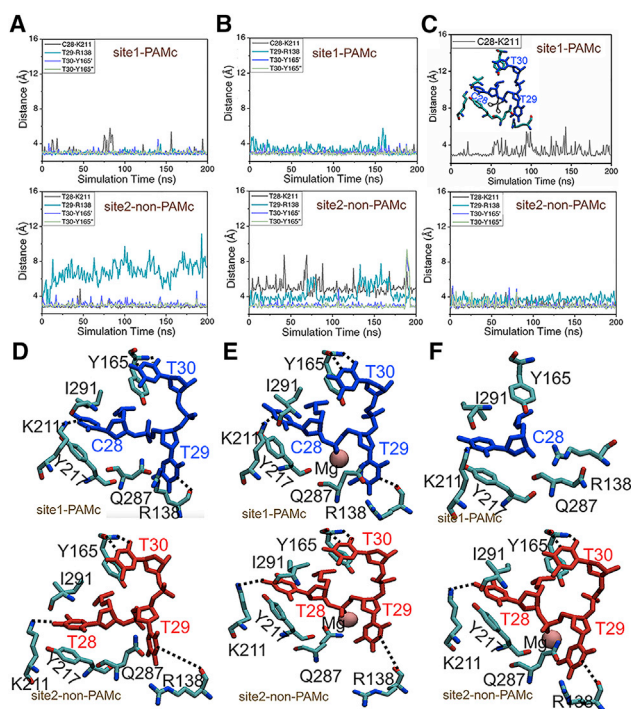
distance between them reached up to  $\sim 12$  Å (Fig. 2 B). The expected stacking between Q287 and T29 at binding site2 also became unstable after  $\sim 100$  ns as the distance between the centers of mass of residues Q287 and T29 reached above  $\sim 6$  Å (Fig. 2 D). The binding configurations of the active site1 and site2 at the end of the simulation are shown in Fig. 2, E and F, respectively. In the repetitive simulation runs as well as in the larger simulation box, site2 remains more or less destabilized (see Figs. S6 and S7). These observations indicate that the two active sites largely maintain asymmetrical binding patterns to PAMc, even though the crystal structure shows twofold symmetry and with two identical PAMc association with site1 and site2. The asymmetrical PAMc-binding dynamics of the two active sites of Cas1-Cas2 also implies that it can easily accommodate binding and recognition to psDNA containing only one PAMc considering that two PAMcs separated by  $\sim 30$  bp is a rare configuration to be found on psDNA. Below, we focus on the system that is commonly incurred, i.e., Cas1-Cas2 binding to psDNA with only one PAMc.

### The site2-non-PAMc cannot be stabilized until after the PAMc cleavage conducted at the site1

Except for the original dual-PAMc psDNA binding complex of Cas1-Cas2, the following systems were all modified to site1-PAMc (CTT) and site2-non-PAMc (TTT), and we modeled the system from the psDNA binding to the pre-catalytic state and then to the half-site postcatalytic state (see Supporting materials and methods). Similarly, we checked the expected protein-DNA HBs and stacking characteristics as such for the stabilized PAMc binding and recognition. First, we found that the HB between residue R138 and T29 from non-PAMc (TTT) at site2 was absent in both the binding and pre-catalytic states during respective simulations (see Fig. 3, A and B). Second, the expected HB between residue K211 and T28 from TTT at site2 in the pre-catalytic state was also absent (Fig. 3 B). In addition, the stacking interaction between Q287 and T29 at site2 was also unstable, in particular at the binding state (see Fig. S8, A and B). In contrast, such expected HBs and stacking interactions were stably maintained at site1 bound with PAMc. Hence, the asymmetrical binding configuration of site1 and site2 does fit well with the PAMc and non-PAMc association, respectively.

Remarkably, for the half-site postcatalytic state, i.e., because the PAMc was cleaved artificially at site1 (by preparing and modeling the site1-PAMc cleaved for the corresponding simulation system), the site2 in association with non-PAMc soon became stabilized within the 200-ns equilibrium simulation, according to the same HBs and stacking characteristics expected for the stable binding configuration (see Fig. 3 C; Fig. S8 C). The molecular views toward the end of each equilibrium MD simulation, from the binding to the pre-catalytic state and to the postcatalytic state, for both the site1-PAMc and site2-non-PAMc are also provided (see Fig. 3, D–F). From the molecular views, one can directly see that the expected HB between residues R138 and T29 from TTT at site2, for example, which was absent in both the binding and pre-catalytic state equilibration simulation, soon became present in the simulation of the half-site postcatalytic state, i.e., upon the PAMc cleavage at site1. In the two sets of repetitive equilibrium MD simulation runs ( $2 \times 3$  simulations) for the site1-PAMc and site2-non-PAMc (TTT) systems from the psDNA binding to the pre-catalytic state and to the half-site postcatalytic state, consistent results show that the site2-non-PAMc cannot be stabilized until after the site1-PAMc cleavage, though detailed HB or stacking dynamics vary slightly in different simulation runs (see Figs. S9 and S10).

These results consequently suggest that the two active sites can exhibit allosteric communication and consequently negative cooperativity with each other, and such cooperativity only allows one active site to be stabilized in close association with the psDNA for binding and recognition on one PAMc. As a result, site2 can bind stably with the non-PAMc only after the PAMc is cleaved at site1 as the close



**FIGURE 3** The association patterns at the site1-PAMc and site2-non-PAMc in the equilibrium MD simulation of the Cas1-Cas2 psDNA complexes from the binding to pre-catalytic and to half-site post-catalytic state. Shown are the expected HBs with distances measured between residues Y165, R138, K211 at site1-PAMc (CTT; *top*), and similarly at site2-non-PAMc (TTT; *bottom*), in the binding state (A), the pre-catalytic state (B), and the half-site post-catalytic state with PAMc cleaved at site1 (C). Shown is the molecular view at the end of the equilibrium MD simulation for the site1-PAMc (*top*) and site2-non-PAMc (*bottom*), from the site1 binding (D) to pre-catalytic (E) and to half-site post-catalytic state (F), with site2 changing from nonstabilized (D and E) to stabilized (F). The coloring scheme is the same as Fig. 2, E and F. To see this figure in color, go online.

association between site1 and psDNA becomes lost. We thus infer that the cleavage to the PAMc at the corresponding binding and recognition site of Cas1-Cas2 can be necessary for the non-PAMc to be bound favorably by the other site and likely to be cleaved fast as well.

Besides, we have also examined another two Cas1-Cas2 psDNA complex systems: 1) with site1-non-PAMc (TTT) and site2-PAMc (CTT), and 2) with both site1-non-PAMc (TTT) and site2-non-PAMc (TTT) (see Fig. S11). For the system 1) we found that site2 with PAMc became stabilized, whereas site1 with non-PAMc was unstable. For the system, 2) both site1 and site2 became highly unstable with the non-PAMc, e.g., the expected HBs between residues K211 and T28 and between residues R138 and T29 were absent.

### The allosteric propagation from the site1-PAMc to the site2-non-PAMc upon the site1-PAMc cleavage

For each equilibrium simulation of the Cas1-Cas2-psDNA complex, from the binding to pre-catalytic and to half-site

post-catalytic state (for PAMc-bound site1), the dynamics correlation within the protein-DNA complex was calculated (see Supporting materials and methods), in particular, between the active site1-PAMc and the rest part of the protein-DNA complex (see Fig. 4 A). Note that high self-correlation values ( $>30$ ) show for Cas1a where the active site1 locates and for the PAMc regions in close association with the active site1. Residues 93–108, residues 200–220, and residues 275–295 in Cas1b around the active site1 also show high correlation values ( $\sim 20$ ) with the active site1 due to local interactions. For nonlocal regions (Cas1a' and Cas1b', the rest part of DNA aside from PAMc, and Cas2 and Cas2', labeled in Fig. 4 A), the site2-non-PAMc, located in Cas1a', demonstrates an elevated correlation with the active site1 into the half-site post-catalytic state (i.e., up to 5–10; see Fig. 4 C); psDNA and Cas2 and Cas2' also show slightly increased correlations with the site1 into the post-catalytic state (Fig. 4, B and D).

Indeed, in comparison with the binding and pre-catalytic state, the protein-DNA complex upon the site1-PAMc cleavage shows lowered self or local correlations with the site1-PAMc but enhanced nonlocal correlations with the site1 at remote regions, including Cas1a' and Cas1b' (hosting site2), psDNA (aside from PAMc), and Cas2 and Cas2'. That is, correlation between the two far-apart active site1 (at Cas1a and Cas1b) and site2 (Cas1a' and Cas1b') exists and strengthens upon the PAMc cleavage at site1, which indicate allostery, i.e., action at distance (39,40), is triggered. Therefore, it is inferred from the lowered local correlation and elevated nonlocal correlation with the active site1 (from binding to pre-catalytic and to post-catalytic state) that the allosteric signal propagates from the active site1 to the rest part of the protein-DNA complex, and such allosteric propagation is mediated by psDNA and Cas2 and Cas2', reaching remotely to the site2-non-PAMc ( $\sim 10.3$  nm away from site1), upon the site1-PAMc cleavage. Similarly, we performed two sets of repetitive 200-ns equilibrium simulations on the one-PAMc-binding, pre-catalytic, and half-site post-catalytic states of Cas1-Cas2 (site1-PAMc and site2-non-PAMc) and calculated the protein internal correlations between the active site1 (with PAMc) and the rest part of the protein-DNA (see Fig. S12).

For further comparison, we also examined the protein internal correlations in the original psDNA binding complex of Cas1-Cas2 with dual PAMc, i.e., with site1 stabilized and site2 destabilized in the simulation. The correlations were calculated for both the active site1 and site2 to see how the two sites respectively couple with the rest of the part of the protein-DNA complex. Overall, the correlations in the dual-PAMc system appear larger than those in the single-PAMc system, and one finds that the remote correlations to the stabilized site1 and to the nonstabilized site2 are similarly large (see Fig. S13). Hence, allosteric propagation between the two active sites remote to each other on Cas1-Cas2 appear mutual or bi-directional.

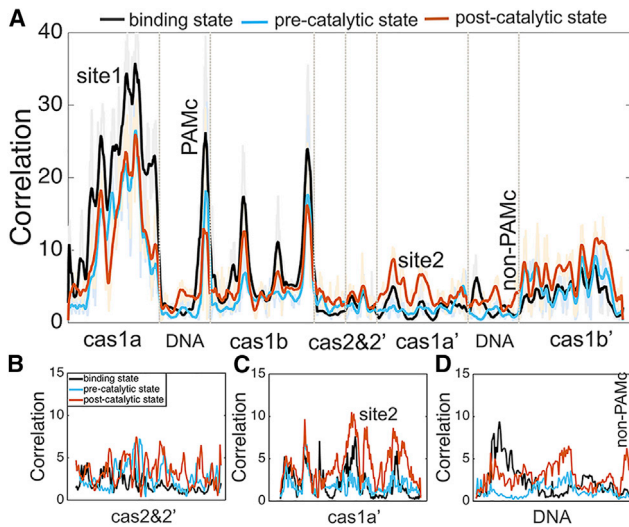


FIGURE 4 The internal correlation strength between the active site1 and the rest parts of the Cas1-Cas2 psDNA complex (site1-PAMc and site2-non-PAMc) during the equilibrium MD simulations. (A) The correlations between the site1 (PAMc) and the rest part, from the binding state (dark line) to precatalytic state (light blue) and to half-site postcatalytic state (orange). (B–D) The zoomed-in views of the correlation patterns between the active site1 and nonlocal residues from Cas2 and Cas2' (B), Cas1a' hosting site 2 (C), and psDNA hosting non-PAMc (D). To see this figure in color, go online.

### Cas1-Cas2 in complex with psDNA of high acquisition efficiency (psH) shows the most prominent allosteric propagation upon the PAMc cleavage

According to synthetic approach to sequence-dependent protospacer acquisition of CRISPR-Cas1-Cas2, different psDNA sequences can lead to different acquisition frequencies or efficiencies (37). In particular, two representative protospacer sequences with particularly low and high acquisition efficiencies identified experimentally were incorporated into our simulation systems (psL and psH) so that we could examine whether allosteric propagation between two active sites persists in the Cas1-Cas2 complex with psL or psH in comparison with the original psDNA.

Similarly, we performed a series of equilibrium simulations on the one-PAMc binding and recognition for both the psL and psH systems at binding, precatalytic, and half-site postcatalytic states of Cas1-Cas2 and calculated the protein internal correlations between the active site1 (with PAMc) or site2 (with non-PAMc) and the rest part of the protein-DNA. Comparing the three sets of systems with the original psDNA, psL, and psH, we found that the overall protein internal correlation is actually highest for the Cas1-Cas2 bound with the psH DNA sequences, particularly in the half-site postcatalytic state (see Fig. 5 for the site1 correlation in the left panel, and site2 correlation in the right panel) as the simulations were extended from 200 to 500 ns. The results thus suggest that the high acquisition

efficiency of the psH system can be possibly explained by significant allosteric communication between the two active sites inside Cas1-Cas2. With a significant allosteric signal propagated from the site1-PAMc upon cleavage to the site2-non-PAMc in the psH system, the non-PAMc stabilization and followed cleavage are expected to happen fast so that the acquisition or integration efficiency can be high in the psH system. The results also support the idea that the allosteric propagation from the site1-PAMc to the site2-non-PAMc is regulated by the psDNA and Cas2 and Cas2', which is in close association with the duplex region of the psDNA.

A color map of the correlation strength (blue, white, and red: high, medium, and low correlation, respectively) on Cas1-Cas2 and psDNA are provided to visualize the allosteric propagation patterns for all the different psDNA systems (original, psL, and psH) at the half-site postcatalytic state (see Fig. 6, top). It becomes quite clear that the psH system is comparatively highly correlated in general upon the first cleavage at PAMc, and the correlation or allosteric propagation proceeds largely via Cas2 and Cas2' and dsDNA regions in the middle of complex. In contrast, psL appears to have the lowest correlation strength going through the psDNA or reaching to the site2-non-PAMc region. In the original psDNA system, noticeable correlation still shows between the site2 and site1, along with a medium level of psDNA correlation propagation compared with psH and psL. Additionally, we have also calculated the Cas2 and Cas2'-dsDNA association energetics for respective Cas1-

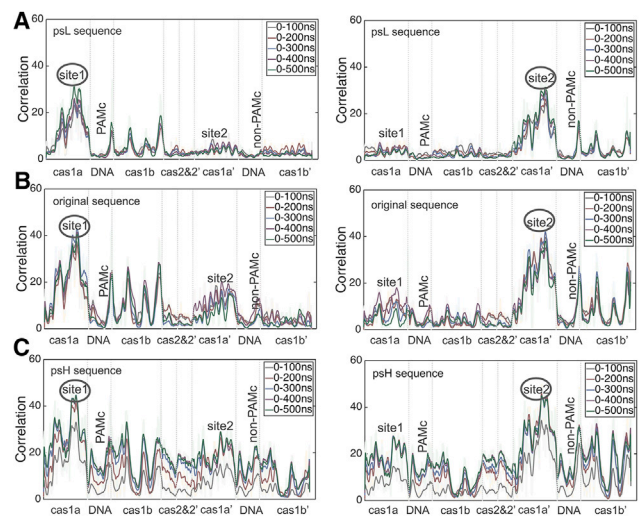
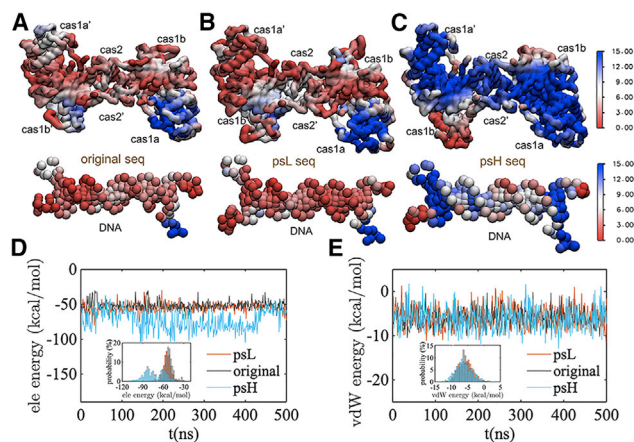


FIGURE 5 The protein internal correlation from extended equilibrium MD simulations (up to 500 ns) between the active site1/site2 and the rest parts of the Cas1-Cas2 psDNA complex, in the half-site postcatalytic state (site1-PAMc cleaved) with different psDNAs. The correlations between site1 and the rest of protein-DNA complex (left) and between site2 and the rest of the protein complex (right) are both shown for the psL psDNA (A), the original psDNA (B), and the psH psDNA (C), with the correlation measurements obtained from 0 to 100, 0–200, 0–300, 0–400, and 0–500 ns. To see this figure in color, go online.

Cas2 psDNA complexes (in the half-site postcatalytic state). Notably, one can see that electrostatic interactions between the Cas2 and Cas2' protein and the duplex part of DNA are strongest in the psH system (see Fig. 6, bottom), which again supports the idea that psDNA contributes directly to the correlation or allosteric communication in the Cas1-Cas2 acquisition complex by associating closely with Cas2 and Cas2'.

We also examined internal correlations for psL and psH and the original psDNA systems in the initial psDNA binding and the precatalytic state (see Fig. S14). Indeed, Cas1-Cas2 in complex with psH produces comparatively low correlation remotely in the initial binding state, i.e., even lower than that in the psL or the original system. In the precatalytic state (see detailed interaction characteristic and active site views in Figs. S15–S17); however, the psH system correlation becomes slightly higher than that in the psL or the original DNA system. In the postcatalytic state or upon the site1-PAMc cleavage, the psH system does show quenched fluctuations or stabilization association at the site2-non-PAMc compared with the psL system (see Fig. S17). The results again suggest that the catalysis or cleavage in the site1-PAMc is necessary to activate the allosteric propagation. In addition, we conducted the principal component analysis on the psL, original, and psH psDNA systems (see Fig. S18), and the results show that the psH system demonstrates comparatively stable motions along the first principal component or PC1, which may support collective motion in the system or facilitate the allostery.



**FIGURE 6** The color maps of correlation strength between the active site1 (bound with PAMc) and the rest of the protein-DNA complex viewed on the structures with different psDNA sequences in the half-site postcatalytic state for the 500-ns equilibrium simulations. Shown is the correlation map on the Cas1-Cas2 structure and dual-fork psDNA (blue, white, and red: high, medium, and low correlation values, respectively) for the original psDNA sequences (A), psL (B), and psH (C). Note the correlation data are the same as that used in Fig. 5, measured from the equilibrium MD simulations. (D and E) The electrostatic and Van der Waals energies between Cas2 and Cas2' and dsDNA for various psDNA complexes: original, psL, and psH (with average values  $-51 \pm 6$ ,  $-55 \pm 6$ , and  $-72 \pm 14$  kcal/mol for the electrostatic energy and  $-5.9 \pm 1.9$ ,  $-5.8 \pm 2.2$ , and  $-6.1 \pm 2.5$  kcal/mol for the Van der Waals energy). To see this figure in color, go online.

In particular, under the PC1 motion, one can find significant loop motions around the site2-non-PAMc in the psL system; not much changes around the site2 but subdomain movements in the original psDNA system and medium level of local loop motions combined with some subdomain movements in the psH system (see Fig. S18 C).

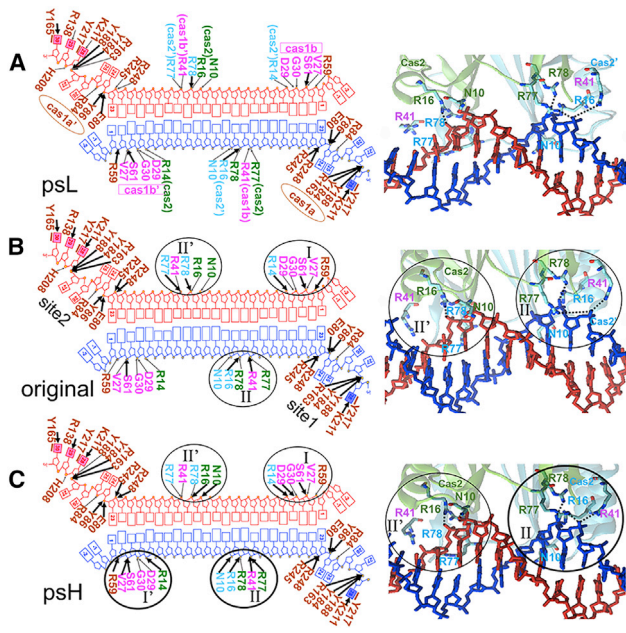
Last, we examined the HBs formed between Cas1-Cas2 and psDNA at the half-site postcatalytic state, comparing the systems with different psDNA sequences, i.e., the original, psL, and psH. In particular, we calculated HB occupancies for these systems. For HB occupancies  $>50\%$  between Cas1-Cas2 and DNA during the simulation, the psH system shows much more stabilized HBs than those in the original and psL systems (see Fig. 7, left panel; Table S1). One notices that there are three protein-DNA binding zones in the original psDNA system that show more HBs formed than that in the psL system: zone I (Cas1a and Cas1b and Cas2' with nts 1–5 on the top strand of psDNA), zone II' (Cas1b' and Cas2 and Cas2' with nts 13–16 on the top strand), and zone II (Cas1b and Cas2 and Cas2' with nts 13–16 on the bottom strand). Furthermore, the regions that show more HBs in the psH system than in the original system include an additional zone I' (Cas1a' and Cas1b' and Cas2 with nt 1–5 on the bottom strand), whereas zone II interaction is also enhanced in the psH system than in the original or psL system. As a result, one can see that Cas2 and Cas2' dominate all the essential protein-DNA HB interaction zones connecting the two active sites, with enhanced HB interactions from the psL to the original and to the psH system. Hence, Cas2 and Cas2' appear critical for the involved allosteric propagation. The key residues at the Cas2 and Cas2'-DNA interface mediating long-range communication are also shown in the right panel of Fig. 7 for structural reviews of respective systems.

Among these HBs, R77, R78, R16, and R14 from Cas2 and Cas2', which interact with DNA substantially in the psH or in the original system, have also been found crucial experimentally. Recent high-resolution structural studies show that the mutants of R77 and R78 by Ala residues can reduce spacer acquisition efficiency; in addition, no spacer acquisition was observed for R16A and R14A mutants (28,29). The HB patterns at the protein-DNA interface and the protein-DNA interaction energies together suggest that the correlation builds up between the two active sites, primarily via the Cas2-Cas2'-psDNA interaction region, and such allosteric communication becomes prominent in the psH system at the postcatalysis state once the PAMc gets cleaved at site1, and the system awaits site2 to be engaged with the non-PAMc.

## DISCUSSION

In this work, we performed multiple sets of atomistic simulations to investigate how CRISPR-Cas1-Cas2 binds, recognizes, and possibly conducts cleavages efficiently on the





**FIGURE 7** The schematics of the Cas1-Cas2 psDNA complex at the half-site postcatalytic state with high HB occupancies obtained from three simulation systems with different psDNA sequences (psL, original, and psH). (A–C) The HB occupancies are compared among the three simulation systems, and those HBs with occupancies  $>50\%$  during the simulation time are labeled with black arrows, those  $<50\%$  are labeled as well (in the *left panel*). Molecular views at Cas2/2'-DNA interface likely involved in the allosteric communications are shown in the right panel for the respective systems. Those contribute to increase HB associations (from psL to original and to psH) are the following: N10, R14, R16, and R77 from Cas2; N10 and R14 from Cas2'; D29 and R41 from Cas1b; and R138, R163, K211, R245, and R248 from Cas1a' (see also [Table S1](#)). The important residues R77, R78, N10, and N16 for the Cas2 long-distance communication are also shown (in the *right panel*) for psL, original, and psH psDNA systems. To see this figure in color, go online.

psDNA in the adaptation stage of CRISPR-Cas immunity process, assuming Cas1-Cas2 played a dominant role in the adaptation. To simulate Cas1-Cas2 consisting of Cas1a and Cas1b, Cas1a' and Cas1b', and Cas2 and Cas2' in stable association with psDNA, i.e., to stabilize at least one active site of Cas1-Cas2a, a particularly large solvent box with a size  $\sim 18$  nm was utilized, which led to a large, solvated simulation system slightly more than  $\sim 0.5$  million atoms. A series of sub-microsecond equilibrium MD simulations were then performed systematically to individual systems 1) on the Cas1-Cas2 associating with psDNA of different PAMc and non-PAMc configurations at two active sites (dual PAMc tested with three water box sizes from  $\sim 15$  to 18 and to 20 nm), two single PAMc, and dual non-PAMc for binding), 2) on the site1-PAMc and site2-non-PAMc configuration, systems from PAMc binding to precatalytic and to a half-site postcatalytic state (i.e., three states with the original psDNA), and 3) with two synthetic psDNA of low and high acquisition efficiency tested (psL and psH, two additional three-state systems), in a total of  $>20$  simulations for over 5  $\mu$ s in aggregation.

Interestingly, we found that even in association with two identical PAMc, the original twofold symmetric Cas1-Cas2 demonstrate an asymmetric binding pattern between two active sites: the active site1 residues form stabilized HBs and stacking interactions with PAMc from one fork region of the psDNA, whereas the active site2 cannot bind stably or specifically to PAMc on the symmetric fork region, and such results are consistently shown in repetitive simulation runs at two large water box sizes (18 and 20 nm;  $2 \times 3$  simulation runs conducted). Therefore, it appears that only one active site of Cas1-Cas2 is capable of binding and recognition of PAMc at a time, which suits for locating only one PAMc on the psDNA for Cas1-Cas2 target search. It then calls into question on how Cas1-Cas2 possibly binds and cleaves the non-PAMc sufficiently fast at site2, e.g., to support formation of a half-site intermediate integration complex (30). In such a complex, psDNA has its non-PAMc side of 3'-OH overhang linking to the CRISPR locus as if it is resulted from the first nucleophilic attack, which is likely only if the non-PAMc is cleaved almost as fast as the PAMc by Cas1-Cas2, presumably in the absence of additional protein factors.

A solution to have non-PAMc on the psDNA cleaved sufficiently fast by Cas1-Cas2 is suggested from our simulation studies of Cas1-Cas2 in complex with dual-forked psDNA, containing one PAMc and one non-PAMc at respective fork regions, i.e., in the site1-PAMc and site2-non-PAMc configuration. We found that even though that the active site2 associates only loosely with non-PAMc at beginning (in the Cas1-Cas2 initial binding state with psDNA), once site1 recognizes PAMc (past the precatalytic state) and cleaves on the PAMc (modeled as the half-site postcatalytic state), site2 immediately becomes stabilized with the non-PAMc to be ready for the second cleavage ( $3 \times 3$  repetitive simulation runs conducted). The mechanism is summarized in schematics in [Fig. 8](#). The two active sites of Cas1-Cas2 can demonstrate negative cooperativity in a “seesaw” manner, with one site being capable of binding specifically on the PAMc for recognition and conducting the first catalytic cleavage at that site; the other site then catches up with stabilized DNA association to cleave nonspecifically immediately after the first site cleavage, taking advantage of the negative cooperativity. Such negative or seesawing type of cooperativity had been reported in other enzyme systems, for example, in the homo-dimeric insulin receptor, the membrane-spanning tyrosine kinase allowing insulin to dock into two binding pockets (41); in the Mo-bisPGD enzyme arsenate oxidase, which impacts on the early life metabolic reactions, with “the redox seesaw” cooperativity induced by the pyranopterin ligands (42); and in the “seesaw” model of enzyme regulation of mTORC1, in which to produce a nonlinear, ultrasensitive responses (43).

The two active sites in the dimeric Cas1-Cas2 system are located symmetrically at two remote regions on the Cas1a-Cas1b and Cas1a'-Cas1b', respectively,  $\sim 10$  nm apart. The

~23-bp dsDNA region of the psDNA is bound primarily with Cas2 and Cas2', in between Cas1a and Cas1b and Cas1a' and Cas1b' containing the two active sites. To find out how the communications between the two remote sites achieve in the Cas1-Cas2 psDNA acquisition complex, we analyzed dynamic correlations within the complex from the MD trajectories, focusing on how the rest of the part of the protein-DNA complex correlates with the active site1 that is bound with PAMc, from the psDNA binding at site1 to the precatalytic state, and to the half-way postcatalytic state. Indeed, one could identify nontrivial correlation between the site2-non-PAMc and site1-PAMc, and such correlation essentially increases into the half-site postcatalytic state right after the site1 cleavage on the PAMc. The comparative correlation analysis, therefore, suggests that the allosteric communication exists that propagates fluctuations at the active site1-PAMc cleavage as a mechanical signal to the remote site2-non-PAMc. Note that such allosteric communication is not triggered by an effector binding (44) but by catalytic cleavage (of PAMc) on the

psDNA. More of the close examinations show that it is primarily the Cas2-Cas2'-psDNA association, i.e., via the protein-DNA electrostatic interactions, as well as the HB interactions formed at the protein-DNA interface (particularly at the Cas2 and Cas2'-dsDNA interaction zones), that are responsible for such allosteric propagation. In such critical Cas2 and Cas2'-dsDNA interaction zones, R14, R16, R77, and R78 residues seem to play important roles to facilitate the allosteric propagation, and the mutants of these arginine residues indeed reduce the protospacer acquisition efficiency significantly (28,29). Correspondingly, we propose that Cas2-Cas2'-dsDNA embedded in the middle of the Cas1-Cas2 dimeric complex mediates the two remotely located active site1 and site2, allosterically. With such mediation, negative cooperativity arises between the two remote sites, which allows the nonspecific non-PAMc cleavages to possibly take place quickly once a specific cleavage of PAMc by Cas1-Cas2 is conducted.

To substantiate such an idea, we further constructed Cas1-Cas2 complexes containing two synthetic DNA protospacers, psL and psH, which were experimentally identified with low and high integration efficiency, respectively (37). Similar MD simulations and comparative correlation analyses were also conducted to the modified Cas1-Cas2 psDNA complexes containing psL and psH, and the correlation patterns between site1-PAMc and site2-nonPAMc were monitored in comparison with the original psDNA system. Notably, one finds that the intersite correlation or allosteric propagation becomes significantly enhanced in the psH system, also in the half-site postcatalytic state, upon the PAMc cleavage at site1. Such correlation or allosteric propagation remains low, however, in the psL system. Hence, such studies further support that allosteric propagation from the site1-PAMc to the site2-non-PAMc can possibly impact on the acquisition efficiency, i.e., by accelerating the non-PAMc cleavage that can be otherwise rate limiting. Such allosteric effect can be particularly significant for certain psDNA sequences that tightly associate with Cas2 and Cas2'. Besides, in the allosteric propagation of Cas1-Cas2 via psDNA, a dsDNA ~23 bp (or 33 bp in full length) actually sets a ruler length for the protospacer acquisition. It is interesting to notice that such a length is right around an upper bound (25–30 bp) for DNA allostery to be effective to regulate binding cooperativity between a pair of proteins at neighboring locations on the DNA (45). Hence, such a psDNA length may indeed be optimized evolutionally to support certain cooperativity between the two active sites on Cas1-Cas2.

Our work also suggests that Cas2-Cas2' plays a key role in the allosteric regulation to support the psDNA acquisition besides its obvious structural role of bridging Cas1a-Cas1b and Cas1a'-Cas1b' and in association with psDNA. Indeed, Cas2-Cas2' appears highly stable in the Cas1-Cas2 psDNA association complex in simulation, whereas psDNA is comparatively flexible. Consistently, Cas2-Cas2' and

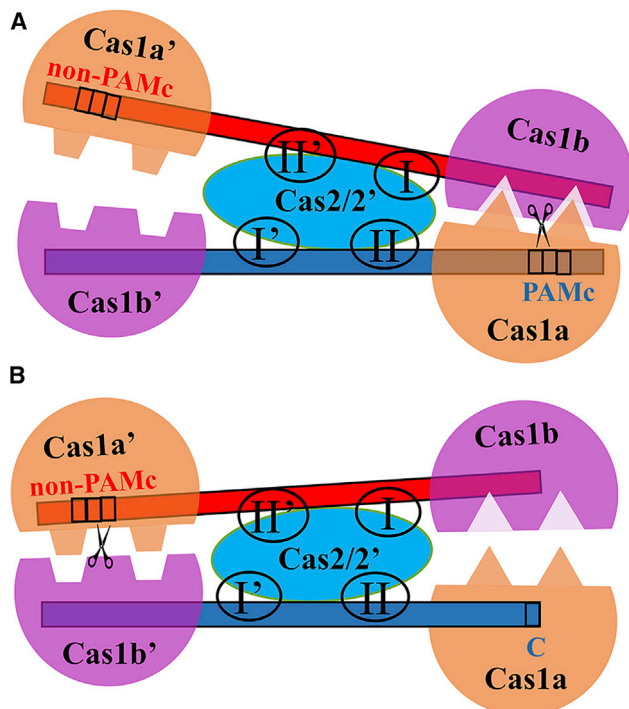


FIGURE 8 The schematics on the “seesaw” type of negative cooperativity identified between the two active sites remotely (separated by ~10 nm) on the Cas1-Cas2 protospacer acquisition complex. The psDNA is represented by red and blue strands. Cas1 and Cas2 consist of dimeric Cas1a-Cas1b, Cas1a'-Cas1b', and Cas2-Cas2', which are colored differently and labeled. (A) The preferred binding to PAMc in the first active site of Cas1-Cas2 in psDNA binding and selection. (B) The consequent non-PAMc binding in the second active site of Cas1-Cas2 immediately after the PAMc cleavage at the first active site from (A). The Cas2 and Cas2' interaction zones with the duplex psDNA are labeled (I and I' and II and II') where essential HBs formed. The interaction zones can play important roles for mediating allosteric propagation from the site1-PAMc to site2-non-PAMc. To see this figure in color, go online.

psDNA become more flexible into the half-site postcatalytic state than in the binding or precatalytic state. Such a role may also explain why Cas2 does not exhibit enzymatic activity in the acquisition process, even though it is enzymatically capable of making cleavages on single-stranded DNA or dsDNA (17,27). Notably, recent studies discovered that two Cas1 dimer alone can still form a mini-integrase that binds psDNA at a shorter length ( $\sim 18$  bp) in the absence of Cas2 (46). Without Cas2, the two sites presumably come closer on the Cas1-psDNA complex (with a distance  $\sim 6$  nm, shortened by 40% as that in the original Cas1-Cas2 system). It would be interesting to find out whether the intersite communication and cooperativity still exist in such a mini-integrase and whether efficient protospacer acquisition is still a part of the ancestral Cas1 function before Cas2 adoption (46). On the other hand, it is important to realize that Cas1-Cas2 protospacer acquisition process can also be supported by Cas3 (6), Cas4 (9), Cas9, and Csn2 in various CRISPR systems (47). Recently, it is also found that DnaQ exonuclease can process the Cas1-Cas2-loaded prespacer precursors into mature prespacers of a suitable size for integration (10). Hence, our suggested scenario of the Cas1-Cas2-efficient cleavages on the protospacer for integration may be one potential mechanism to be employed in certain conditions.

The PAM-specific DNA acquisition of Cas1-Cas2 has actually been implemented for designing molecular recording (37,48). By using the nt content, temporal ordering, and orientation of defined DNA sequences within a CRISPR array, Cas1-Cas2 seems to be able to encode arbitrary information within the genomes and has a potential to record and store DNA information for long period of time. It was in such efforts of integrating synthetic DNA sequences via Cas1-Cas2, the psDNA sequences with particularly low and high acquisition efficiency (psL and psH) were identified (37). To allow Cas1-Cas2 to identify different PAMs in the synthetic approach, many mutants of Cas1-Cas2 have also been generated in the lab directed evolution (37). The physical mechanisms revealed in this work can be further tested, for example, in such a variety of Cas1-Cas2 mutants. It is expected that by combining information from experimental synthetic approaches, computational work would reveal substantial physical mechanisms to enable further rational molecular functional redesign.

## SUPPORTING MATERIAL

Supporting material can be found online at <https://doi.org/10.1016/j.bpj.2021.06.007>.

## AUTHOR CONTRIBUTIONS

J.Y. conceived and designed this computational work. C.L. performed the computation. C.E. and L.-T.D. constructed an initial model. C.L., L.D., and J.Y. analyzed the data. J.Y. and C.L. wrote the article.

## ACKNOWLEDGMENTS

We acknowledge the computational support from the Special Program for Applied Research on Super Computation of the National Natural Science Foundation of China-Guangdong Joint Fund (the second phase) under grant no. U1501501 and from the Beijing Computational Science Research Center. This work has been supported by National Natural Science Foundation of China grants #12005029, #11775016, and #11635002. C.L. has been supported by the Start-up Founding of Chongqing University of Posts and Telecommunications (A2020-029). J.Y. has been supported by the Center of Multiscale Cell Fate Research of University of California, Irvine via Natural Science Foundation, Division of Mathematical Sciences 1763272 and the Simons Foundation grant #594598 and start-up funding from University of California, Irvine.

## SUPPORTING CITATIONS

References (49–69) appear in the Supporting material.

## REFERENCES

- Barrangou, R., C. Fremaux, ..., P. Horvath. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. 315:1709–1712.
- Brouns, S. J. J., M. M. Jore, ..., J. van der Oost. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*. 321:960–964.
- Sorek, R., C. M. Lawrence, and B. Wiedenheft. 2013. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* 82:237–266.
- Barrangou, R., and L. A. Marraffini. 2014. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell*. 54:234–244.
- Xiao, Y., S. Ng, ..., A. Ke. 2017. How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature*. 550:137–141.
- Fagerlund, R. D., M. E. Wilkinson, ..., P. C. Fineran. 2017. Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex. *Proc. Natl. Acad. Sci. USA*. 114:E5122–E5128.
- Fineran, P. C., and E. Charpentier. 2012. Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology*. 434:202–209.
- McGinn, J., and L. A. Marraffini. 2019. Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nat. Rev. Microbiol.* 17:7–12.
- Lee, H., Y. Dhingra, and D. G. Sashital. 2019. The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. *eLife*. 8:e44248.
- Kim, S., L. Loeff, ..., C. Joo. 2020. Selective loading and processing of prespacers for precise CRISPR adaptation. *Nature*. 579:141–145.
- Rollins, M. F., S. Chowdhury, ..., B. Wiedenheft. 2019. Structure reveals a mechanism of CRISPR-RNA-guided nuclease recruitment and anti-CRISPR viral mimicry. *Mol. Cell*. 74:132–142.e5.
- Hale, C. R., P. Zhao, ..., M. P. Terns. 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*. 139:945–956.
- Garneau, J. E., M. E. Dupuis, ..., S. Moineau. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. 468:67–71.
- Yosef, I., M. G. Goren, and U. Qimron. 2012. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* 40:5569–5576.
- Datsenko, K. A., K. Pougach, ..., E. Semenova. 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* 3:945.
- Swarts, D. C., C. Mosterd, ..., S. J. J. Brouns. 2012. CRISPR interference directs strand specific spacer acquisition. *PLoS One*. 7:e35888.

17. Nuñez, J. K., P. J. Kranzusch, ..., J. A. Doudna. 2014. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat. Struct. Mol. Biol.* 21:528–534.
18. Nuñez, J. K., A. S. Lee, ..., J. A. Doudna. 2015. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature*. 519:193–198.
19. Rollie, C., S. Schneider, ..., M. F. White. 2015. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *eLife*. 4:e08716.
20. Makarova, K. S., Y. I. Wolf, ..., E. V. Koonin. 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13:722–736.
21. Mohanraju, P., K. S. Makarova, ..., J. van der Oost. 2016. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science*. 353:aad5147.
22. Sternberg, S. H., H. Richter, ..., U. Qimron. 2016. Adaptation in CRISPR-cas systems. *Mol. Cell*. 61:797–808.
23. Krupovic, M., P. Béguin, and E. V. Koonin. 2017. Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr. Opin. Microbiol.* 38:36–43.
24. Makarova, K. S., Y. I. Wolf, ..., E. V. Koonin. 2020. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 18:67–83.
25. Beloglazova, N., G. Brown, ..., A. F. Yakunin. 2008. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.* 283:20361–20371.
26. Babu, M., N. Beloglazova, ..., A. F. Yakunin. 2011. A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol. Microbiol.* 79:484–502.
27. Nam, K. H., F. Ding, ..., A. Ke. 2012. Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *J. Biol. Chem.* 287:35943–35952.
28. Wang, J., J. Li, ..., Y. Wang. 2015. Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-cas systems. *Cell*. 163:840–853.
29. Nuñez, J. K., L. B. Harrington, ..., J. A. Doudna. 2015. Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature*. 527:535–538.
30. Wright, A. V., J.-J. Liu, ..., J. A. Doudna. 2017. Structures of the CRISPR genome integration complex. *Science*. 357:1113–1118.
31. Xiao, Y., J. Budhathoki, ..., A. Ke. 2018. Spacer acquisition mechanism in type II-A CRISPR system. *Acta Cryst.* 74:A444.
32. Budhathoki, J. B., Y. Xiao, ..., A. Ke. 2020. Real-time observation of CRISPR spacer acquisition by Cas1-Cas2 integrase. *Nat. Struct. Mol. Biol.* 27:489–499.
33. Horvath, P., D. A. Romero, ..., R. Barrangou. 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* 190:1401–1412.
34. Goren, M. G., I. Yosef, ..., U. Qimron. 2012. Experimental definition of a clustered regularly interspaced short palindromic duplcon in *Escherichia coli*. *J. Mol. Biol.* 423:14–16.
35. Deveau, H., R. Barrangou, ..., S. Moineau. 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190:1390–1400.
36. Mojica, F. J. M., C. Díez-Villaseñor, ..., C. Almendros. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading)*. 155:733–740.
37. Shipman, S. L., J. Nivala, ..., G. M. Church. 2016. Molecular recordings by directed CRISPR spacer acquisition. *Science*. 353:aaf1175.
38. El Hage, K., F. Hédin, ..., M. Karplus. 2018. Valid molecular dynamics simulations of human hemoglobin require a surprisingly large box size. *eLife*. 7:e35560.
39. DeLaBarre, B., J. Hurov, ..., L. Dang. 2014. Action at a distance: allostery and the development of drugs to target cancer cell metabolism. *Chem. Biol.* 21:1143–1161.
40. Sharma, N., N. Ahalawat, ..., R. Anand. 2020. Role of allosteric switches and adaptor domains in long-distance cross-talk and transient tunnel formation. *Sci. Adv.* 6:eay7919.
41. Vashisth, H., and C. F. Abrams. 2010. Docking of insulin to a structurally equilibrated insulin receptor ectodomain. *Proteins*. 78:1531–1543.
42. Duval, S., J. M. Santini, ..., B. Schoepp-Cothenet. 2016. The H-bond network surrounding the pyranopterin modulates redox cooperativity in the molybdenum-bisPGD cofactor in arsenite oxidase. *Biochim. Biophys. Acta*. 1857:1353–1362.
43. Rahman, A., and J. M. Haugh. 2017. Kinetic modeling and analysis of the Akt/mechanistic target of rapamycin complex 1 (mTORC1) signaling axis reveals cooperative, feedforward regulation. *J. Biol. Chem.* 292:2866–2872.
44. Guo, J., and H.-X. Zhou. 2016. Protein allostery and conformational dynamics. *Chem. Rev.* 116:6503–6515.
45. Kim, S., E. Broströmer, ..., X. S. Xie. 2013. Probing allostery through DNA. *Science*. 339:816–819.
46. Wright, A. V., J. Y. Wang, ..., J. A. Doudna. 2019. A functional mini-integrase in a two-protein-type V-C CRISPR system. *Mol. Cell*. 73:727–737.e3.
47. Wilkinson, M., G. Drabavicius, ..., D. B. Wigley. 2019. Structure of the DNA-bound spacer capture complex of a type II CRISPR-Cas system. *Mol. Cell*. 75:90–101.e5.
48. Shipman, S. L., J. Nivala, ..., G. M. Church. 2017. CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature*. 547:345–349.
49. Price, D. J., and C. L. Brooks, III. 2004. A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.* 121:10096–10103.
50. MacKerell, A. D., D. Bashford, ..., M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*. 102:3586–3616.
51. Wiedenheft, B., K. Zhou, ..., J. A. Doudna. 2009. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure*. 17:904–912.
52. Berendsen, H. J. C., D. V. Der Spoel, and R. Van Drunen. 1995. GROMACS - a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.* 91:43–56.
53. Abraham, M., D. van der Spoel, ..., B. Hess. 2016. GROMACS User Manual, version 5.1.2: GROMACS Development Team. <http://manual.gromacs.org/5.1.2/ReleaseNotes/index.html>.<span class="role">web
54. Joung, I. S., and T. E. Cheatham, III. 2008. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*. 112:9020–9041.
55. Joung, I. S., and T. E. Cheatham, III. 2009. Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *J. Phys. Chem. B*. 113:13279–13290.
56. Lindorff-Larsen, K., S. Piana, ..., D. E. Shaw. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. 78:1950–1958.
57. Guy, A. T., T. J. Piggot, and S. Khalid. 2012. Single-stranded DNA within nanopores: conformational dynamics and implications for sequencing: a molecular dynamics simulation study. *Biophys. J.* 103:1028–1036.
58. Da, L.-T., C. E., ..., J. Yu. 2017. T7 RNA polymerase translocation is facilitated by a helix opening on the fingers domain that may also prevent backtracking. *Nucleic Acids Res.* 45:7909–7921.
59. Long, C., C. E., ..., J. Yu. 2019. Determining selection free energetics from nucleotide pre-insertion to insertion in viral T7 RNA polymerase transcription fidelity control. *Nucleic Acids Res.* 47:4721–4735.

60. Darden, T., D. M. York, and L. G. Pedersen. 1993. Particle mesh Ewald: an  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
61. Essmann, U., L. Perera, ..., L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.
62. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.
63. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.
64. Nosé, S., and M. Klein. 1983. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* 50:1055–1076.
65. Li, S., W. K. Olson, and X.-J. Lu. 2019. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* 47:W26–W34.
66. Motlagh, H. N., J. O. Wrabl, ..., V. J. Hilser. 2014. The ensemble nature of allostery. *Nature.* 508:331–339.
67. Yu, J., T. Ha, and K. Schulten. 2007. How directional translocation is regulated in a DNA helicase motor. *Biophys. J.* 93:3783–3797.
68. Dai, L., H. Flechsig, and J. Yu. 2017. Deciphering intrinsic inter-subunit couplings that lead to sequential hydrolysis of F1-ATPase ring. *Biophys. J.* 113:1440–1453.
69. Williams, N. 2004. DNA hydrolysis: mechanism and reactivity. *In Artificial Nucleases.* M. A. Zenkova, ed. Springer, pp. 3–17.