


Binding to the conserved and stably folded guide RNA pseudoknot induces Cas12a conformational changes during ribonucleoprotein assembly

Received for publication, September 19, 2022, and in revised form, April 5, 2023. Published, Papers in Press, April 12, 2023,

<https://doi.org/10.1016/j.jbc.2023.104700>

Sruthi Sudhakar^{1,‡}, Christopher L. Barkau^{2,‡} , Ramadevi Chilamkurthy², Halle M. Barber³, Adrian A. Pater⁴, Sean D. Moran⁴ , Masad J. Damha³ , P. I. Pradeepkumar^{1,*}, and Keith T. Gagnon^{2,4,*} 

From the ¹Department of Chemistry, Indian Institute of Technology Bombay, Mumbai, India; ²Department of Biochemistry and Molecular Biology, School of Medicine, Southern Illinois University, Carbondale, Illinois, USA; ³Department of Chemistry, McGill University, Montreal, Quebec, Canada; ⁴Department of Chemistry and Biochemistry, Southern Illinois University, Carbondale, Illinois, USA

Reviewed by members of the JBC Editorial Board. Edited by Craig Cameron

Ribonucleoproteins (RNPs) comprise one or more RNA and protein molecules that interact to form a stable complex, which commonly involves conformational changes in the more flexible RNA components. Here, we propose that Cas12a RNP assembly with its cognate CRISPR RNA (crRNA) guide instead proceeds primarily through Cas12a conformational changes during binding to more stable, prefolded crRNA 5' pseudoknot handles. Phylogenetic reconstructions and sequence and structure alignments revealed that the Cas12a proteins are divergent in sequence and structure while the crRNA 5' repeat region, which folds into a pseudoknot and anchors binding to Cas12a, is highly conserved. Molecular dynamics simulations of three Cas12a proteins and their cognate guides revealed substantial flexibility for unbound apo-Cas12a. In contrast, crRNA 5' pseudoknots were predicted to be stable and independently folded. Limited trypsin hydrolysis, differential scanning fluorimetry, thermal denaturation, and CD analyses supported conformational changes of Cas12a during RNP assembly and an independently folded crRNA 5' pseudoknot. This RNP assembly mechanism may be rationalized by evolutionary pressure to conserve CRISPR loci repeat sequence, and therefore guide RNA structure, to maintain function across all phases of the CRISPR defense mechanism.

Conformational changes are the basis of protein and RNA folding from long biopolymers into biologically active molecules (1). The affinity, specificity, and stability of molecular interactions often scales with buried surface area (or number of molecular contacts) and the degree of surface complementarity (2). For RNA-guided enzymes, coevolution of protein and cognate RNA usually results in burial of significant portions of the RNA in the protein (3–7). Thus, conformational accommodations, primarily for the guide RNA, are an important mechanism for molding and shaping interactions.

Archaeal box C/D ribonucleoprotein (RNP) complexes, for example, have been characterized to undergo substantial remodeling and induced fit during RNP assembly (4, 8–11). Argonaute enzymes undergo minimal, although potentially important, conformational changes while the guide RNA is stretched and splayed through the middle of the protein (7, 12, 13).

CRISPR were first observed in bacteria in 1987 (14) and 1989 (15) as conserved repeat sequences with regularly spaced interrupting sequences. However, the significance of CRISPR loci remained unclear. In 2002, when these repeats were observed in other prokaryotes and CRISPR associated (Cas) protein-coding genes were found nearby, a unifying name of CRISPR was coined for these loci (16). It was later found that CRISPR repeats contained sequences that correspond to bacteriophage, suggesting an origin for the repeats (17–19). In 2006, it was suggested that these repeats act as an adaptive immune system to protect against viral invaders (20). This hypothesis was subsequently shown to be correct (21, 22). The study of CRISPR function and evolution quickly followed, leading to characterization of diverse RNA-guided CRISPR-Cas systems (23).

CRISPR systems act in three phases: adaptation, expression, and interference. During adaptation and expression, foreign DNA sequences are incorporated into CRISPR loci and then expressed as RNA for assembly into CRISPR-Cas RNP effector complexes (24, 25). Cas12a enzymes originate from type V CRISPR systems comprising the genes *cas12a*, *cas4*, *cas1*, *cas2*, and the repeat array (25). The conserved Cas1/Cas2 complex functions in spacer acquisition (26, 27). The adaptation complex comprises dimers of Cas2, which bridge pairs of Cas1 dimers, and binds potential protospacers to coordinate their incorporation into the CRISPR array. Cas1 is thought to have originated from transposon-like elements (termed casposons) which encode Cas1 homologs. Diverse Cas1 homologs exhibit low sequence conservation but retain a similar structure across the Cas1 phylogeny (28). Recently it has been shown that the Cas1/Cas2 complex in the type I-B system of *Haloarcula hispanica* binds the upstream-most repeat and leader at motifs

[‡] These authors contributed equally to this work.

* For correspondence: Keith T. Gagnon, ktgagnon@siu.edu; P. I. Pradeepkumar, pradeep@chem.iitb.ac.in.

Cas12a conformational dynamics during RNP assembly

of the CRISPR locus in a sequence and spacing-specific manner to facilitate new spacer integration (29). Cas12a has been shown to process its own CRISPR RNA (crRNA), cleaving upstream of the repeat sequence, which is a stem-loop structure referred to as the 5' handle. These steps all rely on recognition of the CRISPR array repeat sequence, either at the DNA or RNA level, which serves as a common thread across adaptation, expression, and interference phases of the CRISPR defense mechanism. This offers a rationale for strong conservation of repeat sequences in CRISPR array loci. In contrast, other components, like the Cas12a protein, may be permitted to undergo greater divergence so long as they still recognize CRISPR arrays or crRNA and execute their respective phases.

During the interference phase, RNA-guided CRISPR-Cas effector enzymes find and pair to complementary nucleic acids, most often DNA, and catalyze phosphodiester bond cleavage to induce degradation (25, 30, 31). CRISPR effector enzymes are composed of one or two RNA molecules and an endonucleolytic Cas protein (16, 32, 33). During interference, CRISPR-Cas complexes will first identify candidate targets by transient Cas protein binding to a protospacer adjacent motif (PAM) DNA sequence (31). The crRNA guide sequence will then pair to the complementary protospacer sequence adjacent to the PAM (34). The class 2, type V CRISPR systems use a single effector enzyme, Cas12a, and a single 42 nucleotide long crRNA guide (35). Cas12a appears to have low off-target effects in gene editing applications (36) and possesses a T-rich PAM sequence requirement (35). Interestingly, CRISPR-Cas12a complexes exhibit a nonsequence-specific, multiturn-over single-stranded DNase activity, called *trans* activity, after sequence-specific cleavage of target DNA has occurred (37, 38). For this reason, Cas12a has been used to develop diagnostic assays, such as for COVID-19 detection (38). In addition, CRISPR-Cas12a is being adapted for human therapeutics (39–41).

Conformational changes that occur during CRISPR-Cas RNP assembly and their effects on enzyme activity have been investigated previously. The Doudna laboratory observed global structural changes during RNA binding to Cas9 (42, 43) that coincide with catalytic competence and have been supported by cryo-EM studies (44). Other studies have investigated RNP assembly and conformational dynamics for Cas9 (42, 45, 46). Several Cas enzyme crystal structures have been solved (5, 6, 47–51). Recent cryo-EM studies of Cas9 during precatalytic or target binding steps have revealed some new conformational states, yet these involve substrate DNA, which is difficult to uncouple from guide RNA-induced changes (44, 52). These structures provide important insight into CRISPR-Cas9 structure-function, however the full picture of RNP assembly dynamics may still be incomplete. Cas12a dynamics during RNP assembly with its cognate guide have been less explored, owing in part to the historical lack of available high-resolution structures of Cas12a in the absence of a guide RNA (53, 54). Cas12a dynamics studies have identified large-scale “open” and “closed” conformational states upon crRNA binding that appear to depend on the originating species of the

CRISPR-Cas12a system (51, 53, 55). These studies focused on one Cas12a and large-scale global structural changes with target DNA substrate binding (55, 56).

In this study, we sought to investigate Cas12a RNP assembly through protein and RNA sequence and structural alignments, phylogenetics, molecular dynamics (MD) simulations, and spectroscopic and biochemical experiments. We find that Cas12a sequence diverged during evolution, although protein structure was relatively well-conserved in the RNP state. In contrast, the sequence and structure of the crRNA 5' handle remained highly conserved. MD simulations of *Acidaminococcus* sp. strain BV3L6 (As) Cas12a, as well as Cas12a from *Lachnospiraceae* bacterium ND2006 (Lb) and *Francisella tularensis* subsp. *novicida* (Fn), identified substantial protein structural dynamics that were reduced upon RNA binding, especially in the Wedge domain involved in RNA and DNA binding. In contrast, the crRNA 5' handle structure showed minimal variation between free and bound states. In-solution limited proteolysis followed by MS identified AsCas12a lysine residues with reduced accessibility after RNP assembly. Differential scanning fluorimetry showed improved thermal stability and folding cooperativity of AsCas12a as an RNP. For the crRNA 5' handle, thermal denaturation and CD analyses suggested stable independent folding in the free state. Inter-domain angle calculations from MD simulations predict little or no sampling of the RNP conformational state in apo-Cas12a, suggesting an induced fit over conformational selection model. Together, these results indicate that Cas12a protein undergoes substantial conformational rearrangements induced by binding to the largely prefolded crRNA 5' pseudoknot structure. We propose that this assembly mechanism may be common for CRISPR-Cas systems due to the intrinsic need to conserve CRISPR loci repeat sequences and therefore crRNA structures.

Results

To understand RNP assembly across CRISPR-Cas12a systems, we began by performing a maximum likelihood phylogenetic analysis on an alignment of 147 publicly available Cas12a protein sequences. In our resulting consensus tree (Fig. 1A), most positions in the alignment were poorly conserved (~39% on average at positions homologous to the AsCas12a homolog), leading to low fine-grain phylogenetic resolution. Three of the Cas12a proteins in the alignment have been structurally characterized previously, AsCas12a (6), FnCas12a (57), and LbCas12a (51). The structures of these homologs are shown in Figure 1B with their placement on the phylogenetic tree indicated. The resulting topology strongly suggests that among these three Cas12a homologs, FnCas12a and LbCas12a are more similar to each other than to AsCas12a. Overall, their sequence conservation is moderate based on alignment and BLAST (Fig. S1), with sequence identity/similarity being 40%/59% for LbCas12a to FnCas12a, 34%/50% for LbCas12a to AsCas12a, and 35%/51% for AsCas12a to FnCas12a (35, 51). Their structure, by contrast, is well-conserved, with pairwise RMSD values calculated at

Cas12a conformational dynamics during RNP assembly

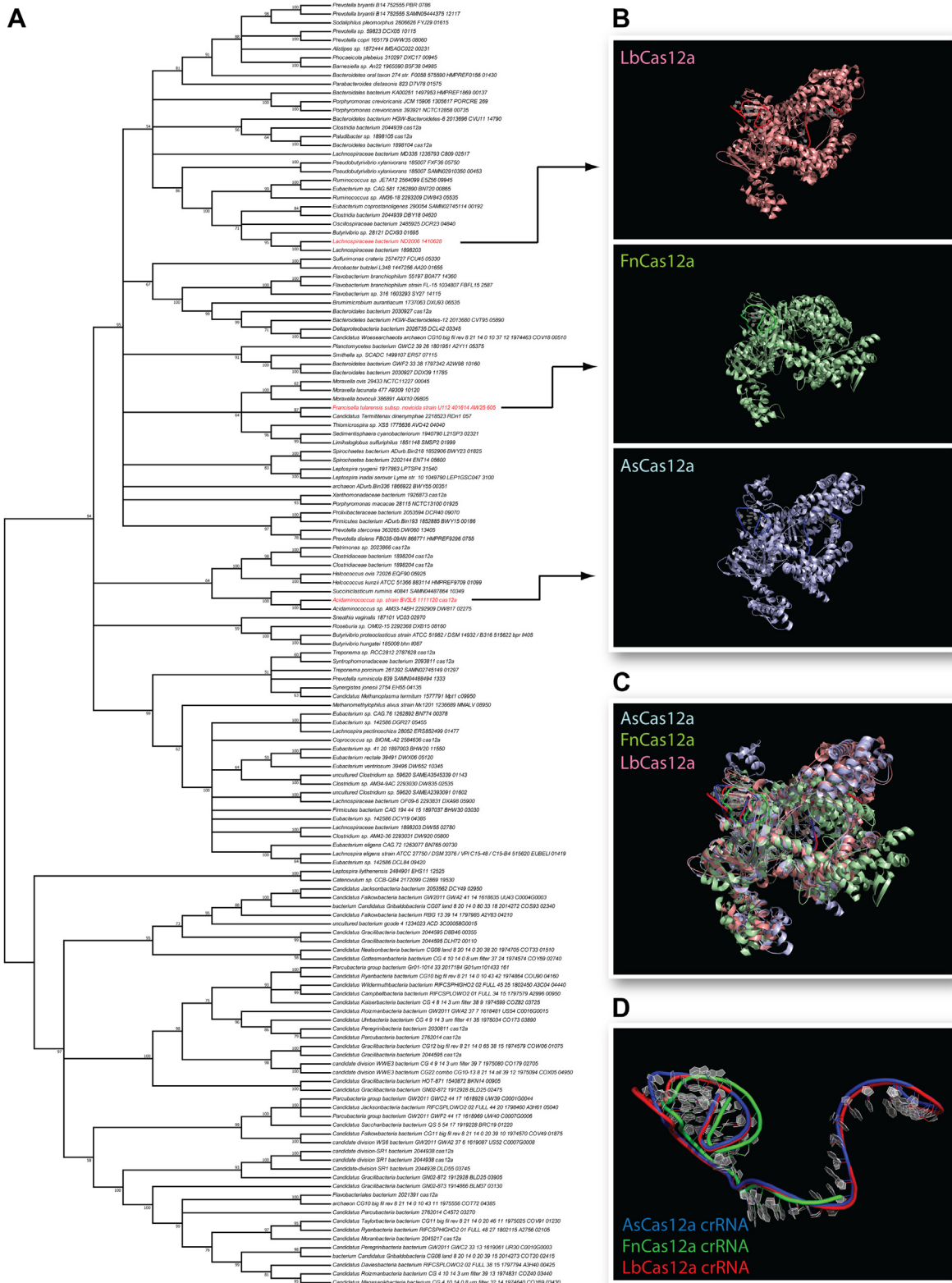


Figure 1. Sequence and structural relationships among Cas12a homologs. A, consensus maximum likelihood tree of available Cas12a protein sequences. Structurally characterized homologs are labeled in red. B, individual crystal structures of Lb, Fn, and AsCas12a corresponding to their location on the tree shown in panel A. Structural alignment of As (PDB ID: 5B43), Fn (PDB ID: 5NG6), and Lb (PDB ID: 5XU5) Cas12a proteins (C) and crRNAs (D). As, *Acidaminococcus* sp; Fn, *Francisella tularensis* subsp. *Novicida*; Lb, *Lachnospiraceae* bacterium.

Cas12a conformational dynamics during RNP assembly

2.56 Å for LbCas12a to FnCas12a, 2.20 Å for LbCas12a to AsCas12a, and 2.78 Å for AsCas12a to FnCas12a when assembled as RNP complexes (FnCas12a) or ternary complexes with guide RNA and target DNA (LbCas12a and AsCas12a) (Fig. 1C). These results suggest that the structure-function of Cas12a RNPs is conserved despite substantial protein sequence divergence.

RMSD values calculated between the crRNAs of each homolog revealed striking structural similarity (Fig. 1D), with the distances of AsCas12a to LbCas12a and FnCas12a crRNAs being 0.12 Å and 0.17 Å, respectively. The distance between

LbCas12a and FnCas12a crRNAs was 0.21 Å. We also aligned repeat sequences from 28 type V CRISPR systems. In contrast to the Cas12a proteins, these sequences were highly conserved in the pseudoknot folding repeat region (35) (Fig. 2, A and B). The relative distance relationship observed among LbCas12a, FnCas12a, and AsCas12a proteins (Fig. 1A) was recapitulated in the topology of the nearest-neighbor tree of their cognate repeat crRNA sequences constructed by LocARNA (58). Secondary structure analysis of this alignment using LocARNA, which does not predict pseudoknots, revealed a conserved hairpin structure (35) (Fig. 2B) consistent with the core of the

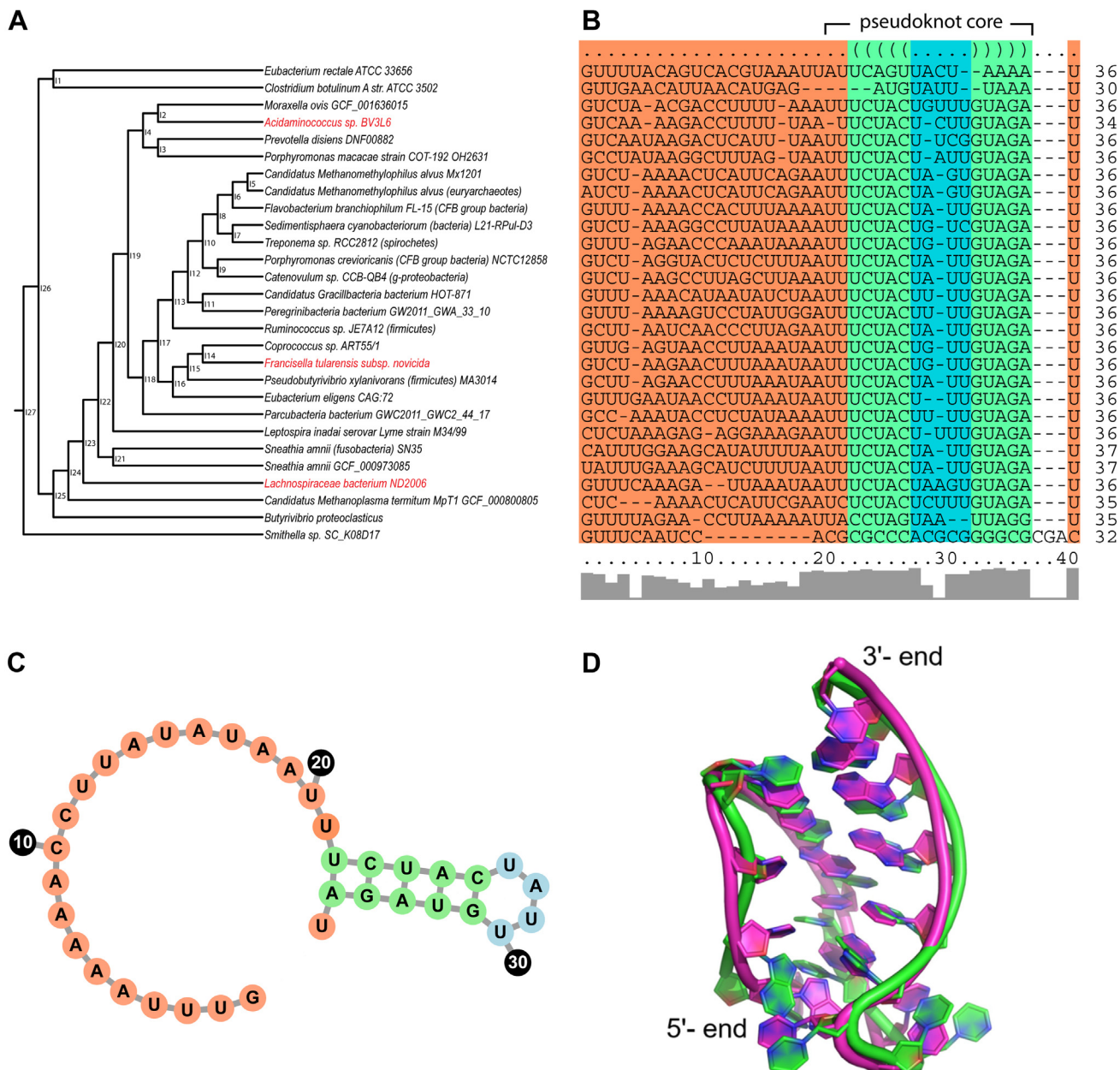


Figure 2. Conservation of Cas12a-associated repeat sequence and folding. A, nearest neighbor joining tree and (B) alignment of 28 type V CRISPR repeats. The LocARNA-predicted secondary structure (right) is indicated at the top of the sequence alignment. The 5' leader sequence is highlighted in orange and the stem and loop of the predicted stem-loop structure is highlighted in green and blue, respectively. C, repeat consensus sequence shown with the secondary structure predicted for the alignment in (A) with the same color coding. D, superimposed image of the representative structures of the major cluster in the free pseudoknot (magenta) and AsCas12a-bound pseudoknot (green) in the AsCas12a RNP complex from 1 μs molecular dynamics simulation trajectories. RNP, ribonucleoprotein.

5' pseudoknot observed in solved Cas12a structures. The predicted structure is shown with the consensus sequence of the aligned repeats (Fig. 2C). These results suggest that the sequence and structure of the crRNA 5' handle has remained remarkably stable.

To gain insights into the structural dynamics of Cas12a-crRNA RNP assembly, MD studies were conducted. Three different systems were initially prepared—AsCas12a protein alone (apo-AsCas12a), the AsCas12a RNA pseudoknot alone (free pseudoknot), and AsCas12a bound to its cognate crRNA (AsCas12a RNP)—and then subjected to 1 μ s of an unrestrained production run in the graphics processor unit accelerated version of Particle Mesh Ewald Molecular Dynamics (59–61) in AMBER 18 (<https://ambermd.org/>) (62).

The analysis of trajectories was carried out in two stages. The pseudoknot structural stability and flexibility were explored in the first stage and the protein structural stability and flexibility were examined in the second stage. The RMSD of the pseudoknot backbone atoms were calculated using the first simulation frame as reference (Fig. S2A). The RMSD plot indicates minimal deviations between the free pseudoknot alone and the pseudoknot in the RNP. This suggests that the simulations are equilibrated and the pseudoknot is stable throughout the 1 μ s production run. The RMSD value is slightly higher in the case of free pseudoknot, mainly due to the fraying of the terminal nucleotides and the absence of interactions with the protein. This is further demonstrated by the RMS fluctuations (RMSF) values (Fig. S2B). The RMSF plot shows that the residues exhibit similar fluctuations, with nearly identical values in the core hairpin of the structure, indicating that the pseudoknot behaves similarly in the free and bound forms.

The end-to-end distances in the pseudoknot can be used to report structural rigidity and stability during the simulation. The terminal nucleotides were not used for the measurement to avoid possible fluctuations due to the terminal fraying of base pairs. When the end-to-end distances were calculated, the values showed negligible deviations in the free and bound forms (Fig. S2C). Thus, the rigidity of the pseudoknot is maintained in both free and bound forms. The 1 μ s simulation trajectories were clustered into ten ensembles. The representative structures of the major cluster of the free and bound pseudoknot were then superimposed to visualize structural fluctuations (Fig. 2D). The structures superimposed with an RMSD value less than 2.5 Å, confirming that the pseudoknot does not undergo any significant structural changes when bound to the Cas12a protein.

In the second stage of analysis, the RMSD of the backbone atoms of AsCas12a protein were measured with respect to the initial frame of the simulation (Fig. S3A). The equilibration of the simulations was observed from the RMSD plots, which showed marginal differences in the protein structure alone without the crRNA and in the RNP complex. The RMSF plots indicate that several regions, which include residues ranging from 150 to 300 and 750 to 85 show significant deviations between the protein alone *versus* in the RNP complex (Fig. S3B). These regions include parts of the PI, wedge, and

REC domains. A previous study reported increased occupancy of a “closed” high fluorescence energy transfer (FRET) state for FnCas12a upon crRNA binding, as indicated by REC and NUC domain distance measurements *via* single molecule FRET (55). The RMSF values for the protein in the RNP complex were less than the protein alone without the pseudoknot, indicating that the protein becomes more ordered in the presence of the crRNA. The distance between the C α atoms of amino acids in close vicinity of the pseudoknot were calculated to identify any rearrangements in the protein that might help accommodate the crRNA pseudoknot. We selected K15-L807, R863-L1022, and H977-D966 distances as representative since they are all in the wedge and RuvC domains and make close contact with either the loop or the terminals of the pseudoknot (Fig. 3A). The distances varied between the systems, and the protein in RNP showed smaller distance fluctuations than the protein alone, especially for K15-L807 and R863-L1022, which suggests conformational sampling by AsCas12a protein to accommodate the crRNA pseudoknot during RNP assembly (Fig. 3, B–D).

The 2D-RMSD values were calculated for each protein domain to detect the regions of high protein mobility. The values clearly showed that the REC domain and WED-III domain showed significant differences between the two systems. The 2D-RMSD values of the wedge domain in the protein without pseudoknot RNA displayed values as high as 10 Å, while the values remained <5 Å in the case of the RNP (Fig. S4, A and B). This confirms our previous observation from the RMSF values that the protein becomes more stabilized in the presence of the crRNA. The representative structures of the major clusters (Fig. 4, A and B) were visually inspected to validate the variations seen in the 2D-RMSD plots. The WED-III domain and REC domain rearranged to incorporate the RNA into the RNP complex. The pseudoknot rearranges the WED-III domain while the single-stranded guide region rearranges the REC domain in the RNP complex for enhanced binding.

To identify the major mode of domain motion (maximum variance), principal component (PC) analysis was carried out (Fig. 4, C and D). The translational motion was initially removed by fitting the trajectory to an average structure. The projection diagrams and pseudo trajectories from PC analysis were analyzed to obtain a pattern of the major movements resulting in the conformational preferences discussed in the cluster analysis. In the protein only system, the REC2 and the WED-III directions allow them to occupy some of the space where the RNA is usually bound. Compared to the RNP complex, the REC2 and WED-III domains show movement directed toward creating space to accommodate the crRNA pseudoknot and guide spacer sequence. The PI and REC1 domains in the RNP complex exhibit a circular motion with one part slightly moving away from the cleft, while the smaller helix moves toward the cleft, which would subsequently bind the DNA target for catalysis (56). This motion is not exactly the same in the absence of crRNA. The amplitude of motion shows minimal variations in domains other than the PI in the first PC. The histogram of the first and second PC modes of

Cas12a conformational dynamics during RNP assembly

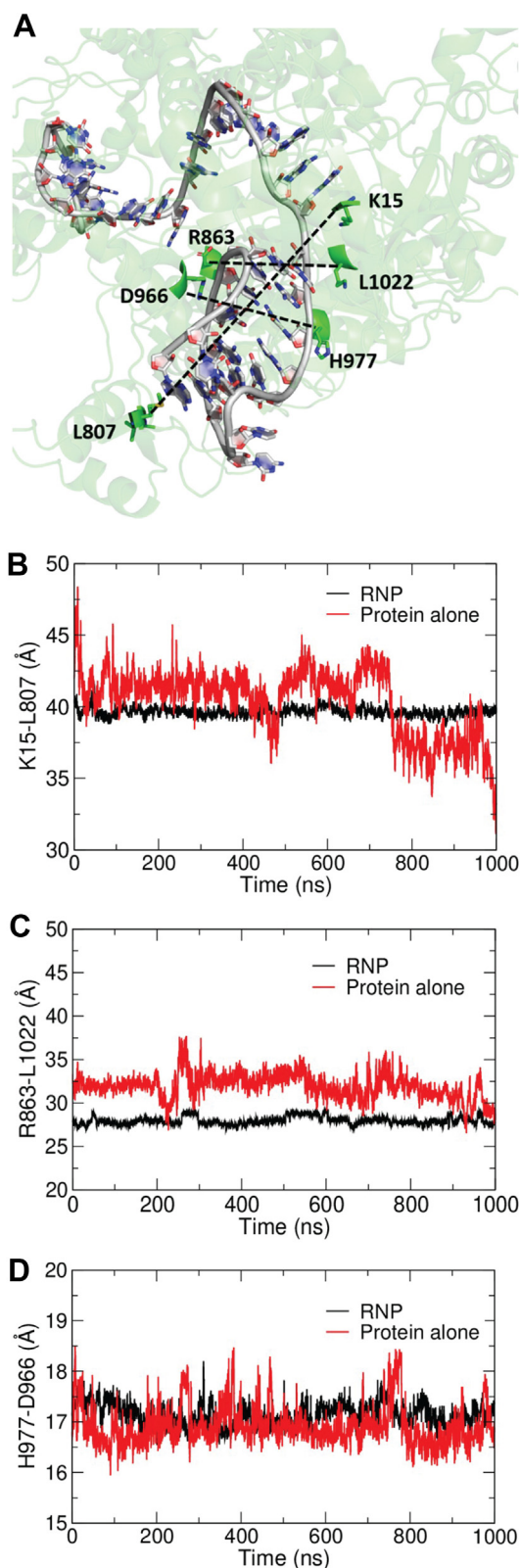


Figure 3. Molecular dynamics simulations indicate that AsCas12a is structurally less dynamic and less solvent accessible when bound to its cognate crRNA. A, representative image illustrating the amino acids considered for distance measurements during molecular dynamic simulations. The distance between the C α atoms of (B) K15 and L807, (C) R863 and L1022, and (D) H977 and D966 from 1 μ s trajectories. The running averages of the distances are represented in the plot.

the RNP and protein alone (Fig. S5A) shows no convergence suggesting the variation in conformational preferences. Though the deviation is not high, the regions spanned by the PC mode 1 and mode 2 of the apo protein and RNP complex are different (Fig. S5B).

Solvent accessible surface area (SASA) of the protein was calculated from the trajectories of protein alone and the RNP complex simulations (Fig. S6). The Δ SASA was found to be ~ 4000 \AA^2 , with the RNP complex having a less value, suggesting the protein becomes more ordered and compact upon RNP assembly. The rearrangements in the protein to accommodate RNA were explored further by comparing the surface electrostatic potential images of the protein (Fig. S7). The positively charged regions are rearranged in the RNP complex, likely to ensure better interaction with the negatively charged backbone of the RNA.

To rule out potential bias by using a manually generated model of apo-AsCas12a from a DNA-bound ternary structure, MD simulations were also carried out on apo form deposited in the AlphaFold database. Though the RMSF values were slightly lower than the manually generated structure, the overall trend in the structural properties (RMSD, RMSF, distances, and SASA) remained the same in both models (Figs. S8–S11).

MD simulations on FnCas12a (PDB ID: 5NG6) and LbCas12a (PDB ID: 5ID6) RNP, apoprotein, and pseudoknot were also carried out to determine whether our observations for AsCas12a might apply more broadly to other CRISPR–Cas12a complexes. For both FnCas12a and LbCas12a, the RNA pseudoknot was stable throughout the simulations with similar values of RMSF (Fig. S12) and end-to-end distances (Fig. S13) in both the free and bound forms. Due to terminal and loop nucleotide fraying, the RMSD is slightly higher for the free forms in both cases (Fig. S14). The major clusters of the free and bound form of pseudoknot superimposed with an RMSD value < 2.7 \AA and < 2.6 \AA in FnCas12a (Fig. S15A) and LbCas12a (Fig. S15B) complexes, respectively. Similar to AsCas12a, the pseudoknot structure also remained stable in these complexes.

The protein alone was more dynamic than the RNP complex for both FnCas12a and LbCas12a, with a higher value for the latter. In the FnCas12a system, the RMSF plots revealed more deviations in the wedge, PI, and RuvC domains, similar to what was observed in AsCas12a (Fig. S16A). The domain designations are different in LbCas12a but greater deviations in the RMSF values were found in the oligonucleotide binding domain (OBD) and RuvC domain (Fig. S16B). These results show that the domains responsible for interacting with RNA exhibit variation in dynamics during RNP assembly. The distance analysis between the C α atoms of the amino acids in the immediate vicinity of the RNA pseudoknot revealed that the Cas12a protein had undergone some rearrangement to accommodate the pseudoknot in both FnCas12a and LbCas12a complexes (Fig. S17). While the values are constant in the RNP complexes, variations in the apo form were observed that indicate a compaction of the protein upon RNA binding.

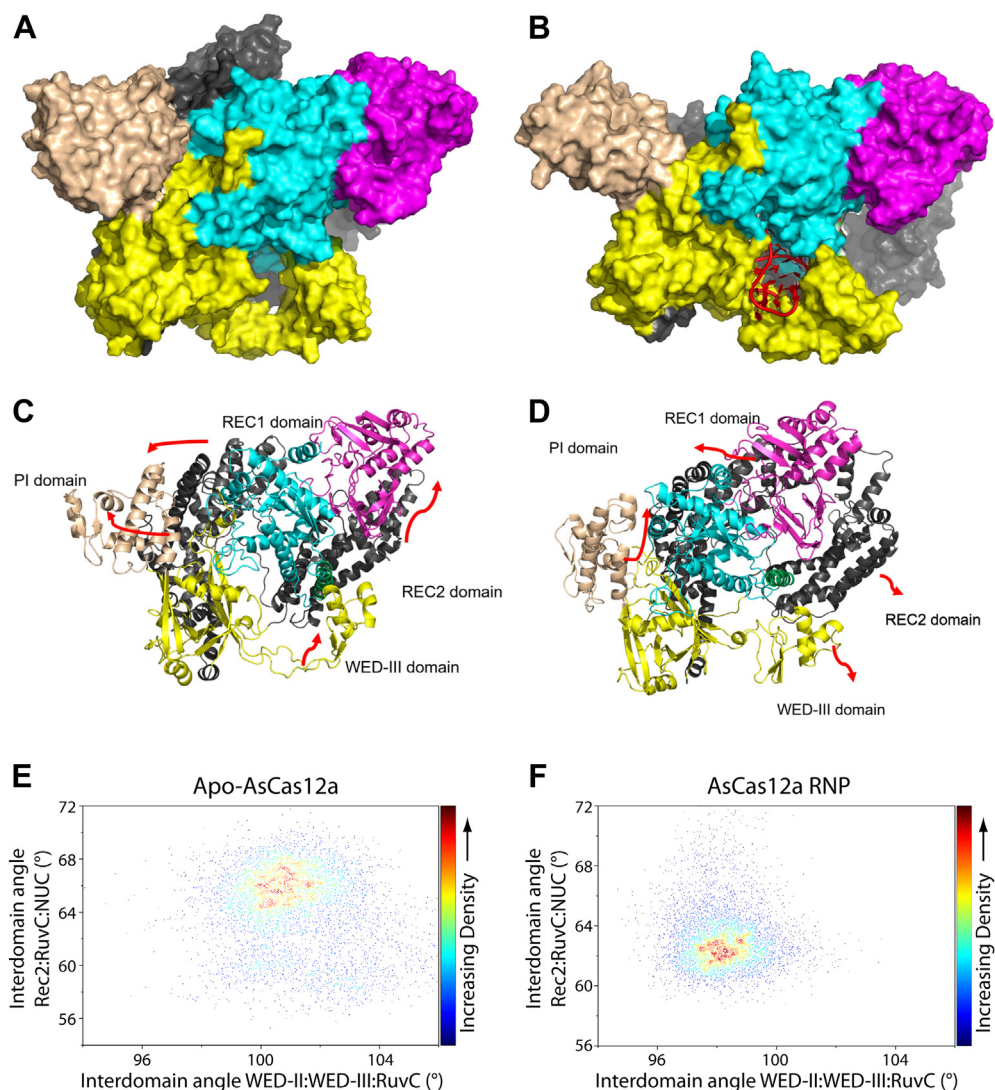


Figure 4. Molecular dynamics simulations predict distinct dynamics and domain motion for AsCas12a protein versus AsCas12a RNP. Representative snapshots of the major clusters of (A) AsCas12a protein system and (B) AsCas12a RNP complex from the 1 μ s trajectories. The protein is shown in surface representation. The PI domain is represented in *beige color*, the wedge domain in *yellow*, RuvC domain in *cyan*, REC domain in *gray*, and the Nuc domain in *magenta*. The pseudoknot is represented as cartoon in *red*. Representative images of the principal motion of protein in (C) AsCas12a protein alone and (D) AsCas12a RNP complex. Domain colors are the same as in panels A and B except the bridge helix (BH) is shown in *green*. The *arrows* are marked according to the pseudo trajectory and do not indicate the amplitude of motion. The 3D perspective of the domain movements might not be clearly visible in the given representation. Density plots of the interdomain angles in (E) apo-AsCas12a protein alone or (F) AsCas12a RNP. The density plot represents the sampling of the angle and the maximum population is represented with the highest density. RNP, ribonucleoprotein.

The superimposition of major clusters of Fn and Lb Cas12a in the apo and RNP complexes revealed changes in the domain indicated by the RMSF values. While the WED-III domain loop showed maximum variation in FnCas12a (Fig. S18A), the OBD, UK, and RuvC domains showed high fluctuation in LbCas12a (Fig. S18B). Irrespective of the Cas12a type, the rearrangement is nearly identical around the RNA pseudoknot. Further, the electrostatic surface potential images also confirmed the rearrangement of positively charged residues to better accommodate the negatively charged backbone of the RNA pseudoknot (Fig. S19). In FnCas12a and LbCas12a, the RNP assembly results in about a 5% reduction in the overall SASA (Fig. S20). Overall, MD studies of FnCas12a and LbCas12a led to results quite similar to those of AsCas12a.

To assess whether conformational changes associated with the Cas12a RNP assembly are a result of induced fit or conformational selection, two related but distinct binding modes, interdomain angles from AsCas12a simulations were sampled (63). The RNA pseudoknot and single-stranded portion of the guide RNA are near the Wedge, REC, RuvC, and NUC domains. Therefore, the orientation of these domains might differ in the apo form and RNP complex. The interdomain angles between WED-II: WED-III: RuvC (Fig. S21A) and REC2: RuvC: NUC (Fig. S21B) were calculated (RuvC-I and RuvC-II were considered together). The density plots with these angles as coordinates revealed that the sampling differs in the apo and RNP forms (Fig. 4, E and F). The apo form exhibits an open structure with a large interdomain angle, while the RNP exhibits a closed structure with a slight

Cas12a conformational dynamics during RNP assembly

interdomain angle. The angle associated with the closed form is only seen in the RNP and not present in the sampling of the apo-protein. The cluster distribution of the apo and RNP also confirmed the same. The major clusters of the apo and RNP forms show high RMSD upon superimposing with each other and no representative structures from the clustering are similar in both of these forms (Fig. 4, A and B). Thus, the closed form is associated only with the RNP and the apo form does not exhibit the flexibility to adopt open/closed conformations. These results suggest that Cas12a RNP assembly proceeds through induced fit rather than conformational selection since the necessary RNP conformation is essentially never sampled in the apo form (63).

To experimentally investigate the structural flexibility of AsCas12a, we subjected the free protein or assembled RNP to limited trypsin proteolysis. The presence of crRNA appeared to reduce cleavage efficiencies at three sites (Fig. 5A). Mass spectrometric analysis of these bands revealed that they corresponded to protection from trypsin proteolysis at K739, K570, and K576 (Fig. S22). Of these, K570 and K576 are found in the OBD-II, which is involved in crRNA and target DNA binding (51) (Fig. 5B). To observe global effects on AsCas12a conformation and stability during RNP assembly, we performed differential scanning fluorimetry (DSF) on AsCas12a with and without crRNA. This method uses a dye specific for protein rather than UV absorbance at 280 nm, eliminating potentially confounding UV signal overlap from the RNA. AsCas12a alone exhibited a biphasic melt profile, suggesting at least two structures that largely unfolded independently (64). Their melting temperature (T_m) values were calculated at $43.10\text{ }^\circ\text{C} \pm 0.01\text{ deg. C}$ and $48.62\text{ }^\circ\text{C} \pm 0.02\text{ deg. C}$. When crRNA was present, AsCas12a unfolded cooperatively as an apparent single structure with a T_m of $46.15\text{ }^\circ\text{C} \pm 0.04\text{ deg. C}$ (Fig. 5C). To assess the unfolding model that best fits the data, we performed fits to a two-state folded–unfolded model and to a three-state sequential folded–intermediate–unfolded model (Fig. S23). While this model fitting is not sufficient to provide a detailed understanding of the protein folding landscape, unfolding *via* a structural intermediate in the three-state model fit the data best. This suggests that significant structural or stability changes to Cas12a occurred upon RNP assembly.

To determine the stability and folding of the crRNA pseudoknot in solution, we performed UV-monitored thermal denaturation and CD analyses. The RNA pseudoknot sequence is composed of the 19 nucleotides of the AsCas12a crRNA 5' end. Control DNA and RNA hairpin sequences were created by replacing the three 5' terminal nucleotides with cytidines to prevent pseudoknot formation (Fig. 5D). Thermal denaturation revealed that the pseudoknot unfolded similarly to the control, model hairpin RNA sequence, with similar T_m values of $61.8\text{ }^\circ\text{C} \pm 0.3\text{ deg. C}$ and $60.7\text{ }^\circ\text{C} \pm 0.3\text{ deg. C}$, respectively (Fig. 5E). The slight increase in T_m for the pseudoknot may be attributable to additional base-pairing interactions in the structure (65). As expected, the DNA hairpin appeared to possess a less stable structure (65), with a T_m of $48.3\text{ }^\circ\text{C} \pm 0.6\text{ deg. C}$. When we performed CD analyses, we observed nearly

identical spectra for the pseudoknot RNA and hairpin RNA, characteristic of A-form structure (66), with a small shift in the 250 to 300 nm range (Fig. 5F). The DNA hairpin generated a distinct spectra indicative of B-form structure (66). These results point to a well-formed crRNA pseudoknot in solution independent of protein binding.

Discussion

Our results indicate that Cas12a proteins undergo conformational rearrangements to accommodate crRNA binding during RNP assembly. In contrast, the crRNA 5' pseudoknot, the primary “handle” that Cas12a binds, appears to undergo very little structural alterations. The highly conserved nature of the crRNA repeat sequence, corresponding to the 5' pseudoknot handle, and its interconnected role across adaptation, expression, and interference phases of CRISPR mechanism, implies that alterations to sequence or structure would likely be deleterious to the system. Cas12a proteins, on the other hand, appear to evolve much more rapidly, sustaining far more sequence changes, and to a lesser extent structural alterations, through evolution. In-solution studies demonstrated conformational changes of AsCas12a upon crRNA binding while the crRNA pseudoknot handle itself appeared quite stable alone, in agreement with MD simulations.

RNA-guided enzymes are a unique class of enzymes that share an intimate relationship with their cognate RNA partners. The protein component is typically inactive when not associated with its RNA guide (67). The guide RNA specifies the substrate through Watson–Crick hybridization and correctly positions it into the protein's active site (7, 68). The guide sequence can vary, allowing a single type of protein to catalyze the same reaction on countless targets of varying sequence. Since RNA-guided enzymes can be programmed to target different sequences, understanding the rules of RNP assembly is important for understanding how these systems evolved and how to better leverage them for biotechnology and biomedical applications.

For engineering Cas12a enzymes, it is important to not introduce modifications that can significantly perturb the folding or stability of the pseudoknot structure. A recent study investigated a diverse class of chemical modifications in the pseudoknot and indeed concluded that smaller modifications known to maintain RNA structure, such as 2'-fluororibose, were the most compatible (39). Likewise, targeted mutagenesis of Cas12a protein should avoid altering the dynamic flexibility or folding of the enzyme, such as increasing rigidity, in regions that must mediate guide RNA contact or else risk weaker binding affinity and poor RNP assembly or stability. Interestingly, unbiased Cas12a mutagenesis screens aimed at improved activity found that favorable mutations lie outside of the wedge domain, which was the most dynamic during our crRNA binding simulations (41, 69).

These findings for Cas12a may also extend to other CRISPR-Cas systems. The need to conserve CRISPR array repeat sequences across all phases of the CRISPR mechanism is a common feature for CRISPR-Cas systems. For Cas9

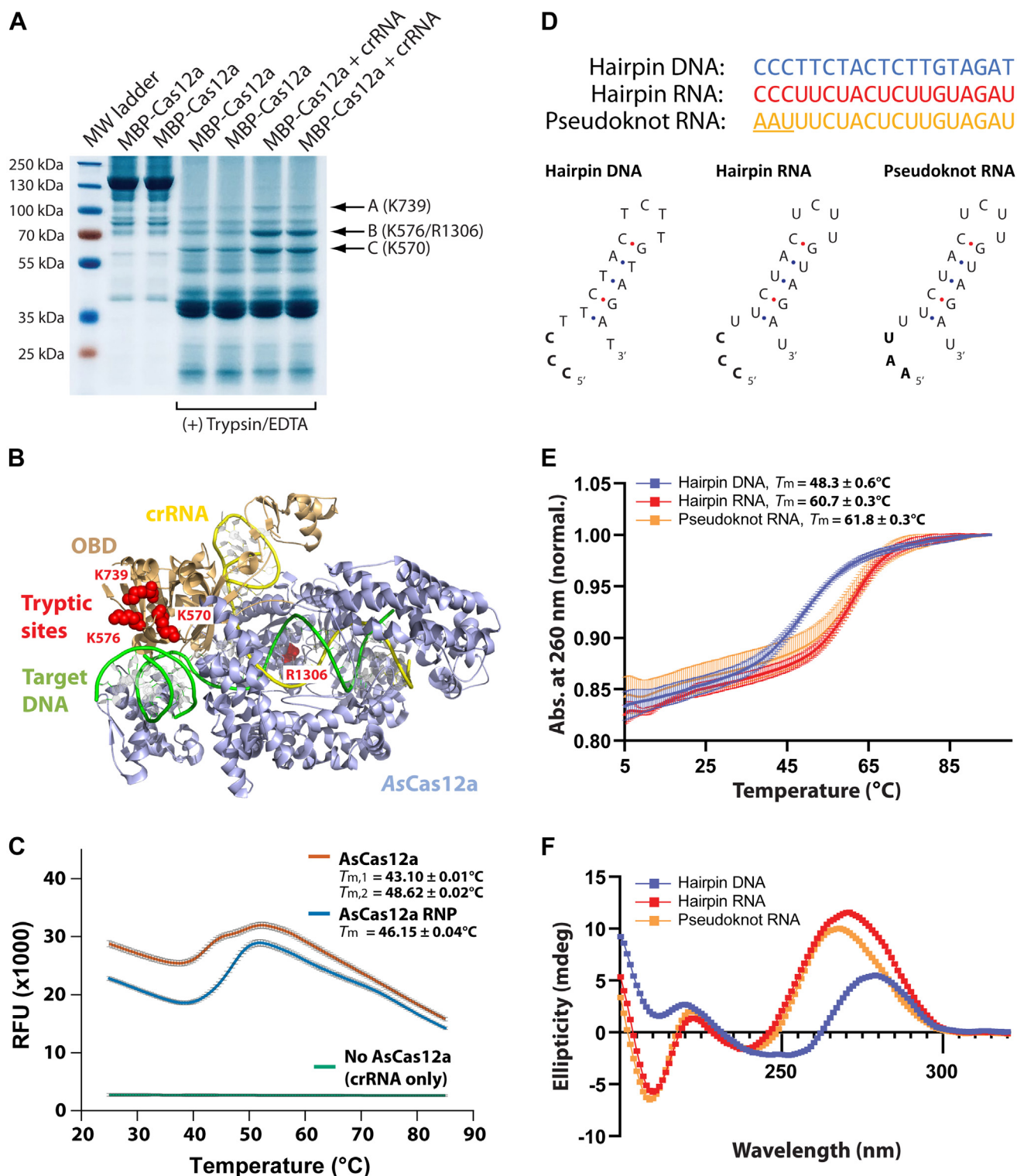


Figure 5. Cas12a undergoes conformational changes upon RNP assembly, whereas the crRNA pseudoknot is independently folded. *A*, limited trypsin proteolysis of AsCas12a with and without its cognate crRNA bound. Digestion products corresponding to differential trypsin accessibility are indicated to the right. *B*, amino acids corresponding to differentially accessible sites from (*A*), determined by MS, are highlighted in red spheres on the structure of AsCas12a. *C*, differential scanning fluorimetry of AsCas12a alone or bound to crRNA. *D*, sequence and predicted 2D structure of RNA and DNA used for thermal denaturation and CD analyses. *E*, UV-monitored thermal denaturation analysis and (*F*) CD spectra of the AsCas12a 5' handle pseudoknot unbound to AsCas12a. DNA and RNA hairpins are comparative controls. Error is shown as SEM of three experimental replicates. RNP, ribonucleoprotein.

systems, the structure of Cas9 from *Streptococcus pyogenes* has a large number of unresolved residues in the absence of a bound guide RNA (43, 70). Upon comparison to Cas9 RNP or

ternary structures, these regions become structured and appear to be heavily involved in RNA binding (5, 43, 48, 70). In addition, the binding of crRNA and *trans*-acting crRNA

Cas12a conformational dynamics during RNP assembly

guides, as well as single-guide RNAs, to Cas9 have been proposed to modulate the conformational dynamics of Cas9 (42, 45, 46).

While our study cannot definitively distinguish between induced fit and conformational selection binding modes during RNP assembly, results are better explained by an induced fit model. In induced fit mechanisms, the ligand binds an “inactive” state of the receptor, usually “loosely” or at low affinity, but then small, sequential conformational changes can occur where the protein creates more ligand contacts. For assembly of RNP complexes, this usually involves induced fit of the RNA rather than the protein or else a mutual induced fit of both protein and RNA (9, 11, 71, 72). Induced fit can be considered a special case of conformational selection but the two are often viewed as competing models (73, 74). Conformational selection posits that the receptor or enzyme occupies all possible states at different likelihoods and toggles between sampling “active” and “inactive” states. Thus, the kinetics of sampling can become important. Only when the protein transiently samples the active state can the ligand bind, which shifts the population of enzyme toward an active form. This creates a “selection” for the active conformation (74).

Some criteria for distinguishing induced fit and conformational selection can be applied (73). First, the range of conformational states can be informative (74, 75). For AsCas12a, the residues and domains that interact with the 5' handle appear to occupy a broad range of dynamic conformations. This would suggest the ability to accommodate many incremental changes that could tune its structure to fit the crRNA ligand (74, 76). Indeed, the protein appears to wrap around and protect the RNA after complex formation, which generates very low nanomolar binding affinity (77) (Fig. 4). Importantly, however, interdomain angle sampling from MD simulations revealed that the angles, and therefore conformational states, primarily occupied by the RNP form of AsCas12a are almost never sampled in simulations of the apo-AsCas12a. These results strongly argue against conformational selection as a likely mechanism of RNP assembly (63). Previous simulations of U1A protein binding to the hairpin II of U1 snRNA argued for induced fit based on the absence of the bound conformer when not assembled into the RNP. Only when the RNP complex was formed was the bound conformer observed and stable (75), similar to our interdomain angle analyses (Fig. 4, E and F).

Second, the timescale of conformational transitions, and therefore kinetics, is important (73, 76). If conformational transitions from an unbound/inactive to bound/active structure are infrequent and slow for Cas12a, then conformational selection could be argued. However, AsCas12a transitions in and out of the free and RNP conformational states are very rapid, on a nanosecond timescale, as observed by K15-L807 and H977-D966 residue distance measurements (Fig. 3). Third, although we did not explicitly focus on the role of the variable guide/spacer sequence of crRNA during binding, it is possible that this contributes an additional binding module. The necessity of distinct binding steps to produce high affinity binding

of U1A to U1 hairpin II RNA has been used to support induced fit mechanisms (78). The first step was electrostatic-mediated binding, referred to as the “lure,” which could be partially fulfilled by the negatively charged phosphate backbone of the crRNA spacer sequence, followed by a more specific “lock” step, potentially fulfilled by the crRNA 5' pseudoknot handle. The observation that modified 5' extensions to the end of the Cas12a crRNA can improve enzyme activity may be partially attributable to improved binding affinity *via* electrostatic interactions with Cas12a (77, 79, 80). Thus, the entire crRNA may participate in stepwise conformational changes that would be indicative of induced fit binding.

In this study, we were limited by the number of Cas12a and crRNA sequences suitable for our analysis. Nonetheless, they were sufficient to generate phylogenetic trees that could assess relatedness and for interpretation of sequence conservation across species. A greater number of sequences could refine our analyses but are unlikely to change the overarching observation of relatively low protein sequence homology compared to high repeat sequence (5' pseudoknot) conservation. For structural comparisons, we were also limited to the three available structures of Cas12a. However, these were fortuitously well-distributed across the inferred phylogenetic tree, which provided higher confidence that they were representative. Importantly, despite sequence divergence, the final folded structures of Cas12a proteins, especially around the guide RNA 5' pseudoknot, were highly similar. The crRNA pseudoknot structures themselves were nearly identical. While this study focused more on AsCas12a, simulations using FnCas12a and LbCas12a were also in agreement with our conclusions. Initial model selection can also be a limitation. Our simulations began with manually generated apo-AsCas12a from the DNA-bound ternary structure using coordinate extraction. Beginning with an apo crystal structure could generate fine variations in the results, though we would not expect these to significantly alter the major conclusions from this study. Indeed, beginning with an alternative *de novo* predicted apo-AsCas12a structure from the AlphaFold database resulted in remarkably similar outcomes in our MD simulations.

We also limited our in-solution experimental studies to AsCas12a as a model system. It is likely that limited trypsin proteolysis of other Cas12a members would yield alternative cleavage patterns as these will depend on surface accessibility of lysine or arginine residues. However, we would still expect to observe altered peptide cleavage patterns. Finer mapping of AsCas12a may also be achieved with other peptide hydrolysis methods or proteases and would be expected to create a more complete picture of conformational changes during RNP assembly. Ultimately, other methodologies to map the conformational states of Cas12a and their occupancy level when free and bound to guide RNA, especially at very fast time scales during RNP assembly, would help further establish the molecular mode of binding. These could include experimental structural, thermodynamic, molecular distance, and kinetic measurements that could further support or refute induced fit mechanisms.

Experimental procedures

Cas12a sequence alignment, phylogenetic analysis, and structure alignment

Cas12a sequences were downloaded from UniProt ([uniprot.org](https://www.uniprot.org)). Sequences less than 800 amino acids and duplicate taxa were removed before alignment with ClustalW (81). This alignment was used for maximum likelihood phylogenetic inference in MEGA11 with 500 bootstrap replicates and the JTT model of protein evolution (82). For simplicity, a consensus tree was produced with nodes having a bootstrap value less than 50 collapsed into polytomies. For comparing Cas12a RNP structures, Cas12a RNP or ternary complexes solved for *Lachnospiraceae bacterium* ND2006 (Lb) Cas12a (PDB ID: 5XUS) (51), *F. tularensis* subsp. *novicida* (Fn) Cas12a (PDB ID: 5NG6) (57), and *Acidaminococcus* sp. strain BV3L6 (As) Cas12a (PDB ID: 5B43) (6) were aligned in PyMOL (<https://pymol.org/2/>) (v1.5.0.5).

crRNA repeat alignment and secondary structure prediction

Type V repeat sequences were downloaded from CRISPRdb (<https://crispr.i2bc.paris-saclay.fr/>). These repeats were aligned and their secondary structure predicted with LocARNA (<https://rna.informatik.uni-freiburg.de/LocARNA>) (58). A consensus sequence for the alignment was generated with a cut-off of 55%. A visualization of the predicted secondary structure from LocARNA with this consensus sequence was constructed with forna (<http://rna.tbi.univie.ac.at/forna/>) (83).

MD structure preparation

FnCas12a, and LbCas12a

The crystal structure of the Cas12a complex (PDB ID: 5B43) (6) FnCas12a (PDB ID: 5NG6) (57) and LbCas12a (PDB ID: 5ID6) (51) were used for MD studies. The missing residues in the protein were filled by using the SWISS-MODEL (84) webserver. To generate the Cas12a RNP model, the DNA part of the crystal structure was manually removed. Similarly, the protein only structure was generated by removing the RNA part. The coordinates of the pseudoknot sequence were manually extracted and used for the preparation of the free pseudoknot RNA. The apo structure of AsCas12a from AlphaFold database (<https://alphafold.ebi.ac.uk/entry/U2UMQ6>) has also been utilized for the simulation. The Amber ff14SB (85) force field for protein, the ff99bsc0+ χ OL3 (86–88) force field for RNA, and TIP3P (89) model for water have been employed. The system was solvated using a 10 Å rectangular water box and Na⁺/Cl⁻ ions were added to neutralize the complexes in the tleap module of AmberTools 19.

MD methodology

The MD protocol used in earlier studies by Palermo and co-workers was used with slight modifications for the protein and RNP systems (56). An initial minimization of 10,000 steps (5000 steps of steepest descent along with 5000 steps of conjugate gradient) with a restraint of 300 kcal/mol·Å² on the biomolecules was carried out followed by another 10,000 steps

(5000 steps of steepest descent along with 5000 steps of conjugate gradient) without any restraints. The heating was done in four stages where systems were heated up from 0 to 50 K and 50 to 100 K by running two simulations of 50 ps each (NVT) imposing harmonic restraints of 100 kcal/mol·Å² on the protein and RNA. The temperature was then increased to 200 K in 100 ps (NVT) with a smaller restraint of 25 kcal/mol·Å² and to 300 K in 500 ps without any restraints (NPT). One nanosecond equilibration followed by a test production run of 10 ns was carried out (NPT).

The MD simulation of the pseudoknot was carried out starting with 1000 steps of minimization (500 steps of steepest descent), followed by a 20 ps of equilibration with a restraint force of 25 kcal/mol·Å² on the pseudoknot, at a temperature of 100 K (NVT). A further 2500 steps of unrestrained minimization with 1000 steps of steepest descent and 100 ps of unrestrained equilibration using a constant pressure periodic boundary of 1 atm was carried out where the temperature was increased from 100 K to 300 K. Finally, production run for 1 μ s was performed on the protein, RNP complex, and pseudoknot at 300 K using the GPU accelerated version of Particle Mesh Ewald Molecular Dynamics of AMBER 18. The time step used in MD simulations was 2 fs and the cut off for long range electrostatic interaction 10 Å. The SHAKE algorithm was applied to the bonds containing hydrogens. Langevin dynamics was used for temperature control (90) and Berendsen barostat (91) was used for pressure control for all the simulations.

MD simulation analysis

All the analyses were carried out using the CPPTRAJ (92) module of the AmberTools 19 (<https://ambermd.org/AmberTools.php>), VMD (<https://www.ks.uiuc.edu/Research/vmd/>) (93), and PyMOL Molecular Graphics System (<https://pymol.org/2/>), version 2.0 (Schrodinger, LLC). The RMSD of the protein backbone and the RNA backbone were calculated using the first frame of the 1 μ s simulation as the reference. All frames were considered for the calculation. RMSF analysis and the distance between the α atoms of the amino acids were calculated using the CPPTRAJ module considering every frame of the simulation. The 2DRMSD studies were carried out using every 1000th frame of simulation. The command 2drms in CPPTRAJ was used for the calculation and only the C, N, O, CA, and CB backbone atoms were considered. The information saved was further plotted using gnuplot (<http://www.gnuplot.info/>). The surf command in CPPTRAJ was used to calculate the SASA, which uses the LCPO algorithm (94). The surface electrostatic potential calculation was carried out using the ABPS electrostatic plugin in PyMOL. This calculation was carried out using the major cluster snapshots of the systems. The trajectories were visualized in VMD and the images were rendered using PyMOL.

PC analysis

PC analysis was carried out using the whole 1 μ s simulation of the MD trajectories. Only the C α atoms were considered for the analysis. Initially, the translational and rotational motion of

Cas12a conformational dynamics during RNP assembly

the protein is removed by fitting the trajectory to the average MD structure. Further coordinate covariance matrix, eigen values, and eigen vectors were calculated. A pseudo trajectory to visualize the first PC was also generated. The principal motions were visualized using the normal mode wizard plugin in the VMD software.

Cluster analysis

Average-linked hierarchical agglomerative clustering algorithm was used for the cluster analysis using CPPTRAJ. The rms distances of the protein backbone was used as the reference distance for the clustering procedure. The centroid of the cluster summary was extracted as the major cluster snapshot. The epsilon value for the analysis was 3 Å and every 15th frame was considered for the analysis.

Limited trypsin hydrolysis and MS of AsCas12a protein and RNP

Plasmid encoding His(6x)-MBP-AsCas12a was obtained from Addgene (79,007) (95) and protein prepared as described (39). crRNA targeting an EGFP sequence (crEGIP) (39) was synthesized and purified by Integrated DNA Technologies. Trypsin proteolysis was performed similarly as described (39). Briefly, 30 µg of AsCas12a was incubated in the presence and absence of crRNA (1:1.5 M ratio) at 23 °C for 5 min. The resulting samples were incubated with Trypsin-EDTA solution (0.05%, Invitrogen) at a mass ratio of 100:1 and incubated at 37 °C for 15 min. The reaction was stopped by the addition of SDS-PAGE loading buffer (New England Biolabs) and incubation at 95 °C for 5 min. The reaction products were analyzed by 12% SDS-PAGE and stained with Coomassie brilliant blue G-250 in 50% (v/v) methanol and 10% (v/v) acetic acid then destained in the same solution without dye. Differential bands of interest were gel extracted and shipped to the UT Southwestern Medical Center Proteomics Core facility for LC/MS/MS mass spectrometric identification using their Gel Band ID service. Data was analyzed using Proteome Discoverer 2.4 (<https://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/proteome-discoverer-software.html>) and searched using the His(6x)-MBP-AsCas12a protein sequence.

Differential scanning fluorimetry of AsCas12a

Thermal-shift assays were performed using a Bio-Rad CFX96 instrument with the FRET channel. AsCas12a alone and AsCas12a:crRNA (1:1) complex were incubated at room temperature for 10 min in a final assay concentration of 1x DSF buffer (100 mM Hepes, pH 7.5, 100 mM NaCl, 1 mM MgCl₂). SYPRO Orange Protein Gel Stain 5000x (Thermo Fisher Scientific) was freshly diluted to 171.25x in nuclease-free water and 2 µl was spotted into the wells of a Hard-Shell 96-well PCR Plate (Bio-Rad). The incubated solution was then added to wells to a final assay volume of 25 µl and AsCas12a or AsCas12a:crRNA (1:1) concentration of 3.3 µM.

The plate was sealed with Microseal “B” PCR Plate Sealing Film (Bio-Rad) and samples were heated from 25 °C to 85 °C at a ramp rate of 1 °C/min. A single fluorescence was collected every 30 s. The melting temperature (T_m) was determined using wTSA-CRAFT (https://bioserv.cbs.cnrs.fr/TSA_CRAFT/) (96).

Temperature-dependent SYPRO Orange emission intensity curves were analyzed to extract midpoint temperatures (T_m) and Van't Hoff enthalpies of folding (ΔH) using two-state ($F \rightleftharpoons U$) and sequential three-state ($F \rightleftharpoons I \rightleftharpoons U$) unfolding models as described (97, 98). Briefly total signals (S_i) in DSF curves were baseline corrected by extrapolating linear regions (99) (e.g., 25–30 °C and 75–80 °C) and fractional signals were fit to each model using unconstrained nonlinear regressions in MATLAB (MathWorks). For each complex, mean values and 95% confidence intervals from three independent trials are reported.

Nucleic acid thermal denaturation monitored by UV absorbance

Experiments were performed on a Varian Cary 100 Bio UV-Visible Spectrophotometer equipped with a Cary Temperature Controller using a 1 cm path length cuvette. Oligonucleotides were purchased from Integrated DNA Technologies. Samples were prepared at a concentration of 2 µM oligonucleotide in 10 mM sodium phosphate buffer, pH 7.0, and 2 mM MgCl₂ with a final sample volume of 500 µl. Samples were slow-annealed by heating to 95 °C for 10 min, slowly cooling to 25 °C over 3 h, and then incubating them at 4 °C for 16 h. Absorbance values were acquired at 260 nm over the range of 5 to 95 °C at a ramp rate of 0.5 °C/min with a data collection interval of 0.5 °C. The background was subtracted using Varian software and measurements performed in triplicate. The maximum value of the first derivative curves were used to determine the T_m values. The average T_m of each oligonucleotide with the corresponding SD is reported along with the normalized absorbance spectra.

Nucleic acid CD

Experiments were performed on a Chirascan VX Spectrophotometer using a 1 mm path length cuvette. Oligonucleotide samples were prepared at a concentration of 20 µM with a final sample volume of 200 µl. The buffer and slow-annealing conditions were the same as for thermal denaturation experiments. Ellipticity spectra were recorded over the range of 320 to 200 nm with a bandwidth of 1 nm, a sampling rate of 0.5 s-per-point, and a data collection interval of 1 nm. Three acquisitions were obtained for each oligonucleotide to create an average spectrum, from which the background spectrum was subtracted. The final reported spectra were smoothed using Chirascan software.

Data availability

The datasets and computer code generated or used in this study are available upon request to P. I. P. Phylogenetic

reconstruction datasets and sequences, as well as original MS data, are available upon request to K. T. G.

Supporting information—This article contains supporting information.

Acknowledgments—We thank Dr Andrew Lemoff (University of Texas Southwestern Medical Center Proteomics Core) for assistance with mass spectrometric analyses and data interpretation. This research was enabled in part by support provided by Calcul Québec (www.calculquebec.ca) and Compute Canada (www.computecanada.ca).

Author contributions—S. S., C. L. B., R. C., H. M. B., A. A. P., S. D. M., M. J. D., and K. T. G. investigation; K. T. G. conceptualization; M. J. D. and K. T. G. methodology; S. S., C. L. B., M. J. D., P. I. P., and K. T. G. writing—original draft; S. D. M., M. J. D., and P. I. P. formal analysis; S. S., C. L. B., R. C., P. I. P., and K. T. G. visualization.

Funding and additional information—This work was supported by a National Institutes of Health grant (1R01GM135646-01) to K. T. G. and M. J. D., an IRCC IIT Bombay research award grant to P. I. P., a National Institutes of Health grant (R35GM119818) to S. D. M., and a Prime Minister's Research Fellowship (PMRF) scheme for the PhD fellowship to S. S. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: DSF, differential scanning fluorometry; FRET, fluorescence energy transfer; MD, molecular dynamics; OBD, oligonucleotide binding domain; PAM, protospacer adjacent motif; PC, principal component; RMSF, RMS fluctuations; RNP, ribonucleoprotein; SASA, solvent accessible surface area.

References

- Moran, L. A., Horton, H. R., Scrimgeour, K. G., and Perry, M. D. (2012) *Principles of Biochemistry*, 5th Ed, Pearson, London, England
- Jones, S., and Thornton, J. M. (1996) Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13–20
- Li, L., and Ye, K. (2006) Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443**, 302–307
- Lin, J., Lai, S., Jia, R., Xu, A., Zhang, L., Lu, J., et al. (2011) Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature* **469**, 559–563
- Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., et al. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949
- Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I. M., Li, Y., et al. (2016) Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **165**, 949–962
- Elkayam, E., Kuhn, C. D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J., et al. (2012) The structure of human argonaute-2 in complex with miR-20a. *Cell* **150**, 100–110
- Gagnon, K. T., Zhang, X., Agris, P. F., and Maxwell, E. S. (2006) Assembly of the archaeal box C/D sRNP can occur via alternative pathways and requires temperature-facilitated sRNA remodeling. *J. Mol. Biol.* **362**, 1025–1042
- Turner, B., Melcher, S. E., Wilson, T. J., Norman, D. G., and Lilley, D. M. (2005) Induced fit of RNA on binding the L7Ae protein to the kink-turn motif. *RNA* **11**, 1192–1200
- Gagnon, K. T., Zhang, X., Qu, G., Biswas, S., Suryadi, J., Brown, B. A., et al. (2010) Signature amino acids enable the archaeal L7Ae box C/D RNP core protein to recognize and bind the K-loop RNA motif. *RNA* **16**, 79–90
- Suryadi, J., Tran, E. J., Maxwell, E. S., and Brown, B. A. (2005) The crystal structure of the methanocaldococcus jannaschii multifunctional L7Ae RNA-binding protein reveals an induced-fit interaction with the box C/D RNAs. *Biochemistry* **44**, 9657–9672
- Kong, R., Xu, L., Piao, L., Zhang, D., Hou, T. J., and Chang, S. (2017) Exploring the RNA-bound and RNA-free human argonaute-2 by molecular dynamics simulation method. *Chem. Biol. Drug Des.* **90**, 753–763
- Pourjafar-Dehkordi, D., and Zacharias, M. (2021) Binding-induced functional-domain motions in the argonaute characterized by adaptive advanced sampling. *PLoS Comput. Biol.* **17**, e1009625
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433
- Nakata, A., Amemura, M., and Makino, K. (1989) Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *J. Bacteriol.* **171**, 3553–3556
- Jansen, R., Embden, J. D., Gaastera, W., and Schouls, L. M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575
- Pourcel, C., Salvignol, G., and Vergnaud, G. (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653–663
- Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182
- Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S. D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561
- Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., and Koonin, E. V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1**, 7
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712
- Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuys, R. J., Snijders, A. P., et al. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964
- Mojica, F. J., and Rodriguez-Valera, F. (2016) The discovery of CRISPR in archaea and bacteria. *FEBS J.* **283**, 3162–3169
- Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., and Doudna, J. A. (2014) Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **21**, 528–534
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736
- Yosef, I., Goren, M. G., and Qimron, U. (2012) Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* **40**, 5569–5576
- Nuñez, J. K., Lee, A. S., Engelman, A., and Doudna, J. A. (2015) Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**, 193–198
- Krupovic, M., Béguin, P., and Koonin, E. V. (2017) Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr. Opin. Microbiol.* **38**, 36–43
- Wang, R., Li, M., Gong, L., Hu, S., and Xiang, H. (2016) DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in *Haloarcula hispanica*. *Nucleic Acids Res.* **44**, 4266–4277
- Jinek, M., Chylinski, K., Fonfara, L., Hauer, M., Doudna, J. A., and Charpentier, E. (2012) A Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821

Cas12a conformational dynamics during RNP assembly

31. Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading)* **155**, 733–740
32. Sander, J. D., and Joung, J. K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355
33. Sapranaukas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011) The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39**, 9275–9282
34. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67
35. Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-cas system. *Cell* **163**, 759–771
36. Kim, D., Kim, J., Hur, J. K., Been, K. W., Yoon, S. H., and Kim, J. S. (2016) Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868
37. Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., et al. (2018) CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439
38. Broughton, J. P., Deng, X., Yu, G., Fasching, C. L., Servellita, V., Singh, J., et al. (2020) CRISPR-cas12-based detection of SARS-CoV-2. *Nat. Biotechnol.* **38**, 870–874
39. Ageely, E. A., Chilamkurthy, R., Jana, S., Abdullahu, L., O'Reilly, D., Jensik, P. J., et al. (2021) Gene editing with CRISPR-Cas12a guides possessing ribose-modified pseudoknot handles. *Nat. Commun.* **12**, 6591
40. Wu, S. S., Li, Q. C., Yin, C. Q., Xue, W., and Song, C. Q. (2020) Advances in CRISPR/Cas-based gene therapy in human genetic diseases. *Theranostics* **10**, 4374–4382
41. Zhang, L., Zuris, J. A., Viswanathan, R., Edelstein, J. N., Turk, R., Thommandru, B., et al. (2021) AsCas12a ultra nuclease facilitates the rapid generation of therapeutic cell medicines. *Nat. Commun.* **12**, 3908
42. Sternberg, S. H., LaFrance, B., Kaplan, M., and Doudna, J. A. (2015) Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110–113
43. Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., et al. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997
44. Zhu, X., Clarke, R., Puppala, A. K., Chittori, S., Merk, A., Merrill, B. J., et al. (2019) Cryo-EM structures reveal coordinated domain motions that DNA cleavage by Cas9. *Nat. Struct. Mol. Biol.* **26**, 679–685
45. Lim, Y., Bak, S. Y., Sung, K., Jeong, E., Lee, S. H., Kim, J. S., et al. (2016) Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nat. Commun.* **7**, 13350
46. Osuka, S., Isomura, K., Kajimoto, S., Komori, T., Nishimasu, H., Shima, T., et al. (2018) Real-time observation of flexible domain movements in CRISPR-cas9. *EMBO J.* **37**, e96941
47. Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573
48. Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J. A. (2015) A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481
49. Jiang, F., Taylor, D. W., Chen, J. S., Kornfeld, J. E., Zhou, K., Thompson, A. J., et al. (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871
50. Nishimasu, H., Cong, L., Yan, W. X., Ran, F. A., Zetsche, B., Li, Y., et al. (2015) Crystal structure of *Staphylococcus aureus* Cas9. *Cell* **162**, 1113–1126
51. Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., et al. (2016) The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526
52. Cofsky, J. C., Soczek, K. M., Knott, G. J., Nogales, E., and Doudna, J. A. (2022) CRISPR-Cas9 bends and twists DNA to read its sequence. *Nat. Struct. Mol. Biol.* **29**, 395–402
53. Min, K., Yoon, H., Jo, I., Ha, N. C., Jin, K. S., Kim, J. S., et al. (2018) Structural insights into the apo-structure of Cpf1 protein from *Francisella novicida*. *Biochem. Biophys. Res. Commun.* **498**, 775–781
54. Li, W., Chan, C., Zeng, C., Turk, R., Behlke, M. A., Cheng, X., et al. (2020) Rational design of small molecules to enhance genome editing efficiency by selectively targeting distinct functional states of CRISPR-cas12a. *Bioconjug. Chem.* **31**, 542–546
55. Wörle, E., Jakob, L., Schmidbauer, A., Zinner, G., and Grohmann, D. (2021) Decoupling the bridge helix of Cas12a results in a reduced trimming activity, increased mismatch sensitivity and impaired conformational transitions. *Nucleic Acids Res.* **49**, 5278–5293
56. Saha, A., Arantes, P. R., Hsu, R. V., Narkhede, Y. B., Jinek, M., and Palermo, G. (2020) Molecular dynamics reveals a DNA-induced dynamic switch triggering activation of CRISPR-cas12a. *J. Chem. Inf. Model.* **60**, 6427–6437
57. Swarts, D. C., van der Oost, J., and Jinek, M. (2017) Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR-cas12a. *Mol. Cell* **66**, 221–233
58. Will, S., Joshi, T., Hofacker, I. L., Stadler, P. F., and Backofen, R. (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA* **18**, 900–914
59. Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born. *J. Chem. Theory Comput.* **8**, 1542–1555
60. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. explicit solvent particle Mesh Ewald. *J. Chem. Theory Comput.* **9**, 3878–3888
61. Le Grand, S., Götz, A. W., and Walker, R. C. (2013) SPFP: speed without compromise—a mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **184**, 374–380
62. Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., III, Cruzeiro, V. W. D., et al. (2015) *Amber 2015 (University of California, San Francisco)*. Available from ambermd.org/doc12/Amber15.pdf
63. Bucher, D., Grant, B. J., and McCammon, J. A. (2011) Induced fit or conformational selection? the role of the semi-closed state in the maltose binding protein. *Biochemistry* **50**, 10530–10539
64. Gao, K., Oerlemans, R., and Groves, M. R. (2020) Theory and applications of differential scanning fluorimetry in early-stage drug discovery. *Biophys. Rev.* **12**, 85–104
65. Kankia, B. I., and Marky, L. A. (1999) DNA, RNA, and DNA/RNA oligomer duplexes: a comparative study of their stability, heat, hydration, and Mg²⁺ binding properties. *J. Phys. Chem. B* **103**, 8759–8767
66. Kypr, J., Kejnovska, I., Renciuik, D., and Vorlickova, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **37**, 1713–1725
67. Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., et al. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441
68. Landthaler, M., Gaidatzis, D., Rothballer, A., Chen, P. Y., Soll, S. J., Dinic, L., et al. (2008) Molecular characterization of human argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA* **14**, 2580–2596
69. Liu, X., Liu, C., Zhou, J., Lv, X., He, Y., Liu, H., et al. (2021) Engineered FnCas12a with enhanced activity through directional evolution in human cells. *J. Biol. Chem.* **296**, 100394
70. Jiang, F., and Doudna, J. A. (2017) CRISPR-Cas9 structures and mechanisms. *Annu. Rev. Biophys.* **46**, 505–529
71. Williamson, J. R. (2000) Induced fit in RNA-protein recognition. *Nat. Struct. Biol.* **7**, 834–837
72. Guo, X., Campbell, F. E., Sun, L., Christian, E. L., Anderson, V. E., and Harris, M. E. (2006) RNA-dependent folding and stabilization of C5 protein during assembly of the *E. coli* RNase P holoenzyme. *J. Mol. Biol.* **360**, 190–203
73. Chakraborty, P., and Di Cera, E. (2017) Induced fit is a special case of conformational selection. *Biochemistry* **56**, 2853–2859

74. Boehr, D. D., Nussinov, R., and Wright, P. E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796
75. Qin, F., Chen, Y., Wu, M., Li, Y., Zhang, J., and Chen, H. F. (2010) Induced fit or conformational selection for RNA/U1A folding. *RNA* **16**, 1053–1061
76. Zhou, H. X. (2010) From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys. J.* **98**, L15–L17
77. Nguyen, L. T., Smith, B. M., and Jain, P. K. (2020) Enhancement of trans-cleavage activity of Cas12a with engineered crRNA enables amplified nucleic acid detection. *Nat. Commun.* **11**, 4906
78. Katsamba, P. S., Myszka, D. G., and Laird-Offringa, I. A. (2001) Two functionally distinct steps mediate high affinity binding of U1A protein to U1 hairpin II RNA. *J. Biol. Chem.* **276**, 21476–21481
79. Lee, K., Mackley, V. A., Rao, A., Chong, A. T., Dewitt, M. A., Corn, J. E., et al. (2017) Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. *Elife* **6**, e25312
80. Park, H. M., Liu, H., Wu, J., Chong, A., Mackley, V., Fellmann, C., et al. (2018) Extension of the crRNA enhances Cpf1 gene editing in vitro and in vivo. *Nat. Commun.* **9**, 3313
81. Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2003) Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics*, 2.3. 1–2.3. 22
82. Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8**, 275–282
83. Kerpedjiev, P., Hammer, S., and Hofacker, I. L. (2015) Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. *Bioinformatics* **31**, 3377–3379
84. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303
85. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713
86. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., III, Laughton, C. A., et al. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.* **92**, 3817–3829
87. Zgarbová, M., Otyepka, M., Šponer, J., Mládek, A. t., Banáš, P., Cheatham, T. E., III, et al. (2011) Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **7**, 2886–2902
88. Banáš, P., Hollas, D., Zgarbová, M., Jurecka, P., Orozco, M., Cheatham, T. E., III, et al. (2010) Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.* **6**, 3836–3849
89. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935
90. Turq, P., Lantelme, F., and Friedman, H. L. (1977) Brownian dynamics: its application to ionic solutions. *J. Chem. Phys.* **66**, 3039–3044
91. Berendsen, H. J., Postma, J. v., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690
92. Roe, D. R., and Cheatham, T. E., III (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095
93. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38
94. Weiser, J., Shenkin, P. S., and Still, W. C. (1999) Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **20**, 217–230
95. Hur, J. K., Kim, K., Been, K. W., Baek, G., Ye, S., Hur, J. W., et al. (2016) Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nat. Biotechnol.* **34**, 807–808
96. Lee, P. H., Huang, X. X., Teh, B. T., and Ng, L. M. (2019) TSA-CRAFT: a free software for automatic and robust thermal shift assay data analysis. *SLAS Discov.* **24**, 606–612
97. Gray, R. D., Buscaglia, R., and Chaires, J. B. (2012) Populated intermediates in the thermal unfolding of the human telomeric quadruplex. *J. Am. Chem. Soc.* **134**, 16834–16844
98. Dignam, J. D., Qu, X., and Chaires, J. B. (2001) Equilibrium unfolding of *Bombyx mori* Glycyl-tRNA synthetase. *J. Biol. Chem.* **276**, 4028–4037
99. Santoro, M. M., and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. I. unfolding of phenylmethanesulfonyl .alpha.-chymotrypsin using different denaturants. *Biochemistry* **27**, 8063–8068

SUPPORTING INFORMATION

Binding to the conserved and stably folded guide RNA pseudoknot induces Cas12a conformational changes during ribonucleoprotein assembly

Sruthi Sudhakar^{1,†}, Christopher L. Barkau^{2,†}, Ramadevi Chilamkurthy², Halle M. Barber³, Adrian A. Pater⁴, Sean D. Moran⁴, Masad J. Damha³, P.I. Pradeepkumar^{1,*}, and Keith T. Gagnon^{2,4,*}

¹ Department of Chemistry, Indian Institute of Technology Bombay, Mumbai, 400076, India.

² Department of Biochemistry and Molecular Biology, Southern Illinois University, School of Medicine, Carbondale, Illinois, 62901, USA.

³ Department of Chemistry, McGill University, 801 Sherbrooke St. West, Montreal, QC H3A 0B8, Canada

⁴ Department of Chemistry and Biochemistry, Southern Illinois University, Carbondale, Illinois, 62901, USA.

[†] Equally contributing authors.

*Corresponding Authors: ktgagnon@siu.edu, pradeep@chem.iitb.ac.in

Lachnospiraceae/1-1228 A A S K L E K F T N C Y S L S K T L R F K A I P V G K T Q E N I D N K R L L V E D E K R A E D Y K G V K K L L D R Y Y L S F I N D V L H S I K L K N - - L N N Y
Francisella/1-1300 - M S I Y Q E F V N K Y S L S K T L R F E L I P Q G K T L E N I K A R G L I L D D E K R A K D Y K K A K Q I I D K Y H Q F F I E E I L S S V C I S E D L L Q N Y
Acidaminococcus/1-1307 - M T Q F E G F T N L Y Q V S K T L R F E L I P Q G K T L K H I Q E Q G F I E E D K A R N D H Y K E L K P I I D R I Y K T Y A D Q C L Q L V Q L D W E N L S A A

Lachnospiraceae/1-1228 I S L F R K K T R T E K E N - - K E L E N L E I N L R K E I A K A F K G - A A G Y K S L F K K D I I E T - - - - - I L P E A A D D - - - - -
Francisella/1-1300 S D V Y F K L K K S D D D N L Q K D F K S A K D T I K K Q I S E Y I K D - S E K F K N L F N Q N L I D A K K G Q E S D L I L W L K Q S K D N G I E L F K A N S D
Acidaminococcus/1-1307 I D S Y R K E K T E E T R N - - - A L I E E Q A T Y R N A I H D Y F I G R T D N L T D A I N K R H A E I Y K G - - - - - L F K A E L F N G K V L K Q L G T

Lachnospiraceae/1-1228 - - - - K D E I A L V N S F N G F T T A F T G F F D N R E N M F S E E A K S T S I A F R C I N E N L T R Y I S N M D I F E E K V D - - - - - A I F D K H E V Q E
Francisella/1-1300 I T D I D E A L E I I K S F K G W T T Y F K G F H E N R K N V Y S N D I P T S I I Y R I V D D N L P K F L E N K A K Y E S L K D K A P E A I N Y E Q I K K D L
Acidaminococcus/1-1307 V T T T E H E N A L L R S F D K F T T Y F S G F Y E N R K N V F S A E D I S T A I P H R I V Q D N F P K F K E N C H I F T R L I T A V P - - - S L R E H F E N V

Lachnospiraceae/1-1228 I K E K I L N S D Y D - - - - - V E D F F E G E F F N F V L T Q E G I D V Y N A I I G G - F V T E S G E K I K G L N E Y I N L Y N A K T - K Q A L P K
Francisella/1-1300 A E E L T F D I D Y K T S E V N Q R V F S L D E V F E I A N F N N Y L N Q S G I T K F N T I I G G K F V N G E N T K R K G I N E Y I N L Y S Q Q I N D K T L K K
Acidaminococcus/1-1307 K K A I G I F V S T S - - - - - I E E V F S F P F Y N Q L L T Q I D L Y N Q L L G G I S R E A G T E K I K G L N E V L N A I Q K N D E T A H I I

Lachnospiraceae/1-1228 F K P - - - - - L Y K Q V L S D R E S L S F Y G E G Y T S D E E V L E V F R N T L N - - - - - K N S E I F S I K K L E K L F K N F D E Y S S - - A G I F V K
Francisella/1-1300 Y K M S V - - - - - L F K Q I L S D T E S K S F V I D K L E D S D V T T M Q S F Y E Q I A A F K T V E E K S I K E T L S L L F D D L K A Q K L D L S K I Y F K
Acidaminococcus/1-1307 A S L P H R F I P L F K Q I L S D R N T L S F I L E E F K S D E E V I Q S F C K Y K T - - - - - L L R N E N V L E T A E A L F N E L N S I D L T - - - H I F I

Lachnospiraceae/1-1228 N G P A I S T I S K D I F G E W N L I R - - - - - D K W N A E Y D D I H L K K K A V V T E K Y E D D R R K S F K I G S F S L E Q L Q E Y A D A D L S V
Francisella/1-1300 N D K S L T D L S Q Q V F D D Y S V I G T A V L E Y I T Q Q I A P K N L D N P S K K E Q E L I A K K T E K A K Y L S L E T I K - L A L E E F N K H R D I D K Q C
Acidaminococcus/1-1307 S H K K L E T I S S A L C D H W D T L R - - - - - N A L Y E R R I S E L T G K I T K S A K E K V Q R S L K H E D I N L Q E I I S A A

Lachnospiraceae/1-1228 V E K L K E I I I Q K V D E I Y K V Y G S S - E K L F D A D F V L E K S L K K - - - - - N D A V V A I M K D L L D S V K S F E N Y I K A F F G E G K E - - - - -
Francisella/1-1300 R F E E I L A N F A A I P M I F D E I A Q N K D N L A Q I S I K Y Q N Q G K K D L L Q A S A E D D V K A I K D L D Q T N N L L H K L K I F H I S Q S E D K A N
Acidaminococcus/1-1307 G K E L S E A F K Q K T S E I L S H A H A A L D Q P L P T T L K K Q - - - - - E E K E I L K S Q L D S L L G L Y H L L D W F A V D E S N - - - - -

Lachnospiraceae/1-1228 - T N R D E S F Y G D F V L A Y D I L L K V D H I Y D A I R N Y V T Q K P Y S K D K F K L Y F Q N P Q F M G G W D K D K E T D Y R A T I L R Y G S K Y Y L A I M
Francisella/1-1300 I L D K D E H F Y L V F E E C Y F E L A N I V P L Y N K I R N Y I T Q K P Y S D E K F K L N F E N S T L A N G W D K N K E P D N T A I L F I K D D K Y Y L G V M
Acidaminococcus/1-1307 - - E V D P E E F S A R L T G I K L E M E P S L S F Y N K A R N Y A T K K P Y S V E K F K L N F Q M P T L A S G W D V N K E K N N G A I L F V K N G L Y Y L G I M

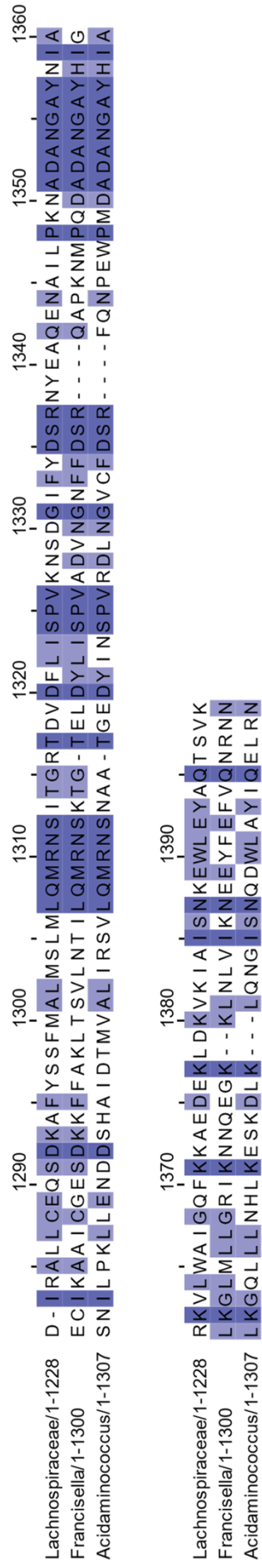


Figure S1. Alignment and conservation of As, Lb, and Fn Cas12a protein sequences. Visualization of sequence alignments and conservation was performed in Jalview. Blue shaded boxes indicate relative degree of sequence conservation.

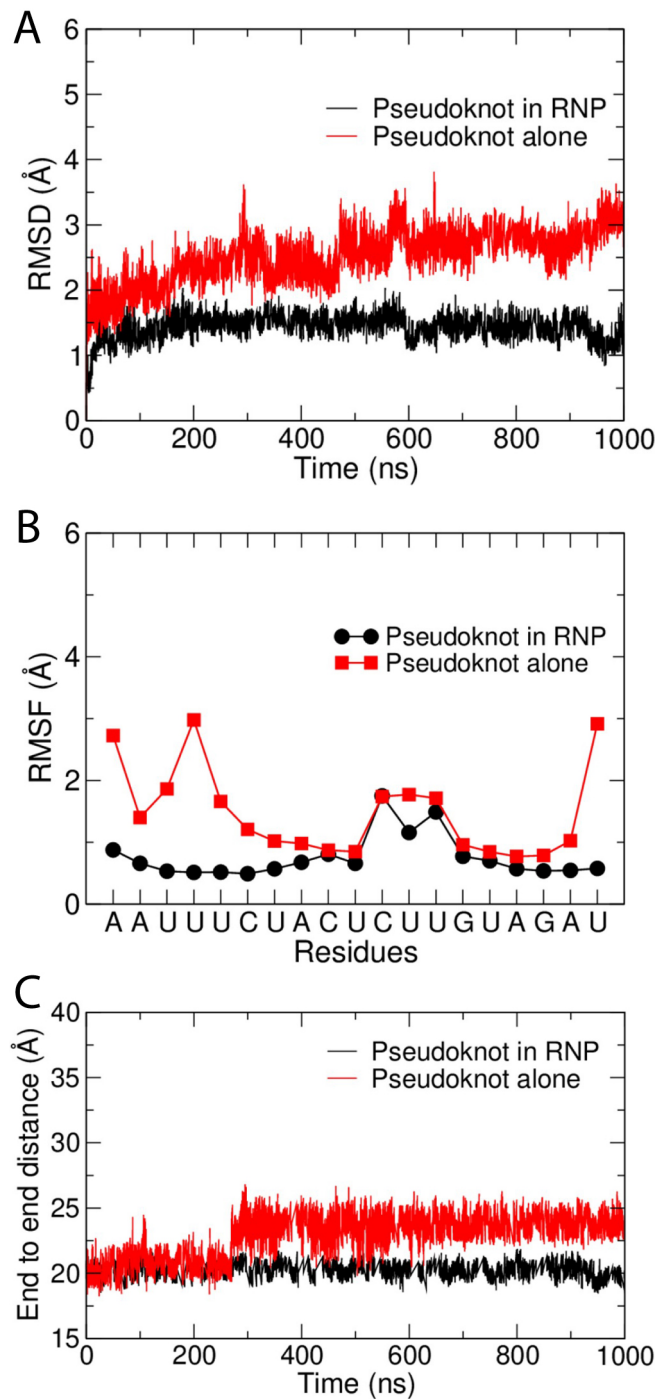


Figure S2. RMSD, RMSF and end-to-end distance graphs of the pseudoknot in free and AsCas12a-bound forms. The free form is represented in red and the pseudoknot bound to RNP complex in black. (A) The RMSD values were calculated from the 1 μ s trajectories. (B) Per nucleotide RMSF of the pseudoknot in the free and bound form. (C) End-to-end distance of the pseudoknot in the free and bound form. The distances were calculated from the 1 μ s trajectories.

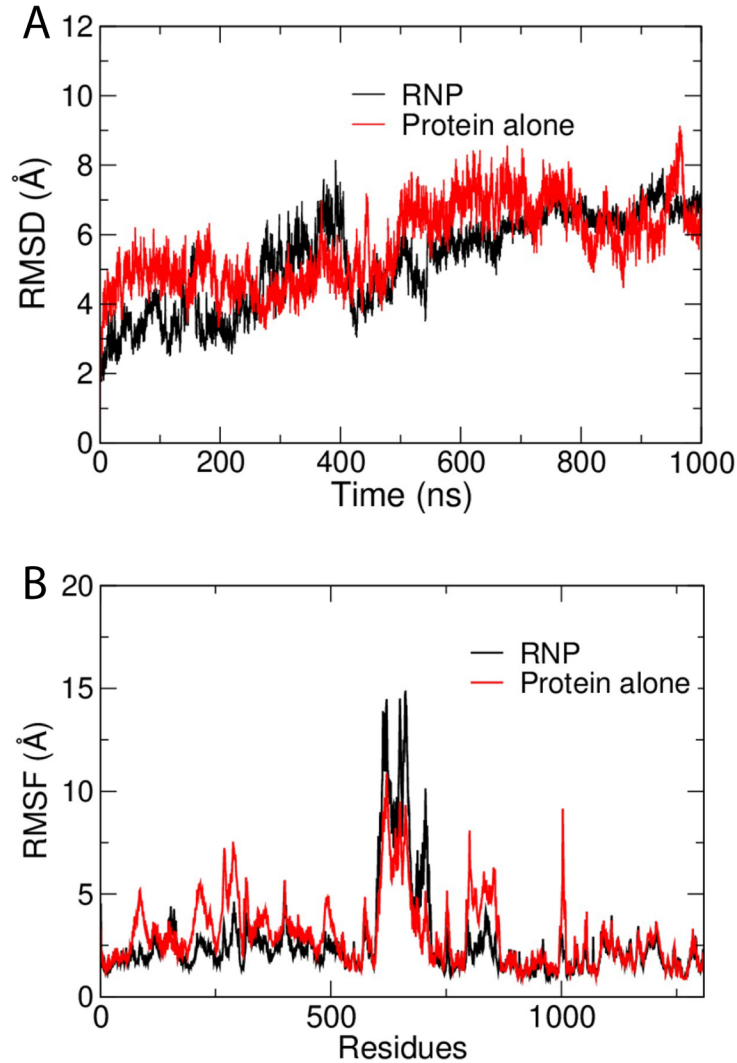


Figure S3. RMSD and RMSF graphs of AsCas12a protein and AsCas12a RNP complex. The free AsCas12a protein is represented in red and the protein in the Cas12a RNP complex is represented in black. **(A)** RMSD graphs of the backbone atoms of protein. RMSD values were calculated from the 1 μ s trajectories. **(B)** Per residue RMSF of the AsCas12a protein system and AsCas12a RNP complex. The AsCas12a protein system is represented in red and protein in the RNP complex is represented in black.

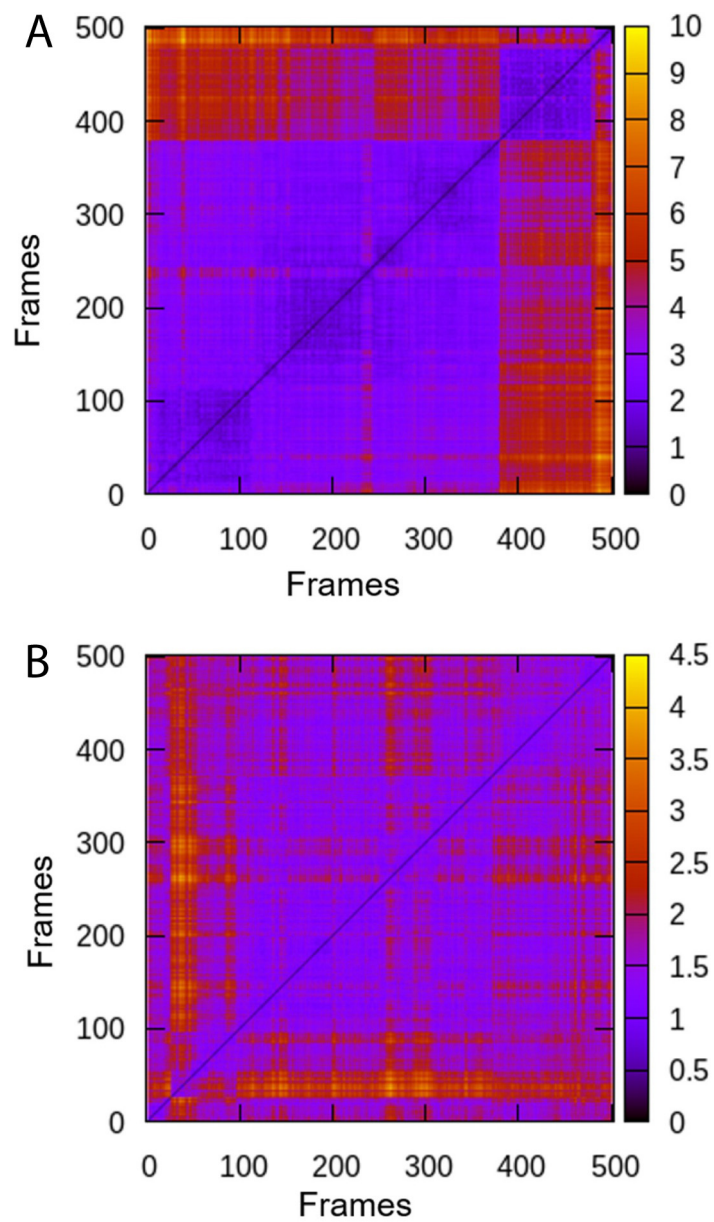


Figure S4. The 2D-RMSD plots of WED-III domain. (A) Cas12a protein alone simulation and **(B)** Cas12a RNP complex simulation. The 2D-RMSD values were calculated from every 500th frame of the 1 μ s trajectories.

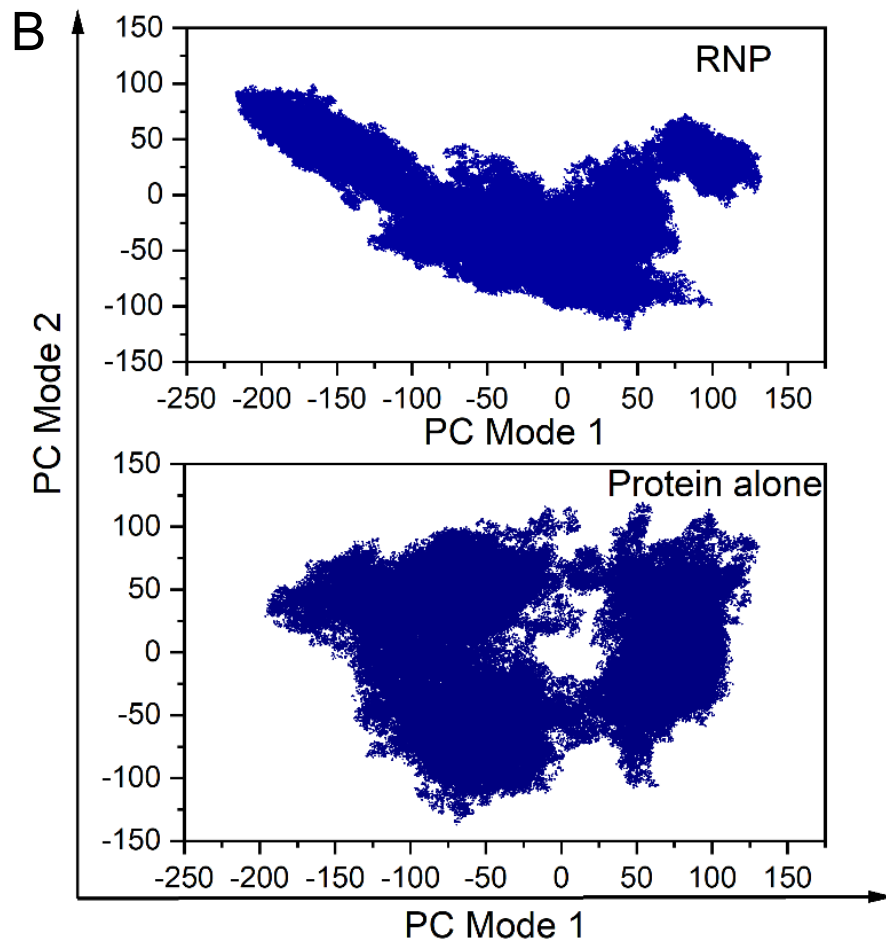
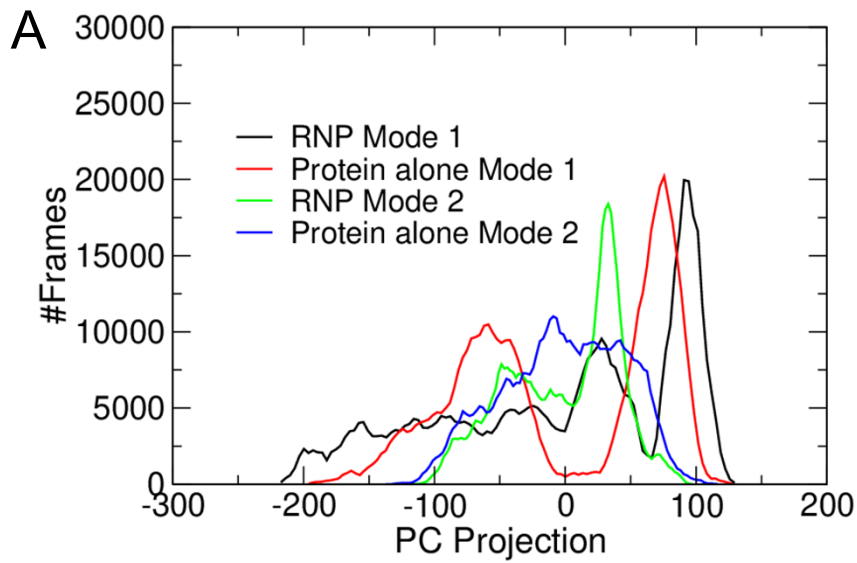


Figure S5. (A) The Principal component projection of the top 2 modes of AsCas12a RNP and protein alone systems. (B) The scatter plot showing PC mode 1 vs PC mode 2 distribution of AsCas12a RNP and protein alone systems.

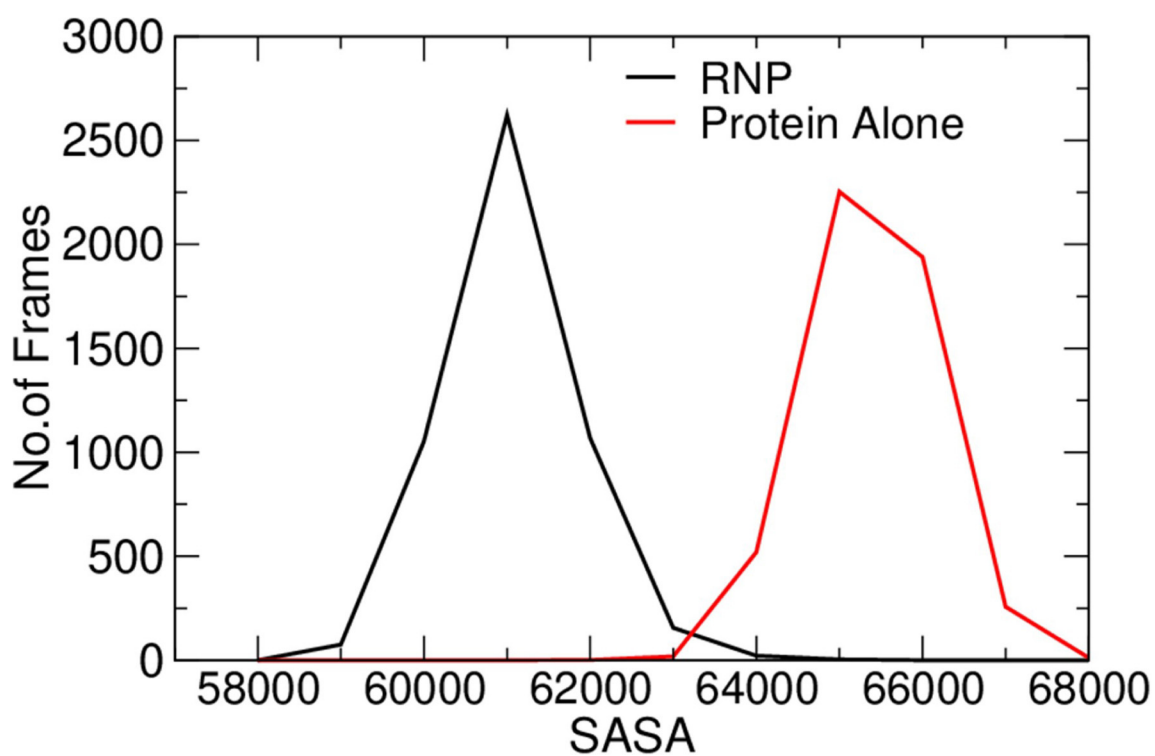


Figure S6. Solvent accessible surface area of the Cas12a protein and Cas12a-crRNA (RNP) complex. The Cas12a protein is represented in red and protein in the Cas12a RNP complex is represented in black.

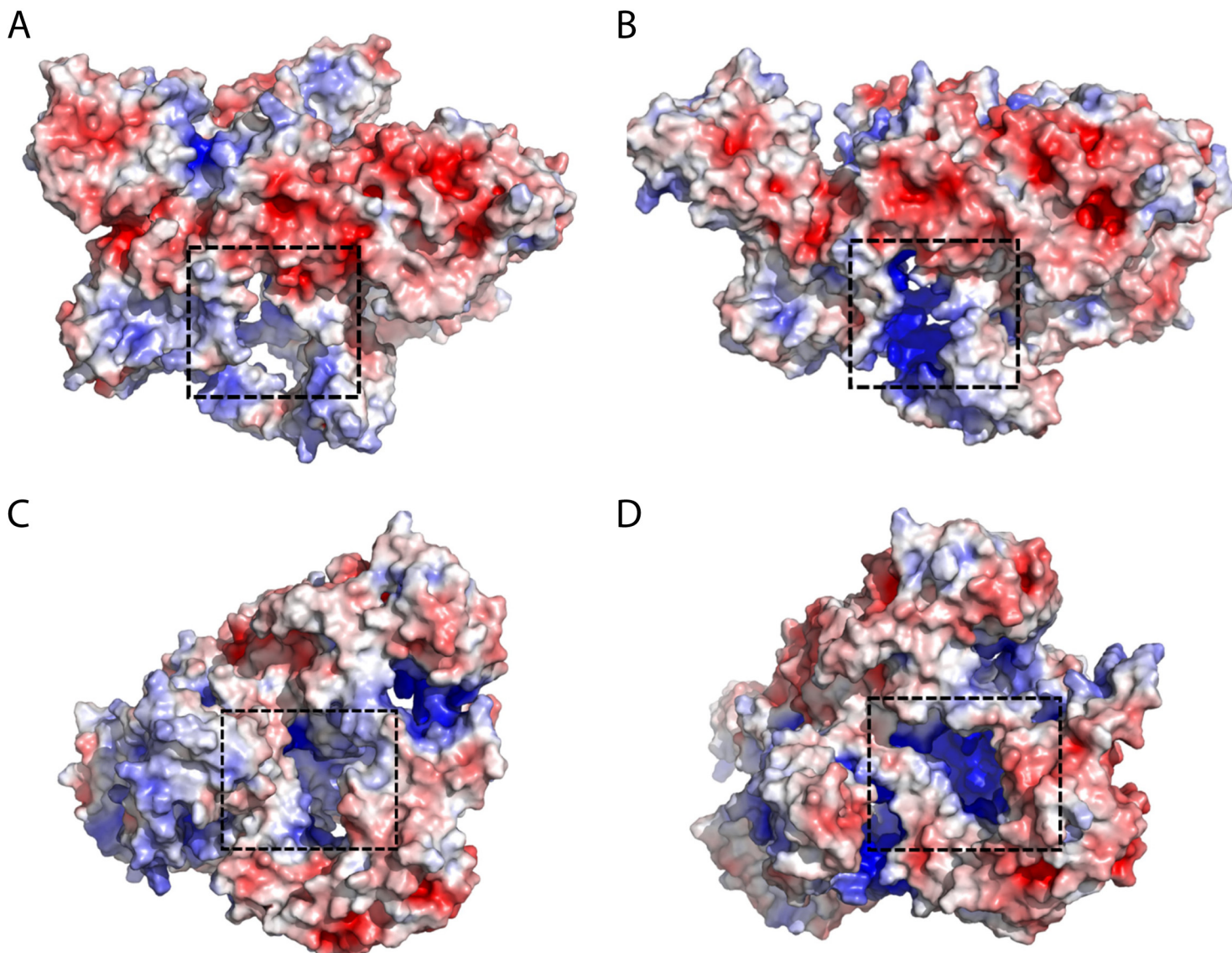


Figure S7. The surface electrostatic potential image of (A) pseudoknot occupied region of the Cas12a protein system; (B) pseudoknot occupied region of the Cas12a RNP complex; (C) ssRNA occupied region of the Cas12a protein system and (D) ssRNA occupied region of the Cas12a RNP complex. The positive charged surface is represented in blue and negatively charged surface is represented in red.

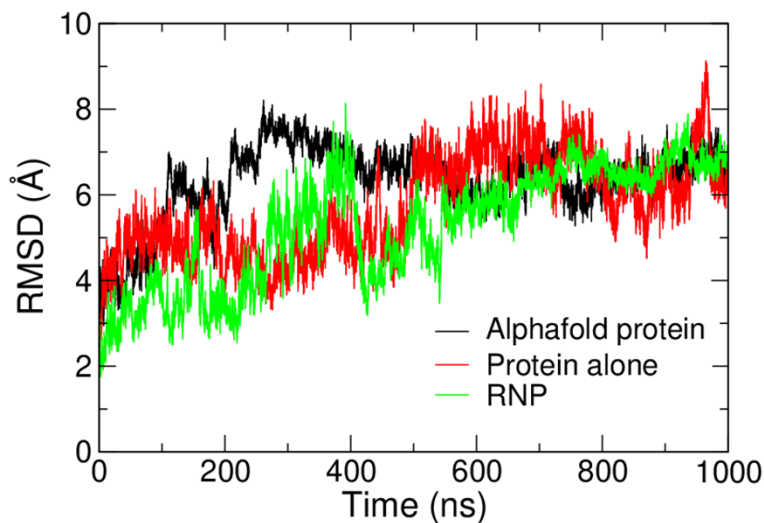


Figure S8. RMSD graphs of the backbone atoms of AsCas12a protein from Alphafold database, protein model generated manually, and AsCas12a RNP complex. The Cas12a protein from the Alphafold database is represented in black, Cas12a generated manually in red, and protein in the Cas12a RNP complex in green. The RMSD values were calculated from the 1 μ s MD trajectories.

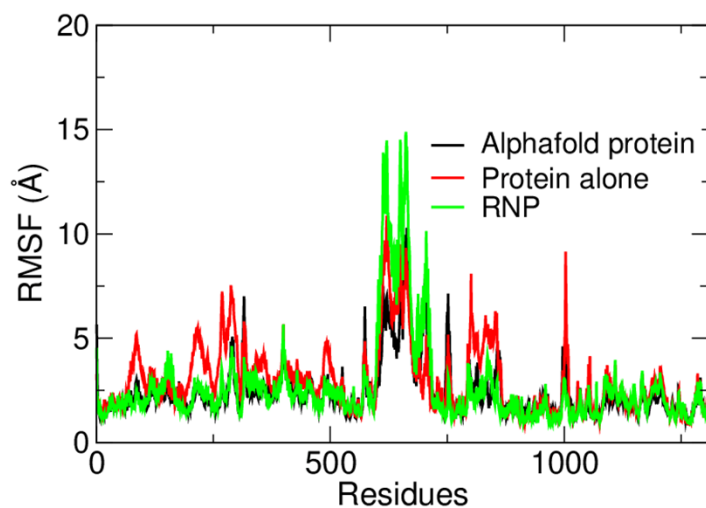


Figure S9. Per residue RMSF of AsCas12a protein from the Alphafold database, protein model generated manually, and AsCas12a RNP complex. The Cas12a protein from the Alphafold database is represented in black, Cas12a generated manually in red, and protein in the Cas12a RNP complex in green.

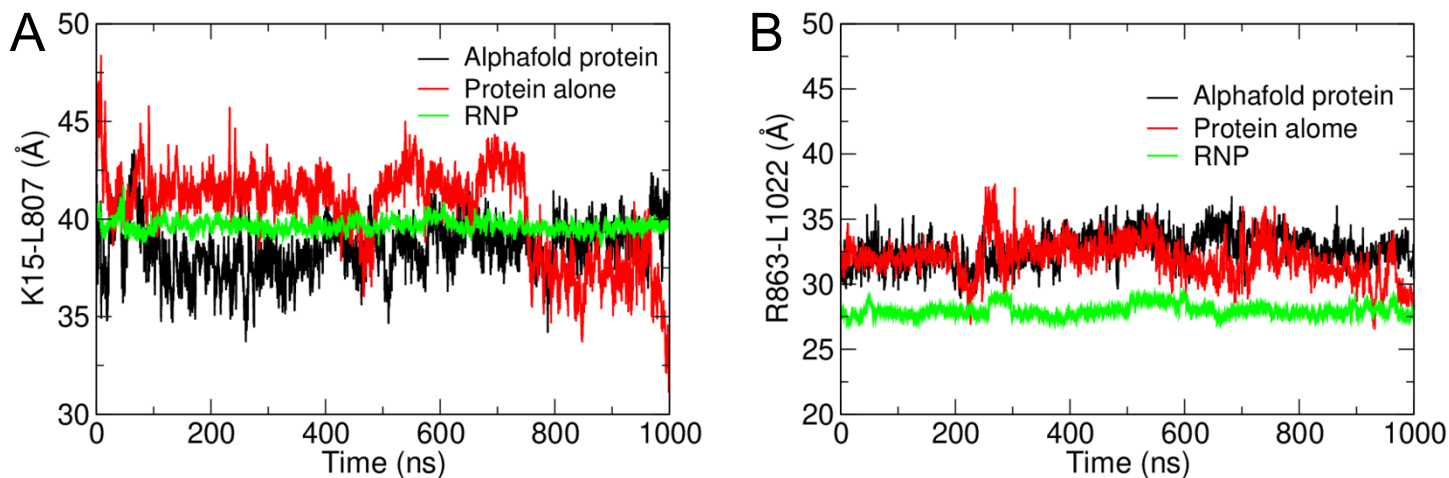


Figure S10. The distance between the C α atoms of (A) K15 and L807; (B) R863 and L1022 from the 1 μ s MD trajectories. The running averages of the distances are represented in the plot. The AsCas12a protein from the AlphaFold database is shown in black, AsCas12a generated manually in red, and protein in the AsCas12a RNP complex in green.

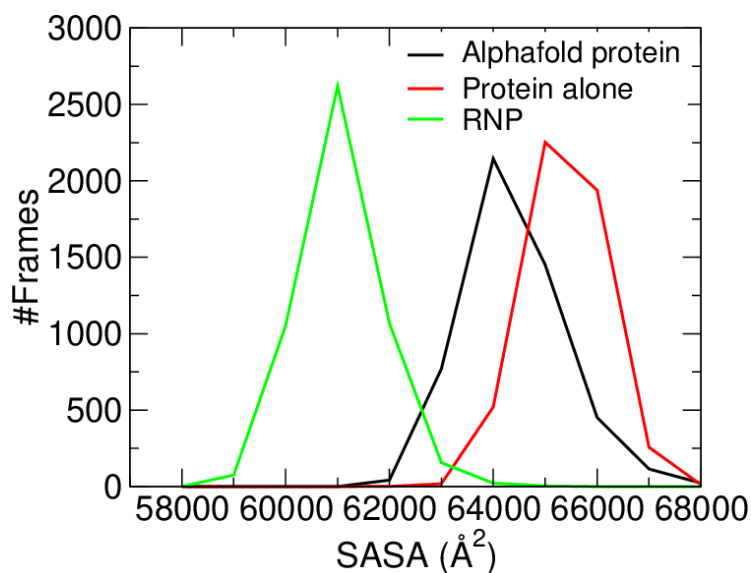


Figure S11. The solvent-accessible surface area of the AsCas12a protein from the AlphaFold database, the protein model generated manually, and the AsCas12a RNP complex. The Cas12a protein from the AlphaFold database is shown in black, Cas12a generated manually in red, and protein in the Cas12a RNP complex in green.

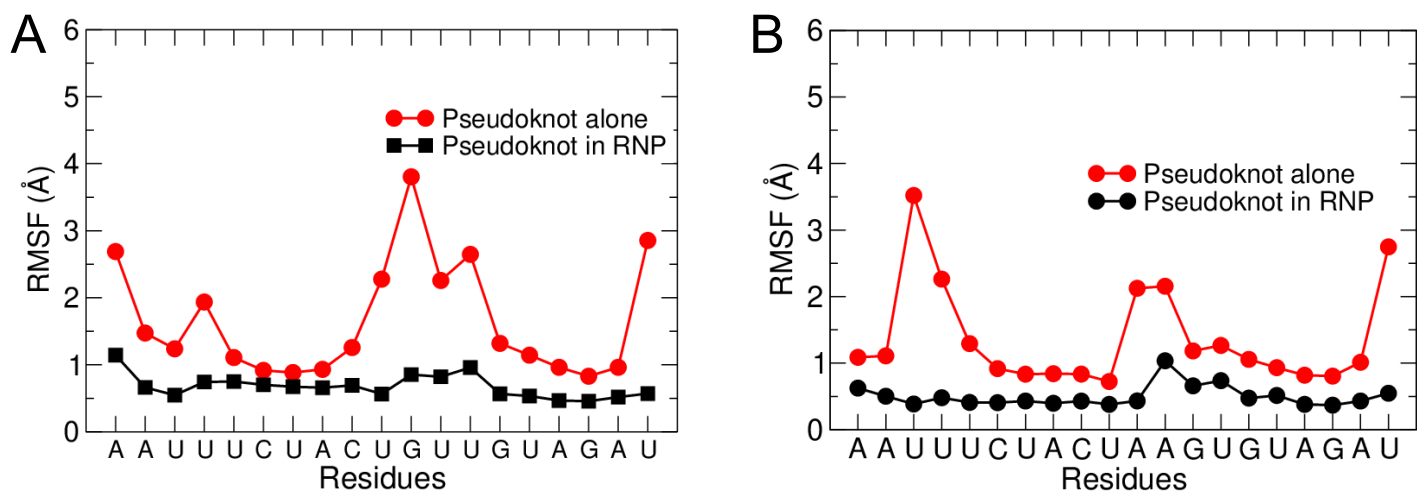


Figure S12. Per nucleotide RMSF of the pseudoknot in the free and bound form of (A) FnCas12a; (B) LbCas12a. The free form is represented in red, and the pseudoknot bound to the RNP complex in black.

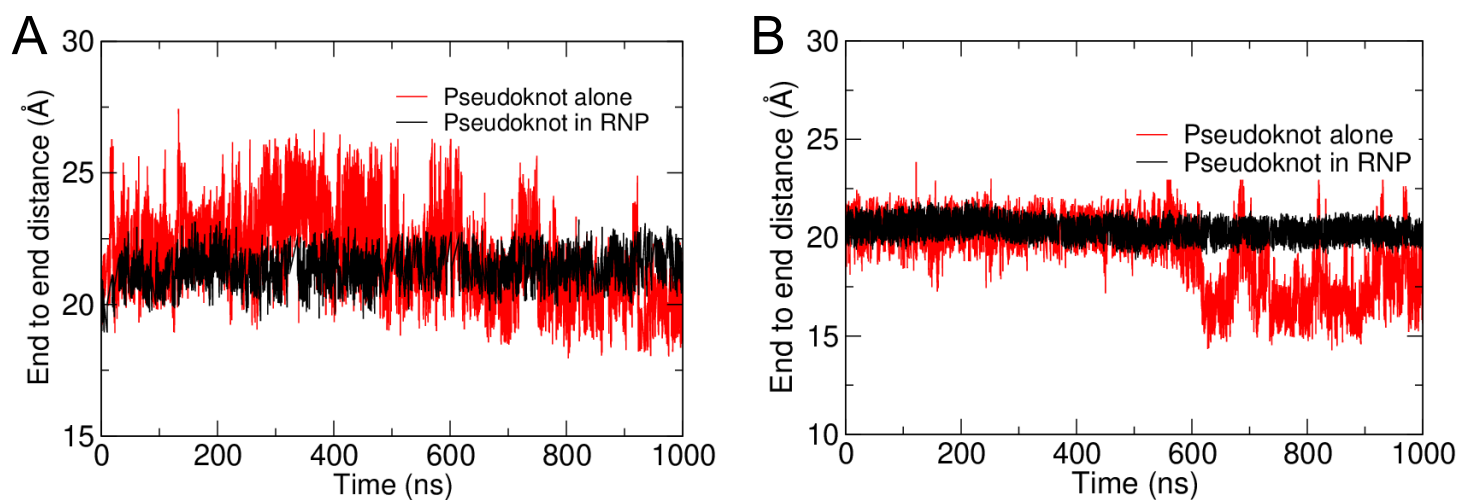


Figure S13. End-to-end distance of the pseudoknot in the free and bound form in (A) FnCas12a; (B) LbCas12a. The free form is represented in red, and the pseudoknot bound to the RNP complex in black. The distances were calculated from the 1 μ s MD trajectories.

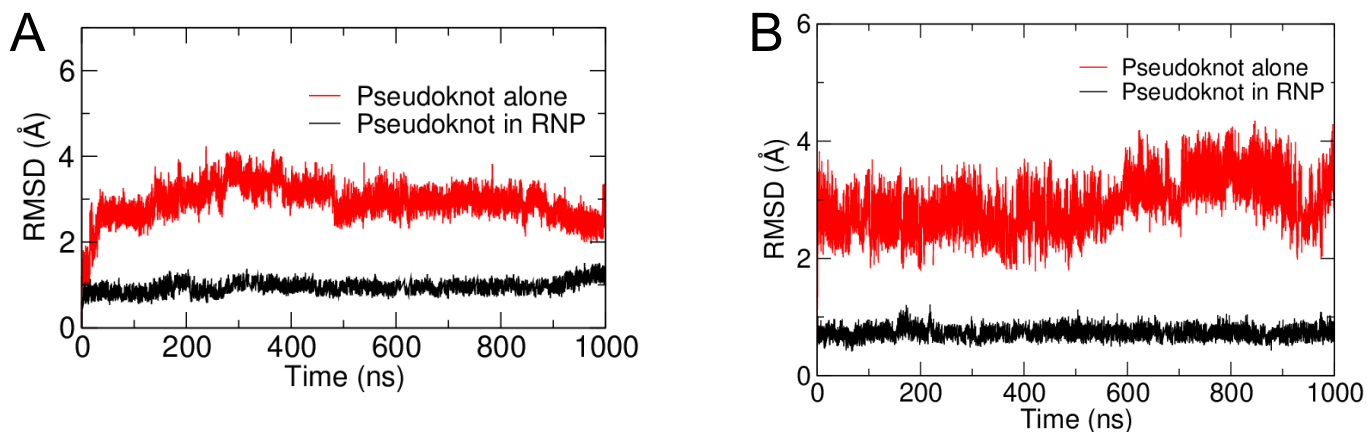


Figure S14. RMSD graphs of the backbone atoms of the free and bound forms of pseudoknot in (A) FnCas12a; (B) LbCas12a. The free form is represented in red, and the pseudoknot bound to the RNP complex in black. The RMSD values were calculated from the 1 μ s MD trajectories.

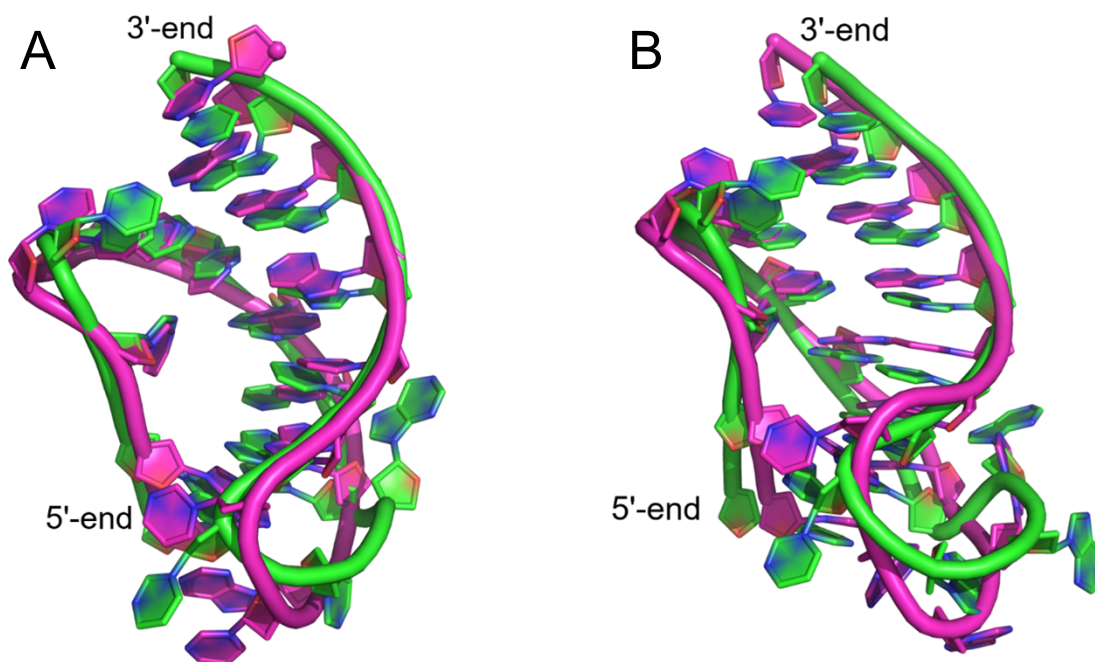


Figure S15. Superimposed image of the representative structures of the major cluster in the free pseudoknot and bound pseudoknot in the RNP complex in (A) FnCas12a; (B) LbCas12a from the 1 μ s MD trajectories. The carbon atoms of the free pseudoknot are represented in magenta, and the carbon atoms of the bound pseudoknot in green, nitrogen atoms in blue and oxygen atoms in red.

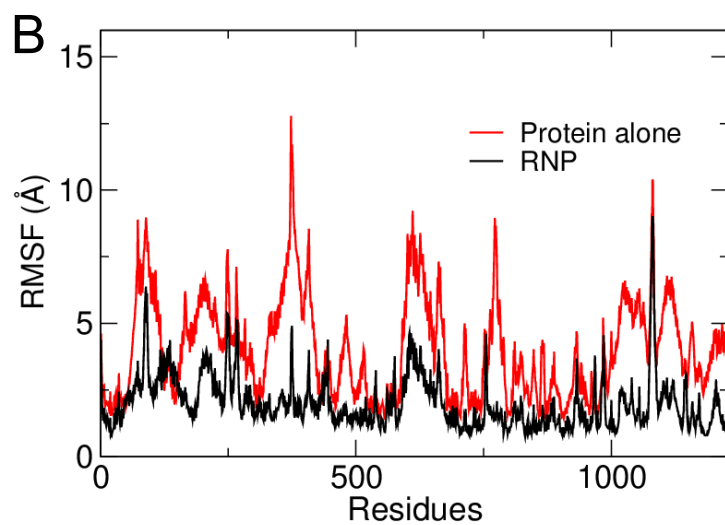
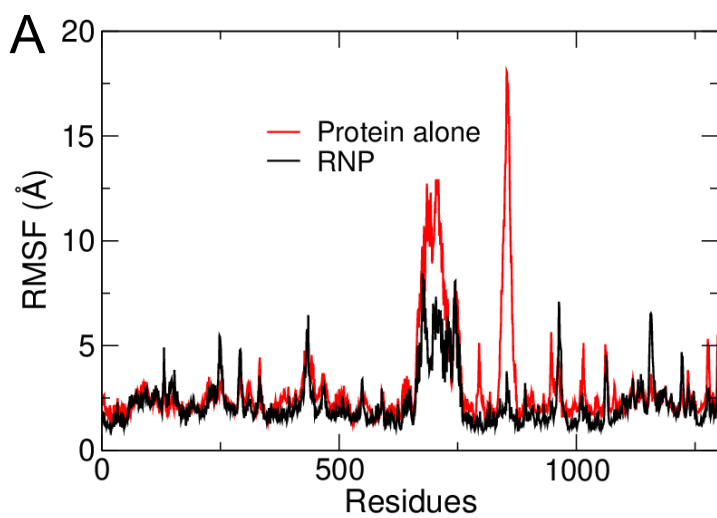


Figure S16. Per residue RMSF of the Cas12a protein system and Cas12a RNP complex in **(A)** FnCas12a; **(B)** LbCas12a. The Cas12a protein system is represented in red, and protein in the Cas12a RNP complex in black. The residues are numbered -1 in FnCas12a and +1 in LbCas12a from the actual crystal structure in the RMSF plots.

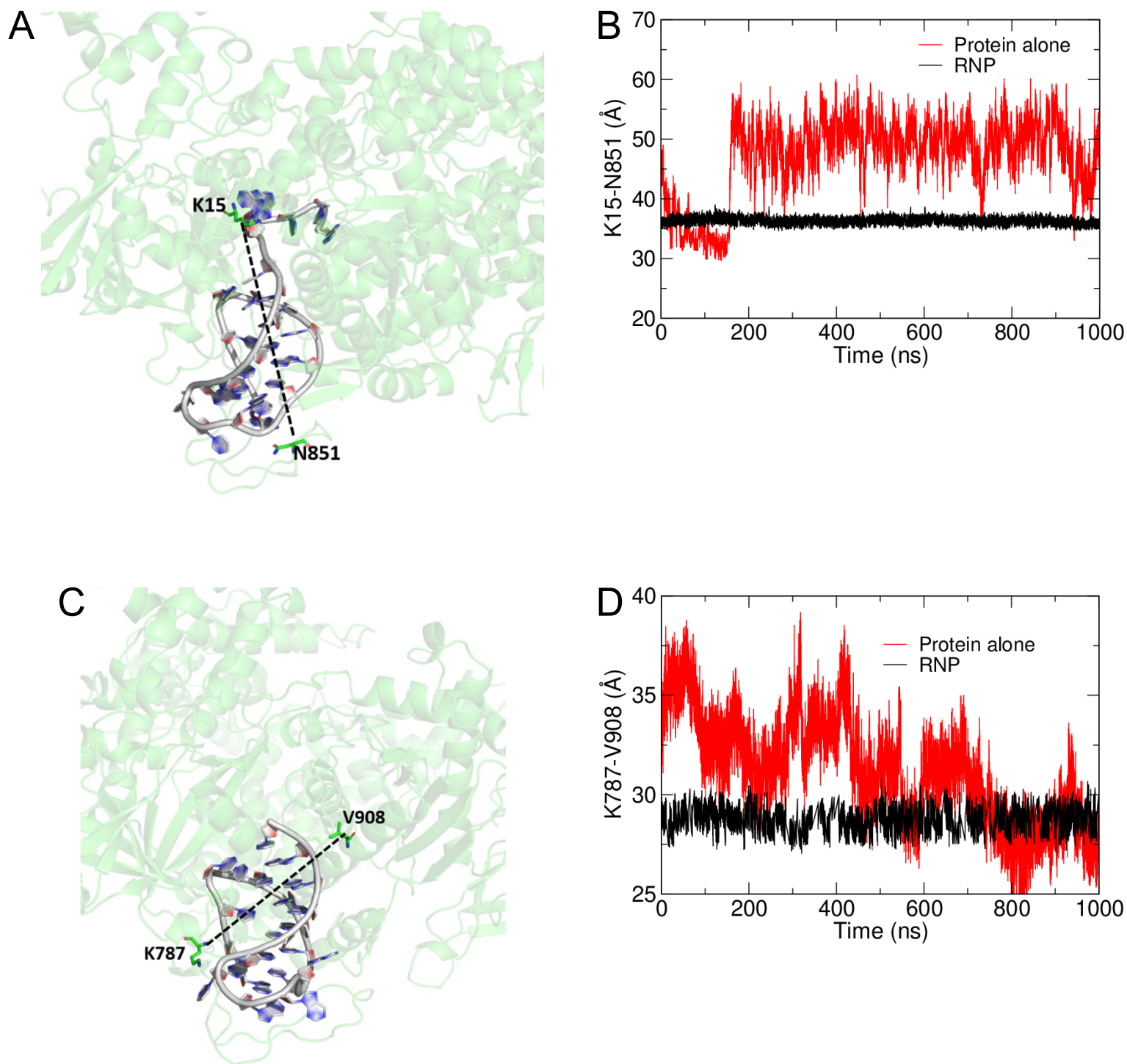


Figure S17. (A) Representative image illustrating the amino acids considered for the distance measurements in FnCas12a. (B) The distance between the $\text{C}\alpha$ atoms of K15 and N851 of FnCas12a. (C) Representative image illustrating the amino acids considered for the distance measurements in LbCas12a. (D) The distance between the $\text{C}\alpha$ atoms of K787 and V908 of LbCas12a. The distances are measured from the 1 μs MD trajectories.

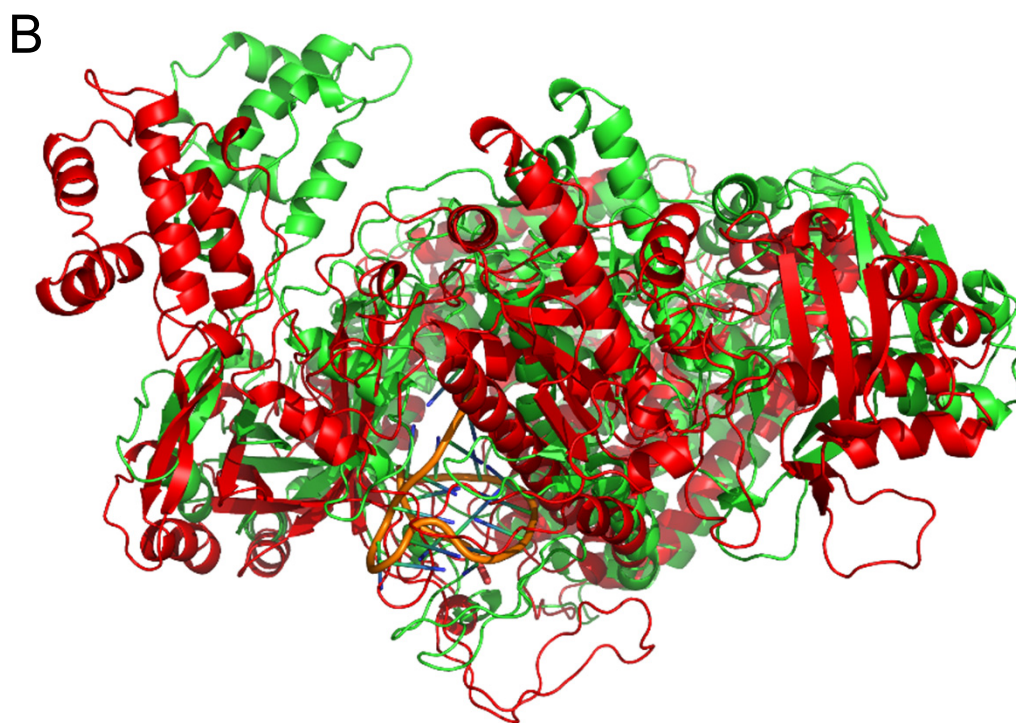
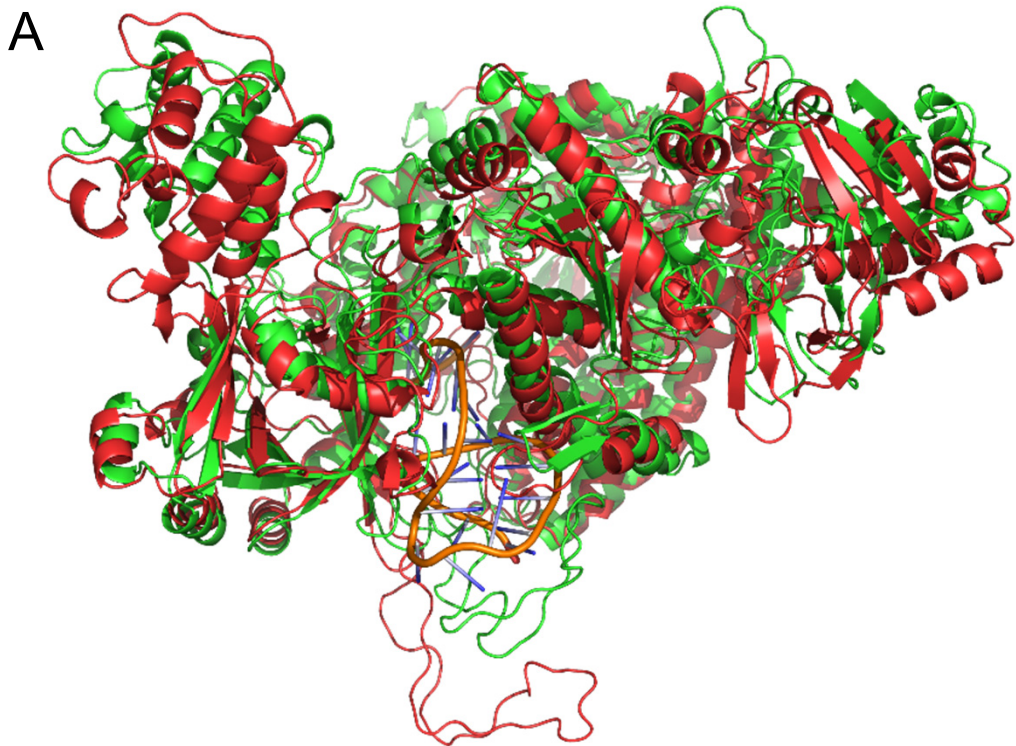


Figure S18. Superimposed images of the major clusters of Cas12a protein and Cas12a RNP complex in (A) FnCas12a; (B) LbCas12a. The Cas12a protein system is represented in red, and the Cas12a RNP complex in green.

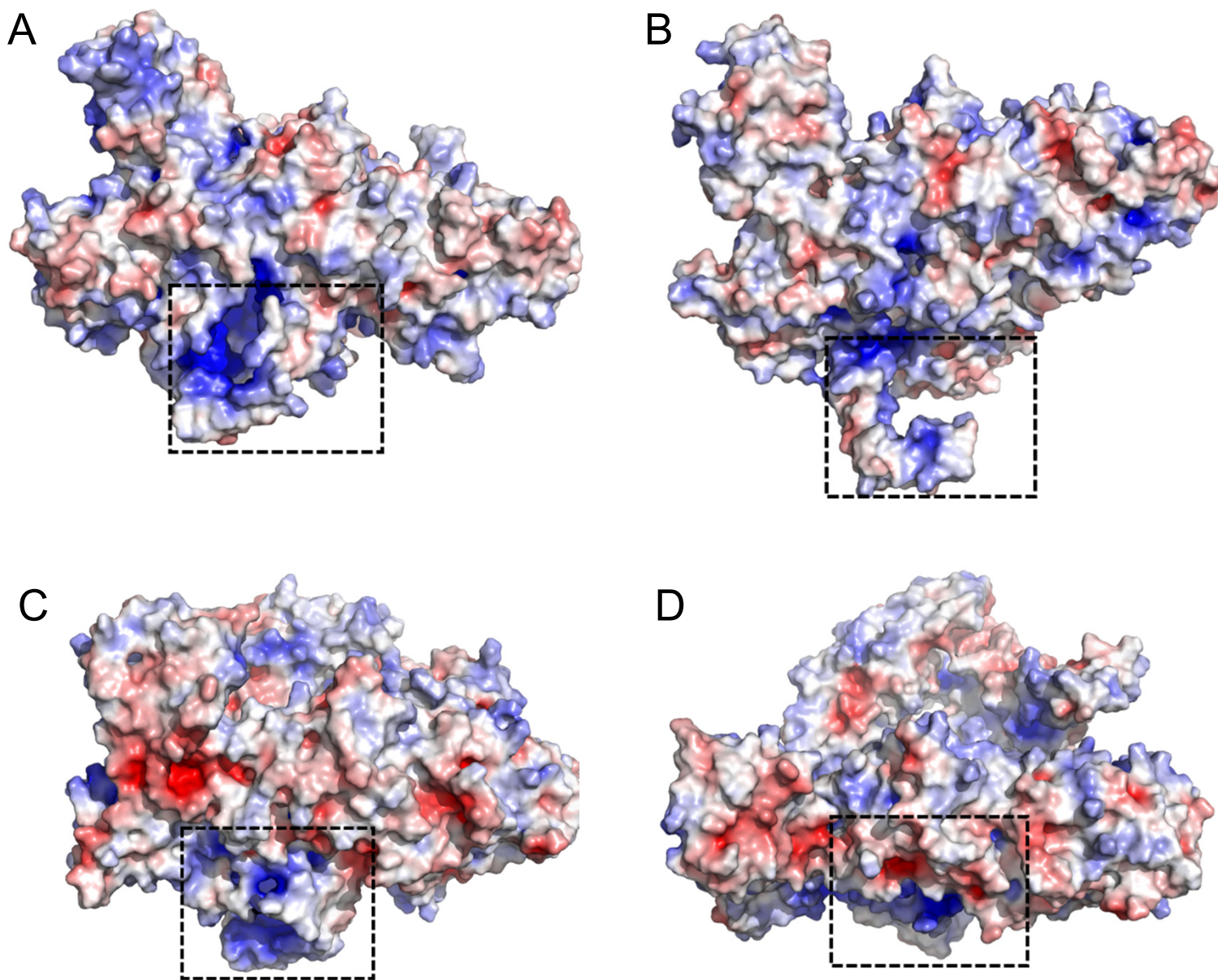


Figure S19. The surface electrostatic potential image of (A) pseudoknot-occupied region of the FnCas12a RNP complex; (B) pseudoknot-occupied region of the FnCas12a protein alone; (C) pseudoknot-occupied region of the LbCas12a RNP complex; (D) pseudoknot occupied region of the FnCas12a protein alone. The positively charged surface is represented in blue and negatively charged surface is represented in red. The location of the pseudoknot and point of distinction are marked by dashed rectangular boxes.

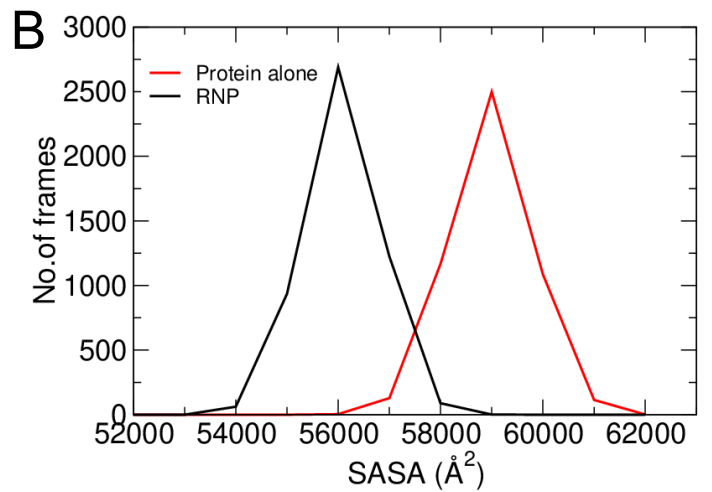
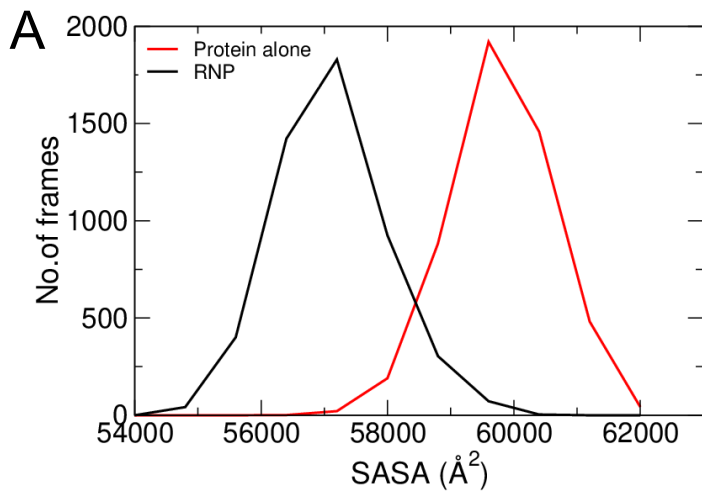


Figure S20. The solvent-accessible surface area of the protein in (A) FnCas12a protein system and RNP complex; (B) LbCas12a protein system and RNP complex. The Cas12a protein system is represented in red, and protein in the Cas12a RNP complex in black.

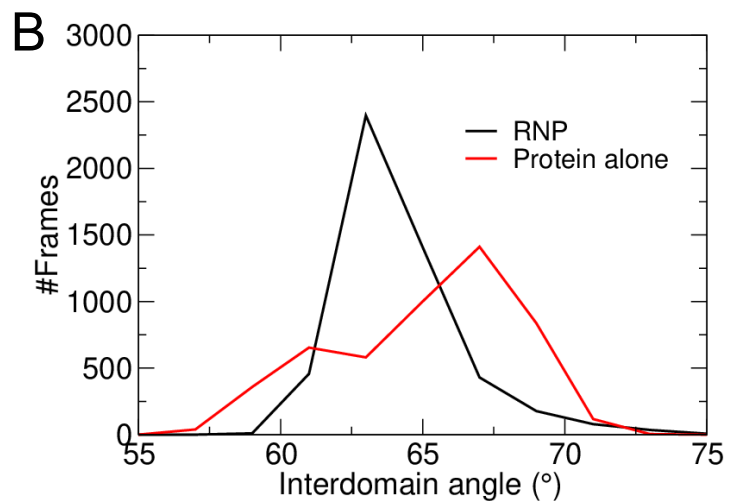
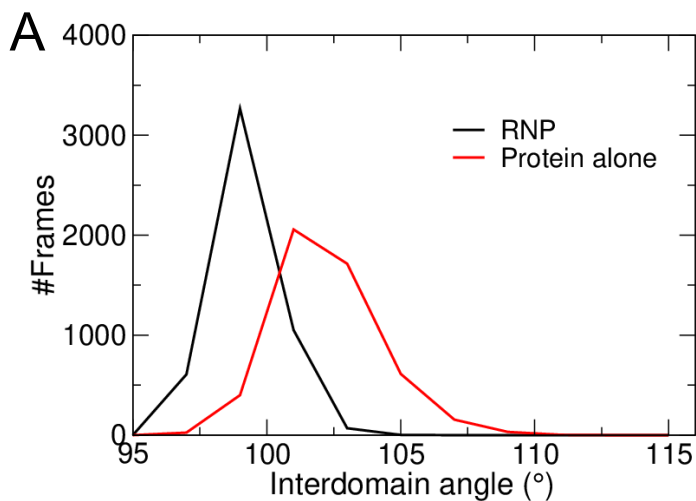


Figure S21. The interdomain angle between (A) WED-II: WED-III: RuvC domain; (B) REC2: RuvC: NUC domain. The histogram is sampled from the 1 μ s MD trajectories. The Cas12a protein alone is shown in red, and the Cas12a RNP in black.

Figure S22. Sequence Analysis of Mass Spec Data for AsCas12a Limited Trypsin Proteolysis

>His6X-MBP-AsCas12a

MKIHVVVVVHHEEGKLVWINGDKGYNGLAEVGKKFEKDTGKIVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGY
AQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSIIYKNDLLPNPPKTWEEIPALDKELKAKGKSALMFNL
QEPYFTWPLIAADGGYAFKYENGYDIKDVGVNAGAKAGLTFVLDIKKNHMNADTDYSIAEAAFNKGETAMTINGP
WAWSNIDTSKVNIGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYE
EELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSITSLYKKAGFMGAPPKK
KRKVGGMNQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDDHYKELKPIIDRIYKTYADQCLQLVQLDWE
NLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDFYGRDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHEN
ALLRSFDKFTTYFSGFYENRKNVSAEDISTAIPHRIVQDNFPKFENKCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIIEEVF
SFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSD
EEVIQSFCKYKTLRNENVLETAELFNEINSIDLTHIFISHKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQ
RSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQDLSLLGLYHLLDWFVAVDESNEVD
PEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPQKQ
GRYKALSFEPTEKTSEGFDMYYDYFPDAKMPKCSQKAVTAHFQTHHTPILLSNFIETITKEIYDLNPEKEPKK
FQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTSSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVE
TGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQ
KTIPTDLYQELYDYNHRLSHDLSDEARALLPNVITKEVSHEIHKDRRFTSDKFFFHVPITLNYQAANSPPKFNQRVNAYL
KEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSNLIQQFDYQKLDNREKERRAARQAWSVVGTIKDLKQGYLSQVIHEI
VDLMIHYQAVVVLENLNGFVKSRGTGIAEKAVYQQFEKMLIDKLNCLVLDYPAEKVGGVLPYQLTDQFTSFAKMG
QSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLGEGDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAW
DIVFEKNETQFDAGTPIAGKRIVPIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLENDSDHAIDTMVALI
RSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPWPMDADANGAYHIALKGQLLNHLKESKDLKLQNGISNQ
DWLAYIQELRNPKKKRKKVGGYPYDVPDYA

Band 1 (A)

MKIHVVVVVHHEEGK

LVWINGDKGYNGLAEVGKKFEKDTGKIVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAF
QDKLYPFTWDAVRYNGKLIAYPIAVEALSIIYKNDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADG
GYAFKYENGYDIKDVGVNAGAKAGLTFVLDIKKNHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNIG
VTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEELAKDPRIAATME
NAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSITSLYKKAGFMGAPPKKKRKKVGG

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDDHYKELKPIIDRIYKTYADQCLQLVQLDWE
NLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDFYGRDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHEN
ALLRSFDKFTTYFSGFYENRKNVSAEDISTAIPHRIVQDNFPKFENKCHIFTRLITAVPSLREHFENVKKAIGIFVSTSI
IEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNT
LSFILEEFKSD EEVIQSFCKYKTLRNENVLETAELFNEINSIDLTHIFISHKLETISSALCDHWDTLRNALYERRISELT
GKITKSAKEKVQRSLKHE DINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQDLSLLGL
YHLLDWFVAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAIL
FVKNGLYYLGIMPQKQGRYKAL

SFEPTSEKTEGFDKMYDYFPDAAKMIPKCSTQLKAVTAHFQTHHTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYA
KKTGDQKGYREALCKWIDFRDLSKYTKTTSIDLSSLRPSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGK

LYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTP
PDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEH
PETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERRAARQAWSVVGTIKDLKQGYLSQVIHEIVDL
MIHYQAVVVLENLNFQKSKRTGIAEKAVYQQFEKMLIDKLNCLVLDYPAEKVGGVLPYQLTDQFTSFAKMGTSQSG
FLFYVPAPYTSKIDPLTGFDVDFVWTKIKNHESRKHFLGFDLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVF
EKNETQFDAQTPFIAGKRIVPVIEHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLENDSSHAIIDTMVALIRSVL
QMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPWPMADANGAYHIALKGQLLNHLKESKDLKLQNGISNQDWL
AYIQELRNPKKRKYGGYPYDVPDYA

Band 2 (B)

MKIHVVVVVHEEGKLVWINGDKGYNGLAEVGKFEKDTGIKVTVEHPDKLEEFQVAATGDGPDIIFWAHRDFGGY
AQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSIIYKDLLPNPKTWEEIPALDKELKAKGKSALMFNL
QEPYFTWPLIAADGGYAFKYENKDYKIDVGVNAGAKAGLTFVLDIKKNHMNADTDYSIAEAFNKGEMAMTINGP
WAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYE
EELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNISITSLYKAGFMGAPPKK
KRKVGGMNTQFEGFTNLYQVSKTLRFELIPQKTLKHIQEQGFIEEDKARNHYKELKPIIDRIYKYADQCLQLVQLDWE
NLSAIDSRYRKEKTEETRNALIEEQATYRNAIHDFYFIGRTDNLTDANKRHAIEYKGLFKAELFNGKVLKQLGTVTTTEHEN
ALLRSFDKFTTYFSGFYENRKNVSAEDISTAIPHRIVQDNFQKFKENCHIFRTRLITAVPSLREHFENVKKAIGIFVSTSEIEVF
SFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSD
EEVIQSFCKYKTLRNENVLETAELFNLNSIDLTHIFISHKLETISSALCDHWDTLRNLYERRISELTGKITKSAKEKVQ
RSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAAALDQPLPTTLKKQEEKEILKSQDLSLLGLYHLLDWFVAVDESNEVD
PEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFLNFQMPTLASGWDVNKEKNNGAILFVKNGLYLLGIMPQKQ
GRYK

ALSFEPTSEKTEGFDKMYDYFPDAAKMIPKCSTQLKAVTAHFQTHHTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTA
YAKTGDQKGYREALCKWIDFRDLSKYTKTTSIDLSSLRPSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLY
LFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTP
DTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHP
ETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERRAARQAWSVVGTIKDLKQGYLSQVIHEIVDL
MIHYQAVVVLENLNFQKSKRTGIAEKAVYQQFEKMLIDKLNCLVLDYPAEKVGGVLPYQLTDQFTSFAKMGTSQSG
FLFYVPAPYTSKIDPLTGFDVDFVWTKIKNHESRKHFLGFDLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVF
EKNETQFDAQTPFIAGKRIVPVIEHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLENDSSHAIIDTMVALIRSVL
QMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPWPMADANGAYHIALKGQLLNHLKESKDLKLQNGISNQDWL
AYIQELR

NPKKRKYGGYPYDVPDYA

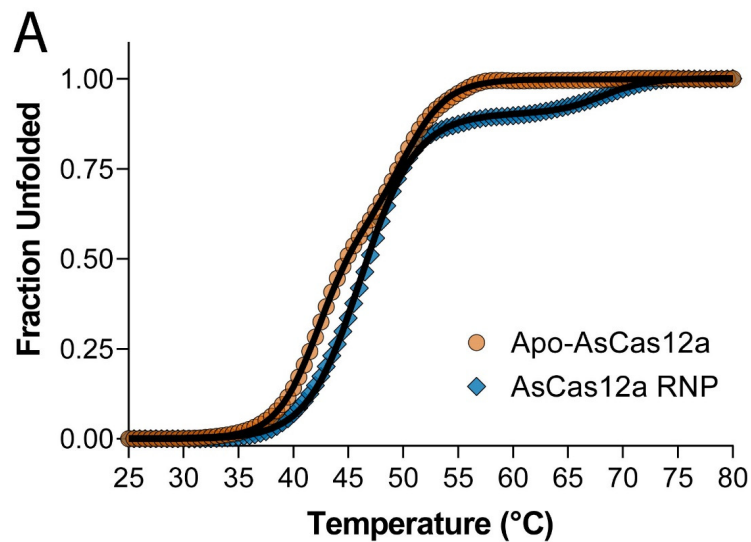
Band 3 (C)

MKIHHEEGKLVWINGDKGYNGLAIEVGGKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDHFRGGY
AQSGLLAEITPDKAFQDKLYPFTWDVRYNGKLIAYPIAVEALS LIYNDKLLPNPPKTWEEIPALDKELKAKGKSALMFNL
QEPYFTWPLIAADGGYAFKYENGYDIKDVGVNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAFNKGETAMTINGP
WAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYE
EELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQNTSITSLYKKAGFMGAPPKK
KRK

VGG

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNHDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAID
SYRKEKTEETRNALIEEQATYRNAIHDFYIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFD
KFTTYFSGFYENRKNVSAEDISTAIPHRIQDNFPKFKENCHIFRTRITAVPSLREHFENVKKAIGFVSTSIIEVFSFPFY
QLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQS
FCKYKTLRNENVLETAELFNELSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITSAKEKVQRSLKHE
DINLQEIISAAGKELSEAFKQKTSEILSHAHAAALDQPLPTTLKKQEEKEILKSQLDSELLGLYHLLDWFVAVDESNEVDPEFSAR
LTGIKLEMPSLSFYNKARNYATKKPYSVEKFLNFQMPTLASGWDVNKEKNNGAILFVKNGLYLLGIMPK

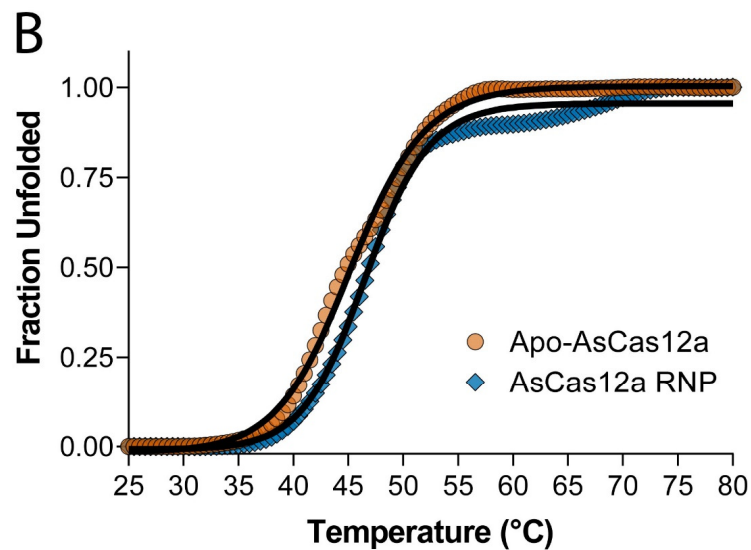
QKGRYKALSFEPTEKTSEGFDKMYDYFPDAAKMIPKCSTQLKAVTAHFQTHHTTPILLSNNFIEPLEITKEIYDLNNPEKEP
KKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDA
VETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAEFYRPKSRMKRMAHRLGKMLNKKLK
DQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIHKDRRFTSDKFFFHVPITLNYQAANSKFNQRVN
AYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVI
HEIVDLMIHYQAVVLENLNFQKSKRTGIAEKAVYQQFEKMLIDKLNCLVLDYPAEKVGGVLPYQLTDQFTSFAKM
GTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLGDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMP
AWDIVFEKNETQFQDAKGTPIAGKRIVPVIEHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLENDSSHAIIDTM
VALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPWPMADANGAYHIALKGQLLNHLKESKDLKQLNG
ISNQDWLAYIQELRNPKKRKKVGGYPYDVPDYA



Fit values for the sequential three state (folded – intermediate – unfolded) model

	$T_{m,1}$ (°C)	$\Delta H_{vH,1}$ (kcal mol ⁻¹)	$T_{m,2}$ (°C)	$\Delta H_{vH,2}$ (kcal mol ⁻¹)
Cas12a	42.2 ± 0.2	-103.3 ± 2.2	50.2 ± 0.3	-105.7 ± 9.8
Cas12a-RNP	46.2 ± 0.2	-82.1 ± 1.4	68.2 ± 1.1	-113.9 ± 4.3

95% confidence intervals (2 x STD) reported



Fit values for the two state (folded -- unfolded) model

	$T_{m,1}$ (°C)	$\Delta H_{vH,1}$ (kcal mol ⁻¹)
Cas12a	42.2 ± 0.2	-103.3 ± 2.2
Cas12a-RNP	46.2 ± 0.2	-82.1 ± 1.4

95% confidence intervals (2 x STD) reported

Figure S23. Model fitting of DSF thermal melt data for apo-AsCas12a and AsCas12a RNP. **(A)** Three-state sequential unfolding model and **(B)** two-state unfolding model.