JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

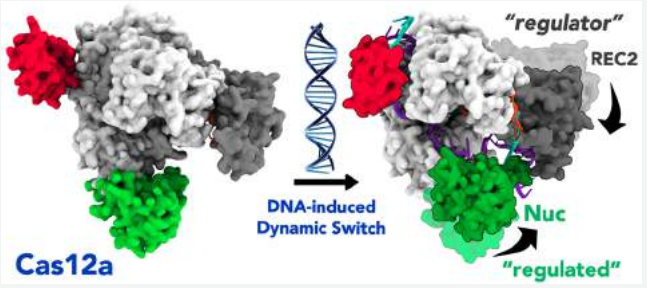
Molecular Dynamics Reveals a DNA-Induced Dynamic Switch Triggering Activation of CRISPR-Cas12a

Aakash Saha,[‖] Pablo R. Arantes,[‖] Rohaine V. Hsu, Yogesh B. Narkhede, Martin Jinek, and Giulia Palermo*

Cite This: https://dx.doi.org/10.1021/acs.jcim.0c00929 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: CRISPR-Cas12a is a genome-editing system, recently also harnessed for nucleic acid detection, which is promising for the diagnosis of the SARS-CoV-2 coronavirus through the DETECTR technology. Here, a collective ensemble of multimicrosecond molecular dynamics characterizes the key dynamic determinants allowing nucleic acid processing in CRISPR-Cas12a. We show that DNA binding induces a switch in the conformational dynamics of Cas12a, which results in the activation of the peripheral REC2 and Nuc domains to enable cleavage of nucleic acids. The simulations reveal that large-amplitude motions of the Nuc domain could favor the conformational activation of the system toward DNA cleavages. In this process, the REC lobe plays a critical role. Accordingly, the joint dynamics of REC and Nuc shows the tendency to prime the conformational transition of the DNA target strand toward the catalytic site. Most notably, the highly coupled dynamics of the REC2 region and Nuc domain suggests that REC2 could act as a regulator of the Nuc function, similar to what was observed previously for the HNH domain in the CRISPR-associated nuclease Cas9. These mutual domain dynamics could be critical for the nonspecific binding of DNA and thereby for the underlying mechanistic functioning of the DETECTR technology. Considering that REC is a key determinant in the system's specificity, our findings provide a rational basis for future biophysical studies aimed at characterizing its function in CRISPR-Cas12a. Overall, our outcomes advance our mechanistic understanding of CRISPR-Cas12a and provide grounds for novel engineering efforts to improve genome editing and viral detection.



INTRODUCTION

CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins) is a part of the bacterial immune system that confers protection against invading viruses. In 2012, the discovery that the CRISPR-associated protein Cas9 is an RNA-programmable endonuclease[1] enabled precise manipulation of nucleic acids, launching an unprecedented genome editing revolution.[2] Recently, a novel CRISPR protein, Cas12a,[3] emerged as a promising tool for innovative applications of the CRISPR technology, such as nucleic acid detection.[4] The CRISPR-Cas12a system is the basis of the DETECTR technology,[5] which allows rapid detection of viruses including SARS-CoV-2, which is spreading across multiple countries.

At the molecular level, the CRISPR-Cas9 and CRISPR-Cas12a systems confer precise double-stranded DNA (dsDNA) breaks by using CRISPR RNAs (crRNA) as a guide for molecular recognition of substrate DNA.[2] Thanks to a short protospacer adjacent motif (PAM) sequence in the viral DNA, these systems can be programmed to recognize any DNA sequence of interest, therefore enabling genome editing.[6] Despite sharing functional similarities with CRISPR-Cas9, Cas12a has a distinct evolutionary history and intriguing mechanistic properties.[7−9] Structural and biochemical studies of Cas12a have revealed a bilobed architecture, similar to what was observed in Cas9. It comprises a recognition (REC) lobe and a nuclease (NUC) lobe connected by a wedge domain (WED (Figure 1).[10−16] The REC lobe includes two α-helical domains (REC1 and REC2), which mediate nucleic acid binding. The NUC lobe consists of the RuvC and Nuc domains, flanked by the PAM-interacting (PI) domain, which binds to the PAM sequence in the DNA. The guide RNA forms a heteroduplex with one DNA strand (the target strand TS), while the other nontarget strand (NTS) is accommodated within a cleft formed by the RuvC and Nuc domains (Figure 1b,c). Unlike Cas9, which cleaves the TS and NTS using two specific catalytic domains, HNH and RuvC, respectively,

Received: August 10, 2020

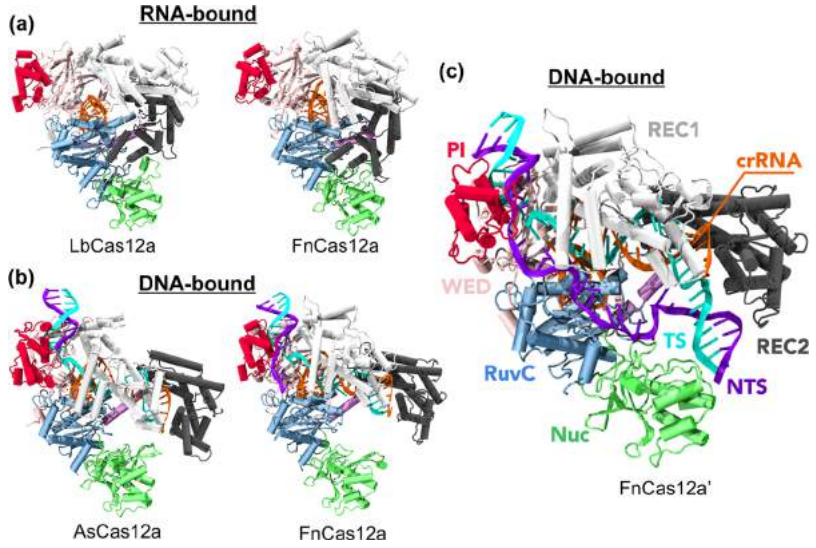


Figure 1. Overview of CRISPR-Cas12a. X-ray structures of Cas12a proteins in complexes with guide RNA (a) or guide RNA and target DNA (b, c) across different species, *viz. Lachnospiraceae* bacterium Cas12a (LbCas12a),[10] *Acidaminococcus sp.* Cas12a (AsCas12a),[11,12] and *Francisella novicida* Cas12a (FnCas12a).[13−16] Cas12a proteins are shown in cartoon format, highlighting individual protein domains in different colors. The guide RNA (orange), the DNA target strand (TS, cyan), and the DNA nontarget strand (NTS, violet) are shown as ribbons. The DNA-bound states include a cleaved NTS (a) and a complete NTS (b).

Cas12a performs cleavage of both the DNA strands using a single active site located within the RuvC domain.[13] This has raised questions on the role of the additional Nuc domain, which was initially thought to cleave the DNA TS, enacting the role of HNH in Cas9.[12] More recent studies have implied that the Nuc domain instead plays a role in NTS and DNA binding.[13,15,16] However, it is unclear how conformational changes of the Nuc domain would activate DNA cleavages or facilitate the exchange of the TS and NTS within the RuvC active site to achieve a dsDNA break.

Atomic resolution structures have captured Cas12a bound to guide RNA alone (for *Lachnospiraceae* bacterium LbCas12a[10] and *Francisella novicida* FnCas12a[13]) and guide RNA and a target dsDNA (for *Acidaminococcus sp.* AsCas12a[12] and FnCas12a,[13] Figure 1). For the DNA-bound states, both AsCas12a and FnCas12a have been determined in complex with a cleaved NTS (Figure 1b),[12−14,16] while FnCas12a has also been obtained including a longer NTS that binds within RuvC and reconciles with the TS (Figure 1c, hereafter referred to as FnCas12a′).[15] Collectively, these atomic-resolution structures offered intricate details about the Cas12a ribonucleoprotein complex, suggesting extensive conformational plasticity. However, it is unclear how this structural plasticity could contribute to the "open-to-closed" conformational change of the protein, which is thought to facilitate substrate DNA binding and subsequent cleavage.[16] In light of these experimental findings, investigating the protein structural plasticity at the atomic level through Molecular Dynamics (MD) is critical for understanding its biological functions and rational engineering of novel Cas12a-based tools for genome editing and viral nucleic acid detection.

Molecular simulations have previously contributed in the understanding of fundamental biophysical aspects of the CRISPR-Cas9 system,[17−22] revealing a striking plasticity of the ribonucleoprotein complex.[17−21] These studies provided predictive insights into the dynamic behavior of Cas9, many of which have been corroborated by single-molecule FRET experiments[23] and cryoEM,[24] providing also a framework for rational design of Cas9 for improved genome editing.[25,26]

Here, we present a multimicrosecond length MD study of CRISPR-Cas12a, to characterize conformational plasticity of the protein and its interplay with the nucleic acids over long time-scales. We collected an overall ensemble of ∼20 μs, carried out in multiple replicates, considering different states and across species. The simulations reveal a "switch" in the conformational dynamics of Cas12a upon DNA binding that results in the activating motions of the peripheral REC2 and Nuc domains. In agreement with previous structural[13,14] and single-molecule experiments,[16] this increased mobility of REC2 and Nuc upon DNA binding could enable the conformational changes associated with DNA cleavage. The simulations also reveal an important role of the Nuc domain, whose large-amplitude motions could enable the conformational activation of the system for completing DNA cleavages. In this process, the REC lobe is critical in aiding the conformational dynamics of Nuc. Indeed, highly coupled dynamical motions of REC2 and Nuc suggest that REC2 could act as a regulator of the Nuc function, as previously observed for the HNH domain in CRISPR-Cas9.[19,23−25,27] Considering the key role of the REC lobe for the specificity of Cas9,[25,28,29] and as recently found in Cas12a,[30] our outcomes now call for future studies aimed at characterizing the functional role of REC and Nuc in CRISPR-Cas12a.

RESULTS

Conformational Flexibility of Cas12a Bound to RNA and DNA. Building on the available structures of Cas12a (Figure 1), we collected ensembles of μs-length MD simulations. We examined LbCas12a, AsCas12a, and FnCas12a to assess differences and similarities in the dynamics across various species and states (i.e., crRNA-bound vs DNA-bound). For each model system of Cas12a, MD simulations were carried out in explicit solvent, obtaining multiple μs-length trajectories (i.e., 4 replicates of ∼1 μs each) and reaching an overall sampling of ∼20 μs. The choice of

https://dx.doi.org/10.1021/acs.jcim.0c00929
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

simulating multiple and independent μs-length trajectories was motivated by the need to achieve solid statistics for the purpose of our analysis and by our previous theoretical investigations of the parent CRISPR-Cas9.[17−19] Those studies had shown that multiple ns-to-μs MD trajectories are critical for describing the interdependent dynamics of the protein domains and their interplay with the nucleic acids. Notably, in the present work, data analysis was performed in analogy to our early multimicrosecond MD investigations of CRISPR-Cas9.[17] This enabled a fair comparison of the dynamical properties between the two CRISPR-Cas systems.

To develop an initial understanding of the overall flexibility of the system when bound to RNA and DNA, we employed root-mean-square fluctuation (RMSF) analysis, which is a traditional mode of measuring protein flexibility (Figure S1). Additionally, to investigate whether the observed changes in fluctuations are maintained across the ns-to-μs runs, we computed a time-dependent RMSF (*t*-RMSF, Figure 2a). As a result, the *t*-RMSF reveals high fluctuations of the PI domain in the RNA-bound LbCas12a and FnCas12a (Figure 2a) complexes, which are conserved along the simulated runs. Upon DNA binding, the flexibility of the PI region is remarkably reduced for both AsCas12a and FnCas12a, as stabilized by the binding of DNA. Simultaneously, the flexibility of REC2 and Nuc shows a remarkable increase. The change in flexibility of the PI, REC2, and Nuc domains upon DNA binding is shown in the 3D structure of FnCas12a, indicating regions of high fluctuations using thicker tubes (Figure 2b). Interestingly, a high flexibility of REC2 and Nuc is preserved in the FnCas12a′ complex, where a dsDNA locates in between REC2 and Nuc (Figure 2b, lower panel). This is notable because the binding of a dsDNA commonly stabilizes the surrounding protein scaffold, as observed at the level of the PI domain. On the other hand, the RMSF of the RuvC domain, which is responsible for both NTS and TS cleavages, reveals low fluctuations (Figure S1). This is in agreement with high structural stability of this conserved domain,[31] also observed in MD simulations of CRISPR-Cas9.[17] The remaining protein domains do not show significant flexibility upon DNA binding. Analysis of the root-mean-square deviation (RMSD) of the protein Cα atoms reveals that the protein backbone reaches a similar stability in both RNA- and DNA-bound states (i.e., the RMSD reaches ∼4−5 Å, Figure S2). This indicates that upon DNA binding, the overall stability of the protein is preserved, but the flexibility of different protein domains changes, as shown by the analysis of the *t*-RMSF (Figure 2a).

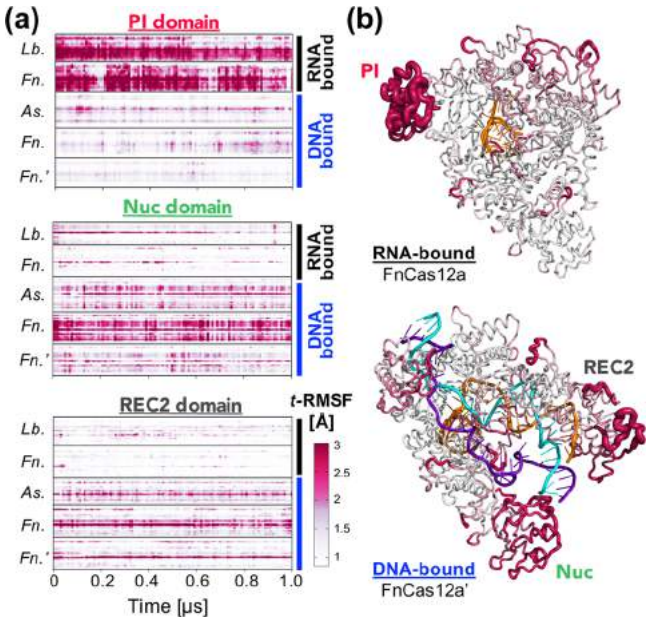


Figure 2. Cas12a flexibility along the dynamics. (a) Time-dependent Root Mean Square Fluctuations (*t*-RMSF, in Å), computed for the PI (top), Nuc (center), and REC2 (bottom) domains of the LbCas12a (*Lb.*), AsCas12a (*As.*), and FnCas12a (*Fn.*) systems. The RNA-bound and DNA-bound systems are indicated using black and blue bars, respectively. *t*-RMSF values are colored from white (low fluctuations) to magenta (high fluctuations), accordingly to the scale on the bottom right. (b) The averaged RMSF values are plotted on the 3D structures of the RNA-bound FnCas12a (top) and DNA-bound FnCas12a′ (bottom), indicating protein regions of high fluctuations using thicker tubes (color-coded according to the *t*-RMSF scale). The RNA (orange) and the DNA TS (cyan) and NTS (violet) are shown as a cartoon.

Large-Amplitude Motions and Conformational Ensemble. Aiming to dissect the large-scale collective motions of the Cas12a-nucleic acid complexes and characterize its essential degrees of freedom, we performed Principal Component Analysis (PCA). With this analysis, the "essential dynamics"[32] of the system is described along the first principal mode of motion (i.e., principal component 1, PC1), providing information on the large-amplitude motions of the complexes and, in turn, on their functional dynamics. PCA has been performed considering all the FnCas12a systems (i.e., the RNA-bound FnCas12a, DNA-bound FnCas12a, and FnCas12a′). In detail, we combined the collected ensembles arising from the compared systems and subjected to RMS-fit to the same reference configuration, ensuring consistency of the eigenbasis and motions of the PCs (details are reported in the Methods section).

In Figure 3a, PC1 is plotted over the 3D structures of FnCas12a in the RNA- and DNA-bound states, where the arrows indicate the direction and relative amplitude of motions. In the RNA-bound state, the PI domain displays large-amplitude motions, which are directed toward the cleft that accommodates the PAM of the DNA. These motions agree well with previous structural analyses, suggesting that the inward movement of the PI domain would accompany PAM binding.[11,14] Upon binding of DNA, the PI domain remarkably reduces the amplitude of its motions, while REC2 and Nuc display substantially increased amplitude motions. The large-amplitude of the motions of REC2 and Nuc is preserved in all DNA-bound states, in the presence of a cleaved NTS (i.e., in the FnCas12a system) and in the FnCas12a′ system, where a complete NTS rehybridizes with the TS to form a duplex in the vicinity of REC2 and Nuc (Figures 1c and 3a). As previously suggested by Stella and co-workers based on single-molecule FRET,[16] an increased flexibility of REC2 and Nuc upon binding of DNA could have a functional role. Indeed, the mobility of these regions could favor the exchange of the NTS and TS, as well as to attain the 5′-3′ polarity of the TS required for cleavage within the RuvC active site.[16] In this respect, the catalytic RuvC domain displays short-amplitude motions, in agreement with the high structural stability of this conserved protein domain.[17,31] Interestingly, REC1 shows motions of smaller amplitude, which arise from strong interactions with

https://dx.doi.org/10.1021/acs.jcim.0c00929
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

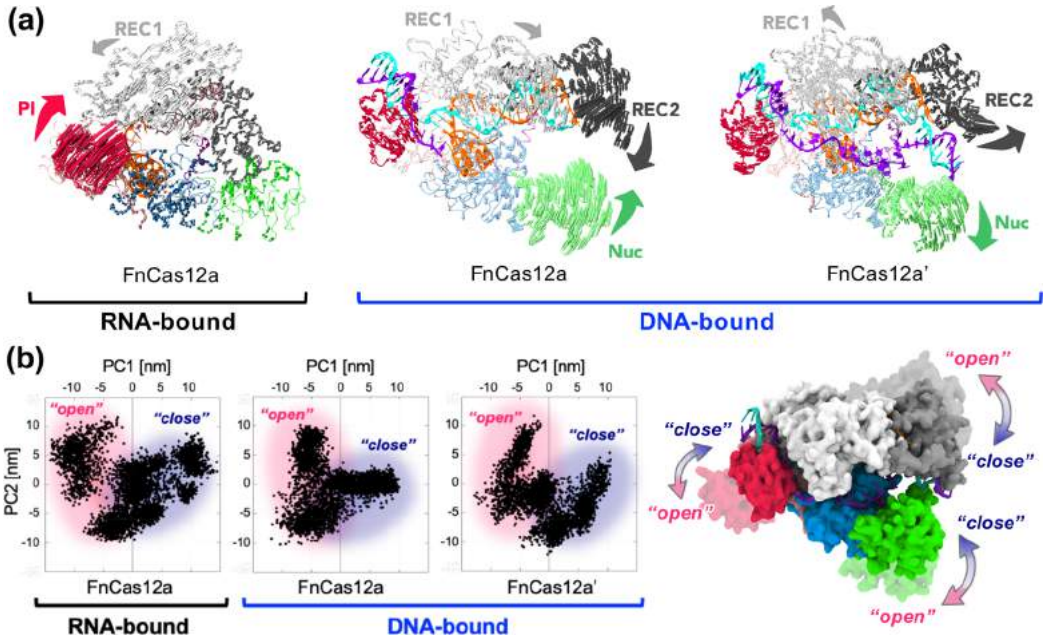


Figure 3. Large-amplitude motions and conformational space adopted by CRISPR-Cas12a. (a) "Essential dynamics", derived from the first principal component (PC1) of the individual protein domains of FnCas12a bound to RNA (left) and DNA (right), shown using arrows of sizes proportional to the amplitude of motions. For the DNA-bound states, the "essential dynamics" is shown for the presence of a cleaved NTS (FnCas12) and complete NTS (FnCas12a′). (b) Projections of the first and second principal components (PC1 vs PC2) from MD simulations of the RNA- and DNA-bound FnCas12a systems. The PC1 vs PC2 plots characterize the conformational space sampled by the FnCas12a into regions in which the protein is "open" (red cloud) and "closed" (blue cloud). A cartoon of FnCas12a indicating the "open-to-closed" conformational transition is shown on the right.

the RNA:DNA hybrid and reflect the function of this domain in anchoring the target DNA strand.[14]

The direction of FnCas12a principal motions upon DNA binding is of particular interest, as differences are observed in the presence of a complete NTS, as opposed to cleaved NTS. In FnCas12a, in which the NTS is cleaved, REC2 and Nuc move toward each other (Figure 3a), and REC1 also moves toward RuvC and Nuc. In the FnCas12a′ system (including a complete NTS), REC1-2 and Nuc preserve opposite motions but directed away from each other. Nuc points its motions out of the protein framework, as also observed for REC1-2. To gain further insights on this observation, we performed volumetric analysis on the equilibrium trajectories (details are reported in the SI). We measured the volume of the cavity between the REC and NUC lobes, which forms the RuvC binding groove. This revealed a contraction of the groove in the FnCas12a system (Figure S6). On the other hand, in the FnCas12a′ system, an expansion of the groove is observed. Considering that Cas12a cuts the NTS first,[13,16,33] the dynamical differences observed in the presence and absence of a complete NTS suggest possible conformational rearrangements of Nuc and REC1-2 upon NTS cleavage, which could allow the subsequent processing of the TS.[16]

Overall, the direction of the motions of these domains (as opposite to each other) indicates the tendency toward the "opening" and "closure" of the protein to accommodate and cleave the nucleic acids.[11,14,16] This is a functional feature shared with CRISPR-Cas9, as observed in previous simulation studies and through structural analyses.[17,34,35] To characterize the conformational space adopted by the ribonucleoprotein, we plotted the first versus the second principal components (i.e., PC1 vs PC2, Figure 3b). As a result, the PC1 vs PC2 plots identify two states, which depict the "open" and "closed" conformations of the protein well (schematically drawn using a cartoon of FnCas12a in Figure 3b), observed in both the RNA-bound and DNA-bound forms of FnCas12a. Notably, a range of conformational states from "open" to "closed" was also observed through PCA of multimicrosecond MD runs on CRISPR-Cas9.[17,35] This indicates that both ribonucleoprotein complexes have a general tendency toward an "open-to-closed" breathing to allow nucleic acid association. In this respect, it is worth noting that DNA binding-induced conformational changes occur at remarkably slower rates than what is possible to simulate using classical MD.[36,37] Yet, our data show a good coverage of the conformational landscape (Figures 3b and S7), implying that the PCA well-represented the large-scale dynamics of the system. It is also notable that in the DNA-bound states, the conformational space explored by the protein is slightly restricted as a result of global stabilization due to the binding of the DNA (FnCas12a) and a complete NTS (FnCas12a′). This also agrees well with previous PCA of the CRISPR-Cas9 complex.[17] In summary, the Cas12a protein preserves a general tendency toward an "open-to-closed" conformational transition in both the RNA-bound and DNA-bound states. However, as described above, prior to DNA binding, large-amplitude motions and high flexibility are observed at the level of the PI domain. On the other hand, upon DNA binding, the largest motions shift to REC2 and Nuc.

Finally, it is noteworthy that the PCs arising from the LbCas12a, AsCas12a, and all FnCas12a systems cannot be directly compared owing to the inconsistency in the eigenbasis. Acknowledging this fact, two independent PCAs have been performed for the RNA-bound LbCas12a and DNA-bound AsCas12a systems to gain insights into their "essential dynamics". We observed that the RNA-bound LbCas12a

https://dx.doi.org/10.1021/acs.jcim.0c00929
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

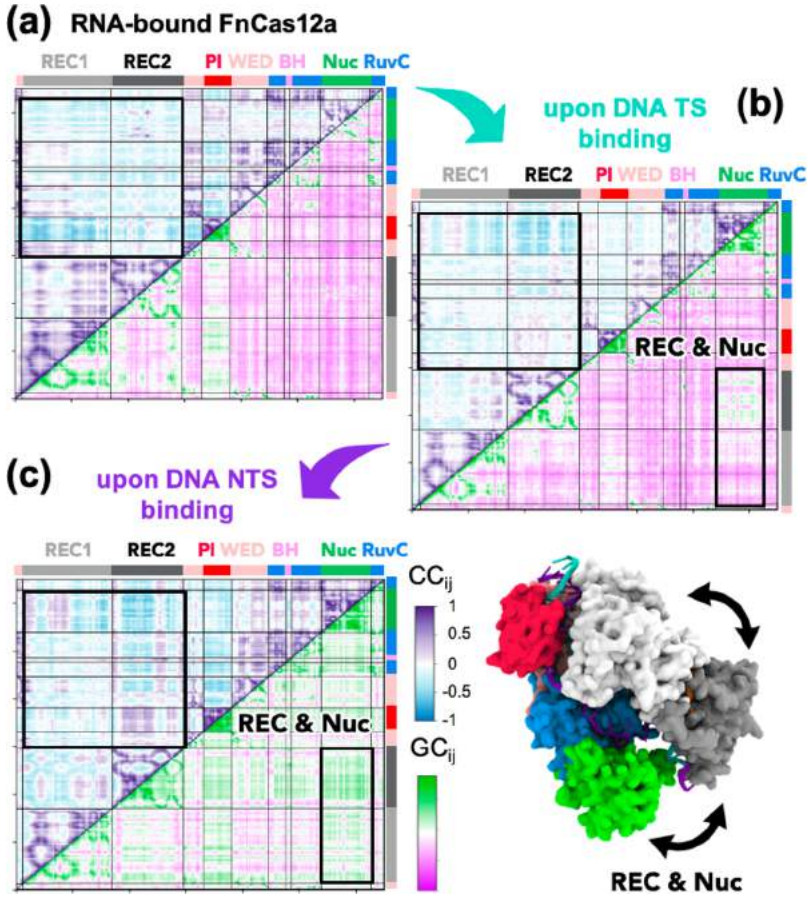


Figure 4. Correlated motions of CRISPR-Cas12a. Cross-Correlation (CC_{ij}, upper triangles) and Generalized Correlations (GC_{ij}, lower triangles) matrices, computed for FnCas12a in the RNA-bound state (a) and upon DNA binding, in the presence of a cleaved NTS (b) and a complete NTS (c). Data are averaged over 4 simulation replicas of ~1 μs each. The strength of the CC_{ij} is colored blue (for $CC_{ij} \geq 0$, lockstep motions) to violet (for $CC_{ij} \leq 0$, anticorrelated motions), while the GC_{ij} are green (correlated) to magenta (not correlated). Color scales are at the bottom. The protein sequence is also shown. Boxes are used to highlight anticorrelated CC_{ij} motions between the REC and NUC lobes and highly coupled GC_{ij} between REC & Nuc, also depicted in the cartoon of FnCas12a (bottom right).

displays large amplitude motions in the PI domain directed inward. On the other hand, AsCas12a bound to a cleaved NTS exhibits the largest motions in the REC2 and Nuc domains, moving toward each other (Figure S7). This is qualitatively consistent with the switch in the "essential dynamics" observed in the FnCas12a systems upon DNA binding.

Coupled Motions of Protein Domains. To investigate the interdependent conformational dynamics among spatially distant protein domains, we performed dynamic correlation analysis. This analysis was performed and averaged over multiple MD trajectories using two different methods. We computed the traditional Pearson cross-correlation (CC_{ij}) coefficients, which measure the collinear correlation between two Cα atoms (i and j), determining whether they tend to move in lockstep (positive CC_{ij}) or show opposed motions (negative CC_{ij}). The CC_{ij} analysis only detects correlations that are collinear with each other, neglecting correlated motions occurring out of phase. Hence, we also employed a generalized correlation (GC_{ij}) scheme.[38] This measures the degree of correlation between Cα atoms based on their mutual information, providing a normalized measure of how much information on one atom's position is provided by that of another atom. The method, however, does not distinguish positive vs negative correlations, neglecting the description of opposite atom's motions. Hence, when employed together, the CC_{ij} and GC_{ij} schemes are powerful in describing the interdependent dynamics of proteins.

The CC_{ij} matrix (i.e., a two-by-two plot of the Cα CC_{ij} coefficients) of FnCas12a shows a conserved pattern of correlated/anticorrelated motions in both RNA- and DNA-bound states (Figure 4, upper triangles), which are also found in the other Cas12a systems (Figures S8−S12). The REC lobe (i.e., REC1-2) preserves anticorrelated motions with the NUC lobe (including RuvC, Nuc, PI, and WED). This indicates the tendency of REC to move in an opposite way with respect to NUC, thereby favoring the "open-to-closed" conformational transition underlying nucleic acid binding.[11,14,16] The GC_{ij} matrix, which goes beyond the reach of a Pearson-like CC_{ij} analysis, captures the overall dependencies of the protein motions (Figure 4, lower triangles). In the RNA-bound state of FnCas12a (Figure 4a), coupled motions are only detected among REC1-2 and PI. On the other hand, in the DNA-bound FnCas12a (Figure 4b) correlated motions of REC2 and Nuc become prominent, while REC1 also displays correlations with Nuc (although at a lower extent). In the presence of a complete NTS (i.e., in the FnCas12a′, Figure 4c), the overall system's GC_{ij} becomes more intense, preserving a high degree of coupling between REC1-2 and Nuc. We note that intense correlations between REC2 and Nuc are found in all simulated replicas of each DNA-bound system (Figures S8−S10). This

https://dx.doi.org/10.1021/acs.jcim.0c00929
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

indicates a shift in the correlated motions from one region of the GC_{ij} matrix (corresponding to REC1-2 and PI, in the RNA-bound state) to the REC1-2/Nuc domains upon DNA binding, finally encompassing the entire protein when bound to a complete NTS. To further evaluate the interdependent coupling between protein domains, we computed the per-domain GC_{ij} scores (Cs), which accumulate (and normalize) the GC_{ij} for each protein domain with each other (details are reported in the SI).[39] As a result, the per-domain Cs matrices highlight the high degree of coupling between Nuc and REC2 upon DNA binding (Figure S13). AsCas12a, which has been simulated only in the DNA-bound state, also shows high GC_{ij} between REC1-2 and Nuc (Figure S11). The RNA-bound LbCas12a preserves a pattern of highly correlated motions (Figure S12). However, for these species, a direct comparison between the RNA- and DNA-bound states is not possible due to the lack of structural information. Finally, it is interesting to note that only in the presence of complete nucleic acids (i.e., in the FnCas12a′, Figure 4c), highly correlated motions are observed across the entire protein. This has also been established through MD simulations of the DNA-bound Cas9,[19,35,40] indicating a mechanism of interdomain allosteric communication.[41,42] This novel hypothesis in CRISPR-Cas12a arising from our data now warrants further computational investigations of protein allostery that are currently being pursued in our laboratory. This will likely also clarify the most important residues sustaining the functional motions and the communication mechanism.[43,44]

DISCUSSION

Dynamic "Switch" upon DNA Binding. MD simulations of Cas12a indicate a change in the structural flexibility and in the conformational dynamics of the protein upon DNA binding. In the RNA-bound states, the PI region displays high flexibility (Figure 2) and large-amplitude motions directed toward the cleft that accommodates the PAM of the substrate DNA (Figure 3a). Upon DNA binding, the flexibility of the PI region is notably reduced, while the flexibility of REC2 and Nuc increases. Furthermore, the PI domain reduces the extent of its motions, while REC2 and Nuc display the largest amplitude motions. This suggests that the binding of DNA, which is initiated at the level of the PI domain in Cas12a, quenches the conformational flexibility of PI and induces the activation of peripheral large-scale motions at the level of REC2 and Nuc. This high flexibility and large-scale breathings of REC2 and Nuc upon DNA binding are also observed in the FnCas12a′ complex, where REC2 and Nuc directly interact with a duplex region of the DNA substrate (Figure 1c). Taken together, analysis of the protein fluctuations (Figure 2) and large-amplitude motions (Figure 3) indicates a switch in the Cas12a dynamics upon DNA binding, which results in quenching of PI domain mobility and activating motions of the peripheral REC2 and Nuc domains. This switch in the Cas12a dynamics rationalizes some functional aspects and previous observations. Indeed, the high flexibility of the PI domain prior to DNA binding could favor its inward conformational change for accommodating the PAM.[11,14] Subsequent DNA binding, which is initiated at the level of the PI region, results in curtailment of the PI domain dynamics. On the other hand, REC2 and Nuc increase in flexibility, also when binding to a complete NTS (as in FnCas12a′, Figures 2 and 3). This is notable because the binding of dsDNA commonly induces the stabilization of the surrounding protein framework, which is indeed observed at the level of the PI domain. The high flexibility of REC2 and Nuc in the presence of complete nucleic acids also agrees with a very recent study revealing that the DNA in this region is intrinsically highly flexible.[45] Moreover, a high degree of flexibility of these regions upon DNA binding and during the formation of an RNA:DNA hybrid has also been observed through cryoEM and single-molecule experiments.[16,46] This increase in flexibility upon DNA binding is possibly a consequence of the need to cleave both DNA strands by a single active site located within the RuvC domain. This requires an exchange of the NTS and TS within the active site, which in turn necessitates conformational changes (and increased mobility) of the interacting protein domains, namely the REC lobe and Nuc.[16,47] Moreover, high conformational plasticity of REC and Nuc is also needed for the release of the product,[48] as well as for enabling the nonspecific cleavage of single-stranded DNA (ssDNA) accessing the RuvC−Nuc interface.[4,15,49] This is a critical property that allows leveraging of Cas12a for the diagnosis of viral nucleic acids in the DETECTR technology.[5]

The switch in the conformational dynamics of Cas12a is also reflected by the analysis of the generalized correlations (GC_{ij}), which capture the overall dependencies of the protein motions. This analysis shows a change in the correlated motions of FnCas12a domains, revealing that correlations between REC1-2 and Nuc increase upon DNA binding (Figure 4, lower triangles). Notably, correlated motions of REC2 and Nuc are preserved across the DNA-bound Cas12a systems (Figures S8−S12), suggesting a coupled dynamical function, where the motions of REC2 allow conformational changes in Nuc. This observation is particularly interesting in light of similar observations previously reported for the Cas9 enzyme. Indeed, in the DNA-bound form of Cas9, REC2 has been shown to play a critical role in regulating the conformational activation of the HNH domain for TS cleavage. Specifically, single-molecule experiments[23,25,27] and recent cryo-EM structures[24] have shown reciprocal conformational changes of REC2 and HNH domains to allow the latter to dock at the TS for cleavage. Accordingly, MD simulations have revealed that the HNH domain approaches the TS in concert with a transition of REC2,[19] while displaying correlated motions similar to what is observed for REC2 and Nuc in Cas12a. This suggests that despite being evolutionarily different than Cas9,[9] REC2 of Cas12a could play a similar regulatory function on the conformational activation of Nuc to allow the cleavage of the TS.

Conformational Activation for TS Processing. Cas12a uses the single RuvC domain to cleave both DNA strands, differing from Cas9, which cleaves the NTS and the TS using the RuvC and HNH nuclease domains, respectively. This has questioned the role and function of the additional Nuc domain, which was initially thought to cleave the DNA TS.[12] It is unclear, in fact, whether and how conformational changes of Nuc would activate DNA cleavages or facilitate the exchange of the TS and NTS within the RuvC active site for cleavage.

Here, molecular simulations of Cas12a upon DNA binding offer interesting insights on the dynamic role of Nuc. We observed significant differences in the direction of the essential motions of Cas12a in the presence/absence of the DNA NTS (Figure 3A). Indeed, in the presence of a complete NTS (i.e., in FnCas12a′), REC1-2 and Nuc display overall outward motions, leading to an expansion of the RuvC binding groove

https://dx.doi.org/10.1021/acs.jcim.0c00929
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

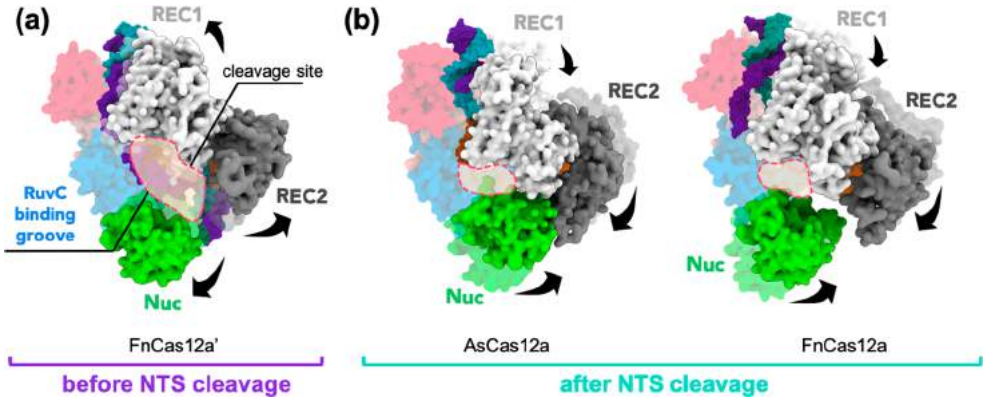


Figure 5. Changes in the RuvC binding groove of CRISPR-Cas12a before and after NTS cleavage. (a) Before NTS cleavage, an expansion in the RuvC binding groove is observed as mediated by outward motions of REC1-2 and Nuc. (b) After NTS cleavage, inward motions of REC1-2 and Nuc dynamics lead to a contraction of the RuvC binding groove. The outward/inward motions of the REC1-2 and Nuc domains are shown in transparent-to-mat colors. Arrows are also used to indicate the conformational transitions. The RuvC binding groove, which is located in between the REC and NUC lobes, is also highlighted.

(Figure 5A). On the other hand, upon cleavage of the NTS (i.e., in AsCas12a and FnCas12a), REC1-2 and Nuc point inward (Figure 5B). This results in a contraction of the groove between the TS and the RuvC active site after NTS cleavage, which is also confirmed through volumetric analysis performed on the equilibrium trajectories (Figure S6). This observation is of particular interest. Considering that Cas12a cuts the NTS first,[13,16,33] this reinforces the outcomes of single-molecule FRET,[16] suggesting that Nuc and REC2 are highly dynamic and could allow the cleavage of the TS. Indeed, given the presence of a single catalytic site within the RuvC domain, after initial cleavage of the NTS, the TS should access the active site with conformational changes of the interacting protein domains. In light of this fact, our atomistic simulations indicate the tendency of Nuc to bend toward REC2, suggesting the narrowing of the groove between the TS and the RuvC active site. Concurrently, the motion of REC2, nearing Nuc and RuvC, shows its propensity to push the DNA TS toward the RuvC active site. REC1 also points toward RuvC, aiding the RuvC active site to access the TS. Hence, the REC lobe and Nuc would cooperate in the activation of the system toward TS cleavage.

Correlation analyses have indicated highly coupled motions between REC2 and Nuc (Figure 4). This suggests a coupled dynamical function, where the motions of REC2 are fundamental to allow the conformational changes of Nuc. This resembles the reciprocal dynamical role of REC2 and HNH in Cas9, where REC2 regulates conformational changes of the catalytic HNH domain to allow TS cleavage.[19,23−25] In light of these observations, the dynamics of REC2 in Cas12a could assist the conformational changes of Nuc and acts as a "regulator" of its function. In support of this hypothesis, single-molecule studies of Cas12a have shown that the REC2 and Nuc domains show the largest conformational rearrangements.[16] Hence, the coupled dynamics observed here between REC2 and Nuc now calls for new biophysical experiments to assess the role of the REC lobe in the conformational activation of CRISPR-Cas12a. In this respect, we note that in the case of Cas9, biophysical studies have shown that conformational changes of the REC lobe are also critical for the enzyme's specificity,[25,27] and mutations in the REC domain can reduce its off-target activity.[25,28,29] Moreover, a recent study has shown that point mutations in the REC lobe of AsCas12a can reduce off-target effects.[30] These findings and the results of the current study thus motivate further investigation on the functional role of REC and Nuc in CRISPR-Cas12a.

CONCLUSIONS

Here, all-atom MD simulations characterize the structural plasticity of CRISPR-Cas12a and the dynamic determinants underlying nucleic acid association. On a collective sampling of ∼20 μs, carried out over multiple states and across different species, we reveal that DNA binding induces a switch in the conformational dynamics of Cas12a, which results in quenching motions of the PAM interacting domain and activating motions of the peripheral REC2 and Nuc domains. This switch in the Cas12a dynamics rationalizes crucial functional aspects. Indeed, the increased flexibility of REC2 and Nuc upon DNA binding could enable the conformational changes associated with DNA cleavage.[16] Considering that the core of the RuvC catalytic domain is highly rigid,[17,31] the observed flexibility of the adjacent Nuc domain and REC2 could contribute to the exchange of the DNA strands within the Cas12a active site for sequential cleavage,[16] rapid release,[48] and subsequent nonspecific cleavage of ssDNAs.[4] This property is at the core of the DETECTR technology for rapid viral nucleic acid detection, especially that of SARS-CoV-2.[5] This suggests that the mutual dynamics of REC2 and Nuc observed here could be critical for the nonspecific binding of ssDNAs and thereby for the underlying mechanistic functioning of the DETECTR technology.

In-depth analysis of the dynamics upon DNA binding also offers mechanistic insights into the role of Nuc, whose function is incompletely understood. Accordingly, the joint dynamics of REC and Nuc shows the tendency to promote the conformational transition of the DNA TS toward the RuvC active site through opposite and concerted motions. REC2 and Nuc also display highly coupled dynamics, suggesting that REC2 could act as a regulator of the Nuc function, as previously observed for the HNH domain in Cas9.[19,23−25,27] In the case of Cas9, the REC lobe plays a critical role in the enzyme's activation and specificity.[25,28,29] Moreover, point mutations in REC of AsCas12a can reduce off-target effects,[30] which limit the applicability of the CRISPR technology. In light of these results, our outcomes motivate future investigations to

https://dx.doi.org/10.1021/acs.jcim.0c00929
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

characterize the functional role of REC and Nuc in CRISPR-Cas12a. It is also notable that our data suggest an unforeseen allosteric communication in CRISPR-Cas12a, which we have previously described in CRISPR-Cas9,[35,40,41] in agreement with experimental data.[6,25] This hypothesis grants now in-depth investigations, which we are currently pursuing building on our interests in the computational determination of protein allostery.

Overall, our work provides an atomic-level characterization of the CRISPR-Cas12a conformational dynamics, with insights into substrate DNA binding and cleavage. The mechanistic understandings arising from molecular simulations are of fundamental importance for further experimental studies aimed at a full characterization of the dynamic features of Cas12a. These outcomes can contribute to engineering efforts aimed at improving the CRISPR-Cas12a technology toward more efficient and specific genome editing and viral detection.

MATERIALS AND METHODS

Structural Models. Molecular simulations have been performed on five model systems of CRISPR-Cas12a, based on the available X-ray structures. The RNA-bound states have been based on two X-ray structures: *Lachnospiraceae* bacterium Cas12a (LbCas12a) solved at 2.38 Å resolution (5id6.pdb)[10] and *Francisella novicida* Cas12a (FnCas12a) solved at 3.34 Å resolution (5ng6.pdb).[13] The DNA-bound states have been based on three X-ray structures. We considered the structures of *Acidaminococcus sp*. Cas12a (AsCas12a) and FnCas12a, in which the NTS is partially cleaved (i.e., 5b43.pdb at 2.38 Å resolution[12] and 5nfv.pdb at 2.50 Å resolution,[13] respectively). The third DNA-bound state has been based on a recent structure of the FnCas12a, including a longer NTS that binds within the RuvC cleft and reconciles with the TS (6I1K.pdb at 2.65 Å resolution).[15] These systems have been embedded in explicit waters, and Na^+ ions were added to neutralize the total charge, leading to orthorhombic periodic cells comprising on average a total number of ∼210,000 atoms, for each system. Full details are reported in the Supporting Information (SI).

Molecular Dynamics (MD) Simulations. MD simulations have been performed employing the Amber ff12SB force field, which includes the ff99bsc0[50] corrections for DNA and the ff99bsc0+χ*OL3*[51,52] corrections for RNA. The Allnér force field[53] has been employed for Mg^{2+} ions, and the TIP3P model[54] has been employed for waters. These force field parameters and the simulation protocol have also been employed in our recent studies of CRISPR-Cas9,[40,55,56] corroborated by NMR experiments and quantum mechanical calculations, enabling a fair comparison. An integration time step of 2 fs has been employed. All bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm. Temperature control (300 K) has been performed via Langevin dynamics, with a collision frequency γ = 1/ps. Pressure control was accomplished by coupling the system to a Berendsen barostat,[57] at a reference pressure of 1 atm and with a relaxation time of 2 ps. The systems have been subjected to energy minimization to relax water molecules and counterions, keeping the protein, RNA, DNA, and Mg^{2+} ions fixed with harmonic position restraints of 300 kcal/mol·Å^2, and then, the systems have been heated up from 0 to 100 K in the canonical ensemble (NVT), by running two simulations of 5 ps each, imposing position restraints of 100 kcal/mol·Å^2 on the above-mentioned elements of each system. The temperature was further increased up to 200 K in ∼100 ps of MD runs in the isothermal−isobaric ensemble (NPT), reducing the restraint to 25 kcal/mol Å^2. Subsequently, all restraints were released, and the temperature of the systems was raised to 300 K in a single NPT simulation of 500 ps. After ∼1.1 ns of equilibration, ∼10 ns of NPT runs were carried out allowing the density of the system to stabilize around 1.01 g/cm^{-3}. Finally, ∼1 μs of MD simulations has been carried out in an NVT ensemble for each system, which has also been simulated in 4 replicates. Independent MD simulation replicas have been obtained starting from different configurations and velocities, initialized accordingly to the Maxwell−Boltzmann distribution at physiological temperature. This approach enabled us to obtain solid statistics for the analysis in our purposes. Considering five simulation systems, we collected a total of ∼20 μs of aggregate sampling (i.e., 5 systems * 4 replicas * ∼1 μs = ∼20 μs). Molecular simulations have been performed using the GPU version of AMBER 18.[58] The analysis of the results has been performed on each simulated MD replica and the ensemble obtained averaging independent ns-to-μs trajectories (details are in the SI). This enabled assessing the reproducibility of our results across independent simulations and also providing a solid statistical ensemble.

Principal Component Analysis (PCA). PCA is a statistical method that can report large-scale collective motions occurring in biological macromolecules undergoing MD simulations. Through this statistical technique, it is possible to reduce the large number of degrees of freedom to an essential subspace set, which captures large-amplitude motions of the system. In PCA, the covariance matrix of the protein Cα atoms is calculated and diagonalized to obtain a new set of coordinates (eigenvectors) to describe the system motions. Each eigenvector−also called Principal Component (PC)−is associated with an eigenvalue corresponding to the mean square fluctuation contained in the system's trajectory projected along that eigenvector. By sorting the eigenvectors according to their eigenvalues, the first PC (i.e., PC1) corresponds to the system's largest amplitude motion, and the dynamics of the system along PC1 is usually referred to as *"essential dynamics"*.[32]

In this work, the principal motions of the protein were captured starting from the mass-weighted covariance matrix of the Cα atoms. In detail, PCA has been performed considering the FnCas12a systems, whereby the collected ensembles (i.e., arising from the compared RNA-bound FnCas12a, DNA-bound FnCas12a, and FnCas12a′ systems; Figure 3) were combined and subjected to RMS-fit to the same reference configuration, removing the rotational and translational motions. This was performed to ensure a consistent eigenbasis and motions of the PCs on all compared systems and to construct the covariance matrices from the atoms' positions. We also performed two independent PCA on the LbCas12a and AsCas12a systems, providing insights into their essential motions. Each element in the covariance matrix is the covariance between atoms i and j, defining the i, j position of the matrix. The covariance C_{ij} is defined as

$$C_{ij} = \langle (\vec{r}_i - \langle \vec{r}_i \rangle)(\vec{r}_j - \langle \vec{r}_j \rangle) \rangle \quad (1)$$

where $\vec{r}_i$ and $\vec{r}_j$ are the position vectors of atoms i and j, and the brackets denote an average over the sampled time period. The two terms in eq 1 represent the displacement vectors for atoms i and j. The covariance matrix was then diagonalized, leading to a complete set of orthogonal collective eigenvectors, each associated with a corresponding eigenvalue. The

https://dx.doi.org/10.1021/acs.jcim.0c00929
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

eigenvalues denote how much each eigenvector is representative of the system dynamics, thus giving a measure of the contribution of each eigenvector to the total variance. Indeed, the eigenvectors with the largest eigenvalues correspond to the most relevant motions. By projecting the displacements vectors of each atom along the trajectory onto the eigenvectors (i.e., by taking the dot product between the two vectors at each frame), the PCs were then obtained. The cumulative variance accounted by all the PCs was calculated for all systems, revealing that the first PCs account for the major contribution (Figures S5 and S6). Full details on the application of this statistical technique on MD simulations of CRISPR-Cas12a are in the SI.

Cross-Correlation Analysis. Correlation analysis has been performed in order to identify the dynamical coupling of the motions between Cα atoms (i and j) in the simulated systems. "Pearson-like" cross-correlation (CC_{ij}) analysis provides a measure of the collinear correlations between the atoms i and j. The CC_{ij} matrix can be computed as a normalization of the covariance matrix

$$CC_{ij} = \frac{\langle(\vec{r}_i - \langle\vec{r}_i\rangle)(\vec{r}_j - \langle\vec{r}_j\rangle)\rangle}{[(\langle\vec{r}_i^{\,2}\rangle - \langle\vec{r}_i^{\,2}\rangle)(\langle\vec{r}_j^{\,2}\rangle - \langle\vec{r}_j^{\,2}\rangle)]} \quad (2)$$

where $\vec{r}_i$ and $\vec{r}_j$ are the position vectors of atoms i and j, considered over the sampled time period (denoted using brackets). Positive values of the CC_{ij} coefficients indicate lockstep motions between atoms i and j, while negative CC_{ij} values are indicative of anticorrelated motions. CC_{ij} values equal to zero evince that the atoms' displacements are independent from each other. The magnitude of CC_{ij} coefficients (i.e., ranging from 0−>1 for lockstep motions and from −1−>0 for anticorrelated motions) indicates strength of the correlation. As noted above, the CC_{ij} neglects the nonlinear contributions between atoms i and j and does not capture correlated motions occurring out of phase with each other. To capture more broadly the dependency of the atomic motions, we also employed a generalized correlation method described below.

Generalized Correlations Analysis Based on Mutual Information. This approach relies on information theory and uses the mutual information (MI) measure to obtain the generalized correlation (GC_{ij}) coefficients.[38] In information theory, two variables, such as the such as the $\vec{r}_i$ and $\vec{r}_j$ position vectors, can be considered correlated when their joint probability distribution, $p(\vec{r}_i,\vec{r}_j)$, is smaller than the product of their marginal distributions, $p(\vec{r}_i) \cdot p(\vec{r}_j)$. The MI is a measure of the degree of correlation between $\vec{r}_i$ and $\vec{r}_j$ defined as a function of $p(\vec{r}_i,\vec{r}_j)$ and $p(\vec{r}_i)\cdot p(\vec{r}_j)$ accordingly to

$$MI[\vec{r}_i, \vec{r}_j] = \int\int p(\vec{r}_i, \vec{r}_j) \ln \frac{p(\vec{r}_i, \vec{r}_j)}{p(\vec{r}_i)\cdot p(\vec{r}_j)} dr_i dr_j \quad (3)$$

Notably, MI is closely related to the definition of the Shannon entropy and can be computed as

$$MI[\vec{r}_i, \vec{r}_j] = H[\vec{r}_i] + H[\vec{r}_j] - H[\vec{r}_i, \vec{r}_j] \quad (4)$$

where $H[\vec{r}_i]$ and $H[\vec{r}_j]$ are the marginal Shannon entropies, and $H[\vec{r}_i,\vec{r}_j]$ is the joint entropy, providing a link between motions' correlations and information content. Based on this definition, and considering that MI varies from 0 to + ∞, the normalized GC_{ij} coefficients, ranging from 0 (independent variables) to 1 (fully correlated variables), can be defined as

$$GC_{ij}[\vec{r}_i, \vec{r}_j] = \{1 - e^{-2MI[\vec{r}_i,\vec{r}_j]/d}\}^{-1/2} \quad (5)$$

where d is the dimensionality of $\vec{r}_i$ and $\vec{r}_j$. This approach has been originally introduced by Lange and Grubmüller,[38] who developed a computationally efficient algorithm. Overall, a GC_{ij} analysis is powerful in capturing nonlinear coupled motions in biomolecular systems. However, the GC_{ij} coefficients do not distinguish positive vs negative motions, giving a normalized measure of how much information on one atom's position is provided by that of another atom. Hence, when employed together, the CC_{ij} and GC_{ij} schemes can provide a more comprehensive understanding of the interdependent dynamics of proteins, with information on whether protein regions move in lockstep or through opposite motions (through CC_{ij}), and provide also more general information on the atoms' interdependence (through GC_{ij}). In the present work, the GC_{ij} coefficients have been computed using the positions vectors of Cα atoms along the simulated trajectories. For each model system, all correlation analyses have been performed over independent MD trajectories and have also been averaged over the aggregate sampling arising from 4 ns-to-μs MD replicas (details are in the SI). To further spotlight relevant correlations among spatially distant domains, the averaged GC_{ij} matrices of the FnCas12a system have also been further processed to compute per-domain GC_{ij} scores (Cs). This measure accumulates and normalizes the GC_{ij} coefficients over each protein domain, resulting in per-domain GC_{ij} matrices that help in identifying the most relevant coupled motions in large biomolecular systems (details are in the SI).[17,39]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.0c00929.

Supplementary Materials and Methods; Figure S1, RMSF profiles; Figures S2−S4, RMSD profiles; Figure S5, cumulative contribution of all principal components; Figure S6, probability distribution; Figure S7, essential dynamics; Figures S8−S12, CC_{ij} and GC_{ij} matrices; Figure S13, GC_{ij} matrices; and references (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Giulia Palermo − *Department of Bioengineering and Department of Chemistry, University of California, Riverside, Riverside, California 92521, United States;* orcid.org/0000-0003-1404-8737; Email: giulia.palermo@ucr.edu

Authors

Aakash Saha − *Department of Bioengineering, University of California, Riverside, Riverside, California 92521, United States;* orcid.org/0000-0003-0776-9771

Pablo R. Arantes − *Department of Bioengineering, University of California, Riverside, Riverside, California 92521, United States;* orcid.org/0000-0003-1946-2750

Rohaine V. Hsu − *Department of Bioengineering, University of California, Riverside, Riverside, California 92521, United States*

Yogesh B. Narkhede − *Department of Bioengineering, University of California, Riverside, Riverside, California 92521, United States*

https://dx.doi.org/10.1021/acs.jcim.0c00929
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Martin Jinek − *Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.0c00929

Author Contributions

‖A.S. and P.R.A. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Łukasz Nierzwicki for his thoughtful comments. This material is based upon work supported by the National Science Foundation under Grant No. CHE-1905374. This work was partially funded by NIH Grant R01 EY027440. This work used resources services and support provided via the COVID-19 HPC Consortium (https://covid19-hpc-consortium.org/), which is a unique private-public effort to bring together government, industry, and academic leaders who are volunteering free compute time and resources in support of COVID-19 research. Computer time for MD simulations has also been awarded by XSEDE via the grant TG-MCB160059. We also acknowledge Microsoft AI for Health (https://www.microsoft.com/en-us/ai/ai-for-health) for granting us Azure cloud computing resources for computations and analysis.

■ REFERENCES

(1) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337*, 816−821.

(2) Doudna, J. A.; Charpentier, E. Genome Editing. The New Frontier of Genome Engineering with CRISPR-Cas9. *Science* **2014**, *346*, 1258096.

(3) Zetsche, B.; Gootenberg, J. S.; Abudayyeh, O. O.; Slaymaker, I. M.; Makarova, K. S.; Essletzbichler, P.; Volz, S. E.; Joung, J.; van der Oost, J.; Regev, A.; Koonin, E. V.; Zhang, F. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **2015**, *163*, 759−771.

(4) Chen, J. S.; Ma, E.; Harrington, L. B.; Da Costa, M.; Tian, X.; Palefsky, J. M.; Doudna, J. A. CRISPR-Cas12a Target Binding Unleashes Indiscriminate Single-Stranded DNase Activity. *Science* **2018**, *360*, 436−439.

(5) Broughton, J. P.; Deng, X.; Yu, G.; Fasching, C. L.; Servellita, V.; Singh, J.; Miao, X.; Streithorst, J. A.; Granados, A.; Sotomayor-Gonzalez, A.; Zorn, K.; Gopez, A.; Hsu, E.; Gu, W.; Miller, S.; Pan, C.-Y.; Guevara, H.; Wadford, D. A.; Chen, J. S.; Chiu, C. Y. CRISPR−Cas12-Based Detection of SARS-CoV-2. *Nat. Biotechnol.* **2020**, *38*, 870−874.

(6) Sternberg, S. H.; Redding, S.; Jinek, M.; Greene, E. C.; Doudna, J. A. DNA Interrogation by the CRISPR RNA-Guided Endonuclease Cas9. *Nature* **2014**, *507*, 62−67.

(7) Garcia-Doval, C.; Jinek, M. Molecular Architectures and Mechanisms of Class 2 CRISPR-Associated Nucleases. *Curr. Opin. Struct. Biol.* **2017**, *47*, 157−166.

(8) Stella, S.; Alcón, P.; Montoya, G. Class 2 CRISPR−Cas RNA-Guided Endonucleases: Swiss Army Knives of Genome Editing. *Nat. Struct. Mol. Biol.* **2017**, *24*, 882−892.

(9) Shmakov, S.; Smargon, A.; Scott, D.; Cox, D.; Pyzocha, N.; Yan, W.; Abudayyeh, O. O.; Gootenberg, J. S.; Makarova, K. S.; Wolf, Y. I.; Severinov, K.; Zhang, F.; Koonin, E. V. Diversity and Evolution of Class 2 CRISPR−Cas Systems. *Nat. Rev. Microbiol.* **2017**, *15*, 169−182.

(10) Dong, D.; Ren, K.; Qiu, X.; Zheng, J.; Guo, M.; Guan, X.; Liu, H.; Li, N.; Zhang, B.; Yang, D.; Ma, C.; Wang, S.; Wu, D.; Ma, Y.; Fan, S.; Wang, J.; Gao, N.; Huang, Z. The Crystal Structure of Cpf1 in Complex with CRISPR RNA. *Nature* **2016**, *532*, 522−526.

(11) Gao, P.; Yang, H.; Rajashankar, K. R.; Huang, Z.; Patel, D. J. Type V CRISPR-Cas Cpf1 Endonuclease Employs a Unique Mechanism for CrRNA-Mediated Target DNA Recognition. *Cell Res.* **2016**, *26*, 901−913.

(12) Yamano, T.; Nishimasu, H.; Zetsche, B.; Hirano, H.; Slaymaker, I. M.; Li, Y.; Fedorova, I.; Nakane, T.; Makarova, K. S.; Koonin, E. V.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **2016**, *165*, 949−962.

(13) Swarts, D. C.; van der Oost, J.; Jinek, M. Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol. Cell* **2017**, *66*, 221−233.

(14) Stella, S.; Alcón, P.; Montoya, G. Structure of the Cpf1 Endonuclease R-Loop Complex after Target DNA Cleavage. *Nature* **2017**, *546*, 559−563.

(15) Swarts, D. C.; Jinek, M. Mechanistic Insights into the Cis- and Trans-Acting DNase Activities of Cas12a. *Mol. Cell* **2019**, *73*, 589−600.

(16) Stella, S.; Mesa, P.; Thomsen, J.; Paul, B.; Alcón, P.; Jensen, S. B.; Saligram, B.; Moses, M. E.; Hatzakis, N. S.; Montoya, G. Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell* **2018**, *175*, 1856−1871.

(17) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. Striking Plasticity of CRISPR-Cas9 and Key Role of Non-Target DNA, as Revealed by Molecular Simulations. *ACS Cent. Sci.* **2016**, *2*, 756−763.

(18) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. CRISPR-Cas9 Conformational Activation as Elucidated from Enhanced Molecular Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 7260−7265.

(19) Palermo, G.; Chen, J. S.; Ricci, C. G.; Rivalta, I.; Jinek, M.; Batista, V. S.; Doudna, J. A.; McCammon, J. A. Key Role of the REC Lobe during CRISPR−Cas9 Activation by 'Sensing', 'Regulating', and 'Locking' the Catalytic HNH Domain. *Q. Rev. Biophys.* **2018**, *51*, e9.

(20) Zuo, Z.; Zolekar, A.; Babu, K.; Lin, V. J.; Hayatshahi, H. S.; Rajan, R.; Wang, Y.-C.; Liu, J. Structural and Functional Insights into the Bona Fide Catalytic State of Streptococcus Pyogenes Cas9 HNH Nuclease Domain. *eLife* **2019**, *8*, e46500.

(21) Zuo, Z.; Liu, J. Cas9-Catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations. *Sci. Rep.* **2016**, *6*, 37584.

(22) Casalino, L.; Nierzwicki, Ł.; Jinek, M.; Palermo, G. Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab-Initio Molecular Dynamics. *ACS Catal.* **2020**, DOI: 10.1021/acscatal.0c03566.

(23) Dagdas, Y. S.; Chen, J. S.; Sternberg, S. H.; Doudna, J. A. A Conformational Checkpoint between DNA Binding and Cleavage by CRISPR-Cas9. *Sci. Adv.* **2017**, *3*, eaao0027.

(24) Zhu, X.; Clarke, R.; Puppala, A. K.; Chittori, S.; Merk, A.; Merrill, B. J.; Simonović, M.; Subramaniam, S. Cryo-EM Structures Reveal Coordinated Domain Motions That Govern DNA Cleavage by Cas9. *Nat. Struct. Mol. Biol.* **2019**, *26*, 679−685.

(25) Chen, J. S.; Dagdas, Y. S.; Kleinstiver, B. P.; Welch, M. M.; Sousa, A. A.; Harrington, L. B.; Sternberg, S. H.; Joung, J. K.; Yildiz, A.; Doudna, J. A. Enhanced Proofreading Governs CRISPR−Cas9 Targeting Accuracy. *Nature* **2017**, *550*, 407−410.

(26) Schmid-Burgk, J. L.; Gao, L.; Li, D.; Gardner, Z.; Strecker, J.; Lash, B.; Zhang, F. Highly Parallel Profiling of Cas9 Variant Specificity. *Mol. Cell* **2020**, *78*, 794−800.

(27) Yang, M.; Peng, S.; Sun, R.; Lin, J.; Wang, N.; Chen, C. The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. *Cell Rep.* **2018**, *22*, 372−382.

(28) Kleinstiver, B. P.; Pattanayak, V.; Prew, M. S.; Tsai, S. Q.; Nguyen, N. T.; Zheng, Z.; Joung, J. K. High-Fidelity CRISPR−Cas9 Nucleases with No Detectable Genome-Wide off-TargetEffects. *Nature* **2016**, *529*, 490−495.

(29) Casini, A.; Olivieri, M.; Petris, G.; Montagna, C.; Reginato, G.; Maule, G.; Lorenzin, F.; Prandi, D.; Romanel, A.; Demichelis, F.; Inga, A.; Cereseto, A. A Highly Specific SpCas9 Variant Is Identified by in Vivo Screening in Yeast. *Nat. Biotechnol.* **2018**, *36*, 265−271.
(30) Kleinstiver, B. P.; Sousa, A. A.; Walton, R. T.; Tak, Y. E.; Hsu, J. Y.; Clement, K.; Welch, M. M.; Horng, J. E.; Malagon-Lopez, J.; Scarfò, I.; Maus, M. V.; Pinello, L.; Aryee, M. J.; Joung, J. K. Engineered CRISPR−Cas12a Variants with Increased Activities and Improved Targeting Ranges for Gene, Epigenetic and Base Editing. *Nat. Biotechnol.* **2019**, *37*, 276−282.
(31) Yang, W.; Steitz, T. A. Recombining the Structures of HIV Integrase, RuvC and RNase H. *Structure* **1995**, *3*, 131−134.
(32) Daidone, I.; Amadei, A. Essential Dynamics: Foundation and Applications. *WIREs Comput. Mol. Sci.* **2012**, *2*, 762−770.
(33) Jeon, Y.; Choi, Y. H.; Jang, Y.; Yu, J.; Goo, J.; Lee, G.; Jeong, Y. K.; Lee, S. H.; Kim, I.-S.; Kim, J.-S.; Jeong, C.; Lee, S.; Bae, S. Direct Observation of DNA Target Searching and Cleavage by CRISPR-Cas12a. *Nat. Commun.* **2018**, *9*, 2777.
(34) Palermo, G.; Ricci, C. G.; Fernando, A.; Basak, R.; Jinek, M.; Rivalta, I.; Batista, V. S.; McCammon, J. A. Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. *J. Am. Chem. Soc.* **2017**, *139*, 16028−16031.
(35) Jiang, F.; Zhou, K.; Ma, L.; Gressel, S.; Doudna, J. A. STRUCTURAL BIOLOGY. A Cas9-Guide RNA Complex Preorganized for Target DNA Recognition. *Science* **2015**, *348*, 1477−1481.
(36) Gong, S.; Yu, H. H.; Johnson, K. A.; Taylor, D. W. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep.* **2018**, *22*, 359−371.
(37) Liu, M.; Sen; Gong, S.; Yu, H. H.; Jung, K.; Johnson, K. A.; Taylor, D. W. Engineered CRISPR/Cas9 Enzymes Improve Discrimination by Slowing DNA Cleavage to Allow Release of off-Target DNA. *Nat. Commun.* **2020**, *11*, 3576.
(38) Lange, O. F.; Grubmuller, H. Generalized Correlation for Biomolecular Dynamics. *Proteins: Struct., Funct., Genet.* **2006**, *62*, 1053−1061.
(39) Casalino, L.; Palermo, G.; Spinello, A.; Rothlisberger, U.; Magistrato, A. All-Atom Simulations Disentangle the Functional Dynamics Underlying Gene Maturation in the Intron Lariat Spliceosome. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 6584−6589.
(40) East, K. W.; Newton, J. C.; Morzan, U. N.; Narkhede, Y. B.; Acharya, A.; Skeens, E.; Jogl, G.; Batista, V. S.; Palermo, G.; Lisi, G. P. Allosteric Motions of the CRISPR−Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J. Am. Chem. Soc.* **2020**, *142*, 1348−1358.
(41) Nierzwicki, Ł.; Arantes, P. R.; Saha, A.; Palermo, G. Establishing the Allosteric Mecanism in CRISPR-Cas9. *WIREs Comput. Mol. Sci.* **2020**, e1503.
(42) Zuo, Z.; Liu, J. Allosteric Regulation of CRISPR-Cas9 for DNA-Targeting and Cleavage. *Curr. Opin. Struct. Biol.* **2020**, *62*, 166−174.
(43) Montefiori, M.; Pilotto, S.; Marabelli, C.; Moroni, E.; Ferraro, M.; Serapian, S. A.; Mattevi, A.; Colombo, G. Impact of Mutations on NPAC Structural Dynamics: Mechanistic Insights from MD Simulations. *J. Chem. Inf. Model.* **2019**, *59*, 3927−3937.
(44) Serapian, S. A.; Colombo, G. Designing Molecular Spanners to Throw in the Protein Networks. *Chem. - Eur. J.* **2020**, *26*, 4656−4670.
(45) Cofsky, J. C.; Karandur, D.; Huang, C. J.; Witte, I. P.; Kuriyan, J.; Doudna, J. A. CRISPR-Cas12a Exploits R-Loop Asymmetry to Form Double-Strand Breaks. *eLife* **2020**, *9*, e55143.
(46) Cui, Y.; Tang, Y.; Liang, M.; Ji, Q.; Zeng, Y.; Chen, H.; Lan, J.; Jin, P.; Wang, L.; Song, G.; Lou, J. Direct Observation of the Formation of a CRISPR−Cas12a R-Loop Complex at the Single-Molecule Level. *Chem. Commun.* **2020**, *56*, 2123−2126.
(47) Strohkendl, I.; Saifuddin, F. A.; Rybarski, J. R.; Finkelstein, I. J.; Russell, R. Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a. *Mol. Cell* **2018**, *71*, 816−824.
(48) Singh, D.; Mallon, J.; Poddar, A.; Wang, Y.; Tippana, R.; Yang, O.; Bailey, S.; Ha, T. Real-Time Observation of DNA Target Interrogation and Product Release by the RNA-Guided Endonuclease CRISPR Cpf1 (Cas12a). *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 5444−5449.
(49) Li, S.-Y.; Cheng, Q.-X.; Liu, J.-K.; Nie, X.-Q.; Zhao, G.-P.; Wang, J. CRISPR-Cas12a Has Both Cis- and Trans-Cleavage Activities on Single-Stranded DNA. *Cell Res.* **2018**, *28*, 491−493.
(50) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007**, *92*, 3817−3829.
(51) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E., 3rd; Sponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6*, 3836−3849.
(52) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E.; Jurecka, P. Refinement of the Cornell et Al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886−2902.
(53) Allnér, O.; Nilsson, L.; Villa, A. Magnesium Ion−Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 1493−1502.
(54) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(55) Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G. Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent. Sci.* **2019**, *5*, 651−662.
(56) Palermo, G. Structure and Dynamics of the CRISPR−Cas9 Catalytic Complex. *J. Chem. Inf. Model.* **2019**, *59*, 2394−2406.
(57) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(58) Case, D. A.; Betz, R. M.; Botello-Smith, W.; Cerutti, D. S.; Cheatham, I. T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Kollman, P. A. *AMBER* 2018; Univ. California: San Francisco, 2018.

https://dx.doi.org/10.1021/acs.jcim.0c00929
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX