Supporting Information

**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*¹,²

1. Department of Bioengineering, University of California Riverside, 900 University Avenue, Riverside, CA 92512, United States

2. Department of Chemistry, University of California Riverside, 900 University Avenue, Riverside, CA 92512, United States

3. Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

*Corresponding authors:

Dr. Giulia Palermo (gpalermo@ucsd.edu)

‡Equally contributed
1. Supplementary Materials and Methods

Structural models. Molecular simulations have been performed on five model systems of CRISPR-Cas12a, based on the available crystallographic structures. The RNA-bound states are based on two X-ray structures: *Lachnospiraceae* bacterium Cas12a (LbCas12a) solved at 2.38 Å resolution (5id6.pdb) and *Francisella novicida* Cas12a (FnCas12a) solved at 3.34 Å resolution (5ng6.pdb). 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., the 5b43.pdb at 2.38 Å resolution and the 5nfv.pdb at 2.50 Å resolution, respectively). The third DNA-bound state has been based on a recent structure of FnCas12a, including a longer NTS that binds within the RuvC cleft and reconciles with the TS (6I1K.pdb at 2.65 Å resolution). Missing residues of the X-ray structures have been added via homology modeling, using SwissModel by Schwede. Missing bases of 6I1K X-ray structure have been modeled by Jinek in the original paper. The simulation systems have been embedded in explicit waters, while Na\(^+\) ions were added to neutralize the total charge, leading to orthorhombic periodic simulation cells of 145 × 135 × 115 Å\(^3\) (RNA-bound LbCas12a, for a total of ~200,000 atoms), ~115 × 130 × 150 Å\(^3\) (RNA-bound FnCas12a, ~200,000 atoms), ~125 × 125 × 150 Å\(^3\) (DNA-bound AcCas12a, ~210,000 atoms), ~120 × 135 × 150 Å\(^3\) (DNA-bound FnCas12a, ~210,000 atoms) and ~120 × 135 × 150 Å\(^3\) (DNA-bound FnCas12a’, ~220,000 atoms).

Molecular Dynamics (MD). The above-mentioned model systems have been object of MD, adopting a well-established simulation protocol for protein/nucleic acid complexes. We employed Amber ff12SB force field, which includes the ff99bsc0\(^7\) corrections for DNA and the ff99bsc0+χ\(OL3\)\(^8,9\) corrections for RNA. The Allnér force field\(^10\) has been employed for Mg\(^{2+}\) ions and TIP3P model\(^11\) for waters. These force field parameters and the simulation protocol have also
been employed in our recent studies of CRISPR-Cas9,\textsuperscript{12–16} 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,\textsuperscript{17} with a collision frequency $\gamma = 1/\text{ps}$. Pressure control was accomplished by coupling the system to a Berendsen barostat,\textsuperscript{18} 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 counter ions, keeping the protein, the RNA, DNA and Mg\textsuperscript{2+} ions fixed with harmonic position restraints of 300 kcal/mol $\cdot$ Å$^2$. 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 $\cdot$ Å$^2$ on the above-mentioned elements of each system. The temperature was further increased up to 200 K in $\sim$100 ps of MD in the isothermal-isobaric ensemble (NPT), reducing the restraint to 25 kcal/mol Å$^2$. Subsequently, all restraints were released and the temperature of the systems was raised up to 300 K in a single NPT simulation of 500 ps. After $\sim$1.1 ns of equilibration, $\sim$10 ns of NPT runs were carried out allowing the density of the system to stabilize around 1.01 g/cm$^3$. Finally, $\sim$1 μs of MD simulations was carried out in NVT ensemble for each system, which has also been simulated in four independent replicates. The MD simulations replicas have been obtained starting from different configurations and velocities, initialized accordingly to the Maxwell-Boltzmann distribution at physiological temperature. Considering that we considered five simulation systems, we collected a total of $\sim$20 μs of aggregate sampling (i.e., 5 systems $\times$ 4 replicas $\times$ $\sim$1 μs $= \sim$20 μs). These simulations have been performed using the GPU version of AMBER 18.\textsuperscript{19}

**Principal Component Analysis (PCA).** PCA is a statistical method that can report on the large-scale, collective motions occurring in biological macromolecules undergoing MD.
simulations. Thus, PCA can provide valuable information on major conformational changes taking place along the trajectory. In fact, 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 to 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 as “essential dynamics”.

Here, the principal motions of the protein have been 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:

$$C_{ij} = \langle \vec{r}_i - \langle \vec{r}_i \rangle \rangle \langle \vec{r}_j - \langle \vec{r}_j \rangle \rangle$$

[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. A positive sign of this product indicates that the two atoms move in
a correlated manner, otherwise, a negative value points to anti-correlated atoms. If the product is zero, then it evinces that the atoms displacements are independent of each other. The covariance matrix was then diagonalized, leading to a complete set of orthogonal collective eigenvectors, each associated to a corresponding eigenvalue. The 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 Principal Components (PC) were then obtained. The cumulative variance accounted by all the PCs was calculated for all system (Figures S5), revealing that the first PCs account for the major contribution. The first PC (i.e., PC1), commonly referred to as “essential dynamics”, is plotted on the 3D structures of the CRISPR-Cas12a complex indicating the largest amplitude motions of the protein (Figure 3a). Upon projecting each structure arising from the MD simulations into the collective coordinate space derived by the first two eigenvectors (PC1 and PC2), plotting PC1 against PC2 allows generation of scatter plots (PC1 vs. PC2) that display how the conformational space defined by the first two modes is sampled through the MD simulations (Figure 3b).

For each simulated system of Cas12a, PCA has been performed considering the aggregate sampling arising from multiple replicas of ~1 μs, each. Indeed, a single MD run is not sufficient to explore the conformational ensemble of large ribonucleoprotein complexes. On the other hand, multiple runs starting from different coordinates and initial velocities can more properly explore the conformational states and provide a better description of the protein conformational ensemble through PCA. Accordingly, PCA analysis over the aggregate multi-μs runs have shown
Volumetric Analysis. To calculate the volume of the groove between the REC and NUC lobes, we performed volumetric analysis by employing the POcket Volume MEasurer (POVME) software (v. 3.0). The volumetric analysis has been performed on the collective sampling arising from the equilibrium MD trajectories of the DNA-bound AsCas12a, FnCas12a and FnCas12a’. To calculate the volume of the groove for a given trajectory frame, we monitored the space within two 20 Å radius spheres, which defined the groove between the REC and NUC lobes. In detail, the two spheres have been centered at the center of mass between the Nuc and REC2 domains and at the center of mass between the REC, WED and RuvC respectively (Figure S6). The pocket volume was calculated from the portion of the sphere that was unoccupied by protein acid atoms. To provide a comparable measure of the change in volume and maintain consistency, nucleic acids have been omitted from volumetric analysis. Only volumetric regions contiguous with the pocket were included. Grid spacing and padding parameters were set to 1.0 and 1.09 Å, respectively.

Cross-Correlation Analysis. Cross-Correlation (CC$_{ij}$) analysis has been performed in order to identify the dynamical coupling of the motions between Ca atoms ($i$ and $j$) in the systems. CC$_{ij}$ analysis is based on Pearson coefficients, which are computed between the fluctuations of the Ca atoms relative to their average positions, according to:

$$
CC_{ij} = \frac{(\langle \vec{r}_i - \langle \vec{r}_i \rangle \rangle (\vec{r}_j - \langle \vec{r}_j \rangle))}{\sqrt{\langle (\vec{r}_i - \langle \vec{r}_i \rangle)^2 \rangle \langle (\vec{r}_j - \langle \vec{r}_j \rangle)^2 \rangle}}
$$

[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). CC$_{ij}$ arise from the normalization of the covariance between atoms.
\(i\) and \(j\) (\(C_{ij}\)) reported in Equation 1. Positive \(CC_{ij}\) values describe lock-step motions between atoms \(i\) and \(j\), while negative \(CC_{ij}\) values describe anti-correlated motions. The magnitude of \(CC_{ij}\) coefficients (i.e., ranging from 0 \(\rightarrow\) 1 for lockstep motions; and from \(-1 \rightarrow 0\) for anticorrelated motions) indicates the strength of the correlation. In this work, the \(CC_{ij}\) coefficients have been plotted as a \(two \times two\) matrix, describing the correlated motions of each residue (i.e., their \(\text{C}^\alpha\) atoms) with each other. For each model system, the \(CC_{ij}\) matrices have been computed over independent \(~1\ \mu\text{s}\) trajectories, and have been subsequently averaged over the aggregate sampling arising from 4 independent MD replicas. The \(CC_{ij}\) matrices computed over the individual trajectories and the averaged \(CC_{ij}\) matrices are reported in Figures S8-12.

**Generalized Correlation (GC\(_{ij}\)) Analysis based on Mutual Information.** The Generalized Correlations (GC\(_{ij}\)) analysis describes correlations independently on the relative orientation of the atomic fluctuations and is able to capture non-linear correlations.\(^{25}\) In this analysis, 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 mutual information (\(MI\)) is a measure of the degree of correlation between \(\vec{r}_i\) and \(\vec{r}_j\) defined as function of \(p(\vec{r}_i, \vec{r}_j)\) and \(p(\vec{r}_i) \cdot p(\vec{r}_j)\) according to:

\[
MI[\vec{r}_i, \vec{r}_j] = \iint p(\vec{r}_i, \vec{r}_j) \ln \frac{p(\vec{r}_i, \vec{r}_j)}{p(\vec{r}_i)p(\vec{r}_j)} \, d\vec{r}_i d\vec{r}_j \tag{3}
\]

Notably, \(MI\) is closely related to the definition of the Shannon entropy, \(H[\vec{r}]\) (i.e., the expectation value of a random variable \(\vec{r}\)), and can be computed as:

\[
MI[\vec{r}_i, \vec{r}_j] = H[\vec{r}_i] + H[\vec{r}_j] - H[\vec{r}_i, \vec{r}_j] \tag{5}
\]

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. The g-correlation tool\textsuperscript{25} implemented in GROMACS v4.6.4\textsuperscript{26} was used to calculate the marginal entropies $H[x_i]$ and $H[x_j]$ and the joint entropy $H[x_i, x_j]$ by means of the $k$-nearest neighbor distances algorithm,\textsuperscript{27} applied to the atomic positions fluctuations from MD simulations. Since $MI$ varies from 0 to $+\infty$, normalized generalized correlation ($GC_{ij}$) coefficients, ranging from 0 (independent variables) to 1 (fully correlated variables), are defined as:

$$GC_{ij} [\vec{r}_i, \vec{r}_j] = \left\{ 1 - e^{-2MI[\vec{r}_i, \vec{r}_j]/d} \right\}^{-1/2} \quad [6]$$

where $d$ is the dimensionality of $\vec{r}_i$ and $\vec{r}_j$. Here, the $GC_{ij}$ coefficients have been plotted as a two-by-two matrix; describing the correlated motions of each residue with each other. For each model system, the $GC_{ij}$ matrices have been computed over independent MD runs, and have been averaged over the aggregate sampling arising from 4 MD replicas. The $GC_{ij}$ matrices computed over the individual trajectories and the averaged $GC_{ij}$ matrices are reported in Figures S8-12.

**Per-domain Correlation Scores (Cs).** In order to evaluate the inter-dependent coupling between the Cas12a protein domains, we computed the per-domain generalized correlations scores (Cs). This measure enables to spotlight the most relevant coupled motions in large biomolecular systems, where the interpretation of correlated motions can be difficult by visual inspection of the correlation matrices. As previously shown, the per-domain Cs helps identifying how specific protein regions mechanistically intervene in the overall correlation network.\textsuperscript{14,28,29} For each protein residue $i$, a $Cs_i$ parameter can be computed as:

$$Cs_i = \sum_{j \neq i}^N GC_{ij} \quad [6]$$

which sums the $GC_{ij}$ established by residue $i$ with the residues $j$, representing a measure of the number and the intensity of the $GC_{ij}$ coefficients displayed by each residue. To filter non-trivial
correlations and eliminate the noise due to uncorrelated motions, per-residue Cs were computed considering only highly positive correlations (GC$_{ij}$ $\geq$ 0.40). To detail the overall inter-domain correlations, the Cs have been accumulated and normalized as follows. First, the Cs were calculated for each residue $i$ belonging to a specific protein domain (e.g., Nuc), with the residues $j$ belonging to another protein domain of interest (e.g., REC2). Then, the Cs were accumulated over all residues $j$ of each specific Cas12a domain and normalized by the number of coupling residues. This resulted in a set of per-domain Cs, ranging from 0 (not-correlated) to 1 (correlated), measuring the strength of the overall correlation that each domain establishes with the others. As mentioned above, by using this approach we captured the most relevant correlations between spatially distant domains in large ribonucleoproteins, such as the spliceosome$^{28}$ and CRISPR-Cas9.$^{14}$ In this study, the per-domain Cs have been computed over the GC$_{ij}$ matrices of all the FnCas12a systems averaged over 4 simulation replicas (Figure S13). This enabled to spotlight differences in the most relevant coupled motions of FnCas12a when bound to RNA (i.e., RNA-bound FnCas12a) and in complex with DNA, considering the effect of binding of a cleaved NTS (i.e., DNA-bound FnCas12a) and a complete NTS (i.e., DNA-bound FnCas12a$'$).
2. Supplementary Figures

Figure S1. Root Mean Square Fluctuations (RMSF) profiles, calculated for the protein C atoms of Cas12a bound to RNA (a) and DNA (b). Data are reported for 5 model systems, namely the RNA-bound LbCas12a and FnCas12a; and the DNA-bound AsCas12a, FnCas12a and FnCas12a’ (full details on the simulated systems are in the SI text). For each system, data are reported for 4 simulation replicas of ~1 μs each. The protein sequence is shown on top of the graphs, highlighting individual protein domains in different colors.
Figure S2. Root Mean Square Deviation (RMSD) profiles, calculated considering the protein C\(\alpha\) atoms of Cas12a bound to RNA (a) and DNA (b). Data are reported for 5 model systems, namely the RNA-bound LbCas12a and FnCas12a; and the DNA-bound AsCas12a, FnCas12a and FnCas12a’ (full details on the simulated systems are in the SI text). For each system, data are reported for 4 simulation replicas of \(~1\ \mu s\) each.
Figure S3. Root Mean Square Deviation (RMSD) profiles, calculated for the RNA backbone of Cas12a bound to RNA (a) and DNA (b). Data are reported for 5 model systems, namely the RNA-bound LbCas12a and FnCas12a; and the DNA-bound AsCas12a, FnCas12a and FnCas12a’ (full details on the simulated systems are in the SI text). For each system, data are reported for 4 simulation replicas of ~1 μs each.
Figure S4. Root Mean Square Deviation (RMSD) profiles, calculated for the DNA target strand (TS) backbone (a) and for the DNA non-target strand (NTS) backbone (b) of Cas12a. Data are reported for 3 model systems of Cas12a bound to DNA, namely AsCas12a, FnCas12a and FnCas12a’ (full details on the simulated systems are in the SI text). For each system, data are reported for 4 simulation replicas of ~1 μs each.
Figure S5. Cumulative contribution (% y-axis) of all the principal components (PCs, x-axis) to the variance of the overall Cas12a motions calculated upon Principal Component Analysis (PCA) of the LbCas12a, AsCas12a and FnCas12a systems (full details on the simulated systems are in the SI text). The contribution of the first three PCs are highlighted in red, blue and green respectively.
Figure S6. Probability distribution of the volume of the groove between the REC and NUC lobes, computed over the equilibrium trajectories of the DNA-bound FnCas12a (a), AsCas12a (b) and FnCas12a’ (c) systems. Vertical bars indicate the crystallographic values. To calculate the volume of the groove for a given trajectory frame, we monitored the space within two 20 Å radius spheres, which defined the groove between the REC and NUC lobes. The two spheres have been centered at the center of mass between the Nuc and REC2 domains (COM1) and at the center of mass between the REC, WED and RuvC (COM2) respectively, as shown in a cartoon of FnCas12a reported on the bottom right. The AsCas12a and FnCas12a’ systems (a, b) display a contraction of the groove between the REC and NUC lobes. On the other hand, in the FnCas12a’ including a complete NTS, an expansion is observed. The POcket Volume MEasurER (POVME) software (v. 3.0) has been used. Full details on volumetric analysis are in the Supplementary Methods.
Figure S7. (a) “Essential dynamics” derived from the first principal component (PC1) of the individual protein domains of the RNA-bound LbCas12a (top) and PC1 vs. PC2 scatter plot (bottom), characterizing the conformational space sampled. (b) “Essential dynamics” derived from PC1 of the individual protein domains of the DNA-bound AsCas12a (top) and PC1 vs. PC2 scatter plot (bottom), characterizing the conformational space sampled. Full details are reported in the Supplementary Methods.
**Figure S8.** Cross-Correlation ($CC_{ij}$, upper triangles) and Generalized Correlations ($GC_{ij}$, lower triangles) matrices for the RNA-bound FnCas12a. Data are reported for four independent MD simulation replicas of ~1 μs each, as well as averaged over 4 simulation replicas (details are reported in the Supplementary Methods). The strength of the computed $CC_{ij}$ is colored blue (for $CC_{ij} \geq 0$, lockstep motions) to violet (for $CC_{ij} \leq 0$, anticorrelated motions), while the $GC_{ij}$ are color-coded green (correlated) to magenta (not correlated). The color scales are on the bottom right. The protein sequence is also shown, highlighting individual protein domains in different colors. A box is used to highlight relevant correlated motions discussed in the main text.
**Figure S9.** Cross-Correlation (CC\(_{ij}\), upper triangles) and Generalized Correlations (GC\(_{ij}\), lower triangles) matrices for the DNA-bound FnCas12a. Data are reported for four independent MD simulation replicas of ~1 μs each, as well as averaged over 4 simulation replicas (details are reported in the Supplementary Methods). The strength of the computed CC\(_{ij}\) is colored blue (for CC\(_{ij}\) ≥ 0, lockstep motions) to violet (for CC\(_{ij}\) ≤ 0, anticorrelated motions), while the GC\(_{ij}\) are color-coded green (correlated) to magenta (not correlated). The color scales are on the bottom right. The protein sequence is also shown, highlighting individual protein domains in different colors. A box is used to highlight relevant correlated motions discussed in the main text.
**Figure S10.** Cross-Correlation (CC$_{ij}$, upper triangles) and Generalized Correlations (GC$_{ij}$, lower triangles) matrices for the DNA-bound FnCas12a’ (including a complete NTS). Data are reported for four independent MD simulation replicas of ~1 μs each, as well as averaged over 4 simulation replicas (details are reported in the Supplementary Methods). The strength of the computed 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 color-coded green (correlated) to magenta (not correlated). The color scales are on the bottom right. The protein sequence is also shown, highlighting individual protein domains in different colors. A box is used to highlight relevant correlated motions discussed in the main text.
**Figure S11.** Cross-Correlation (CC$_{ij}$, upper triangles) and Generalized Correlations (GC$_{ij}$, lower triangles) matrices for the DNA-bound AsCas12a. Data are reported for four independent MD simulation replicas of ~1 μs each, as well as averaged over 4 simulation replicas (details are reported in the Supplementary Methods). The strength of the computed 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 color-coded green (correlated) to magenta (not correlated). The color scales are on the bottom right. The protein sequence is also shown, highlighting individual protein domains in different colors. A box is used to highlight relevant correlated motions discussed in the main text.
Figure S12. Cross-Correlation ($CC_{ij}$, upper triangles) and Generalized Correlations ($GC_{ij}$, lower triangles) matrices for the RNA-bound LbCas12a. Data are reported for four independent MD simulation replicas of ~1 μs each, as well as averaged over 4 simulation replicas (details are reported in the Supplementary Methods). The strength of the computed $CC_{ij}$ is colored blue (for $CC_{ij} \geq 0$, lockstep motions) to violet (for $CC_{ij} \leq 0$, anticorrelated motions), while the $GC_{ij}$ are color-coded green (correlated) to magenta (not correlated). The color scales are on the bottom right. The protein sequence is also shown, highlighting individual protein domains in different colors. A box is used to highlight relevant correlated motions discussed in the main text.
Figure S13 Supplementary Methods. (a) Generalized Correlations (GC$_{ij}$) matrices for FnCas12a bound to RNA (i.e., RNA-bound FnCas12a, left panel) and in complex with DNA, binding a cleaved NTS (i.e., DNA-bound FnCas12a, central panel) and a complete NTS (i.e., DNA-bound FnCas12a, right panel). For each system, the GC$_{ij}$ matrices are averaged over 4 simulation replicas of 1 μs each. The strength of the computed GC$_{ij}$ is colored from green (correlated) to magenta (not correlated, color scale on the right). The protein sequence is also shown. (b) Per-domain GC$_{ij}$ Score (Cs) matrices, which accumulate (and normalize) the per-domain GC$_{ij}$. The per-domain Cs matrices are a “coarse” representation of the GC$_{ij}$ matrices (reported in panel (a)), which can spotlight the most relevant correlations between spatially distant domains, ranging from 0 (not-correlated, white) to 1 (correlated, green). Upon DNA binding, Nuc and REC2 display the highest Cs (a box is used to indicate the REC2/Nuc region on the correlation matrices, highlighting the absence of correlations prior DNA binding). Full details are reported in the Supplementary Methods.
3. Supplementary References

(1) Dong, D.; Ren, K.; Qiu, X.; Zheng, J.; Guo, M.; Guan, X.; Liu, H.; Li, N.; Zhang, B.; Yang, D.; et al. The Crystal Structure of Cpf1 in Complex with CRISPR RNA. Nature 2016, 532, 522–526.

(2) 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.

(3) Yamano, T.; Nishimasu, H.; Zetsche, B.; Hirano, H.; Slaymaker, I. M.; Li, Y.; Fedorova, I.; Nakane, T.; Makarova, K. S.; Koonin, E. V.; et al. Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. Cell 2016, 165, 949–962.

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

(5) Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; et al. SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure Using Evolutionary Information. Nucleic Acids Res. 2014, 42, W252–W258.

(6) Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De
Vivo, M. Catalytic Metal Ions and Enzymatic Processing of DNA and RNA. 

*Acc. Chem. Res.* **2015**, *48*, 220–228.

(7) 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.

(8) 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. Theor. Comput.* **2010**, *6*, 3836–3849.

(9) 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.

(10) Allnér, O.; Nilsson, L.; Villa, A. Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 1493–1502.

(11) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J.
(12) 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.* **2017**, *114*, 7260–7265.

(13) Palermo, G.; Ricci, C. G.; Fernando, A.; Basak, R.; Jinek, M.; Rivalta, I.; Batista, V. S.; McCammon, J. A. Protopacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. *J. Am. Chem. Soc.* **2017**, *139*, 16028–16031.

(14) 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.

(15) 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**, *134*, 1348–1358.

(16) 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*. 
(17) Turq, P.; Lantelme, F.; Friedman, H. L. Brownian Dynamics: Its Applications to Ionic Solutions. *J. Chem. Phys.* **1977**, *66*, 3039.

(18) 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.

(19) Case, D. A.; Betz, R. M.; Botello-Smith, W.; Cerutti, D. S.; Cheatham T. E., I. I. I.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; et al. AMBER **2018**. *Univ. California, San Fr.*

(20) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins Struct. Funct. Genet.* **1993**, *17*, 412–425.

(21) Daidone, I.; Amadei, A. Essential Dynamics: Foundation and Applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 762–770.

(22) Bakan, A.; Meireles, L. M.; Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27*, 1575–1577.

(23) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14*, 33-38.

(24) Wagner, J. R.; Sørensen, J.; Hensley, N.; Wong, C.; Zhu, C.; Perison, T.; Amaro, R. E. POVME 3.0: Software for Mapping Binding Pocket Flexibility. *J. Chem. Theory Comput.* **2017**, *13*, 4584–4592.
(25) Lange, O. F.; Grubmuller, H. Generalized Correlation for Biomolecular Dynamics. *Proteins-Structure Funct. Bioinforma.* **2006**, *62*, 1053–1061.

(26) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(27) Floyd, R. W. Algorithm 97: Shortest Path. *Commun. ACM* **1962**, *5*, 345.

(28) 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.

(29) Ricci, C. G.; Silveira, R. L.; Rivalta, I.; Batista, V. S.; Skaf, M. S. Allosteric Pathways in the PPARγ-RXRα Nuclear Receptor Complex. *Sci. Rep.* **2016**, *6*, 19940.