Supporting Information

PAM-induced allostery activates CRISPR-Cas9

Giulia Palermo,* Clarisse G. Ricci, Amendra Fernando, Rajshekhar Basak, Martin Jinek, Ivan Rivalta, Victor S. Batista and J. Andrew McCammon

Table of contents

Corresponding author:

Dr. Giulia Palermo (gpalermo@ucsd.edu)

S1. Supplementary Materials and Methods

S1.1. Structural models. Molecular dynamics (MD) simulations were based on the high resolution X-ray structures of the ternary Cas9:RNA:DNA complex (4UN3.pdb, solved at 2.59 Å resolution),[1] and on its analogue crystallized lacking the PAM sequence (4OO8.pdb, solved at 2.5 Å resolution).[2] In the 4UN3.pdb and 4OO8.pdb X-ray structures (unit A is considered for 4OO8.pdb), the protein assumes a similar configuration, such representing an ideal starting point for comparative molecular simulations. Missing residues in the X-ray structures have been added via homology modeling, using SwissModel.[3] Two model systems were built, consisting of (1) the ternary Cas9:RNA:DNA complex with the PAM sequence (i.e., Cas9–wPAM, based on 4UN3.pdb) and (2) lacking of PAM (i.e., Cas9–w/oPAM, based on 4OO8.pdb). The obtained model systems were embedded in explicit water molecules, leading to periodic simulation cells of ~144*108*146 $Å^3$, consisting of ~220,000 atoms.

S1.2. Molecular dynamics (MD) simulations. The above-mentioned model systems were equilibrated and production runs were performed using the Amber ff12SB force field, which includes the ff99bsc0 corrections for DNA[4] and the ff99bsc0+χ*OL3* corrections for RNA.[5,6] The TIP3P model was employed for the water molecules.[7] The Åqvist[8] force field parameters for Mg^{2+} ions were employed, as in previous studies and extensive testing on protein/nucleic acids systems.[9-13] A salt concentration of 80 mM of NaCl was considered, in agreement with the experimental conditions of cleavage assays.[1,14] Na^+ counter-ions were used to neutralize the total charge. The systems were simulated with a time step of 2 fs, as detailed below in the description of the simulations protocol. All the simulations were performed using the Gromacs 5.0.2 code.[15] The LINCS algorithm[16] was used to constrain covalent bonds involving hydrogen atoms. Long-range electrostatic interactions were calculated with the particle mesh Ewald method with a real space cutoff of 10 Å. Periodic boundary conditions in the three directions of the Cartesian space were applied. The systems were coupled to a Nosé−Hoover thermostat[17,18] at a reference temperature of 310 K and to an isotropic Parrinello−Rahman barostat[19] at a reference pressure of 1 bar, both with coupling time constants of 1 ps. The following simulation protocol was adopted for each of the systems. First, the systems were subjected to energy minimization using a steepest descent algorithm. Then, the systems were thermalized to the

physiological temperature of 310 K within 1 ns. Then, ~10 ns of equilibration were carried out using a time step of 2 fs, following a protocol that we have employed in our previous studies of protein/nucleic acids complexes.[10,12] Approximately ~1.5 μs of MD simulations were collected in the NPT ensemble under standard conditions, for the Cas9–wPAM and Cas9–w/oPAM systems. Cas9–wPAM was simulated in four replicas, reaching a total of ~6 μs of aggregate sampling. We ran three replicas of the Cas9–w/oPAM system, for a total of ~4.5 μs of MD, and a fourth replica (~2.0 μs) of the Cas9–w/oPAM system in which the nucleotide sequence of the 4OO8 X-ray structure has been substituted with that of the 4UN3.pdb. This simulation shows that the allosteric communication is independent from the nucleotide sequence of the RNA and DNA strands, depending instead on the presence of the PAM sequence, as explained below. Due to the presence of a different nucleotide sequence with respect to the crystallized sequence, the adaptation of the X-ray structure required ~380/400 ns of MD simulations (Fig. S1, panel B). As such, analysis of this system has been performed considering the last ~1.5 μs of MD, as detailed below. Thus, in total ~12.5 μs of MD were carried out. The choice of covering microsecond ranges in CRISPR-Cas9 dynamics is motivated by the fact that critical interactions have been captured over the nano-to-microsecond time scale.[11,20] Moreover, it has become increasingly evident that simulations longer than a microsecond can provide critical information on allosteric mechanisms.[21-23] Finally, independent long-time scale trajectories, obtained starting from different configurations and velocities as initialized according to the Maxwell-Boltzmann distribution at physiological temperature, ensure solid statistics for analysis, as required given the complexity of the system under investigation. Coordinates of the systems were collected every 10 ps, for a total of ~150,000 frames for each run of ~1.5 μs.

S1.3. Electrostatic calculations. Electrostatic calculations were performed using the Adaptive Poisson-Boltzmann Solver (APBS1.4) software.[24] Prior to calculations, structures were converted to the pqr format by using the pdb2pqr software[25,26] with the same Amber force field and protonation states considered in the simulations. Calculations have been performed at 298 K and at a ionic concentration of 150 mM, with the traditional Linearized Poisson Boltzmann Equation (LPBE), on the X-ray structures of CRISPR-Cas9 including PAM (4UN3.pdb)[1] and lacking PAM (4OO8.pdb),[2] as well as on the X-ray structure obtained including PAM (4UN3.pdb),

after computational removal of the PAM segment. We used an internal dielectric constant of 2.0 to represent the non-polar environment of the solute and an external dielectric constant of 78.0 to represent the aqueous environment of the solvent. Because of the large size of the systems, calculations were performed using grid spacings of ~0.7 Å.

S1.4. Analysis of the results. Statistical analysis has been performed on each simulated trajectory, consisting of ~1.5 μs of sampling, as well as on the statistical *ensemble* obtained averaging multiple trajectories (i.e., four replicas for each system). Each simulated trajectory has been obtained upon independent MD simulations initiated from different configurations and velocities, as initialized according to the Maxwell-Boltzmann distribution at physiological temperature. This approach, which overall encompasses the multi-microsecond time scale, allowed solid statistics for the analysis in our purposes.

S1.5. Generalized Correlations (GC_{ij}). The Generalized Correlations (GC_{ij}) analysis has been employed to capture dynamic correlations in the simulated systems. With respect to the more traditional Pearson coefficients analysis, GC_{ij} analysis has the advantage of capturing correlations independently on the relative orientation of the atomic fluctuations, while also being able to capture non-linear correlations.[27] Two variables (x_i,x_j) can be considered correlated when their joint probability distribution, $p(x_i, x_j)$, is smaller than the product of their marginal distributions, $p(x_i) \cdot p(x_j)$. The mutual information (MI) is a measure of the degree of correlation between x_i and x_j defined as function of $p(x_i, x_j)$ and $p(x_i) \cdot p(x_j)$ according to:

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

Notably, MI is closely related to the definition of the Shannon entropy, H[x], i.e., the expectation value of a random variable x, having a probability distribution $p(x_i)$

$$H[x] = \int p(x)\, ln\, p(x) dx \qquad [2]$$

and it can be thus computed as:

$$MI\left[x_i, x_j\right] = H\left[x_i\right] + H\left[x_j\right] - H\left[x_i, x_j\right] \qquad [3]$$

where $H\left[x_i\right]$ and $H\left[x_j\right]$ are the marginal Shannon entropies, and $H\left[x_i, x_j\right]$ is

the joint entropy. In the g_correlation tool,[27] as implemented in Gromacs 3.3,[28] it 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,[29] applied to the atomic positions fluctuations from MD simulations. Since MI varies from 0 to + ∞, normalized generalized correlation coefficients (GC_{ij}), ranging from 0 (independent variables) to 1 (fully correlated variables), are defined as:

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

where d=3, the dimensionality of x_i and x_j.

S1.6. Correlation Score (Cs_i). For each amino acid residue a Cross-Correlation Score (Cs_i) coefficient can be defined:

$$Cs_i = \sum_{j \neq i}^{N} GC_{ij} \quad [5]$$

representing a measure of both the number and the intensity of the generalized correlation coefficients displayed by each residue. To filter the highest correlations, per-residue Cs_i were computed considering only highly positive ($GC_{ij} \geq 0.65$) correlations. To shed light on intra and inter domain correlations, Cs_i were calculated for each residue *i*, with the residues *j* belonging to the same protein domain (i.e., HNH, RuvC, α-helical, Topo and C-term domain, Fig. 1 of the main text) of the residue *i* (Cs_i^{intra}) or else excluding the correlations with residues in the same protein domain to which *i* belongs (Cs_i^{inter}). By detecting the protein residues that are highly correlated, the Cs_i function helps in identifying how specific protein regions mechanistically intervene in the overall correlation network.[30] Finally, Cs_i^{inter} were accumulated over all residues *j* of each specific Cas9 domain and plotted as a two-by-two matrix (Fig. 2 of the main text). Complete analysis of Cs_i^{intra} and Cs_i^{inter} is reported as supplementary results below, as well as in the main text.

S1.7. Principal Component Analysis (PCA). PCA has been employed to capture the essential motions of the simulated systems and characterize conformational transitions. In PCA, the covariance matrix of the protein Cα atoms is calculated and diagonalized to obtain a new set of generalized 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 Principal Component (PC1) corresponds to the system's largest amplitude motion, and the dynamics of the system along PC1 is usually referred as *"essential dynamics"*.[31] In this work, each conformation sampled during the MD trajectories is projected into the collective coordinate space defined by the first two eigenvectors (PC1 and PC2), such allowing the characterization of the essential conformational sub-space sampled by Cas9 during MD. Importantly, in order to identify differences in the essential structural-dynamic properties of Cas9, each system had its trajectory stripped down to the Cα carbons only, and then superposed onto the same reference structure (i.e., considering as a reference the protein framework that does not show relevant conformational differences among the crystallized states). In this way, the simulations (with or without PAM) were all projected onto the same set of eigenvectors, which was obtained by concatenation and subsequent diagonalization of all the Cα trajectories together.

S1.8. Community Network Analysis (CNA). The CRISPR-Cas9 protein communication network was defined as a set of amino acids residues (i.e., nodes of the network) connected by edges (residue pair connection). Edge lengths, i.e., the inter-node distances in the graph, are defined using the GC_{ij} coefficients according to:

$$w_{ij} = -\log GC_{ij} \quad [6]$$

with GC_{ij} being the generalized coefficients (as defined in equation 4). .Two nodes are considered connected if any heavy atom of the two residues is within 5 Å of each other (i.e., *distance cutoff*) for at least the 75 % of the simulation time (i.e., *frame cutoff*). These cutoff are selected according to an extensive convergence study based on the estimation of the Community Repartition Difference (CRD), defined as:

$$CRD\left(c_1, c_2\right) = 1 - \frac{\sum_{n_i,n_j} z(n_i,n_j,c_1)\, z(n_i,n_j,c_2)}{\sum_{n_i,n_j} z(n_i,n_j,c_1)} \quad [7]$$

where $z(n_i, n_j, c_i)$ is defined as 1 if nodes n_i and n_j belong to the same community in a given partition c_i (i.e., the community structure) and 0 otherwise. CRD provides a normalized count of pairs that are grouped together in two community structures, providing a good estimate of the similarities between different network partitions, as in the case of community structures obtained with different

cutoff values. Fig. S10 reports CRD values calculated for both Cas9–wPAM and Cas9–w/oPAM systems showing converged values of CRD for the two selected cutoff parameters.

The resulting *"weighted graph"* defines the CRISPR-Cas9 dynamical network, with information on the critical nodes that are important for the communication within the complex. In the weighted network, a set of *"communities"* can be identified. Communities are groups of nodes in which the network connections are dense but between which they are sparse. These local substructures can be obtained with the Girvan-Newman algorithm, which is a divisive algorithm that uses the *"edge betweenness"* partitioning criterion.[32] Edge betweenness is defined as the number of shortest pathways that cross the edge, being a parameter that favors edges that *inter*-connect communities and disfavors edges that lie within communities. The edges with the highest betweenness connect many pairs of nodes and form the link between different communities. High edge betweenness associates with pairs of residues that are important for the communication flow within the weighted network. The Girvan-Newman algorithm for finding communities is an iterative process, where the edge with the highest betweenness is removed from the network (*edge cutting* procedure) and the betweenness of the remaining edges is recalculated, with communities being progressively isolated up to the point where each node will represent a community. The optimal division of the network has to be obtained in such a way that each community contain nodes that are highly *intra*-connected while different communities are poorly *inter*-connected through few critical nodes. The parameter of *"modularity"*, Q, measures the strength (or the quality) of the community structure and it is used for determining the optimal division of the network. The modularity Q represents the difference in probability of *intra*- and *inter*-community connections for a given network division and is defined as:

$$Q = \sum_i (e_{ii} - a_i^2) \qquad [8]$$

where e_{ij} is the fraction of edges that links nodes in community i to nodes in community j, while $a_i = \sum_j e_{ij}$ is the fraction of edges that connects to nodes in community i. The modularity value falls in the range of 0 to 1, with larger values indicating higher community structure quality. The optimum community structures selected for the Cas9–wPAM and Cas9–w/oPAM systems have the highest

modularity (i.e., Q = 0.78 and Q = 0.77, respectively, Fig. S11), in agreement with standard modularity values found in 3D structures of proteins (0.4–0.7).[30,33,34]

In the weighted dynamic network, the *"shortest pathways"* connecting pairs of catalytic residues of the RuvC and HNH domains were calculated using the Floyd-Warshall algorithm,[29] which sums the lengths (w_{ij}) of all edges involved in different paths of nodes connecting two distant residues and identifies the pathway displaying the shortest total length. Specifically, *"shortest pathways"* were computed considering as *"source"* the catalytic residues of the RuvC domain (E762, D986, D10, H983, S15) and as a *"sink"* the catalytic residue of the HNH domain (H840), such tracking the information transfer from the *"source"* to the *"sink"*. Moreover, we have also computed all possible pathways (not only the shortest) among the above-mentioned residues within the weighted protein-network. The communication pathways could be very close in length to the shortest pathway (i.e., "*sub-optimal pathways"*), where the *path length* (*PL*) is defined as the sum of the edge lengths involved in that pathway. This likely happens when the shortest pathways involve highly correlated residues within the same or nearby secondary structures. For simplicity, we consider as sub-optimal pathways only those whose lengths are not larger than 2% of the shortest pathway length. The number of such sub-optimal pathways defines the *pathway degeneracy* (*PD*).

S1.9. Volumetric analysis. To calculate the volume of the PAM binding cleft, we performed volumetric analysis by employing the POcket Volume MEasurer (POVME) software.[35,36] To calculate the volume of the binding pocket for a given trajectory frame, we monitored the space within a 20 Å radius sphere centered on the central PAM base G2 (Fig. S2). The pocket volume was calculated from the portion of the sphere that was unoccupied by protein/nucleic acid atoms. Only volumetric regions contiguous with the pocket were included. Grid spacing and padding parameters were set to 1.0 and 1.09 Å, respectively.

S2. Supplementary Results

S2.1. PAM induces a conformational transition in CRISPR-Cas9. MD simulations of the Cas9 systems were performed in four replicas, whose RMSD's with respect to the proteins' □ carbons and all atoms are displayed in Fig. S1. In the presence of PAM (Fig. S1A), Cas9 equilibrates within ~100 ns in a configuration relatively close to the X-ray structure (with RMSD of Cα atoms around 3.2-4.0 Å) for all replicas but it undergoes a conformation transition in replicas #1, #2 and #4 – reaching RMSD values above ~5.5 Å after ~750 ns – while it remains in the initial conformational state in replica #3, plateauing at RMSD values of ~4 Å. These results clearly indicate that the conformational change observed in replicas #1, #2 and #4 does not consist of a simple relaxation from the X-ray structure, which occurs in all replicas within ~100 ns and generates a stable configuration in replica #3, but rather refers to a transition between states that belong to the equilibrium regime in the sub-μs time scale. In the absence of PAM (Fig. S1B), the system remains in roughly the same conformational state in all simulated replicas, which is characterized by a plateau of the RMSD values of around 4.5-5Å. Moreover, by substituting the nucleotide sequence of the 4OO8 X-ray structure with that of the 4UN3.pdb (replica #4 in Fig. S1B), we observe that, upon initial equilibration of the X-ray structure adapting to the docked nucleotides (~380/400 ns of MD), the RMSD reaches similar values to the other Cas9–w/oPAM replicas, suggesting that these systems have similar conformational flexibility, regardless to the presence of a different RNA/DNA sequences.

The presence of PAM induces a contraction of the complex structure, which we have measured by computing the volume of the PAM-interacting region (Fig. S2). Specifically, in the Cas9–wPAM system, we observe a decrease of the volume of the PAM binding cleft with respect to the crystallographic value, which is observed in replicas #1, #2 and #4 that are characterized by the PAM-induced conformational transition. In the case of replica #3, in which the system mainly remains in the initial conformational state (Fig. S1), the pocket volume fluctuates around the crystallographic value. In Cas9–w/oPAM, the pocket volumes assume higher values, fluctuating around the crystallographic values.

We then used Principal Component Analysis (PCA) to detail the conformational transitions undergone by Cas9 and to explore how the presence of

PAM re-shapes Cas9 conformational landscape. Fig. S3A displays the projection of CRISPR-Cas9 dynamics onto the subspace formed by the first and second principal components (PC1 *vs*. PC2), for Cas9–wPAM and Cas9–w/oPAM. Data from four independent simulations are reported, using different colors to differentiate them. The dynamics of CRISPR-Cas9 in the presence or absence of PAM is clearly distinguished by the first principal component, with PC1 < 0 in the presence of PAM (left panel, Fig. S3A) and PC1 < 0 in the absence of PAM (right panel, Fig. S3A). Interestingly, this difference is already present in the X-ray structures, but is intensified during MD simulations, as evidenced by the temporal evolution of the systems in the essential sub-space (Fig. S3B). PAM induces a shift in the conformational dynamics of CRISPR-Cas9, which leads it to explore systematically larger values of PC1 and a wider range of values along PC2 as compared to Cas9-w/oPAM. Fig. S3C shows the motion associated to the PC1, revealing a *"breathing"* of the whole complex, accompanied by a reorientation of the HNH domain. When contracting, the HNH domain moves closer to the other nuclease domain (RuvC), as well as to the Topo and C-term domains, which are involved in DNA recognition. Clearly, PAM induces a significant contraction of the complex, tightening up the interaction with target and non-target DNA strands.

S2.2. PAM stably interacts with R1333 and R1335. Two arginine residues (R1333 and R1335) mediate the recognition of the *'5*-TGG-*3'* PAM sequence in CRISPR-Cas9. Specifically, R1333 H-bonds G6, while R1335 engages in interactions with G7. Fig. S4 reports the time evolution of these interactions along four simulation replicas (rep. #1–4) of the Cas9–wPAM system. We found that both interactions are overall stable during the dynamics, with R1335–G7 establishing stronger interactions.

S2.3. Correlation analyses. The GC_{ij} matrices calculated for the replicas of Cas9–wPAM and Cas9–w/oPAM systems are reported in Fig. S5 and S6, respectively. In the Cas9–wPAM system, GC_{ij} are intense in replicas #1, #2 and #4, as a sign of the fact that conformational transition occurring in the presence of the PAM sequence is associated with an overall increase of correlated motions in the complex, while GC_{ij} are weakened in replica #3, where the system remains in an *"open"* conformational state. In the Cas9–w/oPAM system, visual inspection of the GC_{ij} matrices reveals weaker correlated motions in all replicas with respect to the Cas9–wPAM system. We have computed the probability distribution of the GC_{ij} coefficients for each simulated system, quantifying how the correlation network is altered in the presence of PAM

(Fig. S7). In Cas9–wPAM, the GC_{ij} coefficients show a shift in their probability distribution from GC_{ij} ~0.4 (replica #3) to more correlated states characterized by GC_{ij} ~0.5/0.6. Noteworthy, more intense correlations are detected for all the three replicas in which the system departed from its original conformation (#1, #2, #4), confirming a direct relation between the PAM–induced conformational transition and the increase of correlated motions in the system. This indicates that the system undergoes concerted conformational changes, during which spatially distant protein components move jointly. In the simulations of Cas9–w/oPAM, where the system mainly assumes the initial configuration throughout the dynamics (Fig. S1), the probability distribution of GC_{ij} coefficients peaks at ~0.4.

In order to compare the extent of the correlated motions in the Cas9–wPAM and Cas9–w/oPAM systems, GC_{ij} matrices were averaged over the replicas of each system. Fig. 2 of the main text (panel C) reports the averaged matrices, revealing intense correlations in the Cas9–wPAM system, with disruption of the correlation network in the absence of PAM. A shift of the probability distribution of the GC_{ij} coefficients toward higher values (Fig. 1D, main text) quantifies the increase of correlations in the presence of PAM. The averaged matrices, which take into account the overall statistical *ensemble* and conformational transitions of the systems, were used for all further analyses.

S2.4. Community Network Analysis (CNA). CNA was performed to understand the communication structure of the system. By describing the system in a set of nodes and edges, weighted via GC_{ij}, CNA was employed to identify groups – or *"communities"* – of residues closely correlated to each other and how these latter engage in long-range inter-correlations (details are in the Supplementary Methods). In the community network representation (Fig. 3A of the main text), different communities of highly *intra*-correlated nodes (i.e., residues) are represented as spheres connected by bonds. The width of the bonds connecting the communities is proportional to the sum of edge betweenness of the edge connecting them (i.e., the *inter*-communities total betweenness) and, as such, is a measure of the strength of the communication between different communities.

The communities resulting from CNA of the CRISPR-Cas9 system are reported in Fig. 3A of the main text. In Cas9–wPAM, the HNH and RuvC domain are identified as two strongly *inter*-connected communities (i.e., with thick inter-

communities bonds between them). As well, strong inter-communication is observed between HNH and the α-helical lobe. This latter is divided in three communities, which accurately reflects the three subdomains in the secondary sequence of the α-helical domain, which overall mediates RNA binding.[37] The interconnection between the HNH and RuvC domains indicates strong signal transduction and well agrees with the tight correlation experimentally found between the catalytic activity of these domains, which provides evidence of allosteric communication.[38] In the absence of PAM, the system's communities are fragmented. This fragmentation particularly affects the communities that refer to the RuvC domain, the α-helical lobe and the Topo domain, which constitute the cleft allocating the DNA including PAM. As such, PAM binding results into an increase of system's organization in signal communication, which supports the claim that PAM creates a strong allosteric signaling in CRISPR-Cas9.[21,30,34]

Node betweenness, defined as the number of shortest pathways that cross the node, is a measure of the importance of a single amino acid residue for signal communication within the network. In Cas9–wPAM, polar and charged residues display the highest node betweenness (Fig. S8), indicating that a strong flow of allosteric information passes through these key node residues. In the systems lacking of the PAM sequence, charged residues with high node betweenness are depleted in favor of non-polar residues. This indicates the key role of medium/long-chain charged and polar residues in conveying the allosteric information, when in the presence of PAM, in well agreement with previous studies of other protein/nucleic acid systems characterized by the presence of functional allosteric effects.[30] Analogously to node betweenness, edge betweenness is defined as the number of shortest pathways passing through a pair nodes and high edge betweenness indicates a crucial role of a pair of amino acids in the communication network. Edge betweenness are associated to couples of residues that connect the different communities. Generally, the highest edge betweennesses are associated with pair of residues having specific interactions, such as salt-bridges and strong H-bonding. Analysis of edge betweenness shows that in the presence of PAM, community #8 of the HNH domain is connected to the communities #1 (RuvC) and #4 (α-helical) via the L1 (residues 765−780) and L2 (residues 906−918) loops (Fig. 3C of the main text and S8B). Indeed, R905 of the L2 loop (community #8) engages in ionic interactions with E584 of the α-helical lobe

(community #5), while Q771 and K775 (both belonging to the L1 loop) enable the communication between communities #1 and #8, composed of the RuvC and HNH domains, respectively. As such, the L1 and L2 loops, which connect HNH and RuvC, are shown here to enable the information transfer between the two catalytic domains and the a-helical lobe, acting as *"allosteric transducers"*.[38,39] In Cas9–w/oPAM, the information transfer between HNH and the α-helical lobe is maintained, since K212 of the α-helical lobe (community #4') forms high edge betweenness with E786 of the HNH domain (community #8'). However, edge betweenness are not detected between residues of the communities corresponding to HNH and RuvC, indicating a loss of communication between them. This indicates that signal transmission between HNH and RuvC occurs via L1 and L2 and that the presence of PAM is necessary for the information transfer. Taken together, these results strongly encourage site-specific point mutations of residues with high node/edge betweenness. Accordingly, structure-based and rational engineering of Cas9 has shown that the mutation of charged residues within the groove allocating DNA leads to sensible changes in the functional activity, as well as to the decrease of off-target effects.[40]

In the weighted dynamic network, we computed the *"shortest pathways"* connecting pairs of catalytic residues of the RuvC and HNH domains (details are reported in the Supplementary Methods). Specifically, *"shortest pathways"* were computed considering as *"source"* the catalytic residues of the RuvC domain (E762, D986, D10, H983, S15) and as a *"sink"* the catalytic residue of the HNH domain (H840), such tracking the information transfer from the *"source"* to the *"sink"* (Fig. S8). We found that in Cas9–wPAM, the information flows through the L1 loop, which connects HNH and RuvC and enables the information transfer. Indeed, as mentioned above, the L1 and L2 loops have been suggested to be key *"allosteric transducers"* between HNH and RuvC.[38,39] The L2 loop has also been shown to activate the HNH domain toward the catalysis by establishing H-bond interactions with the DNA, the exact role of L1 has not been fully clarified so far.[11,39] Interestingly, in Cas9–wPAM, the identified *"shortest pathway"* display shorter *"pathway length"* (*PL*) with respect to that identified in Cas9–w/oPAM. In particular, the pathway connecting S15–H840, which includes the L1 loop only in the presence of PAM, shows a ~20 % difference between the two systems. This indicates that, in Cas9–wPAM, the residues involved in this pathway display higher correlations with each

other and that the information transfer is more likely to flow through. Moreover, in Cas9–wPAM the identified *shortest pathway* systematically display smaller *"pathway degeneracy"* (*PD*) with respect to Cas9–w/oPAM. *PD* is defined as the number of *sub-optimal pathways* having pathway length (*PL*) similar to that of the corresponding *shortest pathway*. PD is thus associated with the probability that the information pass through a specific shortest pathway or a set of "analogously short" *sub-optimal pathways*. As such, the smaller *PD* observed in Cas9–wPAM with respect to Cas9–w/oPAM (Fig. S9) indicates higher specificity of the communication pathways in the presence of PAM. In particular, the S15–H840 (including the L1 loop) and the D986–H840 shortest pathways have *PDs* close (or equal) to zero in Cas9–wPAM while several ($PD \geq 7$) *sub-optimal pathways* are present in absence of PAM. Overall, the analysis of *"shortest pathways"* connecting pairs of catalytic residues of RuvC and HNH domains reveals a strong and specific flow of allosteric information in Cas9–wPAM, supporting the key role of PAM in inducing allosteric signaling in CRISPR-Cas9.

As a final remark, CNA analysis has been performed using the Community Repartition Difference (CRD), as a criterion for the choice of the parameters of *"distance cutoff"* and *"frame cutoff"* (details are in the Supplementary Methods). Fig. S10 reports CRD values calculated for both Cas9–wPAM and Cas9–w/oPAM systems.

S2.5. Most intense correlations and inter-domain communication. The difference between the most intense correlations (i.e., $GC_{ij} \geq 0.65$) in the Cas9–wPAM and Cas9–w/oPAM systems ($\Delta GC_{ij} = GC_{ij}^{\text{wPAM}} - GC_{ij}^{\text{w/oPAM}}$) identifies correlated motions that are significantly weakened ($\Delta GC_{ij} < -0.3$) or increased ($\Delta GC_{ij} > 0.3$) by the presence of PAM. Fig. S12 shows the correlations that are decreased (magenta) or increased (cyan) by PAM plotted on the 3D structure. We observe that the correlations decreased by PAM are mostly *intra*-domain correlations of the □-helical and C-term domains, whereas the correlations increased by PAM are mainly interconnecting the Cas9 domains. More specifically, the majority of the *inter*-domain correlations triggered by PAM connect the central HNH domain to other domains (i.e., □-helical, RuvC and the C-term).

To shed light on the specific per-residue correlations, we employed Correlation scores (Cs_i), which are a measure of both the number and the intensity of

the generalized correlations displayed by each residue (full details in the Supplementary Methods). Cs_i were calculated for each residue i, with the residues j belonging to the same protein domain of the residue i ($\mathrm{Cs_i^{intra}}$) or excluding the protein domain to which i belongs ($\mathrm{Cs_i^{inter}}$), such giving detailed information on the *intra*- and *inter*-domain correlations. Fig. S13 reports the $\mathrm{Cs_i^{intra}}$ (panel A) and $\mathrm{Cs_i^{inter}}$ (panel B), as calculated from the averaged GC_{ij} matrix of the Cas9–wPAM and Cas9–w/oPAM systems. Both $\mathrm{Cs_i^{intra}}$ and $\mathrm{Cs_i^{inter}}$ track intense correlations in the presence of PAM. Moreover, the loss of correlations observed in the absence of PAM is particularly evident in the case of the *inter*-domain correlations. Fig. S14 reports the probability distribution of the *inter*-domain $\mathrm{Cs_i^{inter}}$, discriminated for each domain. Clearly, for each protein domain, $\mathrm{Cs_i^{inter}}$ are increased in the presence of PAM. This effect is particularly evident for the *inter*-domain correlations off the HNH domain, which results the most correlated in the system.

In order to understand how *inter*-domain correlations are established among different Cas9 domains, per-residue Cs_i^{inter} of each protein domain were calculated with respect to the other specific protein domains. The accumulated Cs_i^{inter} over each specific Cas9 domain were plotted in a two-by-two matrix, which reveals the interdependence of the protein domains in the presence of PAM (Fig. 3B, main text).

S2.6. Electrostatic properties. Electrostatic calculations were performed with the aim of understanding the role of electrostatics in the PAM-induced conformational transition that, as discussed in the main text and in this Supporting Information, results in the increase of system's correlations and in the transmission of allosteric signaling. The electrostatic potential was calculated and mapped on the 3D structures of CRISPR-Cas9 using the Adaptive Poisson-Boltzmann Solver (APBS) software (details in the Supplementary Methods).[24] APBS calculations were performed for the X-ray structures of CRISPR-Cas9 including PAM (4UN3.pdb)[1] and lacking PAM (4OO8.pdb).[2] For comparison, we also calculated the electrostatic potential for the X-ray structure originally obtained with PAM (4UN3.pdb), after removal of the DNA segment that contains the PAM sequence. All calculations represent the electrostatic potential before the systems underwent any conformational transition observed in MD simulations, allowing us to understand the causal role of electrostatics. Fig. S15 shows the electrostatic potential (red, negative (–5 *kT/e*); blue, positive (+5 *kT/e*)), mapped on the 3D structures of CRISPR-Cas9. Field lines (in purple) are used to

indicate the electrostatic pathways along which point charges are most likely to travel. A strong electrostatic field is created in the presence of PAM, as indicated by a series of field lines attracting positively charged regions from the α-helical, C-terminal and HNH domains towards the negatively charged PAM segment (panel B). The electrostatic field is weakened in the absence of PAM (as indicated by the depletion of field lines in panels A and C), which supports the hypothesis that the PAM-induced electrostatics is the driving force behind the *"open–to–close"* conformational transition observed in the presence of PAM. Fig. S16 shows the protein surface color-coded highlighting different protein domains (as in Fig. 1 of the main text), while field lines are color-coded from blue (positive) to red (negative), indicating the direction that a positive charge would accelerate toward a negative charge if placed upon that line. This representation further indicates that, in the absence of PAM, positive regions in the C-term are attracted to negative regions in the α-helical domain, while in the presence of PAM, field lines sensibly increase and include the HNH domain, conveying towards the DNA including PAM.

S2.7. Key interactions and PAM-mediated conformational transition. The structural contraction, observed in the presence of PAM, is accompanied by an increase of H-bonding interactions and salt bridges involving the HNH domain, which enter in action in the PAM-mediated *"open–to–close*" conformational transition (Fig. 2 of the main text) and explain the highly correlated motions displayed by the HNH domain. In detail, during MD simulations of Cas9–wPAM, we observe a gradual increase of H-bonding interactions between the HNH domain and the remaining protein regions, with respect to the system without PAM (Fig. S17). During MD simulations of Cas9–wPAM, we also detect a gradual increase of H-bonding between the protein and DNA, as opposite to their plateau formation along MD simulations of Cas9–w/oPAM. In the presence of PAM, we also observed an increase of the number of salt bridges that the HNH domain establishes with the remaining protein domains (i.e., *inter*–salt bridges, Fig. S18B), which is accompanied by an increase of *inter*–salt bridges of its neighboring domains (mainly α-helical and also C-term, which are the HNH counterparts). Interestingly, in replica #3 of Cas9–wPAM, in which the system mainly assumes the *"open"* configuration (Fig. S1), the HNH and α-helical domains show a decrease in the number of *inter*–salt bridge interactions. Overall, the increase of H-bonding and ionic interactions are clearly associated with the PAM-mediated

"open–to–close" conformational transition observed in CRISPR-Cas9. Along these lines, we recall that electrostatic calculations have shown that, upon site-specific binding of PAM, a strong electrostatic field is created (and is detected in the X-ray structure, Fig. 3D of the main text and S15-S16), attracting positively charged regions of the α-helical, C-term and HNH domains towards the negatively charged DNA backbone. While this drives the *"open–to–close"* conformational transition, H-bonding and salt-bridges enter in action as a final "click" stabilizing the PAM-induced conformational states.

S3. Supplementary Figures

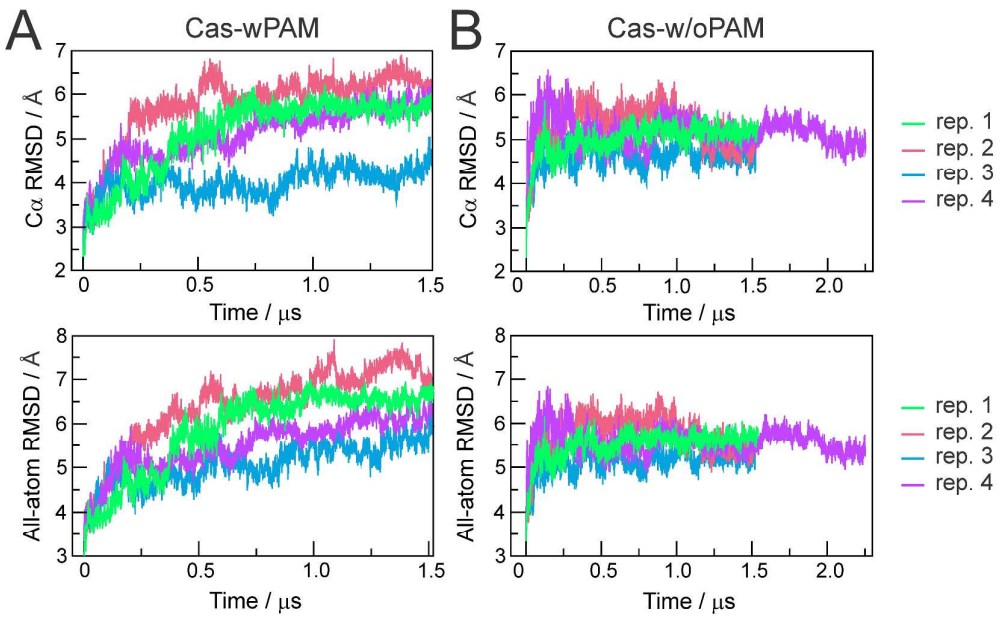


Figure S1. Root Mean Square Deviation (RMSD) profiles, calculated considering the protein Cα atoms (top panels) and all protein atoms (bottom panels), of the Cas9–wPAM **(A)** and Cas9-w/oPAM trajectories **(B)**, simulated as independent replicas of the 4UN3[1] (Cas9–wPAM) and 4OO8[2] (Cas9–w/oPAM) X-ray structures. The Cas9–w/oPAM system, in which the nucleotide sequence of the 4OO8 X-ray structure has been substituted with that of the 4UN3.pdb (i.e., replica #4), was simulated for >2 μs, allowing structural adaptation of the system (~380/400 ns). See the Supplementary Text (paragraph S2.1) for details.

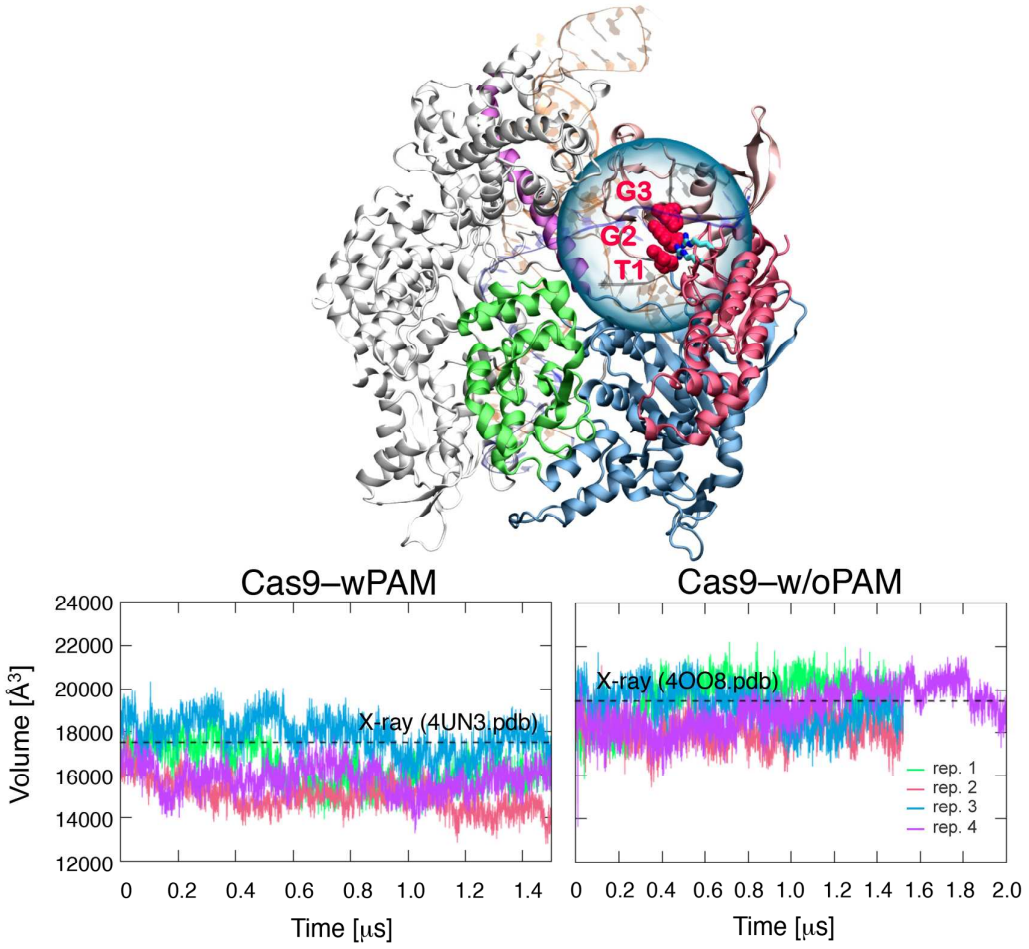


Figure S2. Time evolution of the volume of the PAM-interacting cleft along four simulation replicas (rep. #1–4) of the of Cas9–wPAM (left panel) and the Cas9–w/oPAM (right panel) systems. The volume of the PAM-interacting cleft was calculated by using the POcket Volume MEasurer (POVME) software.[35,36] Crystallographic values, as calculated on the 4UN3 (Cas9–wPAM) and 4OO8 (Cas9–w/oPAM), are indicated by dashed lines. During the simulations, the space within a 20 Å radius sphere centered on the central PAM base G2 (shown on top, as plotted on the 3D structure of CRISPR-Cas9) was monitored and the volume was calculated from the portion of the sphere that was unoccupied by protein/nucleic acid atoms. See the Supplementary Text (paragraph S2.1) for details.

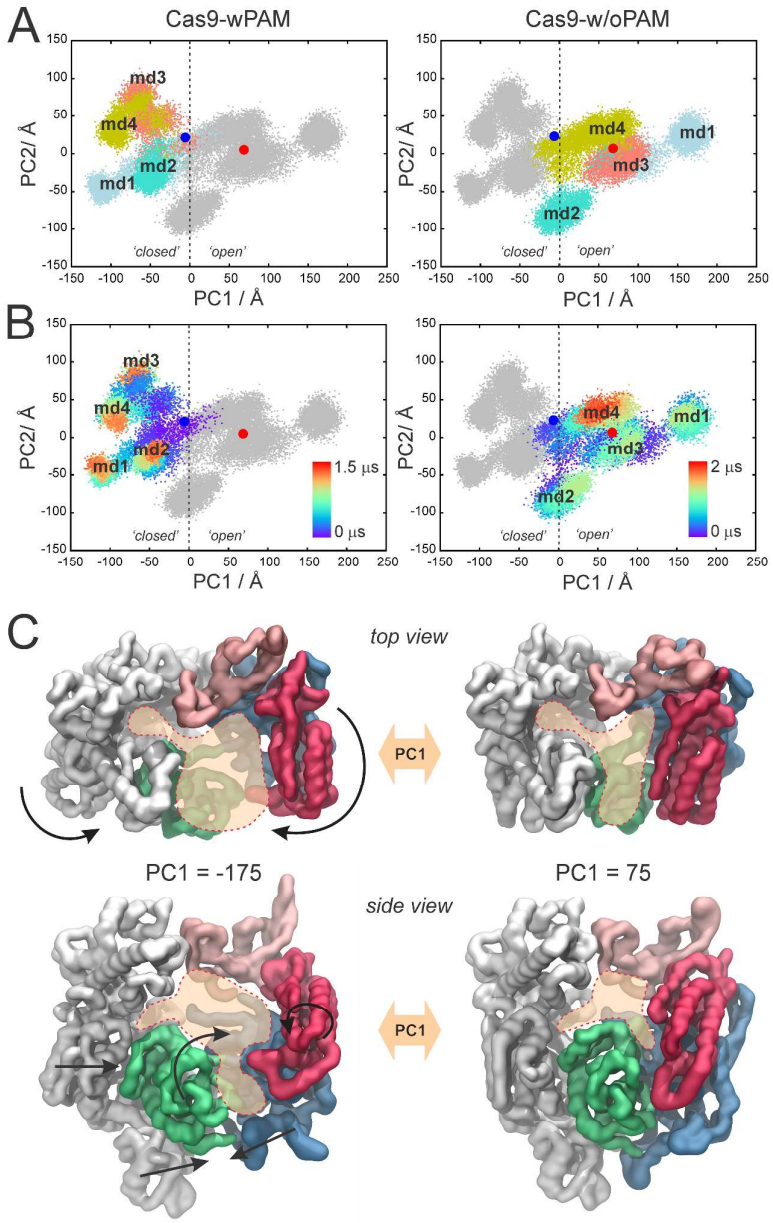


Figure S3. Essential dynamics of CRISPR-Cas9. **(A)** Conformational space explored by Cas9 in the essential subspace of its principal components (PC1 *vs.* PC2), in the presence (Cas9–wPAM, left) or absence (Cas9–w/oPAM, right) of the PAM sequence, as from four independent MD runs of each system (i.e., total of 8 simulations of ~1.5 µs, each). The initial X-ray conformations are also shown as reference (blue circle: 4UN3.pdb; red circle: 4OO8.pdb). **(B)** Time evolution of Cas9 in the essential subspace along four independent MD runs of Cas9–wPAM (red–to–blue) and Cas9–w/oPAM (yellow–to–green). **(C)** Principal mode of motion (PC1) displayed by CRISPR-Cas9 corresponds to a contraction of the system, as viewed from top and side views. See the Supplementary Text (paragraph S2.1) for details.

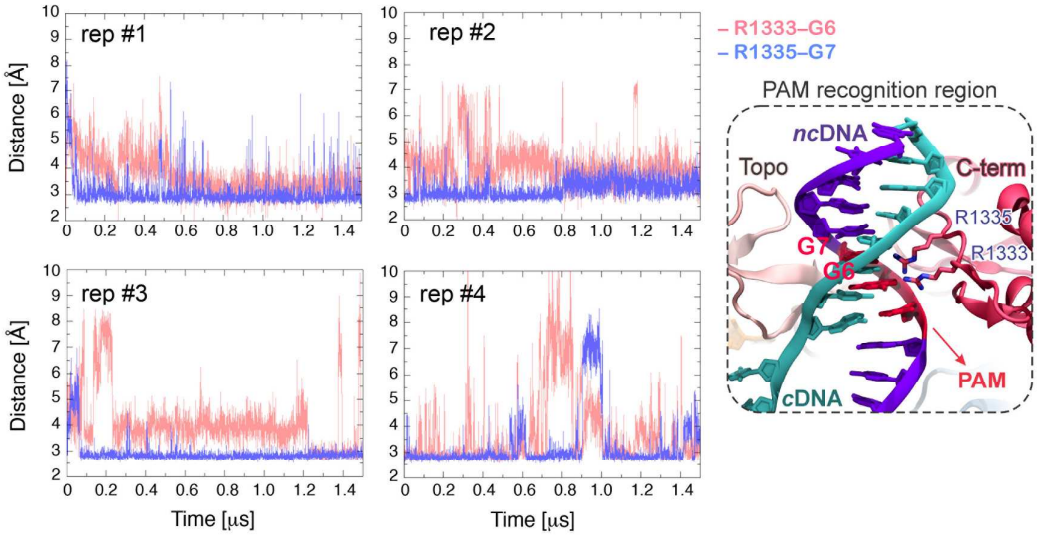


Figure S4. Time evolution of the H-bond interactions established by the PAM guanine bases with the interacting protein residues R1333 and R1335 along four simulation replicas (rep. #1–4) of the Cas9–wPAM system. Specifically, the graphs report the distance (expressed in Å) between the nitrogen atom of the arginine side-chains (one of the two equivalent terminal nitrogen atoms) and the carbonyl oxygen of the guanine bases, for the R1333–G6 (pink) and R1335–G7 (blue) interactions.

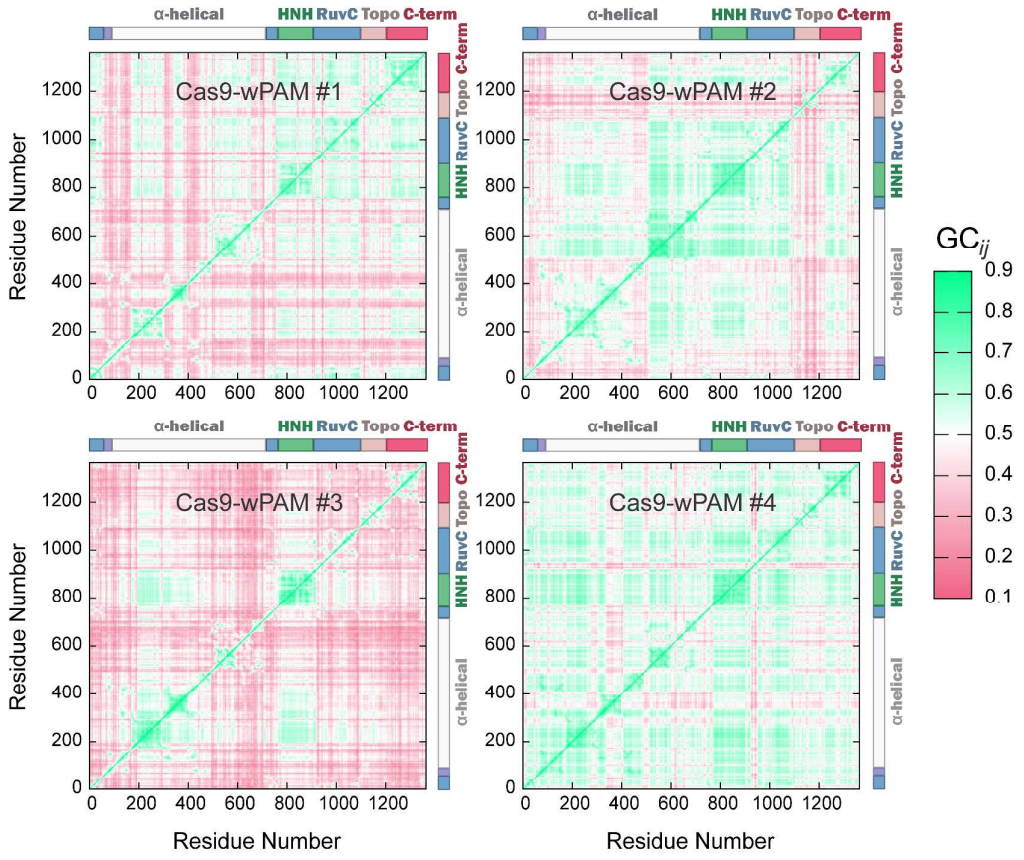


Figure S5. Generalized Correlation (GC_{ij}) matrices of the Cas9–wPAM system, calculated over four MD runs (i.e., replicas) of ~1.5 μs. The strength of the computed correlations is color-coded from green (highly correlated motions) to violet (poorly correlated motions).

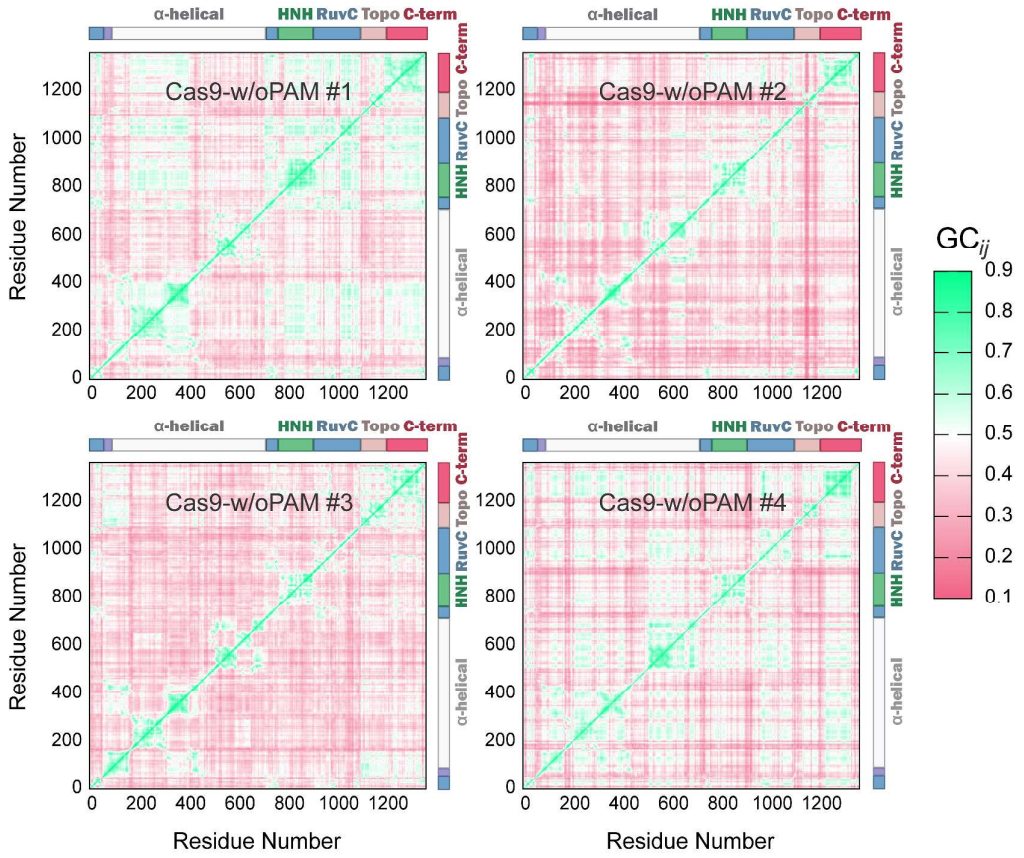


Figure S6. Generalized Correlation (GC_{ij}) matrices of the Cas9–w/oPAM system, calculated over four MD runs (i.e., replicas) of ~1.5 μs. The strength of the computed correlations is color-coded from green (correlated motions) to violet (poorly correlated motions).

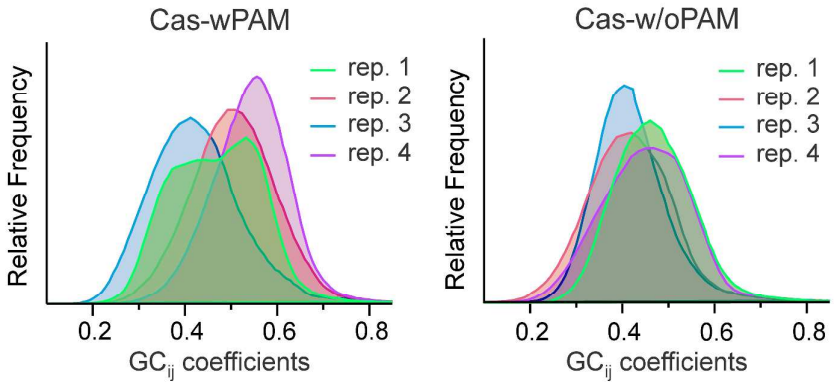


Figure S7. Probability distributions of the Generalized Correlation (GC_{ij}) coefficients of the Cas9–wPAM **(left)** and Cas9–w/oPAM **(right)** systems, calculated over four MD runs (i.e., replicas) of ~1.5 μs.

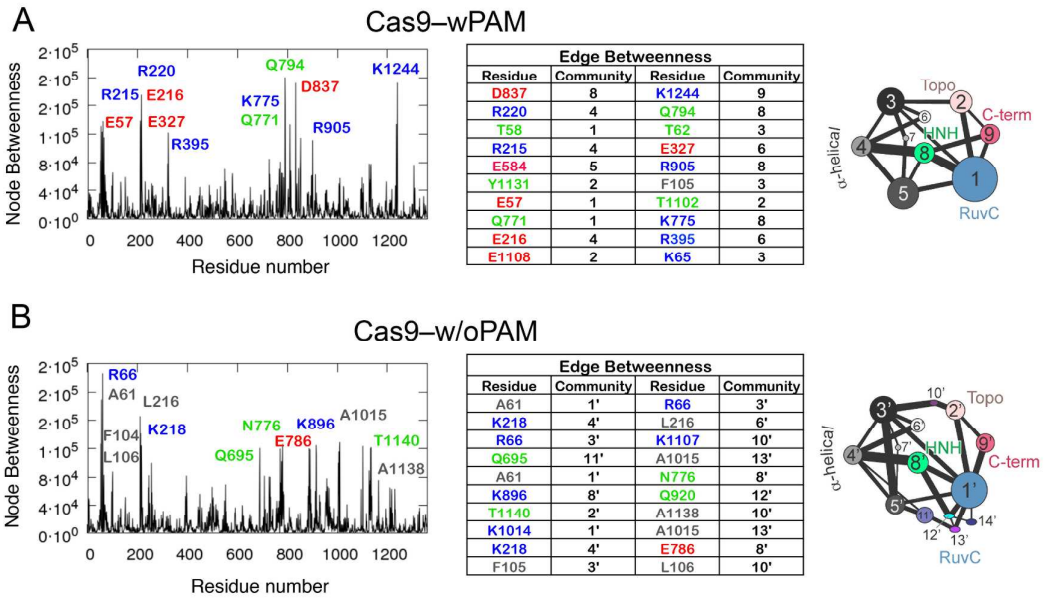
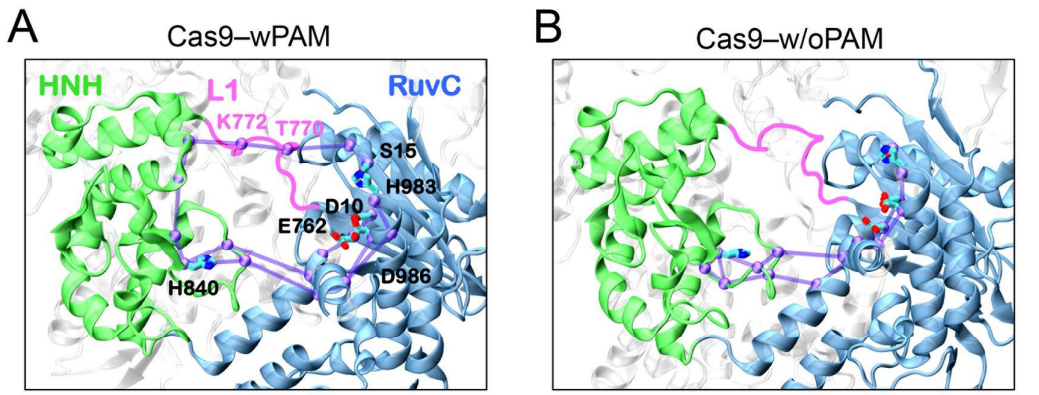


Figure S8. Per-residue node betweenness (graphs, left panel) and edge betweenness of pair nodes connecting different communities (tables, right panel), calculated for the Cas9–wPAM (A) and Cas9–w/oPAM (B) systems. The tables report pairs of residues displaying the highest edge betweenness and their respective community, color-coded per residue type (i.e., blue: basic, red: acidic, green: polar, gray: non polar). The community structure of Cas9–wPAM and Cas9–w/oPAM is also shown. Full details are reported in the Supplementary Text (paragraph S2.4).



Source (RuvC) –> Sink (HNH)								PL	PD
E762	S960	K961	D849	H840				0.604	1
D986	R967	D965	K848	D849	H840			0.717	0
D10	E762	S960	K961	D849	H840			0.735	1
H983	H985	R967	D965	K848	D849	H840		0.851	21
S15	N14	T770	K772	Q807	E809	K855	H840	0.855	0

Source (RuvC) –> Sink (HNH)								PL	PD
E762	K959	Q920	L847	V842	H840			0.754	7
D986	R967	S964	R961	K848	Q844	H840		0.910	10
D10	E762	K959	Q920	K847	V842	H840		0.884	7
H983	D986	R967	S964	K961	K848	Q844	H840	1.037	10
S15	D10	E762	K959	Q920	L847	V842	H840	1.047	15

Figure S9. *"Shortest pathways"* connecting pairs of catalytic residues of the RuvC and HNH domains of the Cas9–wPAM (A) and Cas9–w/oPAM (B) systems. *"Shortest pathways"* were computed considering as a *"source"* the catalytic residues of the RuvC domain (E762, D986, D10, H983, S15) and as a *"sink"* the catalytic residue of the HNH domain (H840), such tracking the information transfer from the *"source"* to the *"sink"*. The identified *"shortest pathways"* are represented on the 3D structure of Cas9 using lines, while spheres are used to identify the pathway residues (Cα atom). *"Source"* to the *"sink"* residues are labeled, as well as the key residues K722 and T700 (magenta) of the L1 loop. Cas9 is shown as cartoons, highlighting the HNH (green) and RuvC (blue) domains, as well as the L1 loop (magenta). The bottom tables detail the identified *"shortest pathways"* between the catalytic residues of the RuvC (*"source"*) and HNH (*"sink"*) domains. *"Pathway length"* (*PL*) and *"degeneracy"* (*PD*) are also reported.

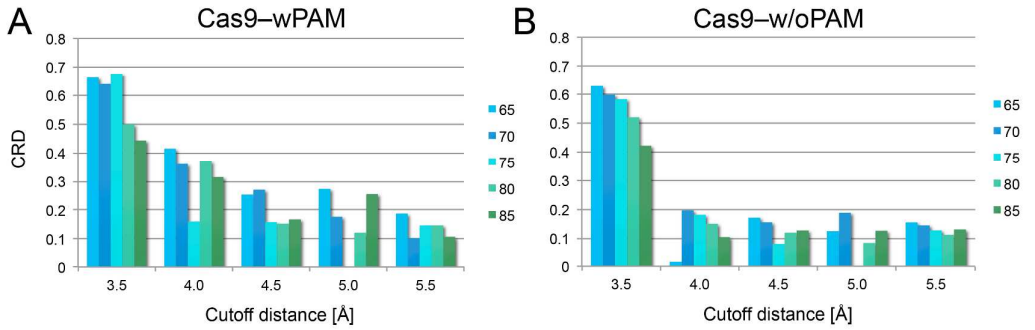


Figure S10. Community Repartition Difference (CRD) for the Cas9–wPAM (A) and Cas9–w/oPAM (B) systems, calculated for the complex community structures. The *"distance cutoff"* (reported on the X axis, ranging from 3.5 to 5.5 Å) defines a contact in each frame of the simulation, while the *"frame cutoff"* (ranging from 65 to 85 %, as in the legend) is the percentage of frames the contact is formed. CRD equals to 0 if two community structures are identical, while it assumes value 1 if the two structures are totally different. For the CNA performed in this study, we have chosen a 5 Å *"distance cutoff"* and 75 % *"frame cutoff"*.

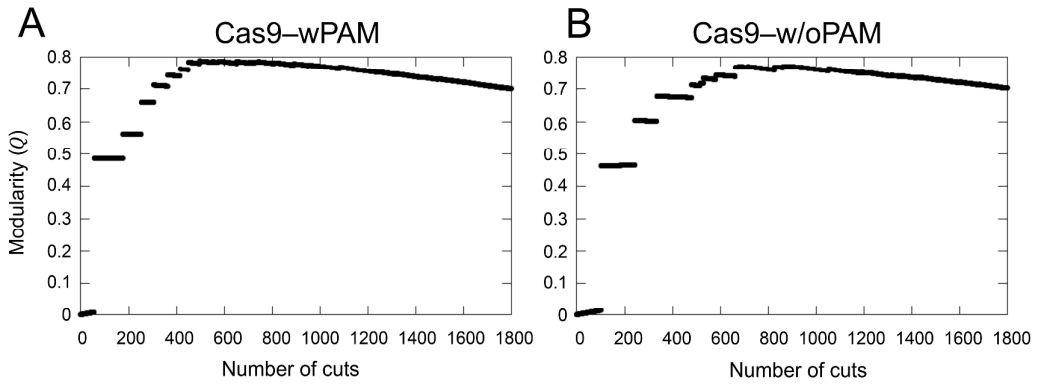


Figure S11. Modularity (Q) as a function of the number of cuts for the Cas9–wPAM (A) and Cas9–w/oPAM (B) systems, as obtained from the edge cutting procedure of the Girvan-Newman algorithm.

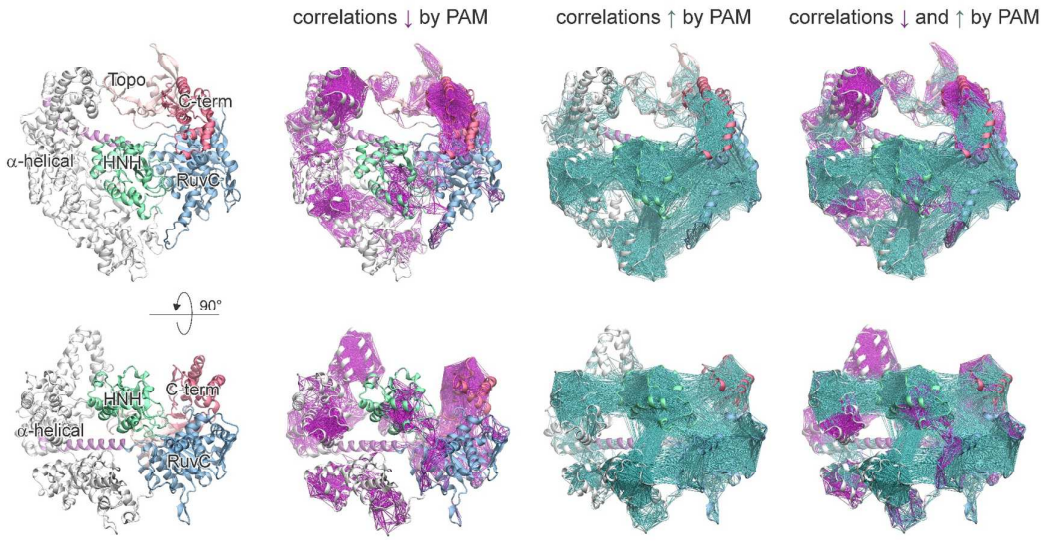


Figure S12. Correlated motions that are weakened (magenta) or increased (cyan) by the presence of PAM are plotted on the 3D structure of Cas9 (detail in the main text). Cas9 is shown on front (top) and rotated by 90° (bottom). Full details and discussion are reported in the Supplementary Text (paragraph S2.5).

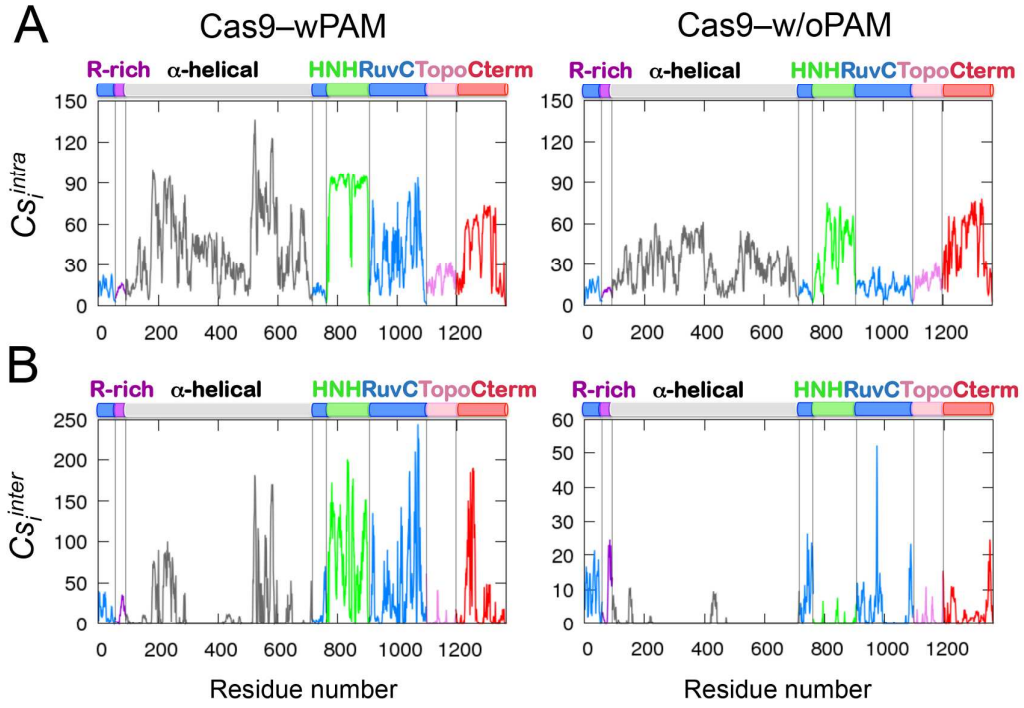


Figure S13. Intra (Cs_i^{intra}, A) and Inter (Cs_i^{inter}, B) per-residue Correlation Score (Cs_i), as calculated from the averaged GC_{ij} matrices of the Cas–wPAM (left column) and the Cas–w/oPAM (right column) systems.

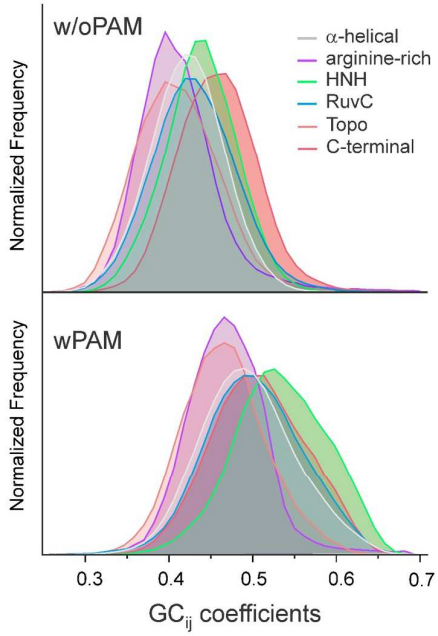


Figure S14. Probability distribution of the *inter*-domain (Cs_i^{inter} , B) Correlation Score (Cs_i), shown for each domain of the Cas–w/oPAM (top) and the Cas–wPAM (bottom) systems.

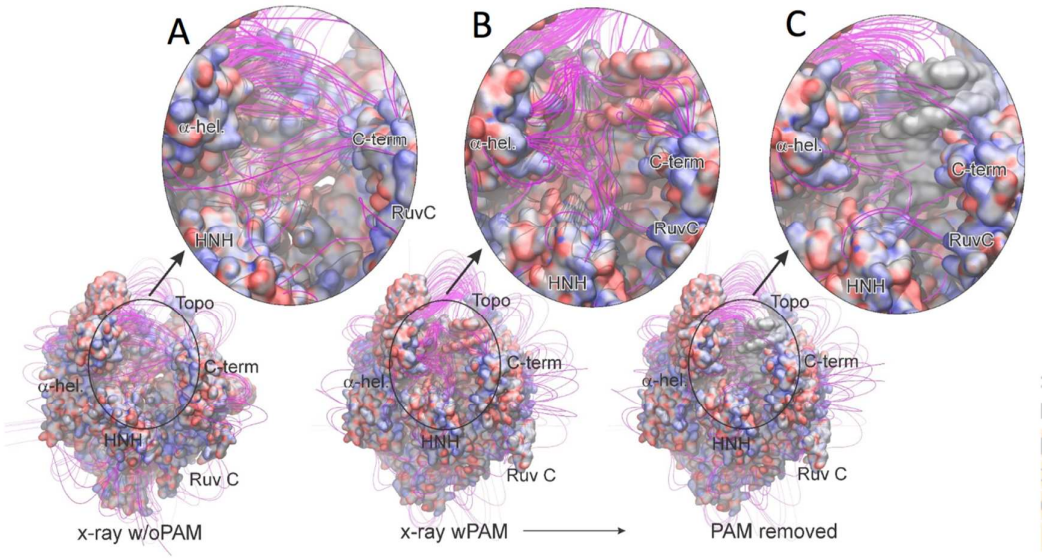


Figure S15. Electrostatic potential (red, negative (–5 *kT/e*); blue, positive (+5 *kT/e*)), mapped on the 3D structures of CRISPR-Cas9, as crystallized in the absence of the DNA including PAM (4OO8.pdb[2] (A)), in the presence of the PAM sequence (4UN3.pdb[1] (B)), as well as upon deletion of the PAM sequence from the crystallized 4UN3.pdb structure (C). Field lines (in purple) are used to indicate the electrostatic pathways along which point charges are most likely to travel (i.e., the direction of the electrostatic field). Field lines were filtered according to a gradient magnitude > 4.9 *kT/e*/Å and length > 13.5 Å using the Visual Molecular Dynamics (VMD) package.[41]

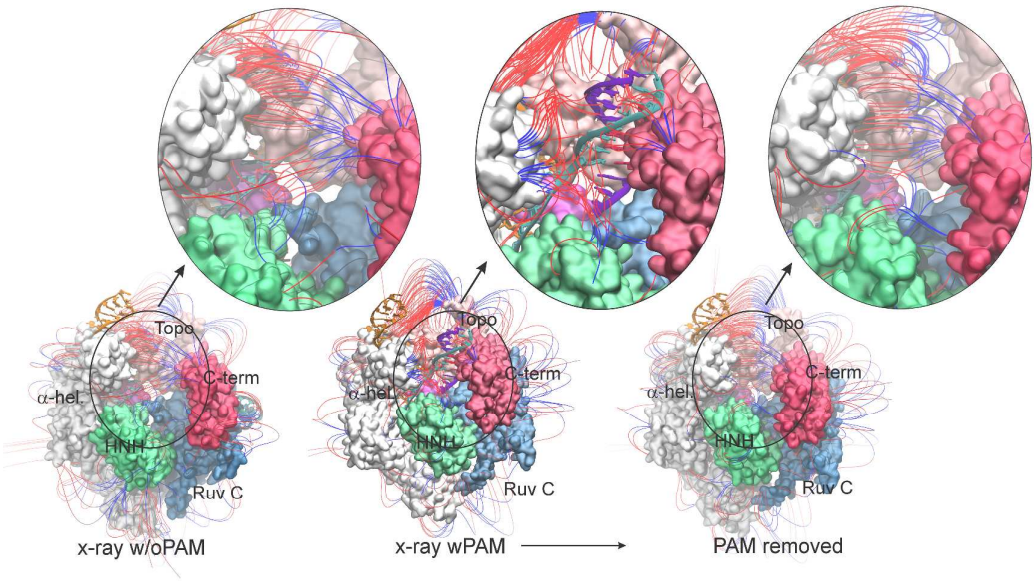


Figure S16. Electrostatic properties of the CRISPR-Cas9 system, as crystallized in the absence of the DNA including PAM (4OO8.pdb[2] (A)), in the presence of the PAM sequence (4UN3.pdb[1] (B)), and upon deletion of the PAM sequence from the crystallized 4UN3.pdb structure (C). Electrostatic properties are represented using field lines color-coded from blue (positive) to red (negative), indicating the direction that a positive charge would accelerate toward a negative charge if placed upon that line (i.e., the direction of the electrostatic field). Field lines were filtered according to a gradient magnitude > 4.9 *kT/e/*Å and length > 13.5 Å using the Visual Molecular Dynamics (VMD) package.[41] Cas9 is shown as a molecular surface colored by domain (as in Fig. 1 of the main text). The DNA segment including PAM (panel B) is shown as ribbons.

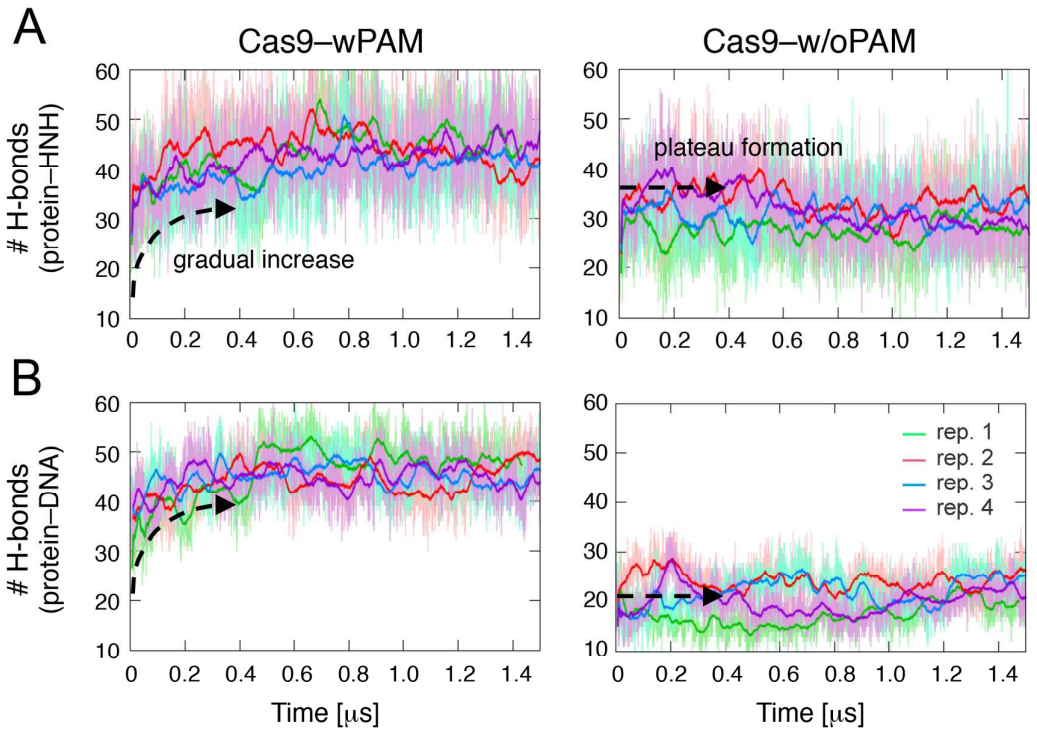


Figure S17. Time evolution of the number of H-bonds established by the Cas9 protein (not including the HNH domain) and the HNH domain (A) and by the Cas9 protein and the DNA (B), along four simulation replicas (rep. #1–4) of the Cas9–wPAM (left panel) and the Cas9–w/oPAM (right panel) systems. Two dashed arrows are used to indicate the gradual increase of H-bonding in Cas9–wPAM, as opposite to their plateau formation starting MD simulations of Cas9–w/oPAM. H-bonds were considered formed if all possible donors and acceptors laid within a distance cutoff radius of 3.5 Å and an angle cutoff of 30°.

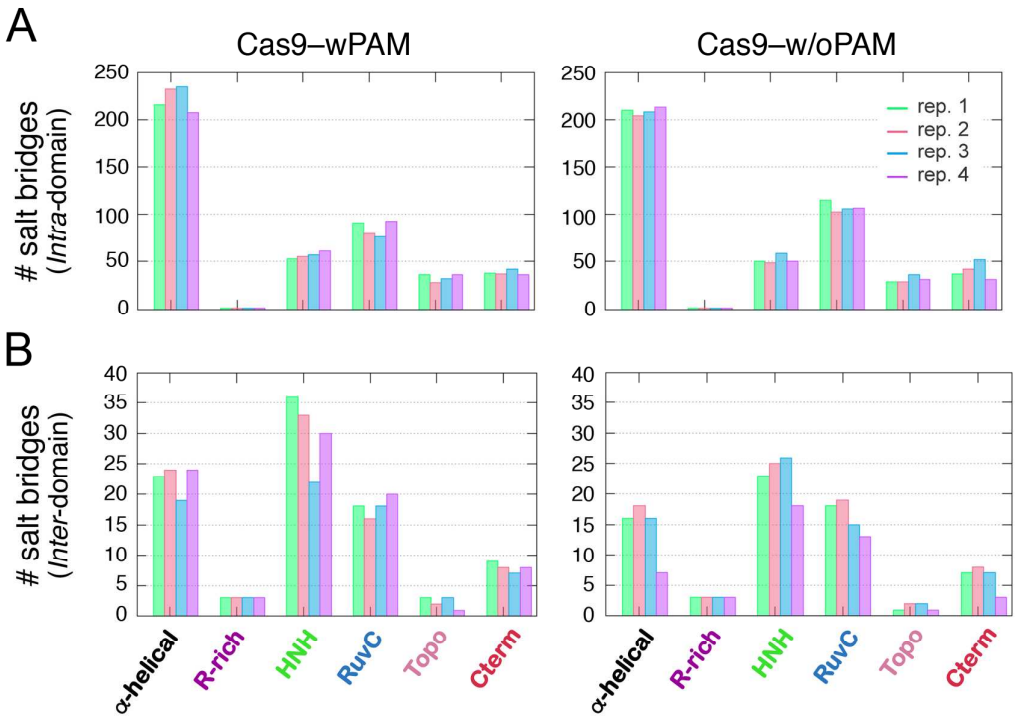


Figure S18. Number of salt-bridges formed by each of the Cas9 protein domains within itself (*Intra*–domain (A)) and with the remaining protein domains (*Inter*–domain (B)), calculated for the four simulation replicas (rep. #1–4) of the Cas9–wPAM (left panel) and the Cas9–w/oPAM (right panel) systems. Cas9 protein domains are listed on the X-axis, while the four simulations replicas are indicated with bars of different colors. Salt bridges were considered formed if a distance cutoff between any of the oxygen atoms of acidic residues and any of the nitrogen atoms of basic residues laid within 3.2 Å.

S4. Supplementary Movies

Movie S1. The movie shows the CRISPR-Cas9 system represented in molecular surface, with individual domains color-coded as in Fig. 1. After zooming on the PAM interacting region, the movie shows the PAM-induced electrostatics. Field lines indicate the direction of the electrostatic field from positive (blue) to negative (red) regions, highlighting that, upon site-specific binding of PAM, positively charged groups of the α-helical, C-term and HNH domains are attracted by the negatively charged DNA backbone, resulting in the *"open–to–close"* conformational transition. The movie concludes showing the *"open–to–close"* conformational transition of the Cas9 protein, as revealed by the first principal mode of motion (i.e., Principal Component – PC1).

S5. Supplementary References

(1) Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M. *Nature* **2014**, *513*, 569.

(2) Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O. *Cell* **2014**, *156*, 935.

(3) Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; Schwede, T. *Nucleic Acids Res* **2014**, *42*, W252.

(4) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. *Biophys J* **2007**, *92*, 3817.

(5) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Sponer, J.; Otyepka, M. *J Chem Theory Comput* **2010**, *6*, 3836.

(6) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E.; Jurecka, P. *J Chem Theory Comput* **2011**, *7*, 2886.

(7) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J Chem Phys* **1983**, *79*, 926.

(8) Aqvist, J. *J Phys Chem* **1990**, *94*, 8021.

(9) L. Casalino; G. Palermo; U. Rothlisberger; Magistrato, A. *J Am Chem Soc* **2016**, *138*, 10374.

(10) Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De Vivo, M. *Acc Chem Res* **2015**, *48*, 220.

(11) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. *ACS Cent Sci* **2016**, *2*, 756−763.

(12) Palermo, G.; Stenta, M.; Cavalli, A.; Dal Peraro, M.; De Vivo, M. *J Chem Theory Comput* **2013**, *9*, 857.

(13) Casalino, L.; Palermo, G.; Abdurakhmonova, N.; Rothlisberger, U.; Magistrato, A. *J Chem Theory Comput* **2016**.

(14) O'connell, M. R.; Oakes, B. L.; Sternberg, S. H.; East-Seletsky, A.; Kaplan, M.; Doudna, J. A. *Nature* **2014**, *516*, 263.

(15) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J comput Chem* **2005**, *26*, 1701.

(16) Hess, G.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J Comput Chem* **1997**, *18*, 1463.

(17) Nose, S. *Mol Phys* **1986**, *57*, 187.

(18) Hoover, W. G. *Phys Rev A* **1985**, *31*, 1695.

(19) Parrinello, M.; Rahman, A. *J Appl Phys* **1981**, *52*, 7182.

(20) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. *Proc Natl Acad Sci U S A* **2017**, in press. doi: 10.1073/pnas.1707645114.

(21) Guo, J.; Zhou, H. X. *Chem Rev* **2016**, *116*, 6503.

(22) Malmstrom, R.; Kornev, A.; Taylor, S.; Amaro, R. E. *Nat Commun* **2015**, *6*, 7588

(23) Hertig, S.; Latorraca, N. R.; Dror, R. O. *PLoS Comput Biol* **2016**, *12*, e1004966.

(24) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *P Natl Acad Sci USA* **2001**, *98*, 10037.

(25) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Res* **2007**, *35*, W522.

(26) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Res* **2004**, *32*, W665.

(27) Lange, O. F.; Grubmuller, H. *Prot: Struct Funct Bioinf* **2006**, *62*, 1053.

(28) Lindahl, E.; Hess, B.; Van Der Spoel, D. *J Mol Model* **2001**, *7*, 306.

(29) Floyd, R. W. *Commun Acm* **1962**, *5*, 345.

(30) Ricci, C. G.; Silveira, R. L.; Rivalta, I.; Batista, V. S.; Skaf, M. S. *Sci Rep-Uk* **2016**, *6*.

(31) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins* **1993**, *17*, 412.

(32) Girvan, M.; Newman, M. E. *Proc Natl Acad Sci U S A* **2002**, *99*, 7821.

(33) Newman, M. E.; Girvan, M. *Phys Rev E* **2004**, *69*, 026113.

(34) Rivalta, I.; Sultan, M. M.; Lee, N. S.; Manley, G. A.; Loria, J. P.; Batista, V. S. *Proc Natl Acad Sci U S A* **2012**, *109*, E1428.

(35) Durrant, J. D.; de Oliveira, C. A.; McCammon, J. A. *M Mol Graph Model* **2011**, *29*, 773.

(36) Durrant, J. D.; Votapka, L.; Sorensen, J.; Amaro, R. E. *J Chem Theory Comput* **2014**, *10*, 5047.

(37) Jinek, M.; Jiang, F.; Taylor, D. W.; Sternberg, S. H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S.; Kaplan, M.; Iavarone, A. T.; Charpentier, E.; Nogales, E.; Doudna, J. A. *Science* **2014**, *343*, 1247997.

(38) Sternberg, S. H.; LaFrance, B.; Kaplan, M.; Doudna, J. A. *Nature* **2015**, *527*, 110.

(39) Jiang, F. G.; Taylor, D. W.; Chen, J. S.; Kornfeld, J. E.; Zhou, K. H.; Thompson, A. J.; Nogales, E.; Doudna, J. A. *Science* **2016**, *351*, 867.

(40) Slaymaker, I. M.; Gao, L.; Zetsche, B.; Scott, D. A.; Yan, W. X.; Zhang, F. *Science* **2016**, *351*, 84.

(41) Jiang, F. G.; Taylor, D. W.; Chen, J. S.; Kornfeld, J. E.; Zhou, K. H.; Thompson, A. J.; Nogales, E.; Doudna, J. A. *Science* **2016**, *351*, 867.

(42) Humphrey, W.; Dalke, A.; Schulten, K. *J Mol Graph* **1996**, *14*, 33.