Supporting Information Appendix

Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics

Kyle W. East,[1] Jocelyn C. Newton,[1,a] Uriel N. Morzan,[2,b] Yogesh Narkhede,[3] Atanu Acharya,[2,c] Erin Skeens,[1] Gerwald Jogl,[1] Victor S. Batista,[2] Giulia Palermo[3*] and George P. Lisi[1*]

1. Department of Molecular Biology, Cell Biology & Biochemistry, Brown University, Providence, RI 02903
2. Department of Chemistry, Yale University, New Haven, CT 06520
3. Department of Bioengineering, University of California Riverside, Riverside, CA 92521

[a]Present address: The Hospital for Sick Children Research Institute, Toronto, ON, Canada

[b]Present address: The Abdus Salam International Center for Theoretical Physics, Condensed Matter Section, Strada Costiera 11, 34151, Trieste, Italy

[c]Present address: School of Physics, Georgia Institute of Technology, Atlanta, Georgia, 30332

*Corresponding authors:

George P. Lisi (george_lisi@brown.edu), Giulia Palermo (giulia.palermo@ucr.edu)

Key Words:

CRISPR-Cas9, protein dynamics, allostery, NMR, MD simulations

Table of Contents

1. Supplementary Materials and Methods

1.1. Protein Expression and Purification

The HNH domain of *S. pyogenes* Cas9 (residues 775-908) was engineered into a pET15b vector with an N-terminal His_6-tag. Isotopically enriched HNH was expressed in Rosetta(DE3) cells in M9 minimal medium containing MEM vitamins, $MgSO_4$ and $CaCl_2$. Cells were induced with 0.5 mM IPTG after reaching an OD_{600} of 0.8 – 1.0 and grown for 16 – 18 hours at 22 ºC post induction. The cells were harvested by centrifugation, resuspended in a buffer containing 20 mM HEPES, 500 mM KCl, and 5 mM imidazole at pH 8.0, and lysed by ultrasonication. HNH was purified on a Ni-NTA column and its His_6-tag was removed by incubation with TEV for 4 hours at 20 ºC as confirmed by SDS-PAGE. NMR samples were dialyzed into a buffer containing 20 mM HEPES, 80 mM KCl, 1 mM DTT and 7.5% (v/v) D_2O at pH 7.4.

1.2. X-ray Crystallography

Following TEV cleavage, samples for crystallization were subsequently purified with a HiPrep 16/60 Sephacryl 100 S-100 HR gel filtration column equilibrated with a buffer of 10 mM HEPES and 100 mM KCl at pH 8.0. Crystals were obtained with sitting drop vapor diffusion at room temperature by mixing 48 mg/mL HNH 1:1 with the Molecular Dimensions Morpheus I Screen condition E4 (0.1 M mixture of [imidazole and MES] pH 6.5, 25% (v/v) mixture of [2-methyl-2,4-pentanediol, PEG1000, and PEG3350], and 0.3 M mixture of [diethylene glycol, triethylene glycol, tetraethylene glycol, and pentaethlyene glycol]). Crystals were flash frozen in liquid nitrogen and diffraction data were collected on a Rigaku MicroMax-003i sealed tube X-ray generator with a Saturn 944 HG CCD detector. HNH crystals diffracted beyond the limit of the detector. Data were processed and scaled using XDS[1] and Aimless in the CCP4 program suite.[2] The HNH domain from full-length *S. pyogenes* Cas9 was used for molecular replacement (PDB: 4UN3)[3] with Phaser in the PHENIX software package.[4] Iterative rounds of manual building in Coot[5] and refinement in PHENIX yielded the final HNH domain structure. Alignment

coordinates of the full-length *S. pyogenes* Cas 9 and predicted NMR structures with the isolated HNH domain crystal structure yielded an RMSD of 0.549 Å and 0.479 Å, respectively.

1.3. NMR Spectroscopy

Amide backbone resonances of ^{2}H, ^{13}C, ^{15}N-HNH were assigned at 25 ºC on a Varian Inova 600 MHz spectrometer with TROSY-HNCA, -HNcoCA, -HNcaCB, -HNcocaCB, -HNcaCO, and -HNCO experiments and deposited into the BMRB. All NMR spectra were processed with NMRPipe [6] and analyzed in SPARKY.[7] Backbone chemical shift data was uploaded to the CS23D server for secondary structure calculations.

NMR spin relaxation experiments were carried out at 600 and 850 MHz on Bruker Avance NEO and Avance III HD spectrometers, respectively. Carr-Purcell-Meiboom-Gill (CPMG) NMR experiments probing millisecond timescale motions of backbone amide groups were performed at 25 ºC. The carbon-decoupled CPMG pulse sequence, adapted from the report of Palmer and coworkers,[8] used a constant relaxation period of 40 ms, a 2.0 second recycle delay, and τ_{cp} points of 0.555, 0.625, 0.714, 0.833, 1.0, 1.25, 1.5, 1.667, 2.5, 5, 10, and 20 ms. Transverse relaxation rates (R_2) were determined from peak intensities of each resonance at multiple τ_{cp} delays with an in-house curve-fitting script. Relaxation dispersion profiles were generated by plotting R_2 vs. $1/\tau_{cp}$ and exchange parameters were obtained from fits of these data carried out with in-house scripts and in RELAX under the R2eff, NoRex, Tollinger (TSMFK01), and Carver-Richards (CR72 and CR72-Full) models.[9,10] Two-fields relaxation dispersion data were fit simultaneously and uncertainty values were obtained from replicate spectra.

Picosecond-to-nanosecond dynamics were measured with relaxation times of 0(x2), 40, 80, 120, 160(x2), 200, 240, 280(x2), 320, 360, and 400 ms for T_1 and 4.18, 8.36(x2), 12.54, 16.72, 20.9(x2), 25.08(x2), 29.26, 33.44, 37.62, and 41.8 ms for T_2. Peak intensities were quantified in Sparky and the resulting decay profiles were analyzed in Mathematica with errors determined from the fitted parameters. Steady-state ^{1}H-[^{15}N] NOE were measured with a 5 second relaxation delay followed by a 3 second saturation (delay) for the saturated (unsaturated) experiments. All relaxation experiments were carried out

in a temperature-compensated interleaved manner. Model-free analysis using the Lipari-Szabo formalism was carried out on dual-field NMR data in RELAX with fully automated protocols.[9]

1.4. Computational Structural Models

MD simulations were performed on several model systems of the entire CRISPR-Cas9 complex and the isolated HNH domain. MD simulations of the CRISPR-Cas9 complex were based on the X-ray structure of the full-length wild-type (WT) Cas9 protein, solved at 2.58 Å resolution (PDB code: 4UN3).[3] This system comprises a total of 1368 amino acids. Missing residues in the X-ray structure were added via homology modeling, as in our previous papers,[11,12] by using the SWISS-MODEL software by Schwede.[13] This nucleoprotein complex was used to simulate the CRISPR-Cas9 complex as WT, and with alanine mutations. Two mutated model systems were built by substituting alanine at key residues forming the allosteric pathway identified in this work. The first mutant system substituted alanine at residues belonging to the computational pathway that also show ms dynamics via CPMG relaxation dispersion (G792, Q794, K797, E798, L816, Q817, N818, R820, V838, H840, I841, and K855; Figure 4), while the second mutant system introduces alanine mutations at all residues that experimentally show slow dynamics (Table S2). MD simulations of the isolated HNH domain were based on both the NMR model derived from the ^{1}H-^{15}N-HSQC data (Figure 1A, close-up view) and on the X-ray structure of HNH solved in this work at 1.3 Å resolution (Figure 2). Both systems are composed of residues 775–908 and align well to the HNH domain in the X-ray structure of the WT CRISPR-Cas9 (PDB code: 4UN3).[3] Indeed, with respect to this WT structure, the NMR structure displays a Cα RMSD of 0.688 Å, while the X-ray structure of HNH shows a Cα RMSD of 0.549 Å. Notably, the NMR and X-ray structures of the isolated HNH domain also align very well with respect to each other (Cα RMSD of 0.479 Å). All model systems were embedded in explicit water, adding Na+ counter-ions to neutralize the total charge. A total of ~220,000 atoms and a box size of ~145 x 110 x 147 Å^3 has been reached for the CRISPR-Cas9 complex; while ~25,000 atoms and a box size of ~72 x 62 x 60 Å characterize the isolated HNH domain from both NMR and X-ray crystallography experiments.

1.5. Molecular Dynamics (MD) Simulations

The above-mentioned model systems were equilibrated through conventional MD using a well-established protocol for protein/nucleic acid complexes,[14] We employed the Amber ff12SB force field, which includes the ff99bsc0 corrections for DNA[15] and the ff99bsc0+χ*OL3* corrections for RNA.[16,17] The Åqvist force field parameters have been employed for the Mg^{2+} ions.[18] This force field model has been shown to properly describe the conformational dynamics of CRISPR-Cas9 during extensive MD simulations, performed using both classical[11,12,19] and accelerated MD.[20,21] It has also been employed to perform multi–μs continuous MD simulations,[22] preserving the overall fold of the structure and capturing critical conformational changes. The TIP3P model was employed for waters.[23] MD simulations were carried out using an integration time step of 2 fs. Hydrogen atoms were added assuming standard bond lengths and constrained to their equilibrium position with the SHAKE algorithm. Temperature control (300 K) was performed via Langevin dynamics,[24] with a collision frequency γ = 1 per picosecond. Pressure control was accomplished by coupling the system to a Berendsen barostat,[25] at a reference pressure of 1 atm and a relaxation time of 2 ps. All simulations were carried out as follows. First, the systems were subjected to energy minimization to relax the water molecules and counter ions, keeping the protein (and in the case of the Cas9 complex, also the RNA, DNA, and Mg^{2+} ions) fixed with harmonic position restraints of 300 kcal/mol · $Å^2$. Then, the systems were heated from 0 - 100 K in the canonical ensemble (NVT), by running two NVT simulations of 5 ps each, imposing position restraints of 100 kcal/mol · $Å^2$ on the above-mentioned elements of these systems. The temperature was further increased to 200 K in ~100 ps of MD in the isothermal-isobaric ensemble (NPT), in which the restraint was reduced to 25 kcal/mol · $Å^2$. Subsequently, all restraints were released and the temperature of the system was raised to 300 K in a single 500 ps NPT simulation. After ~1.1 ns of equilibration, ~10 ns of NPT production was carried out allowing the density of the system to stabilize around 1.01 g/cm^{-3}. Finally, MD simulations were carried out in the NVT ensemble, collecting ~100 ns for each system (obtaining a total of ~500 ns of production

runs). These well-equilibrated systems have been used as the starting point for Gaussian accelerated MD (GaMD, details below). All simulations were performed with the GPU accelerated pmemd code of AMBER 16.[26]

1.6. Gaussian Accelerated MD Simulations (GaMD)

Accelerated MD (aMD) is an enhanced sampling method that adds a boost potential to the Potential Energy Surface (PES), effectively decreasing the energy barriers and accelerating transitions between low-energy states.[27] The method extends the capability of MD simulations over long timescales, providing routine access to the millisecond dynamics.[28] Indeed, by running hundreds of nanosecond trajectories, aMD simulations can capture motions occurring over slow timescales (μs and *ms*), and have been shown to provide excellent comparability with solution NMR experiments.[29-31] Here, we applied a novel and robust aMD method, namely a Gaussian aMD (GaMD),[32] which uses harmonic functions to construct a boost potential that is adaptively added to the PES, enabling unconstrained enhanced sampling and simultaneous reweighting of the canonical ensemble. The method has been shown to capture long timescale events in a variety of biological systems,[33-36] including protein/nucleic acid complexes, such as CRISPR-Cas9.[37-39]

Considering a system with *N* atoms at positions $\vec{r} = \{\vec{r_1} \ldots \vec{r_N}\}$, when the system potential $V(\vec{r})$ is lower than a threshold energy *E*, the energy surface is modified by a boost potential as:

$$V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r}), \qquad V(\vec{r}) < E, \tag{1}$$

$$\Delta V(\vec{r}) = \frac{1}{2}k\left(E - V(\vec{r})\right)^2, \tag{2}$$

where *k* is the harmonic force constant. The two adjustable parameters *E* and *k* are automatically determined by applying the following three criteria. First, for any two arbitrary potential values $V_1(\vec{r})$ and $V_2(\vec{r})$ found on the original energy surface, if $V_1(\vec{r}) < V_2(\vec{r})$, ΔV should be a monotonic function that does not change the relative order of the biased potential values, *i.e.* $V_1^*(\vec{r}) < V_2^*(\vec{r})$. Second, if $V_1(\vec{r}) < V_2(\vec{r})$, the potential difference observed on the smoothed energy surface should be smaller than that of

the original, *i.e.* $V_2^*(\vec{r}) - V_1^*(\vec{r}) < V_2(\vec{r}) - V_1(\vec{r})$. By combining the first two criteria with Eqn [1] and [2], we obtain:

$$V_{max} \leq E \leq V_{min} + 1/k, \quad [3]$$

where V_{min} and V_{max} are the system minimum and maximum potential energies. To ensure that Eqn. [4] is valid, k must satisfy $k \leq 1/V_{max} - V_{min}$. By defining $k \equiv k_0\, 1/V_{max} - V_{min}$, then $0 < k \leq 1$. Lastly, the standard deviation of ΔV must be narrow enough to ensure accurate reweighting using cumulant expansion to the second order: $\sigma_{\Delta V} = k(E - V_{avg})\sigma_V \leq \sigma_0$, where V_{avg} and σ_V are the average and standard deviation of the system potential energies, $\sigma_{\Delta V}$ is the standard deviation of ΔV and σ_0 as a user-specified upper limit (e.g., 10 k_BT) for accurate reweighting. When E is set to the lower bound, $E = V_{min}$, according to Eqn. [4], k_0 can be calculated as:

$$k_0 = \min(1.0, k_0') = \min\left(1.0, \frac{\sigma_0}{\sigma_V} \cdot \frac{V_{max} - V_{min}}{V_{max} - V_{avg}}\right). \quad [4]$$

Alternatively, when the threshold energy E is set to its upper bound $E = V_{min} + 1/k$, k_0 is:

$$k_0 = k_0'' \equiv \left(1 - \frac{\sigma_0}{\sigma_V}\right) \cdot \frac{V_{max} - V_{min}}{V_{avg} - V_{min}}, \quad [5]$$

if k_0'' is calculated between *0* and *1*. Otherwise, k_0 is calculated using Eqn. [4], instead of being set to 1 directly as described in the original paper.[32] GaMD yields a canonical average of an ensemble by reweighting each point in the configuration space on the modified potential by the strength of the Boltzmann factor of the bias energy, $e^{\wedge[\beta \Delta V(rt(i))]}$ at that particular point.

Based on extensive tests on the CRISPR-Cas9 system,[37-39] the system threshold energy is $E = V_{max}$ for all GaMD simulations. The boost potential was applied in a *dual-boost* scheme, in which two acceleration potentials are applied simultaneously to the system: *(i)* the torsional terms only and *(ii)* across the entire potential. A timestep of 2 fs was used. The maximum, minimum, average, and standard deviation values of the system potential (V_{max}, V_{min}, V_{avg} and σ_V) were obtained from an initial ~12 ns

NPT simulation with no boost potential. GaMD simulations were applied to the CRISPR-Cas9 complex and our HNH domain construct. Each GaMD simulation proceeded with a ~50 ns run, in which the boost potential was updated every 1.6 ns, thus reaching equilibrium. Finally, ~400 ns of GaMD simulations were carried in the NVT *ensemble* out for each system in two replicas. Considering that we have simulated three model systems of CRISPR-Cas9 (one system as WT and two systems including alanine mutations, as described above) and two model systems of the isolated HNH domain (as derived from NMR and X-ray crystallography), a total of ~2.4 μs of GaMD for the CRISPR-Cas9 complex and ~1.6 μs of GaMD for the isolated HNH domain were completed. The choice of a simulation length of ~400 ns (for each replica) has been shown to exhaustively explore the conformational space of the CRISPR-Cas9 system.[37,38] All simulations have been performed with GPU accelerated pmemd code of AMBER 16.[26]

1.7. Determination of the Allosteric Pathways across the HNH domain

The allosteric pathway for information transfer within the HNH domain in the full-length Cas9 has been investigated by employing a computational protocol that harnesses correlation analysis and graph theory, and is composed of three stages.[40-43] First, the generalized correlations (GC_{ij}), which capture both linear and non-linear correlations between pairs of residues *i* and *j*, are computed and used as a metrics to define the optimization problem.[44] In this correlation analysis, 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:

$$MI\ [x_i,\ x_j] = \iint p(x_i,\ x_j)\ ln \frac{p(x_i, x_j)}{p(x_i) \cdot p(x_j)} dx_i dx_j \qquad [6]$$

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 $p(x_i)$

$$H[x] = \int p(x)\ ln\ p(x) dx \qquad [7]$$

and it can be thus computed as:

$$MI\left[x_i,\, x_j\right] = H\left[x_i\right] + H\left[x_j\right] - H\left[x_i, x_j\right] \quad [8]$$

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. 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:

$$GC_{ij}\left[x_i,\, x_j\right] = \left\{1 - e^{-\frac{2MI\left[x_i, x_j\right]}{d}}\right\}^{-\frac{1}{2}} \quad [9]$$

where d=3 is the dimensionality of x_i and x_j. GC_{ij} have been computed using a code developed within our group, utilizing the MI defined by Lange.[44] In a second phase, the GC_{ij} are used as a metric to build a dynamical network model of the protein.[42] In this model, the protein amino acids residues constitute the nodes of the dynamical network graph, 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:

$$w_{ij} = -\log GC_{ij} \quad [10]$$

In the present work, two nodes have been 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 70 % of the simulation time (i.e., *frame cutoff*). This leads to the definition of a set of elements w_{ij} of the graph. In the third phase of the protocol, the optimal pathways for the information transfer between two nodes (i.e., two amino acids) are defined using the Dijkstra algorithm,[45] which finds the roads, composed by inter-node connections, that minimize the total distance (and therefore maximize the correlation) between amino acids. Briefly, the Dijkstra algorithm defines a starting and a destination point, and its goal is to optimize iteratively a path from the former to the latter. In each iteration, the closest unvisited node is designated as the current node. From this current node, the distances to the remaining unvisited nodes are updated by determining the sum of the distance between the unvisited node and the value of the current node, if this value is less than the

unvisited intersection's current value, the distance is updated. This process continues until the destination is visited. In the present study, this protocol was applied on the trajectories of the full-length Cas9 simulated for ~400 ns of GaMD simulations and averaged over two replicas. The Dijkstra algorithm was applied between the amino acids 789/794 and 841/858, which belong to HNH and are located at the interface with RuvC and REC2, respectively. As a result, the routes that maximize the correlation between amino acids 789/794 and 841/858 are identified, providing residue–to–residue pathways that optimize the correlations (i.e., the momentum transport). Noteworthy, the momentum transmission between amino acids may not happen exclusively through the optimal path. Indeed, several alternative or *sub-optimal* paths may also contribute to the communication between distant sites. Therefore, together with the optimal motion transmission pathway the following 10 *sub-optimal* information channels where computed. **Fig. 4B** shows the 10 *sub-optimal* paths merged together in one, representing the most likely channel for motion transmission spanning HNH from the interface with RuvC to REC2. Residues belonging to the 10 *sub-optimal* pathways are G792*, Q794*, K797*, E798*, Y812, L813, Y814, L816*, Q817*, N818*, G819, R820*, D825, I830, V838*, D839, H840*, I841*, V842, P843†, Q844†, N854, K855*, V856, L857, T858, R859†, S860†, D861†,K862†; where the asterisk indicates the residues characterized by slow dynamics (experimentally characterized via CPMG relaxation dispersion and R_1R_2 (+1.5σ) measurements, **Tables S1-S2**), and the dagger indicates residues unassigned by NMR.

1.8. Principal Component Analysis (PCA)

PCA has been employed to capture the essential motions of the simulated systems. 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, which denotes how much each eigenvector is representative of the system dynamics. 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 PC can be obtained. 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"*.[46] In this work, each structure arising from the GaMD trajectories has been projected into the collective coordinate space defined by the first two eigenvectors (PC1 and PC2), such allowing the characterization the conformational space sampled by the isolated HNH construct and by the HNH embedded in the full-length Cas9 (***Fig. S5***). For consistency, each simulated system has been superposed onto the same reference structure (i.e., considering as a reference the HNH domain) and aligned, such allowing the projection into the same collective coordinate space. The Normal Mode Wizard plugin[47] of the Visual Molecular Dynamics (VMD) software[48] was used to visualize the essential dynamics along the principal eigenvectors and to draw the arrows highlighting their direction (***Fig. S4***).

1.9. NMR Chemical Shift Predictions

The structural ensemble derived from GaMD simulations was used to predict the $^{1}H^{15}N$ HSQC NMR spectrum of HNH, through the SHIFTX2 code,[49] which employs ensemble machine learning combined with a mixed-sequence, structure-based method. Its algorithm has been trained and tested with high-resolution X-ray structures (< 2.1 Å) with verified chemical shifts assignments. Building on the protein dataset, different machine learning techniques are applied to construct a multiple-regression model that predicts protein chemical shifts from coordinate data. SHIFTX2 combines two ensemble machine learning alghoritms: Bagging and Boosting. Bagging algorithm trains ''base learners'' (*i.e.* the individual learning algorithms of the ensemble) from a random sample of the original dataset and then averages the predictions of all the individual base learners. In contrast, Boosting algorithm trains subsequent base learners on mistakes of the previous base learner. SHIFTX2 further implements a feature selection alghorithm to select the optimal set of features (e.g., χ_2 and χ_3 angles, solvent accessibility, H-bond geometry, pH, temperature). The estimated correlation coefficients between experiments and predicted backbone chemical shifts are 0.9800 (^{15}N), 0.9959 ($^{13}C_\alpha$), 0.9992 ($^{13}C_\beta$), 0.9676 ($^{13}C'$), 0.9714 (^{1}HN), 0.9744 ($^{1}H_\alpha$) and RMS errors of 1.1169, 0.4412, 0.5163, 0.5330, 0.1711, and 0.1231 ppm, respectively. In this work, the prediction of the $^{1}H^{15}N$ HSQC NMR spectrum of the HNH domain has

been performed on the configurations obtained from GaMD simulations of the X-ray structure of the CRISPR-Cas9 complex and the HNH domain structure derived from NMR experiments and presented as ensemble averaged chemical shifts. The prediction of the chemical shifts through the SHIFTX2 code was performed over the structural ensemble derived from GaMD simulations, as reported in Figure 5 of the main text, and in Figure S12. To understand how the computed chemical shifts can be affected by the changes in the trajectories, we also computed the carbon chemical shifts over a set of static models, derived from the trajectories upon cluster analysis. For the aggregated trajectories of both isolated HNH and full-length Cas9, we first performed cluster analysis using the method described by Daura et al. in which for each point (i.e., MD frame)[50], the algorithm calculates the number of other frames for which the RMSD is lower than a given cutoff (which has been set to of 2.5 Å, accordingly to our previous study on CRISPR-Cas9 dynamics).[11] Then, the carbon chemical shifts were computed for each system over the 10 most populated clusters, which account for ~90 % of the overall population. Results from this analysis are reported in Figure S11, showing no significant difference in the computed carbon chemical shifts among the three structures. We note a slightly broader range of standard deviations for the backbone carbon chemical shifts, suggesting that these particular chemical shifts are slightly more sensitive to changes in the trajectories. It is worth noting, however, that both the SHIFTX and the experimentally measured chemical shifts are sensitive to the phi-psi angles of the backbone, while being relatively insensitive to the sidechain conformational changes. A comparison between the NMR chemical shifts computed over the ensemble of structures derived from cluster analysis and the experimental data is also reported in Figure S12.

2. Supplementary Figures

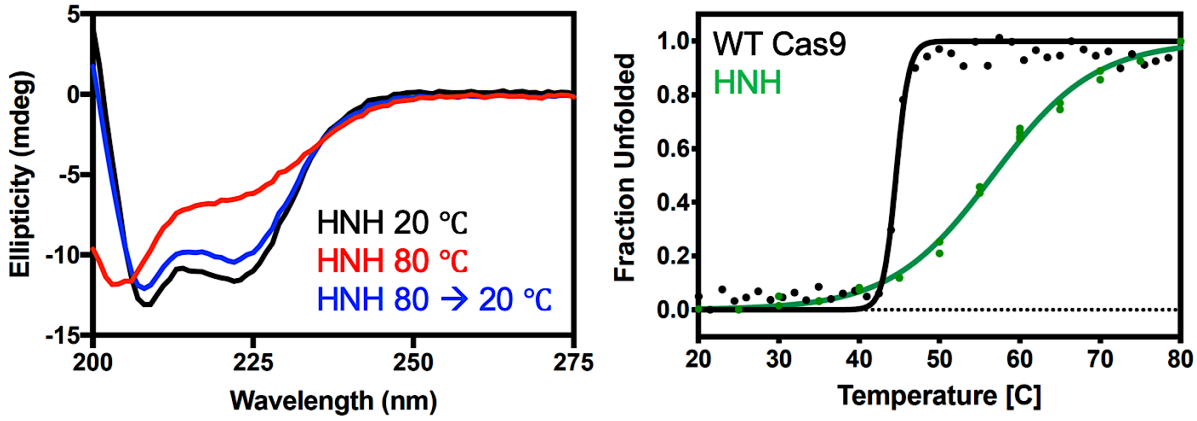

Fig. S1. Circular dichroism spectra of HNH collected at 20 °C (black, α-helical structure), 80 °C (red), and following cooling of denatured HNH back to 20 °C (blue). The blue spectrum highlights ~90% reversibility of the HNH unfolding transition. Temperature-dependent measurement of the WT (full-length) Cas9 and HNH CD spectra results in well-defined unfolding transitions, shown by monitoring the spectrum at 218 nm. The steep and abrupt unfolding transition of full-length Cas9 is indicative of a two-state (cooperative) process that is dictated by the thermal properties of its least stable domain. HNH, by contrast, is a very thermostable portion of the Cas9 structure in isolation.

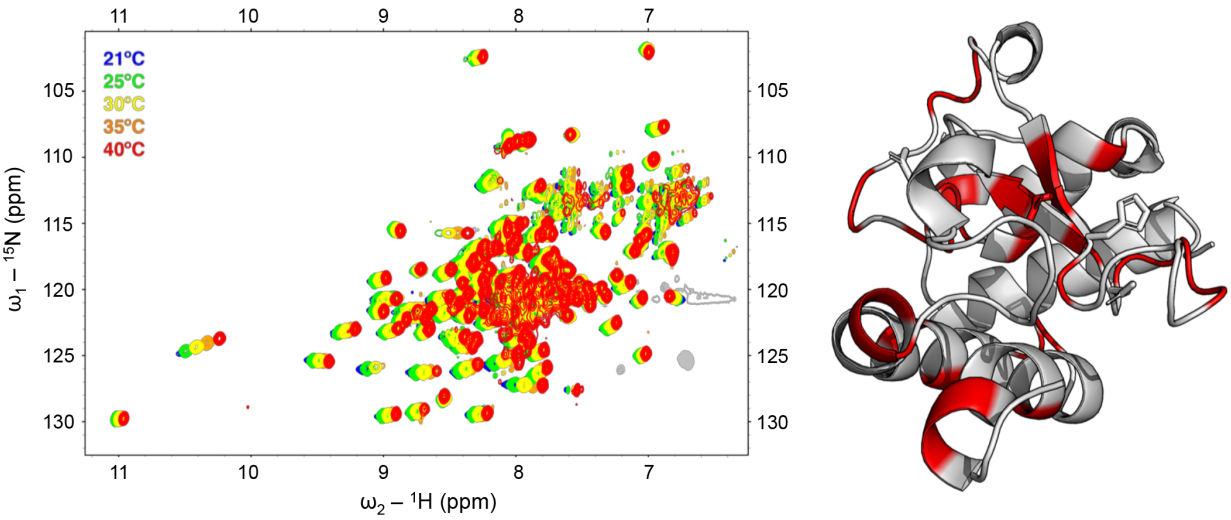

Fig. S2. Effect of temperature on the $^{1}H^{15}N$ HSQC NMR spectrum of Cas9. Identical experiments carried out at 21 (blue), 25 (green), 30 (yellow), 35 (orange), and 40 (red) °C show several resonances that respond strongly to temperature. The corresponding residues are highlighted on the HNH structure in red, and indicate likely sites of conformational flexibility. Several resonances are also only visible at specific temperatures within this gradient, complicating assignment of resonances in a small number of instances.

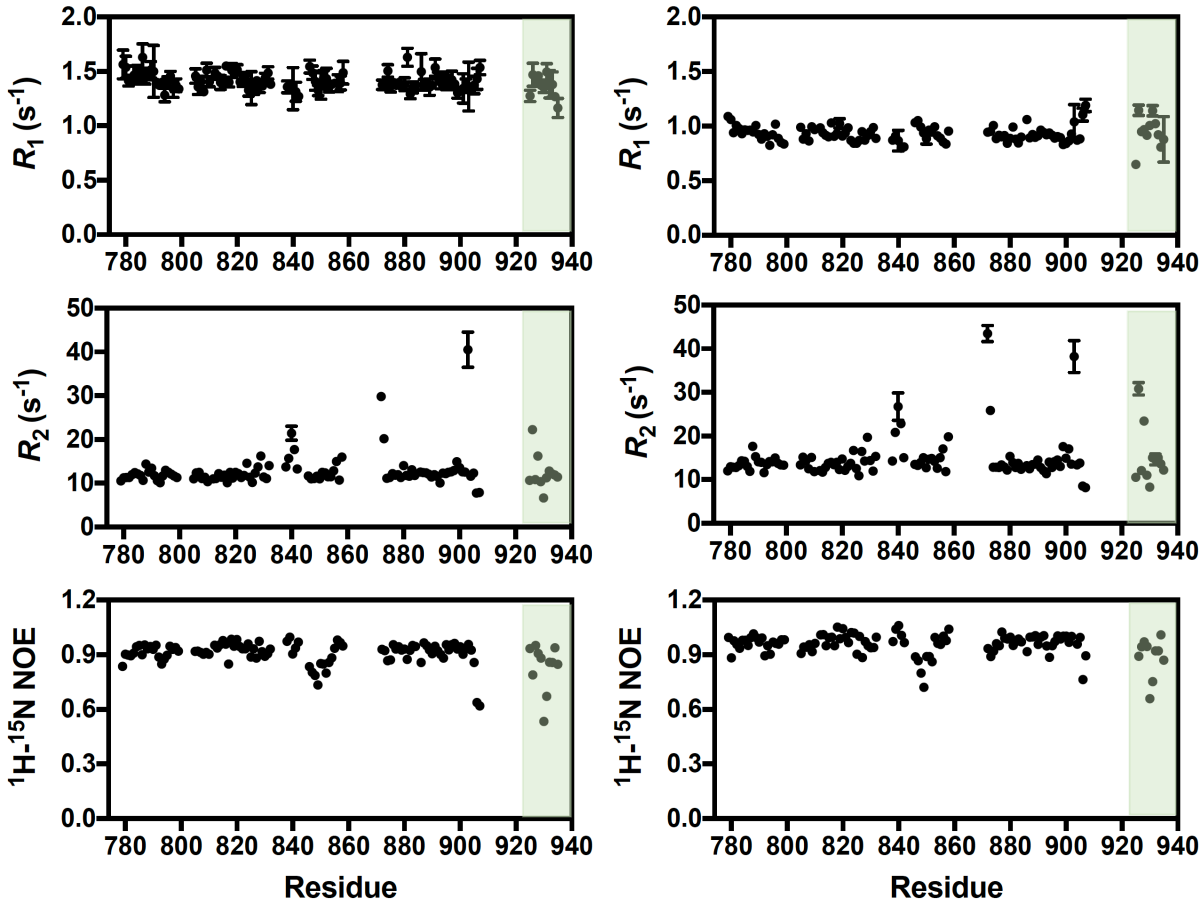

Fig. S3. R_1, R_2, and ^{1}H-^{15}N NOE values for HNH measured by NMR at 600 MHz (left panels) and 850 MHz (right panels). Residues with parameters outside of ±1.5σ of the 10% trimmed mean (red lines) are plotted onto the HNH structures in **Fig. 3** of the main text. Unassigned resonances are included as arbitrarily numbered spin systems in the green highlighted portion of the plot (residue '150' onward). Eight additional sites of flexibility (five *μs–ms*, three *ps–ns*) are observed in unassigned residues based on R_1R_2 products (related to **Fig. 3** of the main text).

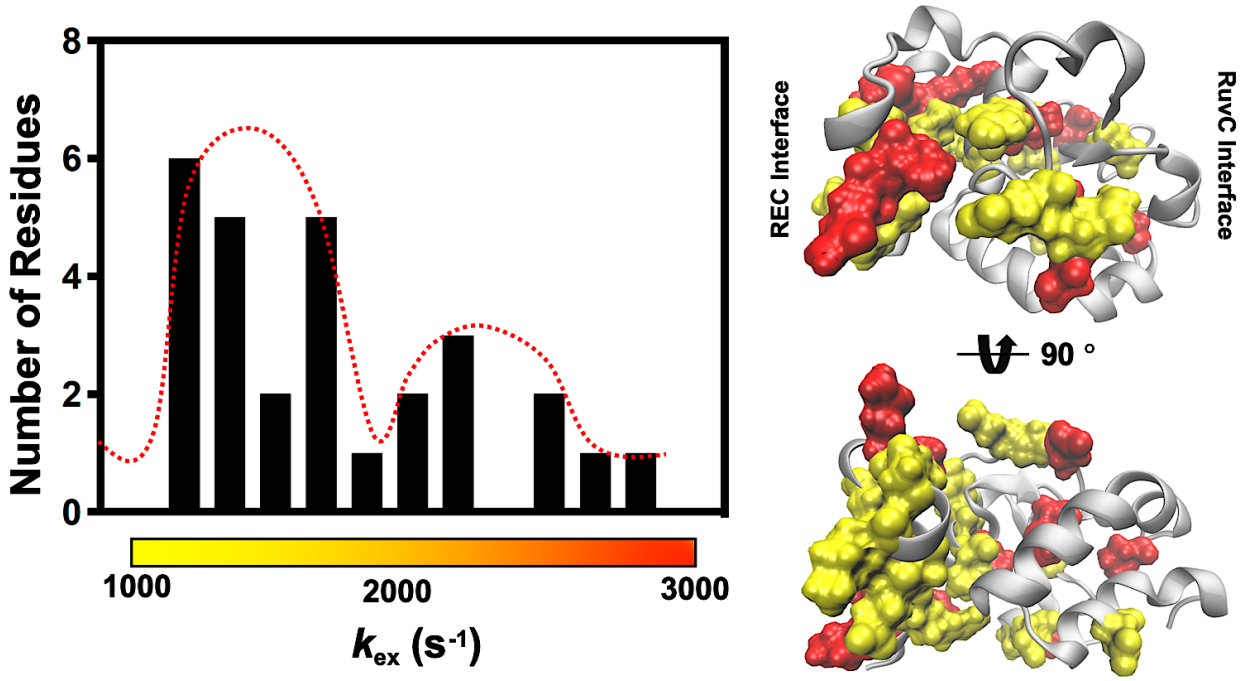

Fig. S4. The apparent bimodal distribution of k_{ex} values determined from CPMG relaxation dispersion is plotted in a residue-specific manner on the HNH structure. Residues with $1000 \leq k_{ex} \leq 2000$ s^{-1} are shown in yellow, while those with $2000 \leq k_{ex} \leq 3000$ s^{-1} are shown in red. Bin sizing for data cutoffs was determined with a protocol described by Scott in *Biometrika*, **1979**, *66*, 605.

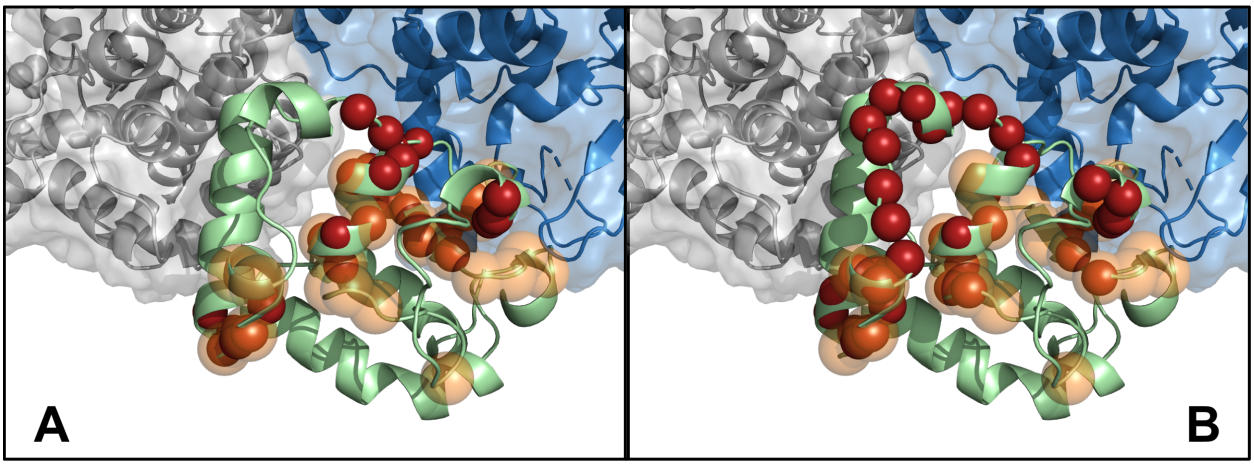

Fig. S5. Optimal pathways across the HNH domain computed from MD simulations of the CRISPR-Cas9 complex upon alanine mutation of key residues. Two systems were considered, namely alanine mutation of **(A)** residues in the WT system belonging to the computational pathway that also show *ms* dynamics through CPMG relaxation dispersion (G792, Q794, K797, E798, L816, Q817, N818, R820, V838, H840, I841, and K855, Figure 4 in the main text), and **(B)** all residues that experimentally show *ms* dynamics (Table S2). The communication pathways were computed between amino acids 789/794 and 841/858, over two simulation replicas of ~400 ns for each system, analogously to the WT CRISPR-Cas9 complex (Figure 4), enabling proper comparison. Solid red spheres represent the new pathway and the transparent orange spheres represent the original pathway (Figure 4). As a result, in both systems, the communication pathway is disrupted, loosening the communication in the core of HNH.

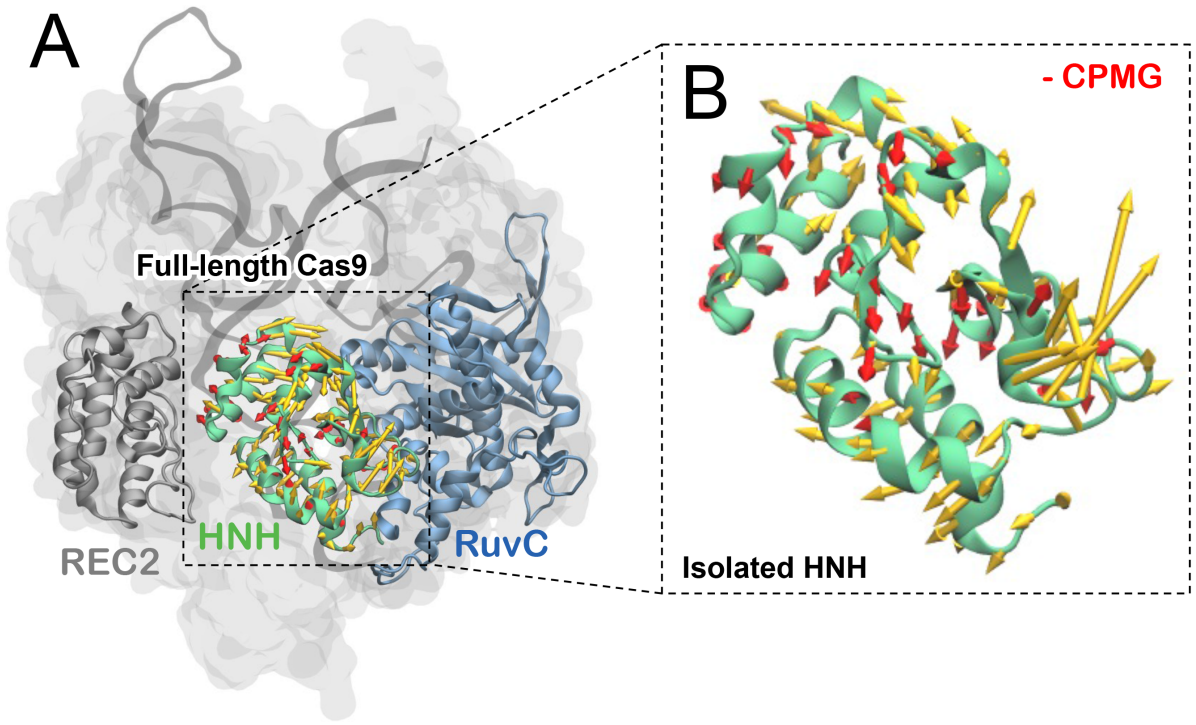

Fig. S6. *"Functional motions"* derived from the first principal component (PC1) of the HNH domain within the full-length Cas9 **(A)** and in the isolated NMR construct **(B)**, shown using arrows of sizes proportional to the amplitude of motions. The motions of the residues that display slow dynamics in solution NMR experiments (i.e., from CPMG relaxation dispersion) are shown using red arrows, whereas the motions of the remaining residues are depicted using yellow arrows. The CRISPR-Cas9 complex is shown in molecular surface, with the HNH (green), RuvC (blue) and REC2 domains as cartoons. Notably, the residues of HNH that experimentally display *ms* dynamics are characterized by short amplitude motions (red arrows) in both the isolated form of HNH and when embedded in the full-length Cas9. In the isolated HNH domain, larger amplitude motions are detected for the terminal residues, due to the lack of the protein stabilization. Apart this difference, a remarkable similarity is observed in the amplitude of PC1 when HNH is embedded in the full-length Cas9 and in its isolated form.

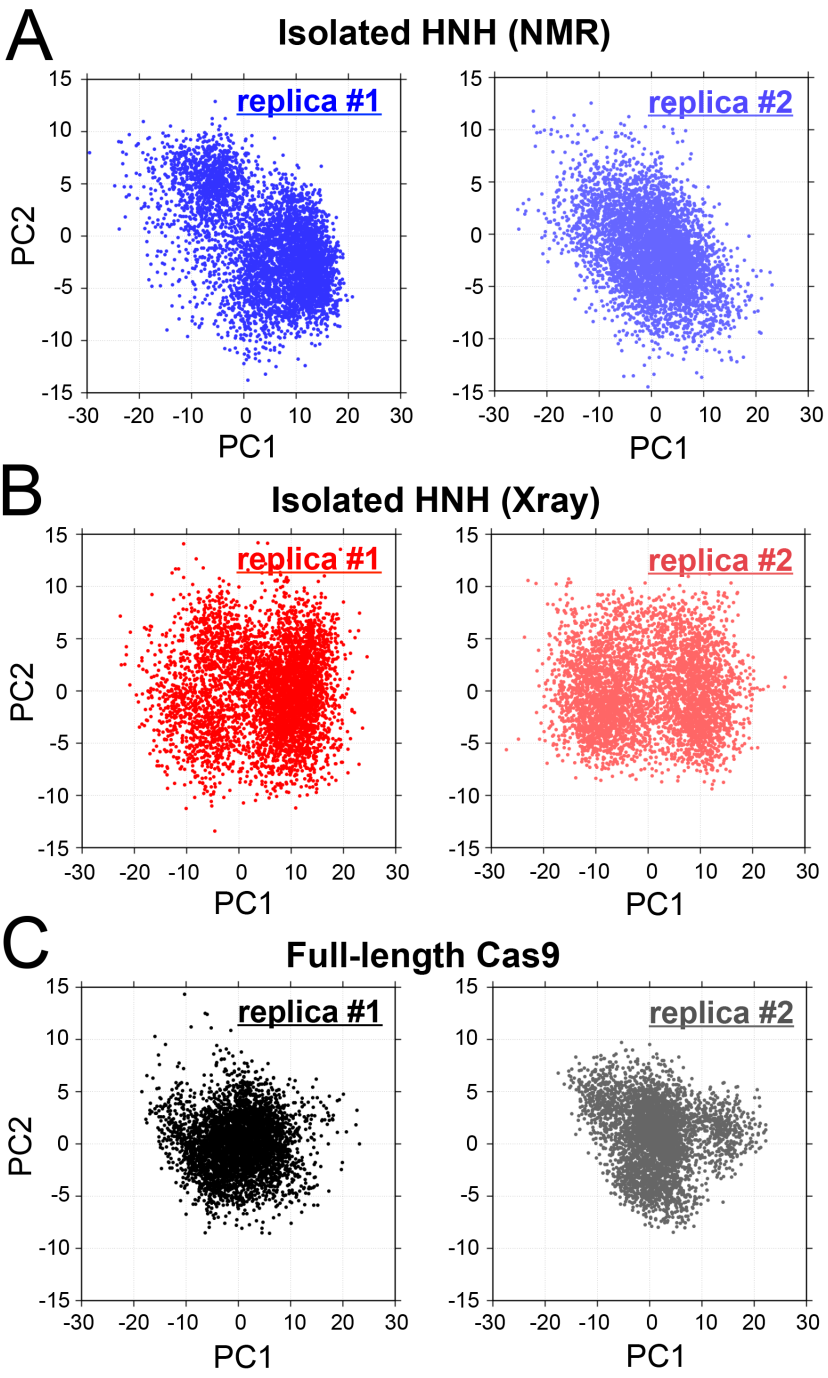

Fig. S7. Principal Component Analysis (PCA) scatter plots. Scatter plots (PC1 vs. PC2) representing the projections of the Cα displacements along the trajectory onto the first principal eigenvector, PC1 (x-axis), vs. the projections onto the second principal eigenvector, PC2 (y-axis), as derived from two GaMD replicas of the isolated HNH domain structure derived from NMR **(A)** and X-ray crystallography **(B)**, and the HNH domain within full-length Cas9 **(C)**. The PCA scatter plots reveal that the isolated HNH explores a wider conformational space, while within the full-length complex, the conformational space explored by HNH results restricted due to the stabilization exerted by the protein/nucleic acid complex.

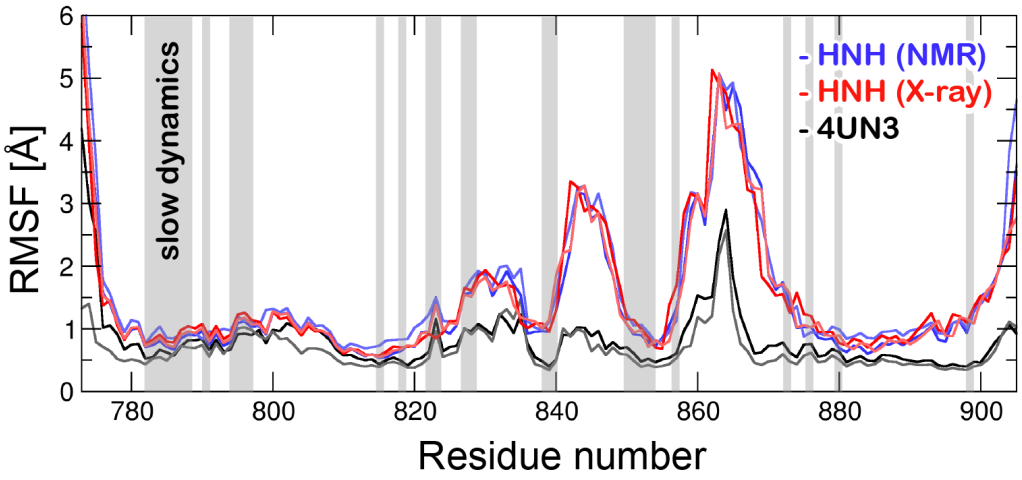

Fig. S8. Comparisons of the Root-Mean-Square Fluctuations (RMSF) of individual Cα atoms of the HNH domain, computed from GaMD simulations of full-length Cas9, versus the isolated HNH domain. RMSF values in full-length Cas9 (black lines) are compared to the NMR structure of HNH derived from the ^{1}H-^{15}N HSQC data (blue lines) and to the X-ray structure of HNH (red lines). Data are reported for two simulation replicas, for each system. Remarkable similarity between the RMSFs of the NMR construct and X-ray structure is observed, arising from the similarity of the two starting structures (RMSD ~0.479 Å). Regions of the sequence characterized by slow dynamics (as experimentally identified via CPMG relaxation dispersion and R_1R_2 (+1.5σ)) are indicated using black vertical bars. In both the isolated HNH and in full-length Cas9, the residues displaying *ms* dynamics are characterized by low fluctuations in the simulations. The magnitude of the fluctuations in certain highly flexible regions is higher in the isolated HNH. This is expected, given the lack of stabilization arising from the absence of the protein/nucleic acid complex. High fluctuations in the isolated HNH are detected at the terminal regions, due to the truncation of the protein, and at the level of residues 840–848. These are flexible loops that in the full-length Cas9 get stabilized by the interaction with the RuvC domain and with the DNA non-target strand, resulting in lower RMSF values. Notably, previous studies employed MD to investigate the stability of the Cas9 protein upon removal of the nucleic acids.[11] The study revealed an increased flexibility of the flexible loops of HNH adjacent to the RuvC domain, also observed here.

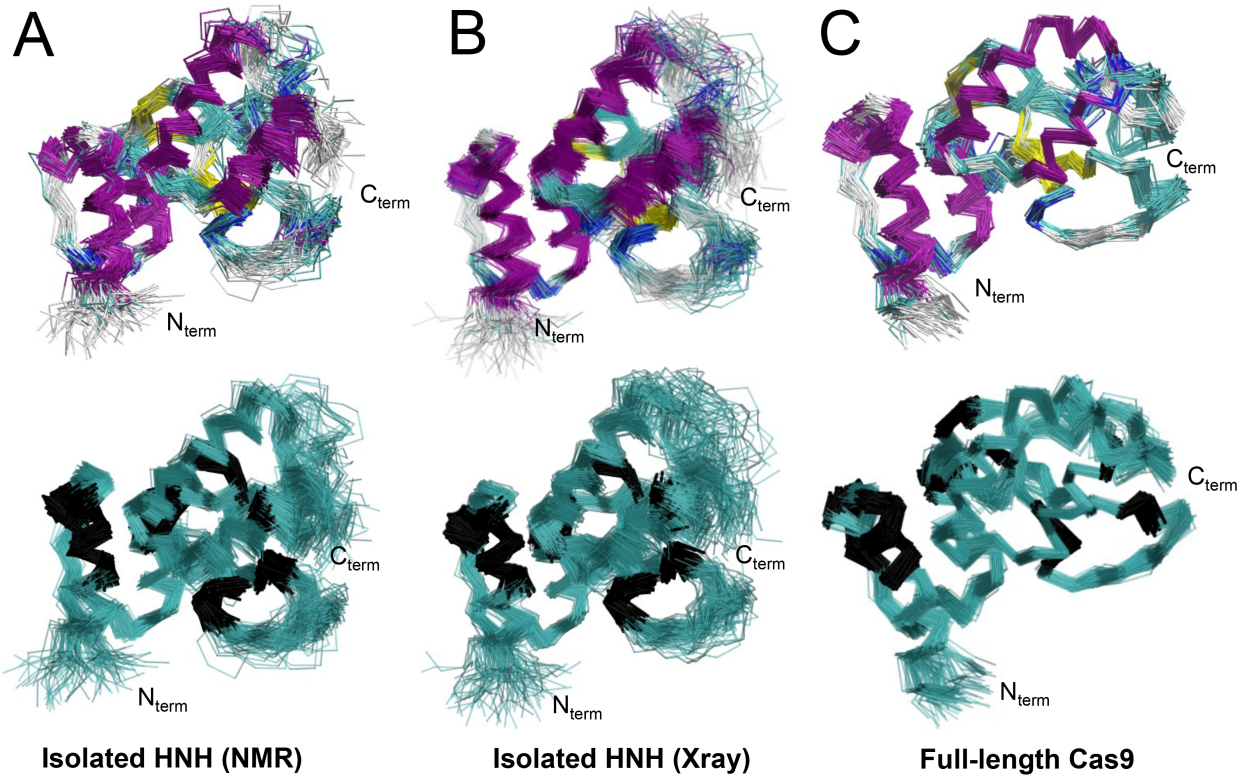

Fig. S9. Representative conformational ensemble accessed during GaMD simulations of the isolated HNH domain as derived from NMR **(A)** and X-ray crystallography **(B)**, and the HNH domain within full-length Cas9 **(C)**. For all systems, 100 structures equally distributed from the GaMD trajectories are superposed. The top panels show the HNH domain color-coded accordingly to the secondary structure. In the lower panels, the regions of the sequence characterized by slow dynamics (as experimentally identified via CPMG relaxation dispersion and R_1R_2 (+1.5σ)) are shown in black. The top and bottom panels report the conformational ensemble for two independent GaMD replicas.

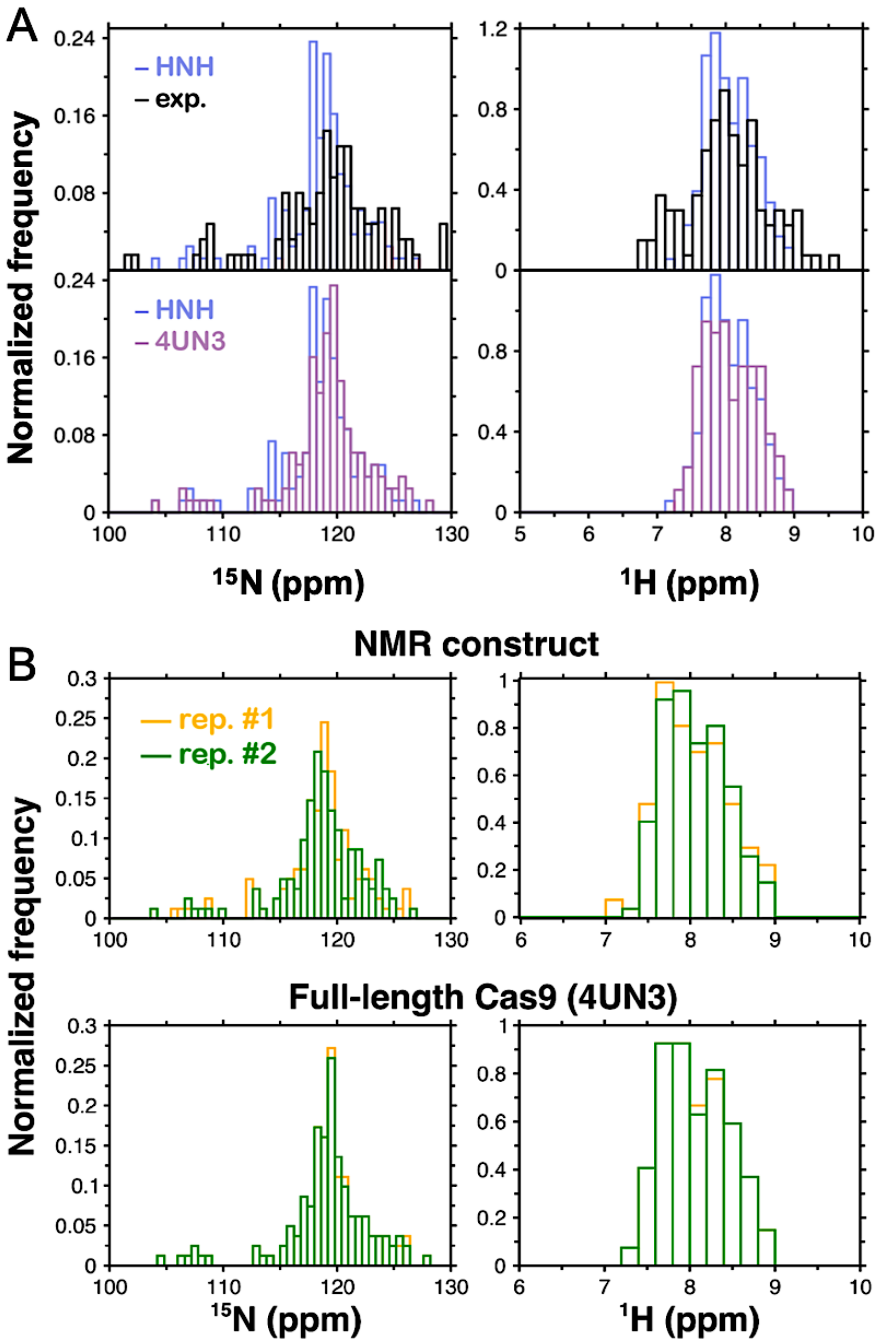

Fig. S10. (A) Experimental and simulated ^{15}N and ^{1}H NMR chemical shifts of the HNH domain, plotted as normalized histograms. The upper panels compare the experimental (black line) and the simulated HNH (light blue) isolated domains. The lower panels compare the simulated HNH domain under two conditions: inside the Cas9 complex (purple) and in isolation (light blue). All simulated spectra were computed as described in Methods utilizing GaMD trajectories. **(B)** Simulated ^{15}N and ^{1}H NMR chemical shifts of HNH plotted as normalized histograms, computed over two GaMD replicas (rep. #1 and # 2, shown in yellow and green, respectively) of isolated HNH (upper panel) and inside full-length Cas9 (lower panel).

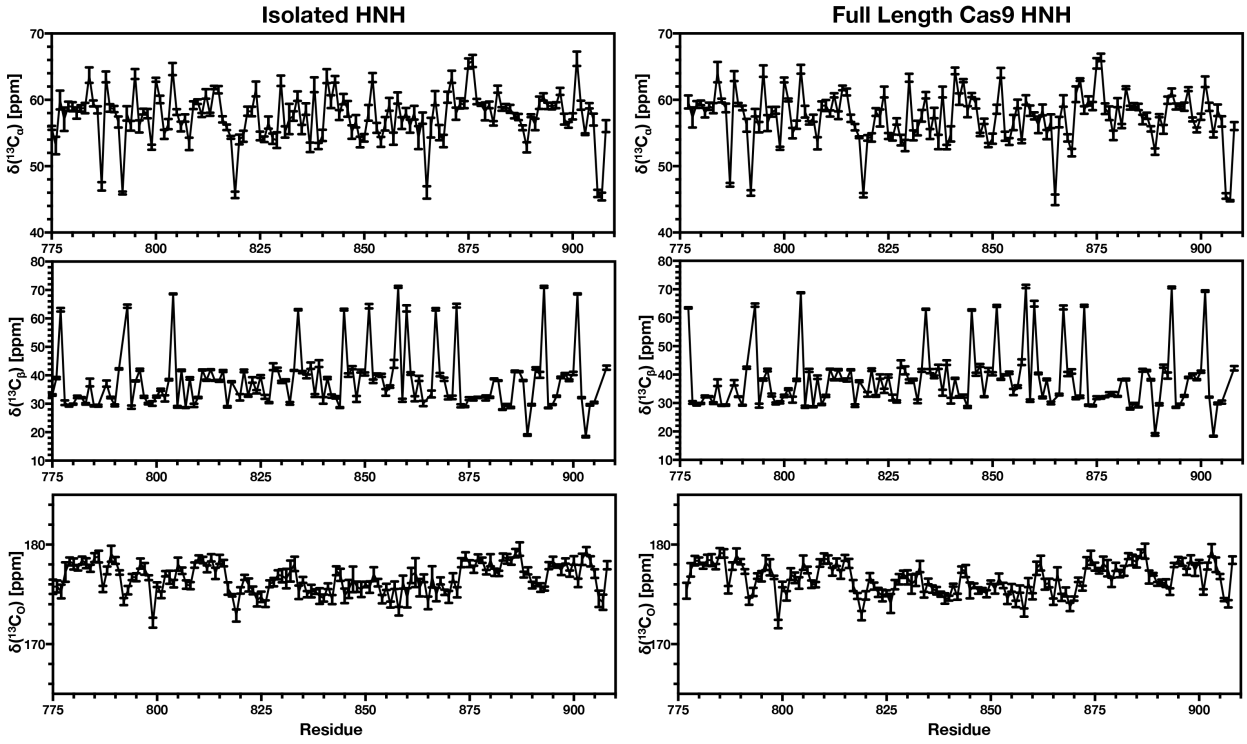

Fig. S11. Simulated chemical shifts of $^{13}C_{\alpha}$, (top graphs) $^{13}C_{\beta}$ (central graphs) and $^{13}C_{o}$ (bottom graphs), plotted for each residue of the HNH domain in its isolated form (left panel) and within the full-length CRISPR-Cas9 system (right panel). Simulated chemical shifts were calculated using SHIFTX2 from an ensemble of structures extracted from MD simulations upon cluster analysis. Specifically, for each system the carbon chemical shifts were computed over the 10 most populated clusters, which account for ~90 % of the overall population (details are reported in the Methods section). Error bars represent the standard deviation of chemical shifts based on the ensemble of structures from the MD trajectories.

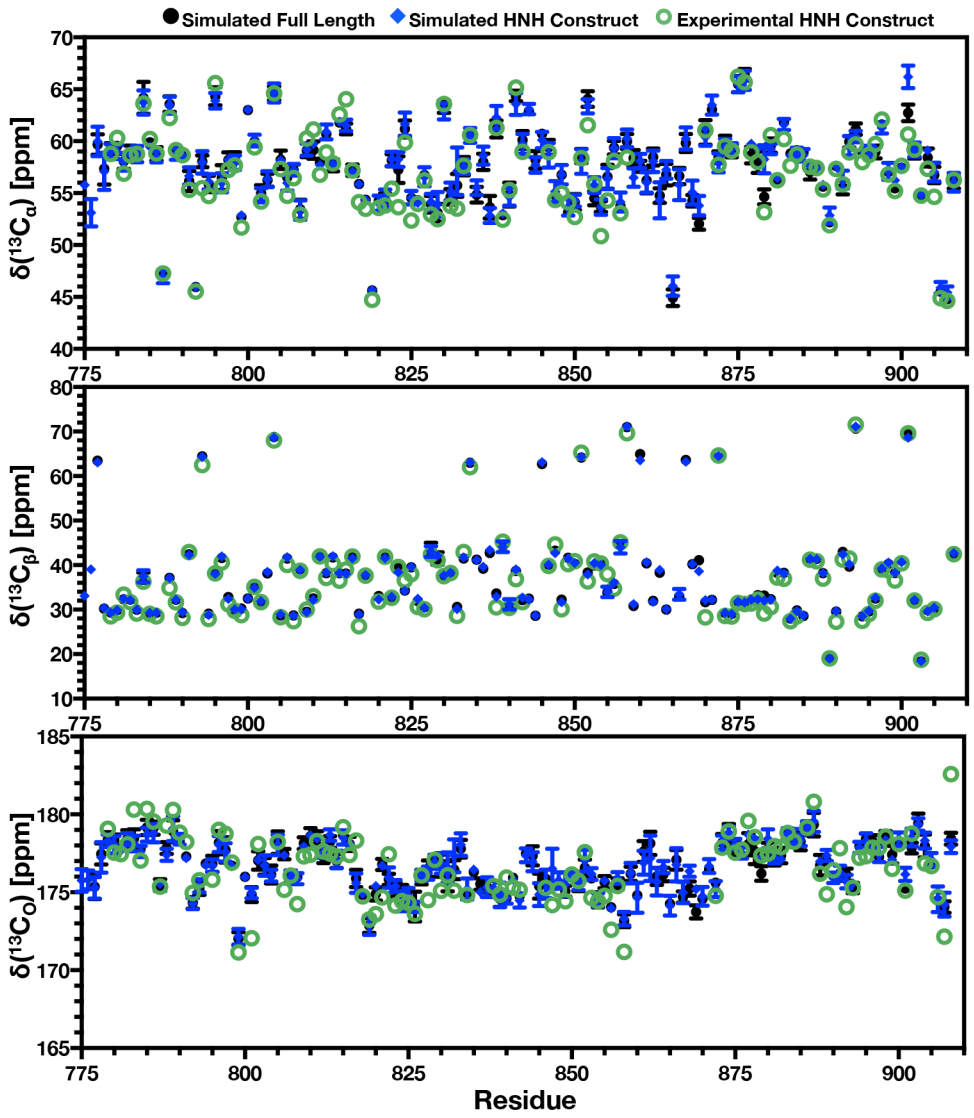

Fig. S12. Experimental and simulated NMR chemical shifts of $^{13}C_{\alpha}$, (top) $^{13}C_{\beta}$ (center) and $^{13}C_{o}$ (bottom) of the HNH domain, plotted for each residue. Experimental values from NMR triple resonance experiments are plotted in green circles. Simulated chemical shifts were calculated using SHIFTX2 from an ensemble of structures extracted from MD simulations of both the full length Cas9 (black) and isolated HNH domain (blue). For both systems, the simulated chemical shifts were computed upon cluster analysis of the MD trajectories, over the 10 most populated clusters, which account for ~90 % of the overall population (details are reported in the Methods section). Error bars represent the standard deviation of chemical shifts based on the ensemble of structures from the MD trajectories.

3. Supplementary Tables

Table S1. R_1R_2 values for the residues in HNH that follow above 1.5σ of the mean (R_1R_2 + 1.5σ) and below 1.5σ of the mean (R_1R_2 – 1.5 σ).

R_1R_2 + 1.5σ	R_1R_2 – 1.5 σ
788	779
789	780
791	795
794	797
818	799
819	805
822	821
823	838
824	875
839	881
840	899
841	906
850	Several Unassigned
852	
855	
858	
873	
876	
880	
899	
905	

Table S2. Conformational exchange parameters determined from CPMG ^{1}H-^{15}N relaxation dispersion analysis of HNH.

Residue	R_2 (s^{-1})	k_{ex} (s^{-1})	$p_a p_b \Delta\omega N$ (s^{-1})
N-H 782	11.3 ± 0.09	2976 ± 679	6.32 ± 1.70
N-H 783	11.9 ± 0.18	2014 ± 203	13.1 ± 4.33
N-H 784	12.4 ± 0.33	1983 ± 718	7.64 ± 2.80
N-H 785	12.2 ± 0.07	980 ± 96	4.28 ± 0.54
N-H 786	11.8 ± 0.15	1542 ± 336	33.4 ± 7.62
N-H 788	14.4 ± 0.30	1322 ± 306	7.25 ± 3.04
N-H 789	12.5 ± 0.16	1106 ± 154	4.61 ± 1.83
N-H 792	11.7 ± 0.10	1463 ±134	36.1 ± 6.80
N-H 794	11.7 ± 0.26	1083 ± 89	1.40 ± 0.32
N-H 795	13.0 ± 0.33	1719 ± 232	2.99 ± 0.68
N-H 796	12.5 ± 0.28	2100 ± 201	1.83 ± 0.51
N-H 797	12.1 ± 0.27	2242 ± 403	54.2 ± 4.59
N-H 798	11.7 ± 0.05	1142 ± 293	1.65 ± 0.40
N-H 815	11.3 ± 0.43	2596 ± 681	25.4 ± 2.49
N-H 816	11.9 ± 0.12	1732 ± 398	28.7 ± 3.91
N-H 817	11.8 ± 0.08	1811 ± 322	4.71 ± 2.40
N-H 818	12.5 ± 0.11	1784 ± 950	3.88 ± 1.59
N-H 820	14.6 ± 0.37	2133 ± 422	2.11 ± 0.78
N-H 827	12.3 ± 0.07	2563 ± 879	6.17 ± 1.42
N-H 828	13.8 ± 0.36	1096 ± 179	9.19 ± 2.86
N-H 829	16.2 ± 0.19	1692 ± 210	12.8 ± 2.37
N-H 840	21.6 ± 0.47	1598 ± 334	31.8 ± 4.20
N-H 841	17.7 ± 0.21	1761 ± 593	18.1 ± 4.54
N-H 851	12.5 ± 0.28	2150 ± 617	5.55 ± 0.39
N-H 853	11.5 ± 0.21	1274 ± 202	11.6 ± 0.97
N-H 855	12.8 ± 0.08	1159 ± 450	50.1 ± 4.21
N-H 873	20.2 ± 0.34	2721 ± 846	8.74 ± 0.68
N-H 900	13.5 ± 0.21	1462 ± 510	23.3 ± 2.24

4. Supplementary References

(1) Kabsch, W. XDS. *Acta Cryst. Sect. D.* **2010**, *66*, 125-132.

(2) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the CCP4 Suite and its Current Developments. *Acta Cryst. Sect. D.* **2011**, *67*, 235-242.

(3) Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M. Structural Basis of PAM-dependent Target DNA Recognition by the Cas9 Endonuclease. *Nature* **2014**, *513*, 569-573.

(4) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: A Comprehensive Python-based System for Macromolecular Structure Solution. *Acta Cryst. Sect. D.* **2010**, 213-221.

(5) Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K.D.; . Reatures and Development of Coot. . *Acta Cryst. Sect. D.* **2010**, *66*, 486-501.

(6) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* **1995**, *6*, 277-293.

(7) Goddard, T. D.; Kneller, D. G. SPARKY 3. **2008**, *University of California, San Francisco*.

(8) Loria, J. P.; Rance, M.; Palmer, A. G., 3rd. A Relaxation-Compensated Carr-Purcell-Meiboom-Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *J. Am. Chem. Soc.* **1999**, *121*, 2331-2332.

(9) Bieri, M.; d'Auvergne, E. J.; Gooley, P. R. relaxGUI: A New Software for Fast and Simple NMR Relaxation Data Analysis and Calculation of ps-ns and us Motion of Proteins. *J. Biomol. NMR* **2011**, *50*, 147-155.

(10) Morin, S.; Linnet, T.; Lescanne, M.; Schanda, P.; Thompson, G. S.; Tollinger, M.; Teilum, K.; Gagne, S.; Marion, D.; Griesinger, C.; Blackledge, M.; d'Auvergne, E. J. relax: The Analysis of Biomolecular Kinetics and Thermodynamics using NMR Relaxation Dispersion Data. *Bioinformatics* **2014**, *30*, 2219-2220.

(11) 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.

(12) 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.

(13) Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; Schwede, T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* **2014**, *42*, W252-W258.

(14) 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.

(15) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E. r.; 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.

(16) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E. r.; 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.

(17) 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. *Journal of Chemical Theory and Computation* **2011**, *7*, 2886-2902.

(18) Aqvist, J. Ion-Water interaction Potentials Derived from Free Energy Perturbation Simulations. *Journal of Physical Chemistry* **1990**, *94*, 8021-8024.

(19) Palermo, G. Structure and Dynamics of the CRISPR-Cas9 Catalytic Complex. *J Chem Inf Model* **2019**, *59*, 2394-2406.

(20) 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.

(21) 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.

(22) 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. *Quarterly Review of Biophysics* **2018**, *51*, e9.

(23) 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.

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

(25) 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.

(26) 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.; Xiao, L.; M., Y. D.; A., K. P. AMBER 2016. *University of California, San Francisco* **2016**.

(27) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919-11929.

(28) Pierce, L. C. T.; Salomon-Ferrer, R.; Oliveira, C. A. F.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theor. Comput.* **2012**, *8*, 2997-3002.

(29) Markwick, P. R. L.; Bouvignies, G.; Blackledge, M. Exploring Multiple Timescale Motions in Protein GB3 using Accelerated Molecular Dynamics and NMR Spectroscopy. *J. Am. Chem. Soc.* **2007**, *129*, 4724-4730.

(30) Mukrasch, M. D.; Markwick, P.; Biernat, J.; von Bergen, M.; Bernado, P.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. Highly Populated Turn Conformations in Natively Unfolded Tau Protein Identified from Residual Dipolar Couplings and Molecular Simulation. *J. Am. Chem. Soc.* **2007**, *129*, 5235-5243.

(31) Salmon, L.; Pierce, L.; Grimm, A.; Ortega, R., J.-L.; Mollica, L.; Jensen, M. R.; van Nuland, N.; Markwick, P.; McCammon, J. A.; Blackledge, M. Multi-timescale Conformational Dynamics of the SH3 Domain of CD2-Associarws Protein using NMR Spectroscopy and Accelerated Molecular Dynamics. *Angew. Chem. Int. Ed.* **2012**, *51*, 6103-6106.

(32) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theor. Comput.* **2015**, *11*, 3584-3595.

(33) Miao, Y.; McCammon, J. A. Mechanism of the G-protein Mimetic Nanobody Binding to a Muscarinic G-protein-coupled Receptor. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 3036-3041.

(34) Chuang, C.-H.; Chiou, S. J.; Cheng, T.-L.; Wang, Y. T. A Molecular Dynamics Simulation Study Decodes the Zika Virus NS5 Methyltransferase Bound to SAH and RNA Analogue. *Sci. Rep.* **2018**, *8*, 6336.

(35) Wang, Y. T.; Chan, Y. H. Understanding the Molecular Basis of Agonist/Antagonist Mechanism of Human Mu Opiod Receptor through Gaussian Accelerated Molecular Dynamics Method. *Sci. Rep.* **2017**, *7*, 7828.

(36) Sibener, L. V.; Fernandes, R. A.; Kolawole, E. M.; Carbone, C. B.; Liu, F.; Mcaffee, D.; Birnbaum, M. E.; Yang, X.; Su, L. F.; Yu, W.; Dong, S.; Gee, M. H.; Jude, K. M.; Davis, M. M.; Groves, J. T.; Goddard, W. A. r.; Heath, J. R.; Evavold, B. D.; Vale, R. D.; Garcia, K. C. Isolation of a Structural Mechanism for Uncoupling T-Cell Receptor Signaling from Peptide-MHC Binding. *Cell* **2018**, *1743*, 672-687.

(37) Palermo, G. Structure and Dynamics of the CRISPR-Cas9 catalytic complex. *J. Chem. Inf. Model.* **2019**.

(38) 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. USA* **2017**, *114*, 7260-7265.

(39) 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.

(40) Rivalta, I.; Sultan, M. M.; Lee, N. S.; Manley, G. A.; Loria, J. P.; Batista, V. S. Allosteric pathways in imidazole glycerol phosphate synthase. *Proc Natl Acad Sci U S A* **2012**, *109*, E1428-1436.

(41) Negre, C. F. A.; Hendrickson, H.; Rhitankar Pal, R.; Rivalta, I.; Ho, J.; Batista, V. S. Eigenvector Centrality Distribution for Characterization of Protein Allosteric Pathways. *P Natl Acad Sci USA* **2018**, *115*, 12201-12208.

(42) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. Dynamical networks in tRNA: protein complexes. *P Natl Acad Sci USA* **2009**, *106*, 6620-6625.

(43) Lisi, G. P.; East, K. W.; Batista, V. S.; Loria, J. P. Altering the allosteric pathway in IGPS suppresses millisecond motions and catalytic activity. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114*, E3414-E3423.

(44) Lange, O. F.; Grubmuller, H. Generalized correlation for biomolecular dynamics. *Proteins-Structure Function and Bioinformatics* **2006**, *62*, 1053-1061.

(45) Dijkstra, E. W. *Numerische Mathematik* **1959**, *1*, 269-271.

(46) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins* **1993**, *17*, 412-425.

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

(48) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **1996**, *14*, 33-38, 27-38.

(49) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR* **2001**, *50*, 43–57.

(50) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Reversible peptide folding in solution by molecular dynamics simulation. *J Mol Biol* **1998**, *280*, 925-932.