

Structural and dynamic insights into the HNH nuclease of divergent Cas9 species

Helen B. Belato^a, Alexandra M. D'Ordine^a, Lukasz Nierzwicki^b, Pablo R. Arantes^b, Gerwald Jogl^a, Giulia Palermo^{b,c,*}, George P. Lisi^{a,*}

^a Department of Molecular Biology, Cell Biology & Biochemistry, Brown University, Providence, RI, USA

^b Department of Bioengineering, University of California Riverside, Riverside, CA, USA

^c Department of Chemistry, University of California Riverside, Riverside, CA, USA

ARTICLE INFO

Edited by Tatyana Polenova

Keywords:

CRISPR-Cas9

HNH

Protein dynamics

NMR

MD simulations

ABSTRACT

CRISPR-Cas9 is a widely used biochemical tool with applications in molecular biology and precision medicine. The RNA-guided Cas9 protein uses its HNH endonuclease domain to cleave the DNA strand complementary to its endogenous guide RNA. In this study, novel constructs of HNH from two divergent organisms, *G. stearothermophilus* (GeoHNH) and *S. pyogenes* (SpHNH) were engineered from their respective full-length Cas9 proteins. Despite low sequence similarity, the X-ray crystal structures of these constructs reveal that the core of HNH surrounding the active site is conserved. Structure prediction of the full-length GeoCas9 protein using Phyre2 and AlphaFold2 also showed that the crystallographic construct of GeoHNH represents the structure of the domain within the full-length GeoCas9 protein. However, significant differences are observed in the solution dynamics of structurally conserved regions of GeoHNH and SpHNH, the latter of which was shown to use such molecular motions to propagate the DNA cleavage signal. Indeed, molecular simulations show that the intradomain signaling pathways, which drive SpHNH function, are non-specific and poorly formed in GeoHNH. Taken together, these outcomes suggest mechanistic differences between mesophilic and thermophilic Cas9 species.

1. Introduction

The CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats – associated protein 9) enzyme machine is an RNA-guided endonuclease central to adaptive immunity in prokaryotes that also has laboratory applications to genome editing and bioengineering (Jinek et al., 2012). Cas9 uses a single-guide RNA (sgRNA) to unwind and cleave a double-stranded DNA target after recognition of its protospacer adjacent motif (PAM), a short sequence of nucleotide bases preceding the cleavage site. The multi-domain Cas9 structure is composed of a recognition (REC) lobe that docks the RNA:DNA hybrid, a PAM interacting (PI) domain, and the HNH and RuvC nucleases that cleave the sgRNA-complementary and non-complementary DNA strands, respectively (Anders et al., 2014; Nishimasu et al., 2014).

The CRISPR-Cas9 system relies on RNA-guided recognition and cleavage of invasive viral and plasmid DNA derived from the adaptive immune systems found of bacteria. Despite the abundance and diversity of these systems (Zhang, 2019), which differ in size, architecture, and

cleavage activities, the vast majority of applications have employed the Cas9 homolog first discovered from *Streptococcus pyogenes* (SpCas9). While SpCas9 provides a robust genome editing tool, it originates from a mesophilic host, making it unsuitable for applications requiring cleavage at higher temperatures or extended protein stability (Harrington et al., 2017).

The DNA recognition and nuclease sites of Cas9 are spatially separated yet functionally coupled, indicative of an allosteric crosstalk throughout the enzyme. The potential for allostery within Cas9 was initially probed by the insertion of exogenous PDZ domains into the native Cas9 structure, highlighting regions of the enzyme as allosteric “hotspots.” (Oakes et al., 2016) Further biochemical studies indicated that allostery synchronizes DNA recognition, binding, and concerted double stranded DNA cleavage (Sternberg et al., 2015). Early computational investigations revealed that allosteric crosstalk between the HNH and RuvC nucleases was essential for concerted DNA cleavage (Palermo et al., 2017b). While both X-ray and computational methods have shown an open-to-closed conformational transition in apo and

* Corresponding authors at: Department of Bioengineering, University of California Riverside, Riverside, CA, USA (G. Palermo). Department of Molecular Biology, Cell Biology & Biochemistry, Brown University, Providence, RI, USA (G.P. Lisi).

E-mail addresses: gpalermo@engr.ucr.edu (G. Palermo), george_lisi@brown.edu (G.P. Lisi).

<https://doi.org/10.1016/j.jsb.2021.107814>

Received 15 August 2021; Received in revised form 15 October 2021; Accepted 29 November 2021

Available online 3 December 2021

1047-8477/© 2021 Elsevier Inc. All rights reserved.

RNA:DNA bound *SpCas9*, it is widely hypothesized that Cas9, like other allosteric proteins, populates a variety of states between those obtained through structural biology (Dagdas et al., 2017; Huai et al., 2017; Jiang and Doudna, 2015; Jinek et al., 2014; Nishimasu et al., 2014; Singh et al., 2018).

Recently, a Cas9 homolog from the thermophilic bacterium *Geobacillus stearothermophilus* (*GeoCas9*) was discovered that is stable and functional in human plasma (Harrington et al., 2017) and enzymatically active at a much wider temperature range than that of *SpCas9* (25–75°C versus 35–45°C, respectively), offering a promising alternative to ongoing bioengineering efforts aimed at optimizing the utility of this molecular machine (Harrington et al., 2017). Expansion of CRISPR-Cas9 technology from mesophiles to thermophiles may be an effective alternative for gene editing in mammalian cells, but an understanding of the molecular motions guiding *GeoCas9* and how this influences the Cas9 mechanism does not exist. Biochemical (Sternberg et al., 2015) and single-molecule experiments (Chen et al., 2017; Dagdas et al., 2017) have suggested that *SpCas9* could be governed by an intricate allosteric mechanism, transferring the information of DNA binding from the REC lobe to the distally located catalytic sites. To establish the molecular details of this crosstalk, extensive molecular dynamics (MD) simulations have been carried out (East et al., 2020a; Nierzwicki et al., 2021; Palermo et al., 2018; Palermo et al., 2017b). As a result, a dynamically driven allosteric signal was shown to span the HNH domain of *SpCas9*, thanks to a “striking plasticity” of the domain that regulates its conformational activation toward DNA cleavage (Palermo et al., 2016; Sternberg et al., 2015). Using a combination of NMR and MD simulations, we recently characterized a millisecond (ms) timescale dynamic pathway within the *SpCas9* HNH domain (East et al., 2020a), providing a specific route for interpreting biochemical findings that tuned the allosteric signaling within *SpCas9* with single point mutations (Chen et al., 2017; Schmid-Burgk et al., 2020; Slaymaker et al., 2016). More generally, CRISPR-Cas proteins have been shown to be “allosteric machines”, where the communication between DNA binding and cleavage is critical for their function (Saha et al., 2020). Here, we investigated the *GeoCas9* HNH domain (*GeoHNH*) with solution NMR and simulations to assess its structure and quantify the timescales of its molecular motions for comparison to the canonical mesophilic *SpHNH*.

2. Results

2.1. Assessment of the atomic structure of *GeoHNH*

The arrangement and molecular weights of the *Sp*- and *GeoCas9* domains are conserved (Fig. 1A), with the exception of REC, which is truncated by 281 amino acids in *GeoCas9*. Due to the lack of experimental structures of full-length *GeoCas9*, Doudna and coworkers generated homology model of the protein (Harrington et al., 2017), which we reproduced here using Phyre2 (Kelley et al., 2015) according to the original paper (Fig. 1B). We then engineered a construct of the *GeoHNH* domain and solved the X-ray crystal structure at 2.0 Å. An overlay of our newly crystallized *GeoHNH* and HNH from the *GeoCas9* model shows high similarity (RMSD 1.32 Å, Figure S1). It is important to note that the Phyre2 homology model of *GeoCas9* used the template of PDB 6JDV (Sun et al., 2019), which is the crystal structure of *Neisseria meningitidis* Cas9 (*Nme1Cas9*) in complex with a sgRNA and target DNA. This *Nme1Cas9* structure was solved in various DNA-bound states, which mainly differ in HNH orientation and could bias the homology modeling of *GeoCas9*. Hence, we performed an alternative modeling harnessing AlphaFold2 (Jumper et al., 2021), a recent deep learning approach that incorporates physical and biological knowledge about protein structure, and leverages multi-sequence alignments to predict three-dimensional protein structures (details are reported in the Methods section, alongside the accuracy of the model, Figure S2). The *GeoHNH* domain of the AlphaFold2 model of *GeoCas9* remarkably superposes to our newly crystallized construct of *GeoHNH* (RMSD 0.4 Å, Fig. 2A), and is also highly similar to the *GeoHNH* domain predicted by the Phyre2 model of *GeoCas9* (RMSD 1.4 Å). This suggests that our construct of *GeoHNH* represents the structure of the HNH domain within the full-length *GeoCas9* protein very well.

A comparison of the full-length *GeoCas9* proteins obtained through Phyre2 and AlphaFold2 also shows that the two models mainly differ in the orientation of the HNH domain with respect to the recognition lobe (Fig. 2B). This is consistent with structural (Jiang and Doudna, 2017) and biophysical (Dagdas et al., 2017; Palermo et al., 2018; Palermo et al., 2017a) studies of *SpCas9*, revealing that the HNH domain can change conformation as a rigid body (i.e., by preserving its overall fold),

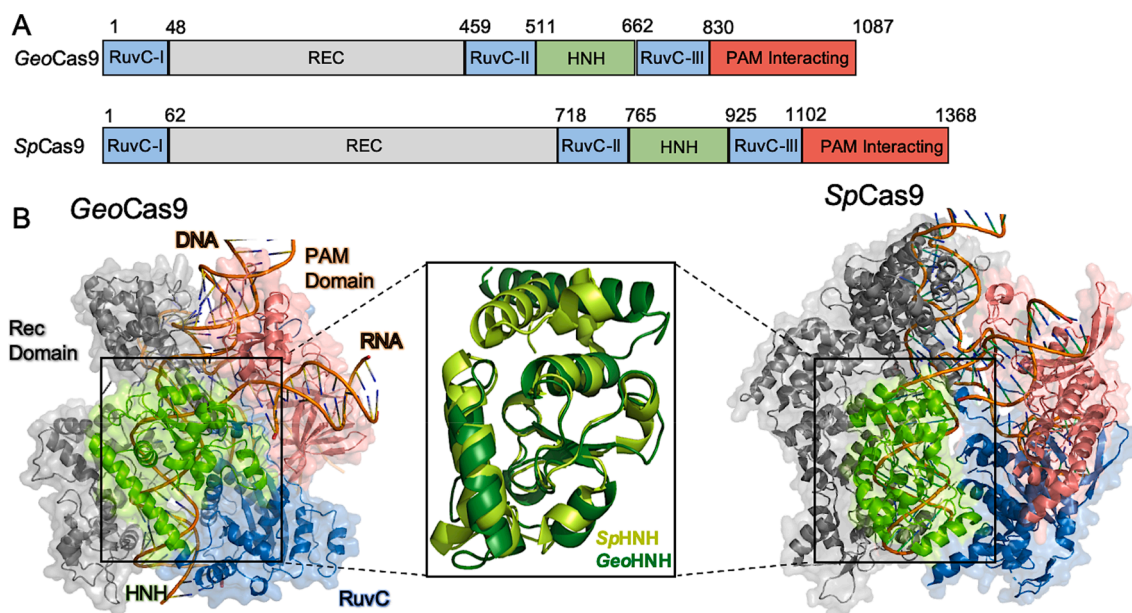


Fig. 1. (A) Arrangement of the *GeoCas9* and *SpCas9* domains and primary sequence. Cartoons are scaled to approximate the molecular weight of each domain. (B) Architecture of the *GeoCas9* model and the *SpCas9* crystal structure (PDB ID 4UN3). A full-length *GeoCas9* homology model was generated with Phyre2 (Kelley et al., 2015). RNA and DNA from PDB 6JDV were docked and aligned into the model. The similarity of the HNH domains of these proteins is shown in the X-ray crystal structures of *SpHNH* (light green, PDB 6O56) and *GeoHNH* (dark green, PDB ID 7MPZ) solved at 1.9 Å and 2.0 Å, respectively.

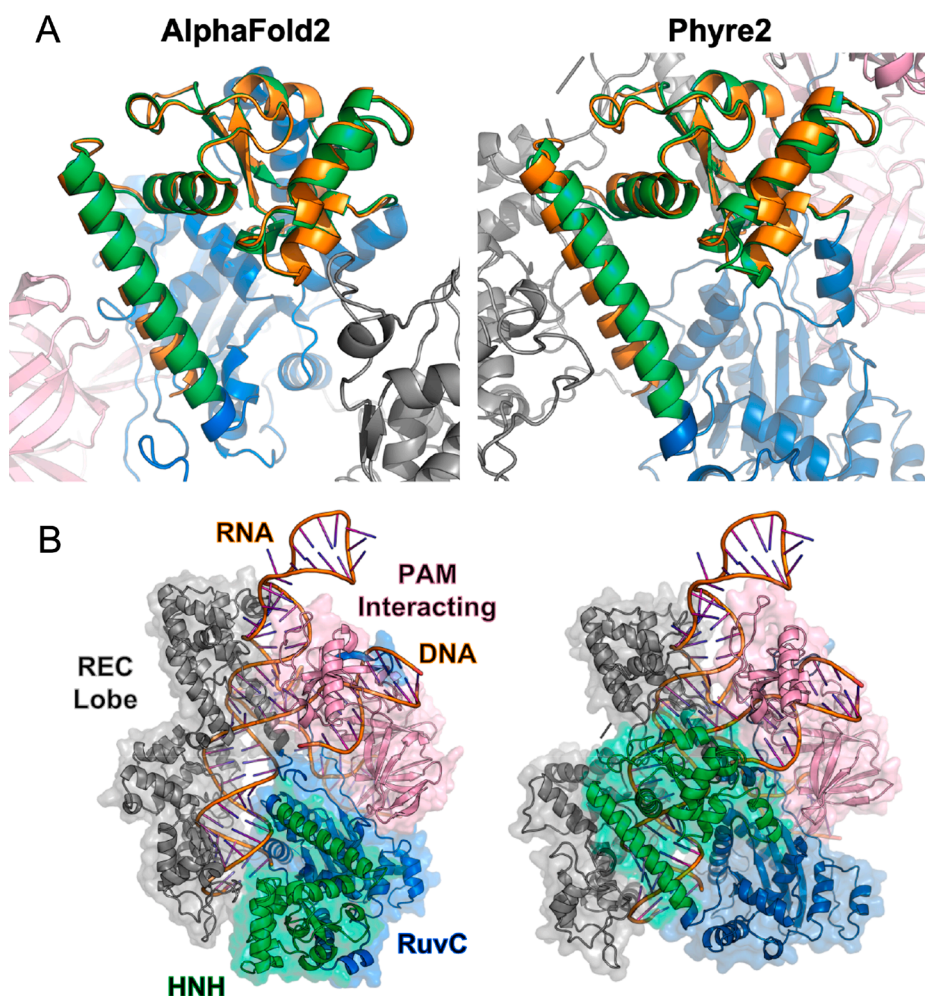


Fig. 2. AlphaFold2 model of *GeoCas9* and comparison with Phyre2. **(A)** Structural alignment of our newly crystallized *GeoHNH* domain (orange) with the model of *GeoCas9* obtained using AlphaFold2 (left, HNH in green) and Phyre2 (right). **(B)** Architecture of *GeoCas9* modeled using AlphaFold2 (left) and Phyre2 (right). RNA and DNA from PDB 6JDV were docked and aligned into the models. The two models mainly differ in the orientation of the HNH domain with respect to the recognition (REC) lobe.

from inactive to active states. One interesting feature of the *GeoHNH* structure is an extended N-terminal α -helix that affects stability and NMR spectral quality in solution. To mitigate this issue, we engineered a second *GeoHNH* construct excluding this N-terminal helix. This construct, hereafter referred as *GeoHNH- Δ N*, is more amenable to NMR studies because it is soluble at concentrations ≥ 1 mM and produces better resolved ^1H - ^{15}N NMR spectra where the active site residues are visible (Figure S3). Backbone NMR resonance assignments (N-H, C_α , C_β , C_O) and NMR-derived secondary structure of *GeoHNH- Δ N* determined from C_α and C_β chemical shifts are also consistent with the core of our X-ray structure of *GeoHNH* and the HNH domain of the *GeoCas9* models (Figure S4). Circular dichroism (CD) spectra of both *GeoHNH* and *GeoHNH- Δ N* are consistent with a predominantly α -helical protein. (Figure S5A). Inspection of the *GeoHNH* structure relative to that of *SphNH* solved at 1.9 Å shows that despite only 33% sequence similarity, the core of the structure surrounding the active site is conserved (Fig. 1B). This suggests that the HNH architecture may be a common thread among Cas9 proteins of different sizes that recognize different PAM sequences.

2.2. Experimental conformational dynamics of the HNH nuclease

To assess the dynamic profile of *GeoHNH- Δ N* relative to that of *SphNH*, we used solution NMR. Given the similar structure of both HNH domains, we wondered whether this thermophilic HNH would harbor a similar pathway of information transfer driven by intrinsic dynamics. Longitudinal (R_1) and transverse (R_2) relaxation rates measured by NMR highlight significant differences in conformational flexibility between

the two domains at 25 °C (Fig. 3) (East et al., 2020b). NMR relaxation rates are plotted as the R_1R_2 product, where the relationship of R_1R_2 values relative to the mean of the data can be used to infer fast (depressed R_1R_2) or slow (elevated R_1R_2) timescale dynamics.

Residues with R_1R_2 values 1.5 σ below the 10% trimmed mean of all rates are suggested to undergo equilibrium fluctuations in bond vectors on the *ps*-*ns* timescale that contribute to entropically driven signaling regulating the population of (in)active state (Frederick et al., 2007; Tzeng and Kalodimos, 2009). In our previous study of *SphNH*, we noted that fast timescale dynamics were not a significant factor (East et al., 2020a). In contrast, *GeoHNH- Δ N* displays several residues with statistically depressed R_1R_2 values (Figure S6). These relaxation rates, along with heteronuclear ^1H - ^{15}N NOEs (Figure S7) were then used to calculate order parameters (S^2) for both HNH proteins. Values of S^2 are notably lower for *GeoHNH- Δ N* than for *SphNH* ($\langle S^2 \text{ SphNH} \rangle = 0.85$; $\langle S^2 \text{ GeoHNH-}\Delta\text{C} \rangle = 0.72$), confirming a greater influence of fast timescale dynamics in *GeoHNH- Δ N* at 25 °C (Fig. 3B). Computed order parameters of *SphNH* agree with those from experiment, where regions of variance occur primarily in residues that have not been assigned in NMR spectra (Figure S8). Agreement between computed and experimental order parameters is quite good for certain regions of *GeoHNH- Δ N* (i.e. residues 600–630) and modest in others, due in part to missing NMR resonance assignments and perhaps also to the difficulty of approximating the complex solution dynamic profile of *GeoHNH- Δ N*, as discussed below. This shift in dynamic profile between *SphNH* from a mesophilic Cas9 and *GeoHNH* from a thermophilic Cas9 is also depicted structurally, where areas of dark purple/red correspond to regions of the protein with the lowest order parameters (Fig. 3C).

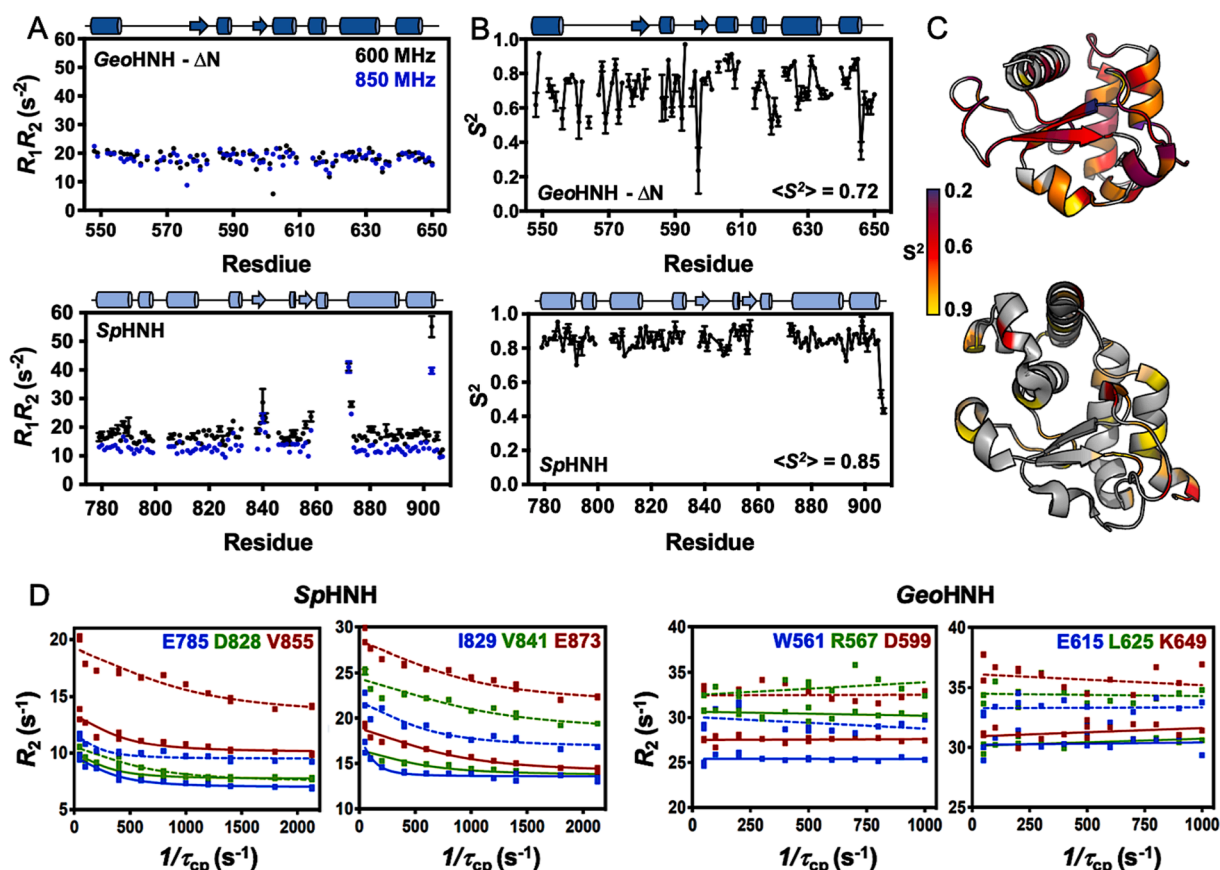


Fig. 3. (A) R_1R_2 values determined from NMR spin relaxation experiments on GeoHNNH- Δ N (top) and SpHNNH (bottom). R_1R_2 parameters were measured at 600 (black) and 850 (blue) MHz. (B) Order parameters (S^2) for GeoHNNH- Δ N (top) and SpHNNH (bottom) determined from Model-free analysis of T_1 , T_2 , and ^1H - ^{15}N NOE measurements. Average S^2 for the entire domain is also indicated. Cartoons of the GeoHNNH- Δ N (dark blue) and SpHNNH (light blue) secondary structures are shown above the plots. (C) Order parameters are mapped onto the GeoHNNH (top) and SpHNNH (bottom) structures, where the heat map legend reports the magnitude of S^2 . (D) Representative CPMG relaxation dispersion profiles for residues in GeoHNNH- Δ N and SpHNNH collected at 25 °C at 600 (solid line) and 850 MHz (dashed line). Flat profiles for GeoHNNH- Δ N indicate the absence of detectable ms motions at these sites.

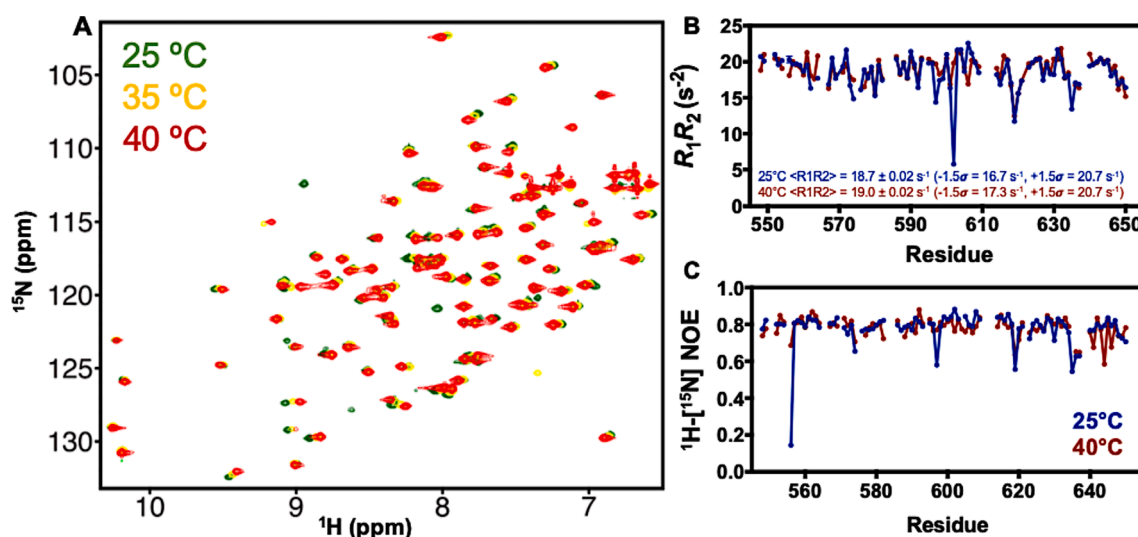


Fig. 4. Temperature dependence of GeoHNNH- Δ N NMR spectra and dynamics. (A) ^1H - ^{15}N HSQC of GeoHNNH- Δ N collected at 25 °C (green), 35 °C (yellow), and 40 °C (red), showing temperature-dependent resonance shifts. (B) R_1R_2 relaxation parameters collected at 25 °C (blue) and 40 °C (maroon), showing very similar overall profiles. Average values of R_1R_2 and $\pm 1.5\sigma$ of the 10% trimmed mean (statistical significance cutoffs) are indicated. (C) ^1H - ^{15}N heteronuclear NOEs collected at 25 °C (blue) and 40 °C (maroon) also appear to be very similar. Small differences in temperature-dependent profiles of (B) and (C) are confined to loops within the protein core and a solvent exposed alpha helix.

R_1R_2 values 1.5σ above the 10% trimmed mean of all rates suggestive of conformational exchange on the μ s-ms timescale (due to the influence of R_{ex}) are more classically associated with allosteric signaling (Kern and Zwietering, 2003) and larger conformational changes (Whittier et al., 2013; Wolf-Watz et al., 2004). Although twenty residues throughout SpHNH have significantly elevated R_1R_2 values, GeoHNH- Δ N lacks residues with relaxation parameters suggestive of μ s-ms motion (Fig. 3A). Consistent with this observation, residues with μ s-ms dynamics detected by Carr-Purcell-Meiboom-Gill (CPMG) (Loria et al., 1999) relaxation dispersion span the entirety of SpHNH, while we observe flat CPMG relaxation dispersion profiles for nearly every residue in GeoHNH- Δ N (Fig. 3D). These data show that the mesophilic SpHNH undergoes extensive μ s-ms motion while its thermophilic counterpart, GeoHNH- Δ N, appears flexible on a faster timescale.

The dynamics of GeoHNH- Δ N were measured at 25 °C to be consistent with published studies on SpCas9 (East et al., 2020a), although GeoCas9 enzymatic activity is optimal at higher temperatures. ^1H - ^{15}N HSQC NMR spectra of GeoHNH- Δ N at 25, 35, and 40 °C show very little evidence of line broadening or altered exchange regimes over the range of higher temperatures (Fig. 4A). To further investigate the dynamics of GeoHNH- Δ N at a temperature closer to its native environment, we measured the R_1 , R_2 , and ^1H - ^{15}N NOE of the system at 40 °C (Fig. 4). Both R_1R_2 values and ^1H - ^{15}N heteronuclear NOEs collected at 40 °C are very similar to those collected at 25 °C (Fig. 4B, 4C, S9). Thus, we find no significant change in dynamic profile in GeoHNH- Δ N at an elevated temperature.

We also tested another aspect of GeoHNH structure, the influence of its N-terminal helix, on the resulting dynamics. We transferred as many NMR assignments from GeoHNH- Δ N to GeoHNH as possible, but due to the poor spectral quality of GeoHNH, the resulting relaxation parameters have larger errors. Although the ^1H - ^{15}N HSQC NMR spectrum of GeoHNH (i.e. containing the N-terminal helix) shows a high degree of peak broadening (Figure S3), we found that the helix does not drastically alter the R_1 , R_2 , ^1H - ^{15}N NOE profiles of GeoHNH (Figure S10). Indeed, R_1 and R_2 values of GeoHNH- Δ N and GeoHNH are similar, though GeoHNH- Δ N has a higher average R_1 . Dual-field CPMG relaxation dispersion profiles overwhelmingly remain flat in GeoHNH (Figure S11). However, R_2^0 values are notably elevated in many cases, implying a μ s-timescale process outside the CPMG regime. Although the

CPMG data is of modest quality, there are three unassigned residues in GeoHNH (presumably part of the N-terminal helix) with CPMG relaxation dispersion curves, suggesting perhaps that the helix itself undergoes ms-timescale motion. Overall, the evidence of a shift in GeoHNH- Δ N dynamic profile due to the addition of its N-terminal helix is modest and qualitative, at best. Though beyond the scope of this work, future studies focused on stabilizing GeoHNH and its N-terminal helix in solution may enhance the spectral quality to enable full resonance assignment and further quantitation of μ s-timescale dynamics.

Sequence alignment analysis of HNH domains from multiple Cas9 species (Fig. 5A) highlight numerous conserved residues that are oriented toward the core of HNH surrounding the catalytic histidine. We find that even though these residues are structurally and sequentially conserved within SpHNH and GeoHNH, their dynamic profiles are quite different (Fig. 3). This finding suggests that despite the similarity of the HNH domains of SpCas9 and GeoCas9, the properties of their full-length proteins exert some measure of influence over the isolated domains. This is perhaps unsurprising for an allosteric system, and is illustrated in temperature-dependent circular dichroism (CD) experiments (Figure S5B) showing that the T_m for SpHNH and GeoHNH are very similar, while those of the full-length proteins are quite different. Here, the stability of Cas9 is controlled by its least stable domain, which does not appear to be HNH in either case, since SpHNH and GeoHNH have similar stabilities when isolated that are greater than their respective full-length Cas9 proteins. These data indicate that despite maintaining a conserved thermodynamic signature in isolation, HNH domains are part of a highly interconnected and cooperative system within full-length Cas9.

3. Molecular dynamics simulations of the HNH nuclease

With the aims of revealing differences in the dynamic behavior of GeoHNH and SpHNH and examining temperature effects on allosteric signal transmission in GeoCas9, we performed MD simulations. We carried out two ~ 3 μ s-long simulations of GeoHNH at temperatures of 25 °C and 40 °C (GeoHNH is $\sim 98\%$ folded at 40 °C). We then performed principal component analysis (PCA) and cluster analysis to examine the conformational flexibility of GeoHNH. MD simulations of GeoHNH at 25 °C revealed a remarkable plasticity of the N-terminal helix, which

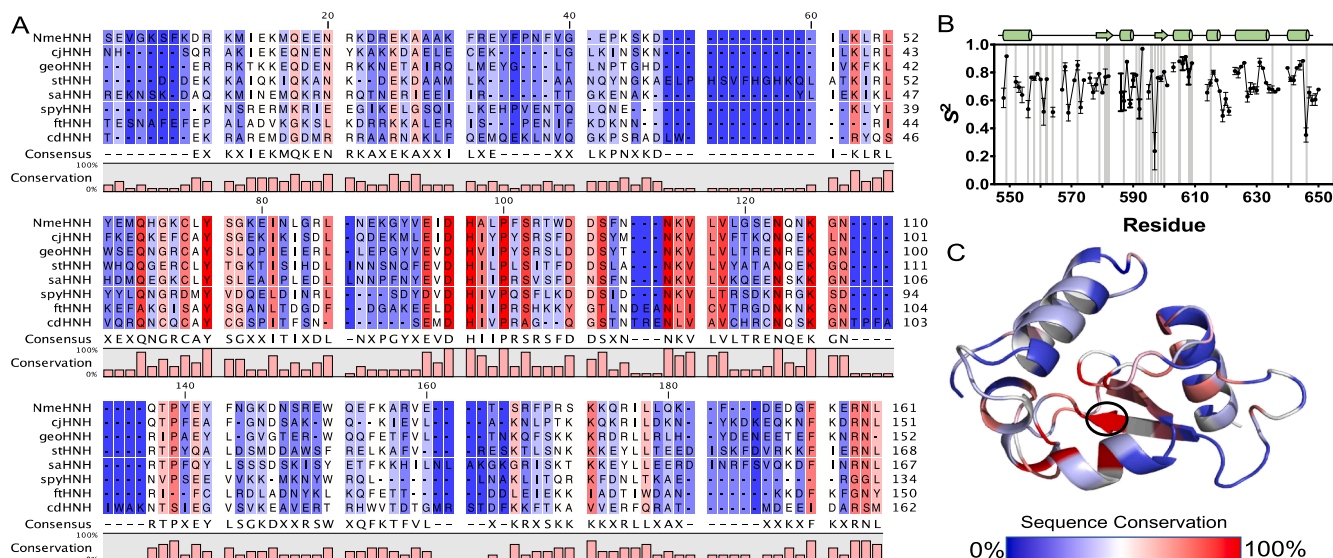


Fig. 5. (A) Sequence alignment of HNH domains from divergent Cas9 species; NmeHNH (*N. meningitidis*), CjHNH (*C. jejuni*), GeoHNH (*G. stearothermophilus*), StHNH (*S. thermophilus*), SaHNH (*S. aureus*), SpvHNH (*S. pyogenes*), FthHNH (*F. novicida*), and CdHNH (*C. diphtheriae*). The level of sequence conservation is indicated by the bars below each row and by color (blue = low conservation; red = high). (B) Plot of NMR order parameters (S^2) for GeoHNH highlighting residues that are $\geq 65\%$ conserved across all Cas9 species with gray bars. (C) HNH sequence conservation within the active site (circled) is mapped onto the GeoHNH structure and colored according to the legend.

indicates a tendency to bend according to the first principal vector (Figure S12). Cluster analysis showed that, although the *Geo*HNH domain populates the crystallographic conformation for $\sim 90\%$ of the simulation time, it adopts multiple states varying the conformation of the N-terminal helix the remaining 10% of the time. Such dynamic behavior of the N-terminal helix is consistent with its role in diminishing the quality of *Geo*HNH NMR spectra, due to its highly dynamic nature. The motions of the N-terminal helix were even more pronounced at 40 °C, where we identified the three most probable states of *Geo*HNH (Fig. 6A): (i) crystal-like ($\sim 35\%$), (ii) with a bent N-terminal helix ($\sim 5\%$) and (iii) with the N-terminal helix shifted relative to the domain core ($\sim 46\%$). Interestingly, large amplitude motions of the *Geo*HNH core were not observed through PCA, nor clustering analysis (Fig. 6A, Figure S12-S13).

To further investigate the dynamics of the *Geo*HNH core, we computed the residue-averaged root-mean-square fluctuation (RMSF), which accounts for the per-residue conformational variability throughout the domain. As we already observed a high flexibility of the N-terminal helix, we ignored this fragment during the fitting procedure, allowing us to properly capture the core dynamics. Surprisingly, the dynamics of the *Geo*HNH core are preserved among the examined temperatures (Fig. 6B), showing that, unlike the N-terminal helix, the core dynamics are affected very little by temperature.

Finally, we compared the possible allosteric routes of communication within *Geo*HNH and *Sp*HNH by applying correlation analysis and network models derived from graph theory, through an approach that we previously implemented for *Sp*Cas9 (East et al., 2020a). This approach builds on the calculation of a generalized correlation (GC) matrix, which captures any type of correlations (including nonlinear motions) between C α atoms. The system's GCs are then used to define a network of nodes and edges, where the nodes represent the C α atoms of the protein connected by the edges, whose length depends on the strength of the correlation between the residues ($w_{ij} = -\log(GC_{ij})$). From this dynamic network, the allosteric pathways between “source” and “sink” nodes can be extracted by maximizing the overall residue-to-residue correlation. Here, the allosteric routes were computed

between residues of *Geo*HNH adjacent to the REC lobe (residue 18, “source”) and RuvC (residues 70/87, “sink”) (Fig. 6C, Figure S14). We found that in case of *Geo*HNH, the identified pathways are composed simply by the residues that are directly between the source and sink amino acids. Such pathways suggest that there is no specificity in the communication through *Geo*HNH, consistent with CPMG results, where no residues with relaxation parameters suggestive of μ s-ms motion were found.

4. Discussion

The HNH protein is found in many organisms, and its prominent role as a nuclease in the Cas9 molecular machine has necessitated investigations into its biophysical properties. We used truncated constructs of Cas9 to study the structure and dynamics of a novel HNH domain from *G. stearothermophilus* and compared these findings to the canonical *S. pyogenes* HNH. Using a combined approach of experimental NMR and MD simulations, we showed previously that *Sp*HNH undergoes extensive conformational dynamics on the ms timescale that form a well-defined “pathway” for biological signal transduction (East et al., 2020a). Here, we asked whether a sequentially and environmentally divergent HNH from a thermophilic bacterium maintained this structure and conformational motion.

Inspection of the NMR order parameters (S^2) and sequence alignments reveal no correlation in *Geo*HNH between conserved residues of the primary sequence and the degree of fast timescale flexibility (Fig. 5B). That is, ps-ns dynamics in *Geo*HNH are not confined to areas of high or low sequence similarity, but instead dispersed throughout its entire structure. While our data represent an unexpected shift in dynamic profile for *Geo*HNH when compared to *Sp*HNH, it is not necessarily uncommon. There are many examples in the literature of fast timescale dynamics dominating in thermophiles, as well as proteins that have distinct dynamic profiles in mesophilic versus thermophilic homologs. This has been illustrated in studies of thermophilic and mesophilic bacterial enzyme 1, which have a strong temperature dependence to μ s-ms timescale motions. Chimeras of the two enzymes impart

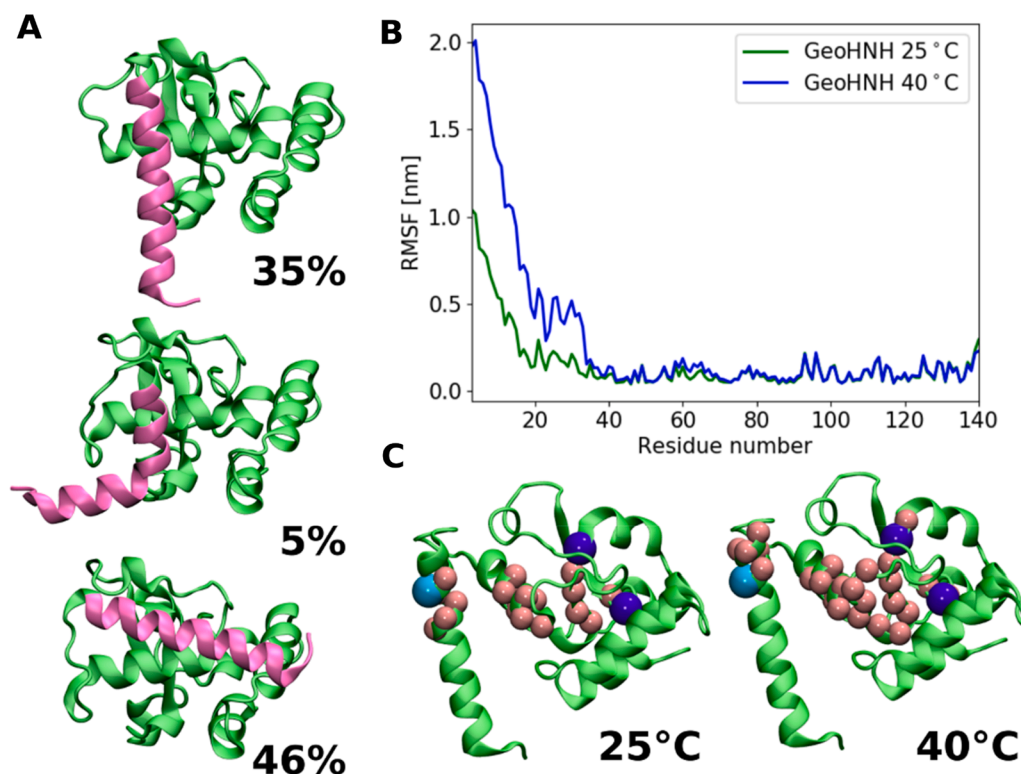


Fig. 6. (A) Representative structures of *Geo*HNH at 40 °C. The flexible N-terminal helix is represented in pink and the core of the domain is shown in green. (B) Residue-averaged root mean square fluctuation for *Geo*HNH at 25 °C (green) and 40 °C (blue). (C) The allosteric pathways identified for *Geo*HNH at 25 °C (left) and 40 °C (right). Cyan and dark blue spheres represent residues proximal to Rec3 (source of the signal) and RuvC (sinks of the signal), respectively. Pink spheres represent the residues that are involved in the allosteric signal transmission.

“hybrid” motional properties that pinpoint the regions of structure most strongly affected by temperature (Dotas et al., 2020). Studies of mesophilic and hyperthermophilic adenylate kinases (Adk) also exhibit different compositions of fast-and-slow dynamics at 20 °C, while motions of the Adk thermophile shift to faster timescales at its physiologically relevant temperature of 70 °C (Wolf-Watz et al., 2004). In contrast, the IGPS hyperthermophile maintains μ s timescale motions over a temperature range of 30–70 °C, though the motions become less concerted and diminish allosteric coupling at high temperature (Lisi et al., 2017, 2018). The motional properties of proteins can be quite different at physiological and non-physiological temperatures, but we find little evidence for this in *Geo*HNH- Δ N relaxation experiments conducted at 40 °C (Fig. 4), consistent with RMSF analysis from simulations. However, our findings are relevant only to the isolated *Geo*HNH domain in this report, while prior work was conducted on intact enzymes. While we have proven our HNH constructs to be excellent representations of HNH domains in full-length Cas9s, we cannot rule out a stronger effect of temperature in experimental studies of larger *Geo*Cas9 portions.

Nearly all of the residues that are conserved in the HNH domains of numerous Cas9 proteins are located on the half of the structure that faces the core of full-length Cas9 and interacts with the DNA strand targeted for cleavage (Fig. 5C). Interestingly, the *Geo*HNH residues suggested by our NMR relaxation analysis to undergo *ps*-*ns* motion (R_1/R_2 values 1.5 σ below the 10% trimmed mean, Fig. 3), are also localized to the interface of the structure that faces the core of the full-length protein. Thirteen of these sites span the length of the HNH structure, bridging the Rec-HNH interface (Figure S6). In contrast, the dynamic pathway bridging Rec-HNH in *Sp*Cas9 consisted of > 25 residues undergoing chemical exchange on a much slower timescale. It should be noted that the dynamic profiles of *Geo*HNH were measured at 25 °C (to be consistent with published studies on *Sp*Cas9) although *Geo*Cas9 enzymatic activity is optimal at higher temperatures (50 – 70 °C). Although studies of other thermophilic enzymes have shown the extent of protein motions and allosteric coupling to be quite different at elevated temperature than that observed near room temperature (Lisi et al., 2018), we find no evidence for this in *Geo*HNH relaxation experiments conducted at 40 °C (Fig. 4), consistent with RMSF analysis from simulations.

We initially speculated that the basic biophysical properties of *Sp*HNH would be shared by *Geo*HNH, as well as the HNH domains of most Cas9s. Such an assertion was supported by the structural similarity observed in Cas9 complexes in the Protein Data Bank. However, the native environment in which these Cas9 proteins exist, evidenced by wildly variable thermal stabilities, appears to modulate the dynamic properties of HNH. These observations may eventually be rationalized by differences in overall architecture of the full-length *Sp*Cas9 and *Geo*Cas9 proteins, but at this point, isolated HNH domains with conserved core structures appear to be “imprinted” with certain biophysical properties stemming from their respective full-length Cas9. That is, the intrinsic protein motions that are suggested to regulate Cas9 may be different in the mesophilic and thermophilic systems, as indicated through studies of their respective HNH domains.

The divergence of *Sp*Cas9 and *Geo*Cas9 within the broader Type II CRISPR class may also be connected to the observed biophysical differences in future studies. Such designations are based on homology and the presence or absence of additional Cas proteins beyond the canonical Cas1, Cas2, and Cas9 systems (Shmakov et al., 2017). *Sp*Cas9 is a Type II-A CRISPR system (Jackson et al., 2017), while *Geo*Cas9 is Type II-C. Type II-C CRISPR systems are the simplest in that there are only three Cas proteins, but also the most diverse in the environments of their host bacteria and archaea (Mir et al., 2018). Type II-A CRISPR systems also include the β -casein protein (Csn2) (Jackson et al., 2017), which is required for adaptation of viral DNA to be inserted into CRISPR loci (Mir et al., 2018). Without the presence of Csn2 in Type II-C CRISPR systems, it is possible that *Geo*Cas9 might also play a role in this adaptation mechanism in addition to its better-known functions of DNA recognition

and cleavage. Interestingly, among the Cas9 proteins described in the literature, the only native Cas9s that have been validated for mammalian genome editing are Type II-C systems (Mir et al., 2018), including *Geo*Cas9 (Harrington et al., 2017), *Nme*Cas9 (Esvelt et al., 2013; Hou et al., 2013; Lee et al., 2016), and *Cj*Cas9 (Fig. 3A) (Kim et al., 2017). Modified versions of *Sp*Cas9 have also been validated for mammalian genome editing (Vakulskas et al., 2018). It is known that *Geo*Cas9 is functional and stable in human plasma (Harrington et al., 2017), while *Sp*Cas9 is not, thus it is possible that Cas9 proteins from Type II-C CRISPR systems are evolutionarily best equipped to function in humans, as seen from the prevalence of Type II-C CRISPR Cas9 proteins in human pathogenic and commensal bacteria (Chylinski et al., 2014). Continued study of this protein family is essential to comment further on this point. Additionally, Type II Cas systems have predominantly conserved endonuclease domains and are delineated by highly varied REC domains (Mir et al., 2018), consistent with our crystallographic, solution, and modeling studies of *Sp*Cas9 and *Geo*Cas9. The biophysical properties of the REC domain, including studies of dynamic signaling in solution, may provide additional insight into the distinct genome editing properties of *Sp*Cas9 and *Geo*Cas9. Further biophysical characterization of the *Geo*Cas9 REC domain will likely increase our understanding of how fast timescale flexibility is related to its ability to function at higher temperatures.

5. Materials and methods

5.1. Protein expression and purification

The HNH domain of *G. stearothermophilus* Cas9 (residues 511–662) was cloned into a pET28a vector with a N-terminal His₆-tag and a TEV protease cleavage site. The plasmid was transformed into BL21 (DE3) cells (New England Biolabs). Isotopically labeled samples were grown in M9 minimal media containing CaCl₂, MgSO₄, and MEM vitamins, and supplemented with ¹⁵N ammonium chloride and ¹³C glucose (1.0 g/liter and 2.0 g/liter, respectively; Cambridge Isotope Laboratories). Cells were induced with 1 mM IPTG after reaching an OD₆₀₀ of 0.8–1.0 and grown for 4 hr at 37 °C post induction. The cells were harvested by centrifugation, resuspended in lysis buffer (20 mM HEPES, 150 mM KCl, 5 mM imidazole, and 1 mM PMSF, pH 8.0), lysed by ultrasonication, and loaded onto a Ni – NTA column. The column was washed with 1 column volume of buffer A (20 mM HEPES, 80 mM KCl, pH 8.0), loaded onto a heparin column and eluted with a linear gradient from 0.08 to 1 mM KCl. The resulting fractions containing *Geo*HNH were further purified on a Hi-load Superdex 75 size-exclusion column and then dialyzed into 20 mM HEPES, 150 mM KCl, and 5 mM imidazole at pH 8.0 overnight at 4 °C with TEV. The sample was again loaded on a Ni-NTA column and pure *Geo*HNH was collected from the flow-through and dialyzed into a buffer of 20 mM HEPES, 80 mM KCl, 1 mM DTT, 1 mM EDTA, 5% glycerol, and 10% (v/v) D₂O at pH 7.4.

5.2. X-ray crystallography

*Geo*HNH protein used for crystallization was purified as described above and stored in a buffer of 20 mM HEPES and 80 mM KCl at pH 8.0. Crystals were obtained using sitting drop vapor diffusion at room temperature by mixing 0.2 μ L of 32 mg/mL *Geo*HNH with 0.4 μ L of crystallizing condition from the Morpheus Screen (Molecular Dimensions) well E3: 0.12 M Ethylene Glycols mixture containing 0.3 M diethylene glycol, 0.3 M triethylene glycol, 0.3 M tetraethylene glycol, and 0.3 M pentaethylene glycol; 0.1 M of Buffer System 1 mixture containing MES and imidazole, pH 6.5; and 30% (v/v) Precipitant Mix 3 containing 40% (v/v) glycerol and 20% (w/v) polyethylene glycol 4000. Diffraction images were collected using a Rigaku MicroMax-003i sealed tube X-ray generator with a Saturn 944 HG CCD detector. Images were processed using XDS (Kabsch, 2010) and Aimless in CCP4 (Winn et al., 2011). Chain A of the *N. meningitidis* HNH domain X-ray structure (PDB ID:

6J9N) was used for molecular replacement with Phaser followed by AutoBuild in Phenix (Liebschner et al., 2019). The GeoHNH structure was finalized through manual building in Coot (Emsley et al., 2010) and refinement in Phenix.

5.3. Generation of the GeoCas9 homology model

The homology model of GeoCas9 was performed using Phyre2 (Kelley et al., 2015) and AlphaFold2 (Jumper et al., 2021; Mirdita et al., 2021). Homology modeling with Phyre2 (Kelley et al., 2015) used information inferred from PDB 6JDV (Sun et al., 2019), which is the crystal structure of *Neisseria meningitidis* Cas9 (Nme1Cas9) in complex with a sgRNA and a target DNA. Due to potential bias in the orientation of HNH from GeoCas9 caused by the DNA-bound configuration of Nme1Cas9, we also carried out an alternative modeling using AlphaFold2 (Jumper et al., 2021) through the ColabFold (Mirdita et al., 2021) notebook environment. The AlphaFold2 model of GeoCas9 (Fig. 2) was obtained using as input the amino acid sequence and a HMMer (Eddy and Pearson, 2011) multiple sequence alignment (default method from DeepMind) (Jumper et al., 2021). In this step, the goal is to learn a rich “pairwise representation” that is informative about which residue pairs are close in 3D space. The next step uses this pairwise representation to directly produce atomic coordinates by treating each residue as a separate object, predicting the rotation and translation necessary to place each residue, and ultimately assembling a structured chain. In the present work, AlphaFold2 was used changing the number of recycles (max_recycle = 12). The accuracy of the obtained model was evaluated using a per-residue measure of local confidence (i.e., a predicted Local Distance Difference Test, pLDDT) on a scale from 0 (low confidence) to 100 (high confidence), revealing high confidence for the GeoCas9 structured domains (Figure S2). A second metric, the Predicted Aligned Error (PAE), reports AlphaFold’s expected position error at residue x , when the predicted and true structures are aligned on residue y . Low values of PAE suggested that AlphaFold’s structure exhibits confidence about the relative domain positions.

5.4. NMR spectroscopy

NMR data for backbone assignments were collected on Bruker Avance III 850 MHz and Bruker Avance NEO 600 MHz spectrometers at 25 °C. The following triple resonance experiments were collected using at 850 MHz: ^1H - ^{15}N HSQC, HNCA, HN(CO)CA, HN(CA)CB, and HN(COCA)CB. The HN(CA)CO and HNCO experiments were collected at 600 MHz. All spectra were processed in NMRPipe (Delaglio et al., 1995) and analyzed in Sparky (Lee et al., 2015). Three-dimensional correlations and assignments were made in CARRA (Keller, 2005) and deposited in the BMRB under accession number 50887.

NMR spin relaxation experiments were carried out at 600 and 850 MHz. Carr-Purcell-Meiboom-Gill (CPMG) experiments were adapted from the report of Palmer and coworkers (Loria et al., 1999) and performed at 25 °C with a constant relaxation period of 40 ms, a 2.0 s 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. Relaxation dispersion profiles were generated by plotting R_2 vs. $1/\tau_{\text{cp}}$ and exchange parameters were obtained from fits of these data carried out with in-house scripts, which are available upon request. Uncertainty values were obtained from replicate spectra. Longitudinal and transverse relaxation rates 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 ^1H - ^{15}N NOE were measured with a 6 s relaxation delay followed by a 3 s 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 (Bieri et al., 2011) with fully automated protocols.

5.5. Molecular dynamics simulations

The MD systems were composed of a single copy of GeoHNH domain embedded in 7.2 nm × 8.3 nm × 7.2 nm dodecahedron box, solvated with 13,279 water molecules and 39Na⁺ and 45Cl[−] to provide physiological ionic strength. The initial structure of GeoHNH was taken from the crystal structure reported in this article. All MD simulations were performed with Gromacs2019.6 (Abraham et al., 2015). The AMBER ff99SBnmr2 force field (Yu et al., 2020), which improves the consistency of the backbone conformational ensemble with NMR experiments, was used to represent the protein and the TIP3P model was used for water. The simulations were performed in NPT ensemble with temperature kept at 298 K or 313 K with v-rescale thermostat (Bussi et al., 2007) and pressure kept at 1 bar with Parrinello-Rahman barostat (Parrinello and Rahman, 1981). Particle Mesh Ewald method (Darden, 1999) was used to compute long-range electrostatic interactions with a real space cut-off of 1 nm. Van der Waals interactions were computed using cutoff of 1 nm. LINCS algorithm (Hess, 2008) was used to restrain all bonds involving hydrogen. The equations of motion were integrated with leapfrog Verlet algorithm with a time step of 2 fs. For both systems at 298 K and 313 K, 3 μs of the continuous trajectory was obtained.

Principal component analysis was performed for the whole GeoHNH domain with gmx covar and gmx anaeig tools, separately for the system at 298 K and 313 K. Cluster analysis was performed with gmx cluster, with cutoff of 0.25 nm and 0.3 nm for GeoHNH at 298 K and 313 K, respectively. Residue-averaged RMSF for the heavy atoms of GeoHNH was computed using gmx rmsf. Previously to RMSF computation, the trajectory was fitted with respect to the position of a domain core (residues 29–140).

The allosteric pathways were investigated using correlation analysis and graph theory. First, generalized correlations (GC), which capture both linear and nonlinear correlations between residues, were computed between C α atoms of GeoHNH. Next, the computed GCs were used to define a network of nodes and edges, where the nodes represent the C α atoms of the protein, while the edges are formed between interacting residues and their length is proportional to a negative logarithm of the correlation strength between the residues ($w_{ij} = -\log(\text{GC}_{ij})$). Finally, the suboptimal pathways between GeoHNH residue adjacent to REC lobe (residue 18, “source”) and RuvC (residues 70 and 87, “sinks”) were computed using subopt tool from Dynamical Network Analysis package with a cutoff for suboptimal pathways of 10 (Sethi et al., 2009).

Author contributions

HBB produced the GeoHNH protein and conducted the NMR experiments; AMD and GJ solved the X-ray crystal structure; LN conducted the MD simulations; PA performed AlphaFold2 modeling; GJ supervised the X-ray crystallography; GP supervised the MD simulations; and GPL conceived the study and supervised the NMR spectroscopy. The manuscript was written through contributions from all authors.

CRediT authorship contribution statement

Helen B. Belato: Methodology, Investigation, Validation, Formal analysis, Writing – original draft, Visualization. **Alexandra M. D’Ordine:** Methodology, Investigation, Validation, Formal analysis, Visualization. **Lukasz Nierzwicki:** Methodology, Software, Investigation, Validation, Formal analysis, Writing – original draft, Visualization. **Pablo R. Arantes:** Methodology, Software, Investigation, Formal analysis, Visualization. **Gerwald Jogi:** Methodology, Resources, Validation, Formal analysis, Data curation, Supervision. **Giulia Palermo:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing – review & editing, Supervision, Project administration,

Funding acquisition. **George P. Lisi:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by funds from the COBRE Center for the Computational Biology of Human Disease (NIH P20 GM109035) and NIH R01 GM136815. This work was also funded by the National Science Foundation Grant CHE-1905374 and NIH R01GM141329. Computer time for MD simulations has been awarded by XSEDE under Grant No. TG-MCB160059 and by NERSC under Grant No. M3807.

Accession codes

Information regarding *SpCas9* and *GeoCas9* can be found under Uniprot IDs Q99ZW2 and A0A178TEJ9, respectively. NMR resonance assignments for the *Geo*HNH domain are deposited into the BioMagRes Bank (<https://bmrb.io/>) under entry number 50887. The X-ray crystal structure and atomic coordinates of *Geo*HNH are available in the Protein Data Bank (<https://www.rcsb.org/>) under accession number 7MPZ.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jsb.2021.107814>.

References

- Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- Anders, C., Niewoehner, O., Duerst, A., Jinek, M., 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513 (7519), 569–573. <https://doi.org/10.1038/nature13579>.
- Bieri, M., d'Auvergne, E.J., Gooley, P.R., 2011. relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and μ s motion of proteins. *J. Biomol. NMR* 50 (2), 147–155. <https://doi.org/10.1007/s10858-011-9509-1>.
- Bussi, G., Donadio, D., Parrinello, M., 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126 (1), 014101. <https://doi.org/10.1063/1.2408420>.
- Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., Doudna, J.A., 2017. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 550 (7676), 407–410. <https://doi.org/10.1038/nature24268>.
- Chylinski, K., Makarova, K.S., Charpentier, E., Koonin, E.V., 2014. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* 42 (10), 6091–6105. <https://doi.org/10.1093/nar/gku241>.
- Dagdas, Y.S., Chen, J.S., Sternberg, S.H., Doudna, J.A., Yildiz, A., 2017. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.* 3 (8), ea00027. <https://doi.org/10.1126/sciadv.a00027>.
- Darden, T., 1999. Ionic Charging Free Energies Using Ewald Summation. In: *Rational Drug Design*, pp. 149–161.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., Bax, A., 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6 (3), 277–293. <https://doi.org/10.1007/bf00197809>.
- Dotas, R.R., Nguyen, T.T., Stewart Jr., C.E., Ghirlando, R., Potoyan, D.A., Venditti, V., 2020. Hybrid Thermophilic/Mesophilic Enzymes Reveal a Role for Conformational Disorder in Regulation of Bacterial Enzyme I. *J. Mol. Biol.* 432 (16), 4481–4498. <https://doi.org/10.1016/j.jmb.2020.05.024>.
- East, K.W., Newton, J.C., Morzan, U.N., Narkhede, Y.B., Acharya, A., Skeens, E., Jogl, G., Batista, V.S., Palermo, G., Lisi, G.P., 2020a. Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J. Am. Chem. Soc.* 142 (3), 1348–1358. <https://doi.org/10.1021/jacs.9b10521>.
- East, K.W., Skeens, E., Cui, J.Y., Belato, H.B., Mitchell, B., Hsu, R., Batista, V.S., Palermo, G., Lisi, G.P., 2020b. NMR and computational methods for molecular resolution of allosteric pathways in enzyme complexes. *Biophys. Rev.* 12 (1), 155–174. <https://doi.org/10.1007/s12551-019-00609-z>.
- Eddy, S.R., Pearson, W.R., 2011. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* 7 (10), e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
- Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66 (Pt 4), 486–501. <https://doi.org/10.1107/s0907444910007493>.
- Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yang, S.J., Church, G.M., 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* 10 (11), 1116–1121. <https://doi.org/10.1038/nmeth.2681>.
- Frederick, K.K., Marlow, M.S., Valentine, K.G., Wand, A.J., 2007. Conformational entropy in molecular recognition by proteins. *Nature* 448 (7151), 325–329. <https://doi.org/10.1038/nature05959>.
- Harrington, L.B., Paez-Espino, D., Staahl, B.T., Chen, J.S., Ma, E., Kyrpides, N.C., Doudna, J.A., 2017. A thermostable Cas9 with increased lifetime in human plasma. *Nat. Commun.* 8 (1), 1424. <https://doi.org/10.1038/s41467-017-01408-4>.
- Hess, B., 2008. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* 4 (1), 116–122. <https://doi.org/10.1021/ct700200b>.
- Hou, Z., Zhang, Y., Propson, N.E., Howden, S.E., Chu, L.F., Sontheimer, E.J., Thomson, J. A., 2013. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc Natl Acad Sci U S A* 110 (39), 15644–15649. <https://doi.org/10.1073/pnas.1313587110>.
- Huai, C., Li, G., Yao, R., Zhang, Y., Cao, M.I., Kong, L., Jia, C., Yuan, H., Chen, H., Lu, D., Huang, Q., 2017. Structural insights into DNA cleavage activation of CRISPR-Cas9 system. *Nat. Commun.* 8 (1) <https://doi.org/10.1038/s41467-017-01496-2>.
- Jackson, S.A., McKenzie, R.E., Fagerlund, R.D., Kieper, S.N., Fineran, P.C., Brouns, S.J., 2017. CRISPR-Cas: Adapting to change. *Science* 356 (6333). <https://doi.org/10.1126/science.aal5056>.
- Jiang, F., Doudna, J.A., 2015. The structural biology of CRISPR-Cas systems. *Curr. Opin. Struct. Biol.* 30, 100–111. <https://doi.org/10.1016/j.sbi.2015.02.002>.
- Jiang, F., Doudna, J.A., 2017. CRISPR-Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* 46 (1), 505–529. <https://doi.org/10.1146/annurev-biophys-062215-010822>.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337 (6096), 816–821. <https://doi.org/10.1126/science.1225829>.
- 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., 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343 (6176). <https://doi.org/10.1126/science.1247997>.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohli, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- Kabsch, W., 2010. XDS. *Acta Crystallogr. D Biol. Crystallogr.* 66 (Pt 2), 125–132. <https://doi.org/10.1107/s0907444909047337>.
- Keller, R.L., 2005. Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment (Doctoral Thesis). ETH Zürich, ETH.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10 (6), 845–858. <https://doi.org/10.1038/nprot.2015.053>.
- Kern, D., Zwieterweg, E.R., 2003. The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* 13 (6), 748–757. <https://doi.org/10.1016/j.sbi.2003.10.008>.
- Kim, E., Koo, T., Park, S.W., Kim, D., Kim, K., Cho, H.-Y., Song, D.W., Lee, K.J., Jung, M. H., Kim, S., Kim, J.H., Kim, J.H., Kim, J.-S., 2017. In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* 8 (1) <https://doi.org/10.1038/ncomms14500>.
- Lee, C.M., Cradick, T.J., Bao, G., 2016. The *Neisseria meningitidis* CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells. *Mol. Ther.* 24 (3), 645–654. <https://doi.org/10.1038/mt.2016.8>.
- Lee, W., Tonelli, M., Markley, J.L., 2015. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* 31 (8), 1325–1327. <https://doi.org/10.1093/bioinformatics/btu830>.
- Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A.J., Moriarty, N.W., Oeffner, R.D., Poon, B.K., Prisant, M.G., Read, R.J., Richardson, J.S., Richardson, D.C., Sammito, M.D., Sobolev, O.V., Stockwell, D.H., Terwilliger, T.C., Urzhumtsev, A.G., Videau, L.L., Williams, C.J., Adams, P.D., 2019. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* 75 (10), 861–877.
- Lisi, G.P., Currier, A.A., Loria, J.P., 2018. Glutamine Hydrolysis by Imidazole Glycerol Phosphate Synthase Displays Temperature Dependent Allosteric Activation. *Front. Mol. Biosci.* 5, 4. <https://doi.org/10.3389/fmolb.2018.00004>.
- Lisi, G.P., East, K.W., Batista, V.S., Loria, J.P., 2017. Altering the allosteric pathway in IGPS suppresses millisecond motions and catalytic activity. *Proc Natl Acad Sci U S A* 114 (17), E3414–E3423. <https://doi.org/10.1073/pnas.1700448114>.
- Loria, J.P., Rance, M., Palmer, A.G., 1999. A Relaxation-Compensated Carr–Purcell–Meiboom–Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *J. Am. Chem. Soc.* 121 (10), 2331–2332. <https://doi.org/10.1021/ja983961a>.
- Mir, A., Edraki, A., Lee, J., Sontheimer, E.J., 2018. Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. *ACS Chem. Biol.* 13 (2), 357–365. <https://doi.org/10.1021/acscmbio.7b00855>.

- Mirdita, M., Ovchinnikov, S., Steinegger, M., 2021. ColabFold - Making protein folding accessible to all. *bioRxiv*. <https://doi.org/10.1101/2021.08.15.456425>.
- Nierzwicki, L., Arantes, P.R., Saha, A., Palermo, G., 2021. Establishing the allosteric mechanism in CRISPR-Cas9. *Wiley Interdiscip Rev Comput Mol Sci* 11 (3). <https://doi.org/10.1002/wcms.1503>.
- Nishimasu, H., Ran, F.A., Hsu, P., Konermann, S., Shehata, S., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O., 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156 (5), 935–949. <https://doi.org/10.1016/j.cell.2014.02.001>.
- Oakes, B.L., Nadler, D.C., Flamholz, A., Fellmann, C., Staahl, B.T., Doudna, J.A., Savage, D.F., 2016. Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.* 34 (6), 646–651. <https://doi.org/10.1038/nbt.3528>.
- Palermo, G., Chen, J.S., Ricci, C.G., Rivalta, I., Jinek, M., Batista, V.S., Doudna, J.A., McCammon, J.A., 2018. Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. *Q. Rev. Biophys.* 51 <https://doi.org/10.1017/S0033583518000070>.
- Palermo, G., Miao, Y., Walker, R.C., Jinek, M., McCammon, J.A., 2016. Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. *ACS Cent. Sci.* 2 (10), 756–763. <https://doi.org/10.1021/acscentsci.6b00218>.
- Palermo, G., Miao, Y., Walker, R.C., Jinek, M., McCammon, J.A., 2017a. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proc Natl Acad Sci U S A* 114 (28), 7260–7265. <https://doi.org/10.1073/pnas.1707645114>.
- Palermo, G., Ricci, C.G., Fernando, A., Basak, R., Jinek, M., Rivalta, I., Batista, V.S., McCammon, J.A., 2017b. Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. *J. Am. Chem. Soc.* 139 (45), 16028–16031. <https://doi.org/10.1021/jacs.7b05313>.
- Parrinello, M., Rahman, A., 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52 (12), 7182–7190. <https://doi.org/10.1063/1.328693>.
- Saha, A., Arantes, P.R., Hsu, R.V., Narkhede, Y.B., Jinek, M., Palermo, G., 2020. Molecular Dynamics Reveals a DNA-Induced Dynamic Switch Triggering Activation of CRISPR-Cas12a. *J. Chem. Inf. Model.* 60 (12), 6427–6437. <https://doi.org/10.1021/acs.jcim.0c00929>.
- Schmid-Burgk, J.L., Gao, L., Li, D., Gardner, Z., Strecker, J., Lash, B., Zhang, F., 2020. Highly Parallel Profiling of Cas9 Variant Specificity. *Mol. Cell* 78 (4), 794–800. <https://doi.org/10.1016/j.molcel.2020.02.023>.
- Sethi, A., Eargle, J., Black, A.A., Luthy-Schulten, Z., 2009. Dynamical networks in tRNA: protein complexes. *Proc. Natl. Acad. Sci.* 106 (16), 6620–6625. <https://doi.org/10.1073/pnas.0810961106>.
- Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., Severinov, K., Zhang, F., Koonin, E.V., 2017. Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* 15 (3), 169–182. <https://doi.org/10.1038/nrmicro.2016.184>.
- Singh, D., Wang, Y., Mallon, J., Yang, O., Fei, J., Poddar, A., Ceylan, D., Bailey, S., Ha, T., 2018. Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat. Struct. Mol. Biol.* 25 (4), 347–354. <https://doi.org/10.1038/s41594-018-0051-7>.
- Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., Zhang, F., 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science* 351 (6268), 84–88. <https://doi.org/10.1126/science.aad5227>.
- Sternberg, S.H., LaFrance, B., Kaplan, M., Doudna, J.A., 2015. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* 527 (7576), 110–113. <https://doi.org/10.1038/nature15544>.
- Sun, W., Yang, J., Cheng, Z., Amrani, N., Liu, C., Wang, K., Ibraheim, R., Edraki, A., Huang, X., Wang, M., Wang, J., Liu, L., Sheng, G., Yang, Y., Lou, J., Sontheimer, E.J., Wang, Y., 2019. Structures of *Neisseria meningitidis* Cas9 Complexes in Catalytically Poised and Anti-CRISPR-Inhibited States. *Mol. Cell* 76 (6), 938–952. <https://doi.org/10.1016/j.molcel.2019.09.025>.
- Tzeng, S.R., Kalodimos, C.G., 2009. Dynamic activation of an allosteric regulatory protein. *Nature* 462 (7271), 368–372. <https://doi.org/10.1038/nature08560>.
- Vakulskas, C.A., Dever, D.P., Rettig, G.R., Turk, R., Jacobi, A.M., Collingwood, M.A., Bode, N.M., McNeill, M.S., Yan, S., Camarena, J., Lee, C.M., Park, S.H., Wiebking, V., Bak, R.O., Gomez-Ospina, N., Pavel-Dinu, M., Sun, W., Bao, G., Porteus, M.H., Behlke, M.A., 2018. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat. Med.* 24 (8), 1216–1224. <https://doi.org/10.1038/s41591-018-0137-0>.
- Whittier, S.K., Hengge, A.C., Loria, J.P., 2013. Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases. *Science* 341 (6148), 899–903. <https://doi.org/10.1126/science.1241735>.
- 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.W., 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., 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* 67 (4), 235–242.
- Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E.Z., Kern, D., 2004. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat. Struct. Mol. Biol.* 11 (10), 945–949. <https://doi.org/10.1038/nsmb821>.
- Yu, L., Li, D.-W., Bruschweiler, R., 2020. Balanced Amino-Acid-Specific Molecular Dynamics Force Field for the Realistic Simulation of Both Folded and Disordered Proteins. *J. Chem. Theory Comput.* 16 (2), 1311–1318. <https://doi.org/10.1021/acs.jctc.9b01062>.
- Zhang, F., 2019. Development of CRISPR-Cas systems for genome editing and beyond. *Q. Rev. Biophys.* 52 <https://doi.org/10.1017/s0033583519000052>.