Atomistic Tuning of the *Geo*Cas9 Recognition Lobe Modulates Allosteric Motions and Guide RNA Interactions

Helen B. Belato[1,#], Alexa L. Knight[1,#], Alexandra M. D'Ordine[1], Zhiqiang Fan[2], Jinping Luo[2], , Gerwald Jogl[1], George P. Lisi[1,*]

[1]Department of Molecular Biology, Cell Biology & Biochemistry, Brown University, Providence, RI USA

[2]Brown University Mouse Transgenic & Gene Targeting Facility, Providence, RI USA

[#]These authors contributed equally to this work

*** Correspondence:**
george_lisi@brown.edu

Keywords: CRISPR-Cas9, guide RNA, protein dynamics, NMR, thermophile

Abstract

The intuitive manipulation of specific amino acids to alter the activity or specificity of CRISPR-Cas9 has been a topic of great interest. As a large multi-domain RNA-guided endonuclease, the intricate molecular crosstalk within the Cas9 protein hinges on its conformational dynamics, but a comprehensive understanding of the extent and timescale of the motions that drive its allosteric function and association with nucleic acids remains elusive. Here, we investigated the structure and multi-timescale molecular motions of the recognition (Rec) lobe of *Geo*Cas9, a thermophilic Cas9 from *Geobacillus stearothermophilus.* Our results provide new atomic details about the *Geo*Rec subdomains (*Geo*Rec1, *Geo*Rec2) and the full-length domain in solution. Two single-point mutants, K267E and R332A, enhanced and redistributed micro-millisecond flexibility throughout *Geo*Rec, and NMR studies of the interaction between *Geo*Rec and its guide RNA showed that mutations reduced this affinity and the stability of the ribonucleoprotein complex. Despite measured biophysical differences due to the mutations, DNA cleavage assays reveal only modest functional differences in on-target activity, and similar specificity. These data highlight

how guide RNA interactions can be tuned in the absence of major functional losses, but also raise questions about the underlying mechanism of *Geo*Cas9, since analogous single-point mutations have significantly impacted on- and off-target DNA editing in mesophilic *S. pyogenes* Cas9. A K267E/R332A double mutant did modestly enhance *Geo*Cas9 specificity, highlighting the robust evolutionary tolerance of Cas9 and species-dependent complexity. Ultimately, this work provides an avenue by which to modulate the structure, motion, and nucleic acid interactions at the level of the Rec lobe of *Geo*Cas9, setting the stage for future studies of *Geo*Cas9 variants and their effect on its allosteric mechanism.

Introduction

The vast majority of Cas systems explored as genome editors originate from mesophilic hosts. The emergence of the thermophilic *Geo*Cas9, with DNA cleavage function up to 85 °C, can expand CRISPR technology to higher temperature regimes and stabilities, but similarities and differences in its mechanism relative to other Cas9s must be established. The widely-studied *Sp*Cas9, as well as *Geo*Cas9, are Type-II CRISPR systems, though only the Type II-C class to which *Geo*Cas9 belongs has been rigorously validated for mammalian genome editing, reinforcing the need to better understand this CRISPR class.[1] The similar domain arrangements of *Geo*Cas9 and *Sp*Cas9 led us to initially speculate that these could share atomic level mechanistic similarities.[2] *Geo*Cas9 utilizes a guide RNA (gRNA) to recognize, unwind, and cleave a double-stranded DNA target after recognition of its 5'NNNNCRAA-3' protospacer adjacent motif (PAM)[3,4] and both of the *Geo*Cas9 nuclease active sites are spatially distinct from the PAM recognition site, necessitating structural and dynamic changes that allosterically propagate DNA binding information to these subdomains. Our prior work revealed a divergence in the timescales of allosteric motions in the *Sp*Cas9 and *Geo*Cas9 HNH domains[2,5] and it is also possible that

docking of the gRNA with *Geo*Cas9, and therefore its interaction with the RNA:DNA hybrid, may differ from the *Sp*Cas9 system, as *Geo*Cas9 contains a truncated recognition (Rec) lobe with only two of the three major subdomains.

The high thermal stability and more compact size of *Geo*Cas9 (it is 281 residues shorter than *Sp*Cas9) can be especially important for *in vivo* delivery applications, since promising viral vectors (*i.e.* adeno-associated virus, AAV) have cargo capacities of ~4.7kb,[6] which prevents *Sp*Cas9-gRNA packaging into a single AAV vector but permits "all-in-one" delivery of *Geo*Cas9-gRNA.[1] Despite this potential, little is known about specific residues that influence *Geo*Cas9 structure, nucleic acid binding, or function. Our recent NMR work with *Sp*Cas9 uncovered pathways of micro-millisecond timescale motions that propagate chemical information related to allostery and specificity through *Sp*Rec and its RNA:DNA hybrid,[5,7] prompting us to investigate this phenomenon in *Geo*Rec.

The structural basis for specificity in protein-nucleic acid complexes is also poorly understood and difficult to quantify. Although dynamic ensembles in DNA repair enzymes have provided some context,[8-10] many efforts to improve Cas9 specificity and reduce off-target activity have relied on large mutational screens or error-prone PCR,[11-14] which are less intuitive. Inter-subunit allosteric communication between the catalytic HNH domain and the Rec lobe is critical to Cas9 specificity, as the binding of off-target DNA sequences at Rec alter HNH dynamics to affect DNA cleavage.[3,11,15] To further probe the fundamental role of protein motions in the function and specificity of *Geo*Cas9, as well as the effect of protein-nucleic acid interactions on its structural signatures, we engineered two mutations in *Geo*Rec (K267E and R332A, housed within *Geo*Rec2). We hypothesized that these variants could enhance *Geo*Cas9 specificity (*i.e.* limit its off-target cleavage) for two reasons. First, the chosen mutation sites are homologous to those of specificity-

enhancing variants of *Sp*Cas9.[16,17] Second, altered Cas9-gRNA interactions have been shown to be a consequence of specificity-enhancement and these charged residues appear to directly interact with the gRNA.[18-21] Balancing these two points is the fact that Type-II Cas systems generally have conserved nuclease domains, but are delineated by highly varied Rec domains.[1] This implies that the structural and dynamic properties of Rec may play an outsized role in differentiating the functions of *Sp*Cas9 and *Geo*Cas9, which may not be identical. Nevertheless, our work provides new insight into the biophysical, biochemical, and functional role of the *Geo*Rec lobe and how mutations modulate the domain itself and its interaction with gRNA in full-length *Geo*Cas9.

Results

The structural similarity of GeoRec1, GeoRec2, and GeoRec facilitates NMR analysis of protein dynamics and nucleic acid affinity

*Geo*Cas9 is a 1087 amino acid polypeptide, thus we employed a "divide and conquer" approach for NMR studies, which we previously showed to be useful for quantifying allosteric structure and motion in *Sp*Cas9.[5,7,22,23] The *Geo*Rec lobe is comprised of subdomains *Geo*Rec1 and *Geo*Rec2, which likely work together to recognize nucleic acids. We engineered constructs of the *Geo*Rec1 (136 residues, 16 kDa) and *Geo*Rec2 (212 residues, 25 kDa) subdomains and solved the X-ray crystal structure of *Geo*Rec2 at 1.49 Å, which aligns remarkably well with the structure of the *Geo*Rec2 domain within the AlphaFold model of *Geo*Cas9 (RMSD 1.025 Å, **Figure 1A**). We were neither able to crystallize *Geo*Rec1 nor full-length *Geo*Cas9, nonetheless *Geo*Rec2 in isolation represents the structure of the subdomain within full-length *Geo*Cas9 very well. Our previous studies of *Geo*HNH also show identical superpositions of X-ray crystal structures with full-length Cas complexes.[2] In addition to the individual subdomains, we also generated a construct of the intact *Geo*Rec lobe (370 residues, 43 kDa).

Despite only 22% sequence identity **(Figure S1A)**, the structure of *Sp*Rec3 and *Geo*Rec2 are highly similar (RMSD 2.00 Å, **Figure S1B)**. The structure of *Geo*Rec1, in contrast, does not align perfectly with *Sp*Rec1, instead, it partially aligns with both *Sp*Rec1 and *Sp*Rec2 **(Figure S1C)**. Thus, the nearly identical *Sp*Rec3 and *Geo*Rec2 architectures and their intrinsic dynamics may be a common thread among Type II Cas9s of different size and PAM preference. To capture atomic-level signatures of *Geo*Rec, we obtained well-resolved ^{1}H-^{15}N NMR fingerprint spectra for all three protein constructs, and assigned the amide backbones (**Figure S2)**. ^{1}H-^{15}N amide (**Figure**

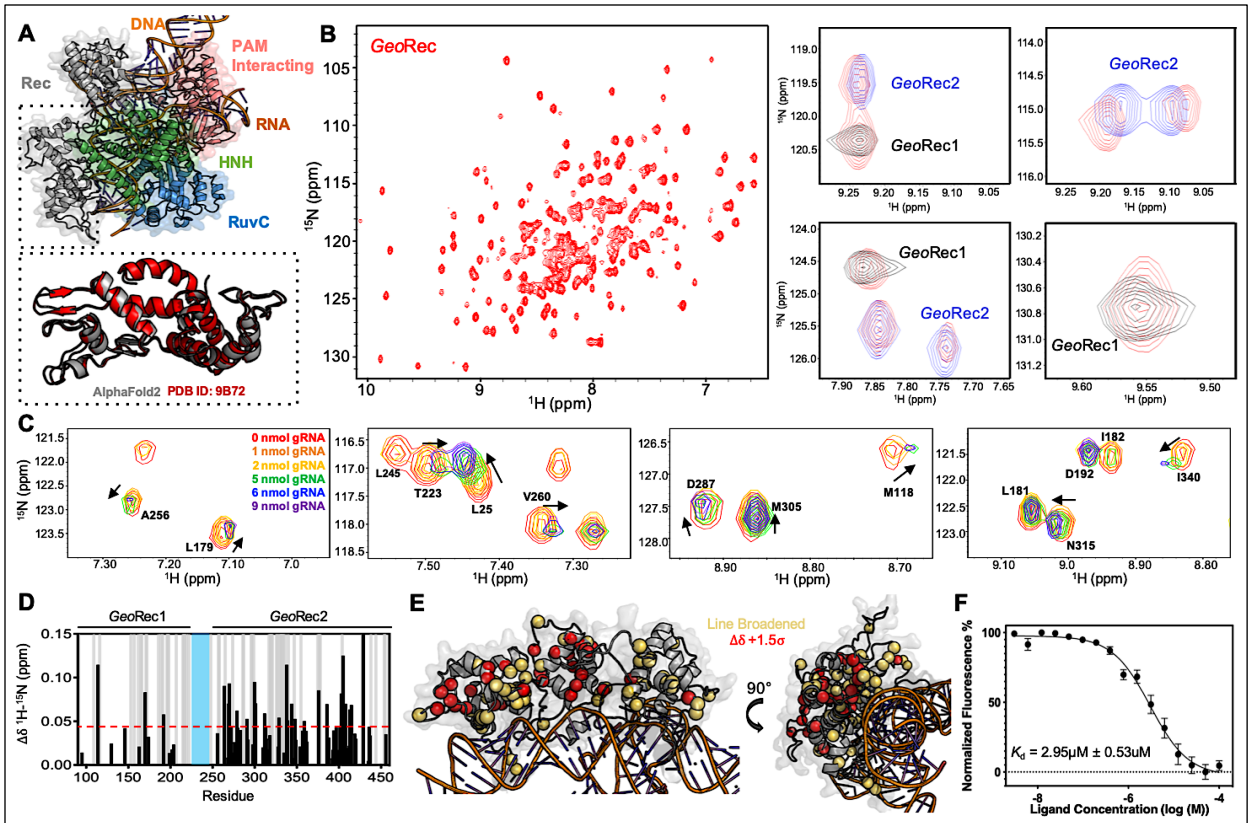

Figure 1. (A) Architecture of *Geo*Cas9 modeled with AlphaFold2. The *Geo*Rec2 domain from the model (grey) is overlaid with an X-ray structure of *Geo*Rec2 (red, PDB ID: 9B72, RMSD 1.03 Å). **(B)** ^{1}H-^{15}N TROSY HSQC NMR spectrum of *Geo*Rec collected at 850 MHz. Overlays of this spectrum with resonances from spectra of *Geo*Rec1 (black) and *Geo*Rec2 (blue) demonstrate a structural similarity between the isolated subdomain and intact *Geo*Rec. **(C)** NMR titration of a 39-nt gRNA into WT *Geo*Rec. Representative resonances are colored by increasing gRNA concentration in the legend. **(D)** NMR chemical shift perturbations caused by gRNA binding to WT *Geo*Rec. Gray bars denote sites of line broadening and the blue bar denotes an unassigned region of *Geo*Rec corresponding to the native Rec1-Rec2 linker. The red dashed line indicates 1.5σ above the 10% trimmed mean of the data. **(E)** Chemical shift perturbations >1.5σ are mapped onto *Geo*Rec (red spheres). Resonances that have broadened beyond detection are mapped as yellow spheres. **(F)** MST-derived binding affinity of *Geo*Rec to the Cy5-labeled 39-nt gRNA yields a K_d = 2.95 ± 0.53 μM.

1B, S2) and ^{1}H-$^{13}CH_3$ Ile, Leu, and Val (ILV)-methyl NMR spectra (**Figure S3**) of *Geo*Rec overlay very well with those of its individual subdomains, suggesting that the linkage of subdomains within the full-length *Geo*Rec polypeptide does not alter their isolated folds. Consistent with this observation, circular dichroism (CD) thermal unfolding profiles of *Geo*Rec1 (T_m ~ 34 °C) and *Geo*Rec2 (T_m = 61.50 °C) are distinct, and occur as separate events in the unfolding profile of *Geo*Rec (**Figure S4**). The dumbbell shape of the *Geo*Rec lobe, where the two globular subdomains are connected by a linker is a likely contributor to these biophysical properties.

The Rec lobe of Cas9 is responsible for orienting the RNA:DNA hybrid, as well as the adjacent nuclease domains, into their active conformations.[1,19,20,24] Thus, the structure, motions, and nucleic acid interactions of Rec represent a critical piece of the Cas9 signaling machinery. Previous studies of *Sp*Cas9 revealed that gRNA binding (*i.e.* ribonucleoprotein complex (RNP) formation) induces a global structural rearrangement of the protein that positions the adjacent HNH into its "proofreading" state,[18] after which target DNA binding positions the nucleases into active conformations for cleavage.[18,25] We wondered if the atomistic details of the apo *Geo*Cas9-to-RNP transition could be captured by NMR using the *Geo*Rec construct and a 101-nucleotide (nt) gRNA derived from the full 141-nt guide containing a 23-nt spacer followed by an 8-nt repeat (PAM complement) and 110-nt guide sequence intrinsic to *Geo*Cas9. An overlay of the ^{1}H-^{15}N HSQC NMR spectra of apo *Geo*Rec and *Geo*Rec-RNP at a 1:1 molar ratio shows extensive line broadening (**Figure S5A**), likely due to the large size of the complex (75.5 kDa). To mitigate this issue, we created a second 39-nt RNA construct maintaining the spacer sequence that is known to bind to the Rec lobe of other Cas9s. When bound to *Geo*Rec, this complex is 55.6 kDa and more amenable to NMR studies. A ^{1}H-^{15}N NMR spectral overlay of apo *Geo*Rec and *Geo*Rec in complex with 39-nt gRNA at a 1:1 molar ratio shows clear, resolved resonances with significant chemical

shift perturbations and line broadening (**Figure 1C, S5B**). Importantly, many resonances are also minimally impacted. The results of gRNA titration experiments were quantified in **Figure 1D** and mapped onto the *Geo*Rec structure in **Figure 1E**. The strongest chemical shift perturbations are localized to the *Geo*Rec2 subdomain that interfaces with the RNA:DNA hybrid at the PAM distal end. Previous studies of specificity-enhancing variants of *Sp*Cas9 identified structural perturbations to *Sp*Rec3, affecting RNA:DNA hybrid binding at the PAM distal end.[7] These data suggest that in the analogous *Geo*Rec2, residues located near the PAM distal end of the nucleic acids play a similarly critical role in binding. Modulation of binding affinity within this region, as in *Sp*Cas9, may lead to predictable functional changes in *Geo*Cas9. In addition, line broadening is evident in both *Geo*Rec1 and *Geo*Rec2, primarily localized to the RNA:DNA hybrid interface. Microscale thermophoresis (MST) experiments quantified the affinity of *Geo*Rec for gRNA, producing a K_d = 2.95 ± 0.53 μM that is consistent with the strong NMR chemical shift perturbations (**Figure 1F**).

GeoRec2 mutants do not substantially impact the GeoRec structure

To understand how the structure and gRNA interactions of *Geo*Cas9 can be modulated at the level of *Geo*Rec, we engineered two charge-altering point mutants in the *Geo*Rec2 subdomain, K267E and R332A. Based on the AlphaFold2 model of *Geo*Cas9, both of these residues are < 5Å from the bound RNA:DNA hybrid and are predicted to interface with the nucleic acids directly (**Figure 2A,B)**. Of course, we must note that due to the lack of an experimental structure of *Geo*Cas9, the interactions between K267 and R332 and the gRNA may be different than predicted. The rationale for our designed mutations was that removal of positive charge would weaken the interactions between *Geo*Cas9 and the gRNA, affecting K_d via the electrostatics or dynamics of the *Geo*Rec lobe. Studies of *Sp*Cas9 revealed that interactions of *Sp*Rec3 (analogous to *Geo*Rec2)

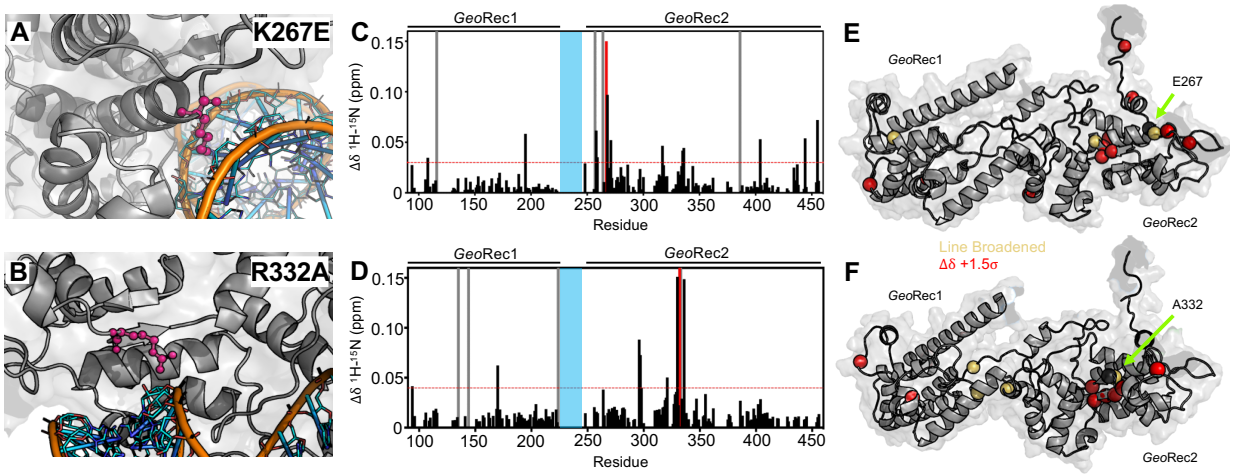

Figure 2. (A, B) Sites of selected mutations within *Geo*Rec2, K267 and R332, are highlighted as purple sticks directly facing the nucleic acids modeled from *Nme*Cas9 (PDB ID: 6JDV), allowing for prediction of the binding orientation within *Geo*Cas9. NMR chemical shift perturbations caused by the K267E **(C)** or R332A **(D)** mutations are plotted for each residue of *Geo*Rec. Gray bars denote sites of line broadening, the blue bar denotes an unassigned region of *Geo*Rec corresponding to the native Rec1-Rec2 linker, and the red bar indicates the mutation site. The red dashed line indicates 1.5σ above the 10% trimmed mean of all shifts. Chemical shift perturbations >1.5σ are mapped onto K267E **(E)** and R332A **(F)** *Geo*Rec (red spheres). Resonances that have broadened beyond detection are mapped as yellow spheres and the mutation sites are indicated by a black sphere and green arrow.

with its RNA:DNA hybrid triggers conformational rearrangements that allow the catalytic HNH domain to sample its active conformation.[19] Thus, *Sp*Rec3 acts as an allosteric effector that recognizes the RNA:DNA hybrid to activate HNH. Nucleotide mismatches (*i.e.* off-target DNA sequences) in the target DNA generally prevent *Sp*Rec3 from undergoing the full extent of its required conformational rearrangements to allosterically activate HNH, leaving HNH in a "proofreading" state with its catalytic residues too far from the DNA cleavage site. Off-target DNA cleavage by Cas9 remains an area of intense study and substantial effort from various groups has gone into mitigating such effects.[18,21,26-28] Indeed, many high-specificity *Sp*Cas9 variants contain mutations within *Sp*Rec3 that increase the threshold for its conformational activation, reducing the propensity for HNH to sample its active state in the presence of off-target DNA sequences.[18,19,28] Studies of flexibility within Rec itself, as well as its gRNA interactions in the presence of mutations, are therefore essential for connecting biophysical properties to function and specificity in related Cas9s.

The K267E *Geo*Rec2 variant is sequentially and structurally similar to a specificity enhancing site in *Sp*Cas9 (K526E), within the evoCas9 system.[16] The *Sp*Cas9 K526E mutation substantially reduced off-target activity alone, but was even more effective in conjunction with three other single-point mutations in *Sp*Rec3.[16] The R332A *Geo*Rec2 variant also mimics one mutation within a high-specificity *Sp*Cas9 variant, called HiFi Cas9-R691A.[17] We assessed mutation-induced changes to local structure in *Geo*Rec via NMR chemical shift perturbations to ^{1}H-^{15}N HSQC backbone amide spectra. Consistent with experiments using *Geo*Rec2 alone, chemical shift perturbations and line broadening are highly localized to the mutation sites. (**Figure 2C-F**). Perturbation profiles of the *Geo*Rec2 subdomain and intact *Geo*Rec also implicate the same residues as sensitive to the mutations (**Figure S6**).

CD spectroscopy revealed that wild-type (WT*)*, K267E, and R332A *Geo*Rec2 maintained similar alpha-helical secondary structure (**Figure S7A**), though the thermostability of both variants was slightly reduced from that of WT *Geo*Rec2 (**Figure S7B)**. The T_m of WT *Geo*Rec2 is ~62 °C, consistent with the T_m of the full-length *Geo*Cas9, while that of K267E *Geo*Rec2 was decreased to ~55 °C. Though the R332A *Geo*Rec2 T_m remains ~62 °C, this variant underwent a smaller unfolding event near 40 °C before completely unfolding. These data suggest that despite small structural perturbations, both mutations are destabilizing to *Geo*Rec2, which implies a change in NMR-detectable protein dynamics.

Mutations enhance and redistribute protein dynamics within GeoRec2

Due to the high molecular weight of the intact *Geo*Rec lobe, decays in NMR signal associated with spin relaxation experiments were significant and hampered data quality. Thus, we focused on quantifying the molecular motions of the *Geo*Rec2 subdomain, where the K267E and R332A mutations reside and the chemical shift perturbations are most apparent. To obtain high-

quality per-residue information representative of *Geo*Rec, we measured longitudinal (R_1) and transverse (R_2) relaxation rates and heteronuclear ^{1}H-[^{15}N] NOEs, then used these data in a Model-free analysis of per-residue order parameters (S^2, **Figure 3A**). Previous measurements of S^2 across the adjacent *Geo*HNH nuclease revealed substantial ps-ns timescale flexibility,[2] leading us to wonder whether a similar observation would be made for *Geo*Rec2, which abuts *Geo*HNH. Such a finding would suggest that HNH-Rec2 crosstalk in *Geo*Cas9 is driven primarily by rapid bond vector fluctuations. However, unlike *Geo*HNH, S^2 values for *Geo*Rec2 are generally elevated, suggesting that the ps-ns motions of this subdomain arise primarily from global tumbling of the protein in solution. We therefore carried out Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion NMR experiments to assess the μs-ms flexibility of *Geo*Rec2, which has been linked to chemical information transfer in the well-studied *Sp*Cas9.[5,7,20,22] Evidence of μs-ms motions (*i.e.* curved relaxation dispersion profiles) is observed in 17 residues within the *Geo*Rec2 core, spanning its interfaces to Rec1 and HNH **(Figure 3B,C)**. Such motions are absent from *Geo*HNH, thus two neighboring domains, *Geo*Rec2 and *Geo*HNH, diverge in their intrinsic flexibility (at least in isolation), raising questions about the functional implications of these motions in *Geo*Rec2. We previously showed that heightened flexibility of *Sp*Rec3 via specificity-enhancing mutations concomitantly narrowed the conformational space sampled by *Sp*HNH, highlighting a "motional trade-off" between the domains, which may be mirrored here. Manipulation of the flexibility of *Sp*Cas9 and *Geo*Cas9 domains by mutagenesis also impacts aspects of nucleic acid binding and cleavage,[18,20,22,29,30] which led us to investigate similar perturbations in *Geo*Rec2.

Since *Sp*Rec3 and *Geo*Rec2 have similar structures and μs-ms flexibility, we speculated that charge-altering mutations would modulate the biophysical properties of *Geo*Rec and the function of *Geo*Cas9, as observed for *Sp*Cas9. We investigated K267E and R332A *Geo*Rec2 with

NMR spin relaxation, as described for WT *Geo*Rec2 (*vide supra*). An analysis of chemical exchange rates, k_{ex}, derived from dual-field CPMG relaxation dispersion show a global k_{ex} for WT *Geo*Rec2 of 147 ± 41 s^{-1}. The K267E mutation shifts the globally fitted k_{ex} to 376 ± 89 s^{-1}, while the R332A variant maintains a k_{ex} similar to that of WT *Geo*Rec2 (142 ± 28 s^{-1}). The global fit of the K267E variant is based on CPMG profiles of 33 residues, while that of R332A is derived from 18 residues (**Table S1**). Interestingly, the sites participating in global motions of both variants are distinct from those of WT *Geo*Rec2, demonstrating that residue-specific flexibility is redistributed throughout *Geo*Rec2, affecting intradomain molecular crosstalk within *Geo*Rec. Indeed, perturbation to NMR-detectable motions in *Sp*Cas9 rewired its allosteric signaling and enzymatic function.[7,22] A similar dynamic modulation of *Geo*Cas9 may fine-tune its DNA cleavage, which has been demonstrated within the *Geo*HNH nuclease[29] and wedge (WED) domains.[26] We also assessed the ps-ns fluctuations of *Geo*Rec2 variants (a negligible contribution to the WT *Geo*Rec2 dynamic profile) and calculated order parameters from R_1, R_2 and ^{1}H-[^{15}N] NOE relaxation measurements (**Figure 3A, S8**). Bond vector fluctuations on the ps-ns timescale are only locally altered, thus the mutation-induced reshuffling of these dynamics is negligible ($\langle \Delta S^2 \rangle \leq 0.1$) and suggests that, like WT *Geo*Rec2, ps-ns motion arises primarily from global tumbling in solution.

Mutations within GeoRec diminish its affinity for guide RNA

To understand how the K267E and R332A mutants impact gRNA binding to *Geo*Rec, we conducted gRNA titration experiments via NMR and observed that gRNA-induced chemical shift perturbations were attenuated in both variants, relative to WT *Geo*Rec. Despite this muted structural effect, the impact from gRNA-induced line broadening remains substantial in the *Geo*Rec1 subdomain. Our NMR data reveal that a three-fold greater concentration of gRNA is required to induce the maximal structural and dynamic effects in the variants than is required for

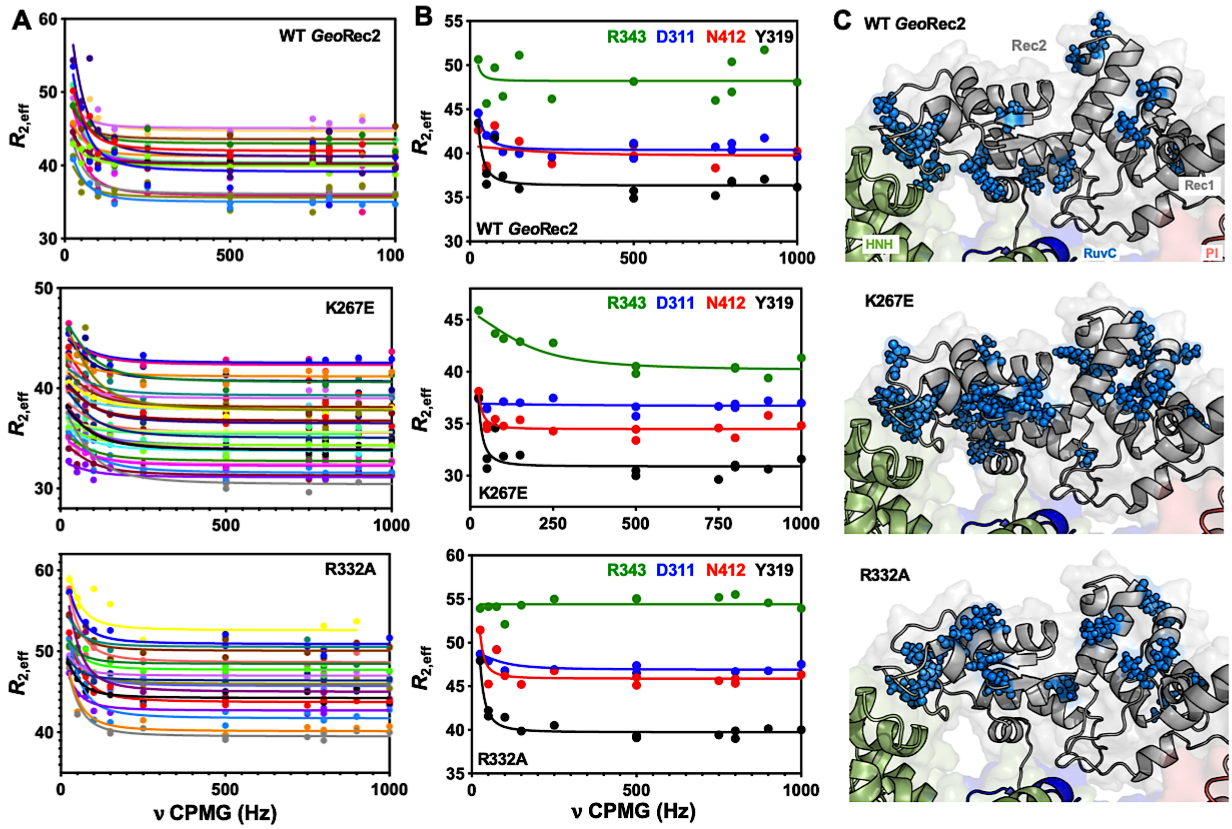

Figure 3. (A) CPMG relaxation dispersion profiles for all sites of µs-ms flexibility, fit to a global k_{ex} of 147 ± 41 s^{-1} (WT *Geo*Rec2, top), 376 ± 89 s^{-1} (K267E *Geo*Rec2, middle), and 142 ± 28 s^{-1} (R332A *Geo*Rec2, bottom). Relaxation dispersion profiles are colored according to **Table S1**. **B)** CPMG relaxation dispersion profiles for representative residues in *Geo*Rec2. A mixture of curved and flat profiles illustrate the changes in and redistribution of µs-ms motions between WT and *Geo*Rec2 variants. **(C)** All sites exhibiting CPMG relaxation dispersion plotted in A are mapped to *Geo*Rec as blue spheres. Adjacent domains within the AlphaFold2 model of *Geo*Cas9 are also shown.

WT *Geo*Rec **(Figure 4A-C**), suggesting that the variants have a reduced gRNA affinity. MST experiments showed modest reductions in gRNA affinity for the K267E and R332A constructs, relative to WT *Geo*Rec (K_d = 2.95 ± 0.53 µM), where K267E *Geo*Rec produced a K_d = 6.67 ± 2.0 µM and R332A *Geo*Rec produced a K_d = 4.02 ± 1.34 µM (**Figure 4E**). Collectively, these data reveal that mutations within *Geo*Rec only subtly alter its structure, but more significantly impact protein dynamics and in turn, the gRNA interaction. NMR experiments also demonstrate that the presence of gRNA impacts both subdomains of *Geo*Rec, providing an avenue for additional molecular tuning.

To more completely understand the mechanism of gRNA binding to *Geo*Rec, we carefully examined the NMR data to identify any residues critical for gRNA binding. When overlaying the gRNA-induced chemical shift perturbations (Δδ) of WT and *Geo*Rec variants (**Figure S9A**), it became clearer that the effect of gRNA binding to both variants was muted and that even at saturating concentrations, the chemical shift perturbations of identical resonances in the K267E and R332A *Geo*Rec2 spectra were weaker than those site in the WT *Geo*Rec spectrum. Plots of the change in Δδ (WT - mutant) indicate site-specifically that residues in the variants are less structurally affected by gRNA binding, noted as positive values in **Figure S9B**. Negative values

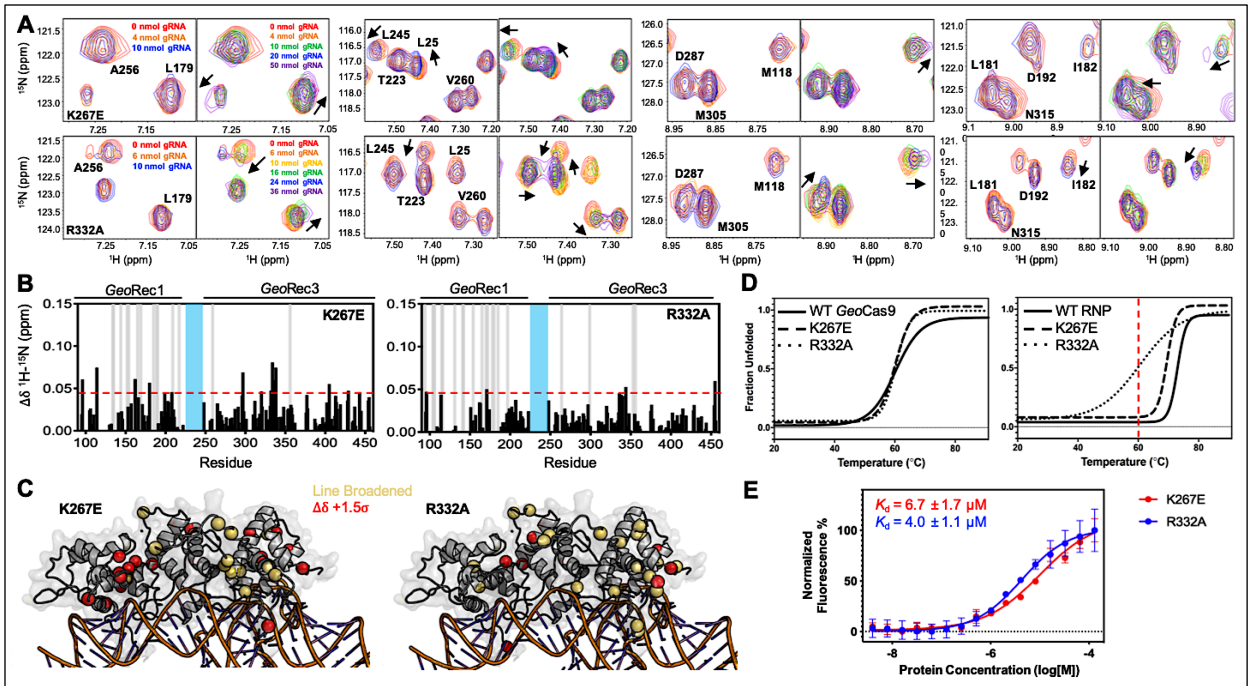

Figure 4. (A) NMR titration of a 39-nt gRNA into K267E (top) and R332A (bottom) *Geo*Rec. The left panel of each pair demonstrates that gRNA concentrations mimicking the WT titration induce minimal change in NMR chemical shift or resonance intensity. The right panel of each pair depicts the titration over a three-fold greater concentration range of gRNA, where shifts and line broadening are visible. Representative resonances are colored by increasing gRNA concentration according to the legend. **(B)** NMR chemical shift perturbations caused by gRNA binding to K267E and R332A *Geo*Rec. Gray bars denote sites of line broadening and the blue bar denotes an unassigned region of *Geo*Rec corresponding to the native Rec1-Rec2 linker. The red dashed line indicates 1.5σ above the 10% trimmed mean of all shifts. **(C)** Chemical shift perturbations >1.5σ are mapped onto *Geo*Rec (red spheres). Resonances that have broadened beyond detection are mapped as yellow spheres. **(D)** Fitted thermal denaturation profiles derived from CD spectra of apo WT (black line), K267E (dashed line), and R332A (dotted line) *Geo*Cas9 are shown on the left. Fitted denaturation profiles of the same proteins, as an RNP in complex with 141-nt gRNA, are shown in the right panel. The red dashed line denotes the T_m of the apo proteins, which are nearly identical. **(E)** MST-derived binding affinities of K267E and R332A *Geo*Rec and a Cy5-labeled 39-nt gRNA yielding K_d = 6.67 ± 2.0 μM and K_d = 4.02 ± 1.34 μM, respectively.

in the same plot highlight residues in the variants that experience greater structural changes than WT *Geo*Rec during gRNA binding. Residues with comparative change in Δδ exceeding one standard deviation above the 10% trimmed mean of all data were identified to represent key residues influencing gRNA affinity. These sites were mapped onto the *Geo*Rec structure (**Figure S9C**) and considered to be allosteric hotspots, as many are distal to the gRNA binding interface. Notably, residues critical for the tight association of *Geo*Rec and gRNA (positive change in Δδ) largely overlap when comparing WT *Geo*Rec to either variant. These sites include F170, R192, H264, R269, L270, L279, H300, D301, D376, E368, D403, E405, E408, and I429. Mutations of these residues in future work will confirm their role as hotspots for precisely tuning the affinity of *Geo*Rec for gRNA.

Having observed a weakened affinity of *Geo*Rec variants for gRNA by NMR and MST, we next quantified the impact of the K267E and R332A mutations on RNP formation and stability in full-length *Geo*Cas9. The thermal unfolding midpoint of full-length WT *Geo*Cas9 determined by CD is ~60 °C and the K267E and R332A mutations do not change the T_m of the apo protein (**Figure S10**). Upon formation of an RNP, the T_m of WT *Geo*Cas9 increases to 73 °C, but the K267E and R332A *Geo*Cas9 variants form less stable RNPs with T_m values of 70 °C and 61 °C, respectively. The trend of these data is consistent with NMR and MST, which highlight that although K267E and R332A mutations within *Geo*Rec have somewhat muted structural effects, these changes alter protein dynamics and the interaction with gRNA.

DNA cleavage assays suggest GeoCas9 is resistant to K267E- or R332A-induced functional changes

The dynamic impact of the *Geo*Rec mutations and their altered gRNA interactions at the biophysical level led us to speculate that when either mutation was incorporated into full-length

*Geo*Cas9, they would also alter DNA cleavage function, especially at elevated temperatures where WT *Geo*Cas9 is most active. Temperature-dependent functional alterations were previously observed for single-point mutations within *Geo*HNH.[29] Although the K267E and R332A mutations slightly diminished on-target DNA cleavage by *Geo*Cas9, the effect was subtle and these overall cleavage activities followed the temperature dependence of WT *Geo*Cas9 quite closely (**Figure S11**).

To assess the impact of these mutations on *Geo*Cas9 specificity, we assayed the propensity for off-target cleavage using DNA substrates with mismatches 5-6 or 19-20 base pairs from the PAM seed site (**Figure S12A, Table S2**). As a control for on- and off-target activity, we also assayed WT *Sp*Cas9 alongside the widely used high-specificity HiFi-*Sp*Cas9 variant[17] (**Figure S12B, Table S3**) and found a lower percent of digested off-target (mismatched) DNA sequences when compared to WT *Sp*Cas9. As expected, WT *Geo*Cas9 was increasingly sensitive to mismatched target sequences closer to the seed site, which has been demonstrated with *Sp*Cas9 and other Cas systems.[3,4,31,32] No significant differences in activity were observed with digestion durations ranging from 1-60 minutes,[4] implying that a 1-minute digestion is sufficient for *in vitro* activity of *Geo*Cas9 with the target DNA template. While these findings generally align with prior investigations of off-target DNA cleavage,[3,4,31] there are nuanced differences. Specifically, a previous study reported ~10% cleavage of off-target DNA with a mismatch 5-6 base pairs from the PAM by WT *Geo*Cas9.[4] Our results showed nearly 50% cleavage for the same off-target mismatch, but still a significant decrease in cleavage from on-target or 19-20 base pair distal mismatches. This could be due to the relatively high RNP concentrations (600-900 nM) in our assay (for clear visibility on the gel), compared to prior studies with RNP concentrations ≤ 500 nM.[4] Our results corresponded closely to those of prior studies with a 19-20 base pair mismatch,

where off-target cleavage is tolerated by WT *Geo*Cas9.[4] Single-point mutants K267E and R332A *Geo*Cas9 have negligible impact on *Geo*Cas9 specificity (both variants follow the trend of WT *Geo*Cas9, **Figure S12A**), which contrasts prior work with *Sp*Cas9 that demonstrated robust specificity enhancement with single-point mutations in Rec.[16,17] However, the *Geo*Cas9 double mutant K267E/R332A modestly enhances its specificity by reducing cleavage proximal and distal to the PAM seed site. The concomitant reduction in on-target cleavage activity of K267E/R332A has also been noted in other Cas systems.[16-18,28,33-35]

Discussion

CRISPR-Cas9 is a powerful tool for targeted genome editing with high efficiency and modular specificity.[4,16-18] Allosteric signals propagate DNA binding information to the HNH and RuvC nuclease domains, facilitating their concerted cleavage of double-stranded DNA.[3,18,20,30] The intrinsic flexibility of the nucleic acid recognition lobe plays a critical role in this information transfer, exerting a measure of conformational control over catalysis.[20] This study provides new insights into the structural, dynamic, and functional role of the thermophilic *Geo*Cas9 recognition lobe. Novel constructs of subdomains *Geo*Rec1 and *Geo*Rec2, as well as intact *Geo*Rec show a high structural similarity to the domains in full-length *Geo*Cas9, facilitating solution NMR experiments that captured the intrinsic allosteric motions across *Geo*Rec2. These studies revealed the existence of μs-ms timescale motions that are classically associated with allosteric signaling and enzyme function, which span the entire *Geo*Rec2 subdomain to its interfaces with *Geo*Rec1 and the adjacent *Geo*HNH.

Based on homology to specificity-enhancing variants of the better studied *Sp*Cas9, the biophysical and biochemical consequences of two mutations were tested in *Geo*Rec2, the larger *Geo*Rec lobe, and full-length *Geo*Cas9. We speculated that removing positively charged residues

with potential to interact with negatively charged nucleic acids could disrupt *Geo*Cas9-gRNA complex formation, stability, and subsequent function by altering the protein or nucleic acid motions. Indeed, CPMG relaxation dispersion experiments revealed that mutations enhanced and reorganized the μs-ms flexibility of *Geo*Rec2. Further, NMR titrations showed the affinity of K267E and R332A *Geo*Rec for gRNA to be weaker than that of WT *Geo*Rec, consistent with MST-derived K_d values. The mutations also diminished the stability of the full-length *Geo*Cas9 RNP complex.

The collective changes to protein dynamics, gRNA binding, and RNP thermostability suggested that mutations could modulate *Geo*Cas9 function, as observed in similar studies of *Sp*Cas9 reporting that gRNA dynamics, affecting the potential for the RNA:DNA hybrid to dissociate, have affected function.[18,19,36] Yet, the functional impact of the single-point mutations in this work was relatively modest, suggesting that multiple additive (or synergistic) mutations within *Geo*Rec are required to fine-tune activity or specificity to a large degree. Although homologous K-to-E and R-to-A single point mutations enhance specificity of the mesophilic *Sp*Cas9, their impact in the thermophilic *Geo*Cas9 system may be muted due to its evolutionary resilience.[37,38] Further, the biophysical impact of the mutations within *Geo*Rec2 and *Geo*Rec may be tempered by the unusually stable neighboring domains in the context of full-length *Geo*Cas9.

It should be noted that the effects of these and other *Geo*Rec mutations may vary *in vivo* or with alternative target cleavage sites and cell types. Such studies will be the subject of future work, as will biochemical assays of homologous mutations across diverse Cas9s, which have contributed to the wide use of CRISPR technology.[39] We also note that despite the homology between *Geo*Rec2 and *Sp*Rec3 and the latter's role in evo- and HiFi-*Sp*Cas9 variants that guided our K267E and R332A mutations, the maximally enhanced *Sp*Cas9 variants contain four mutations

each. Presumably, the individual substitutions all play some role modulating specificity. However, there is no consistent pattern that discerns whether multiple mutations will have additive or synergistic impacts on Cas9 function. NMR and MD studies of high-specificity *Sp*Cas9 variants (HF-1, Hypa, and Evo, each with distinct mutations in the *Sp*Rec3 subdomain) reveal universal structural and dynamic variations in regions of *Sp*Rec3 that interface with the RNA;DNA hybrid. Notably, a recently published variant, i*Geo*Cas9,[26] demonstrated enhanced genome-editing capabilities in HEK293T cells with eight mutations, though none in the Rec2 subdomain. This study highlighted the functional adaptability of i*Geo*Cas9 under low magnesium conditions, a trait beneficial in mammalian cells, distinguishing it from WT *Geo*Cas9. These very recently published data, as well as the findings reported here, advance our molecular understanding of *Geo*Cas9 as a promising avenue for other enhanced variants.

This study marks the first phase of mapping allosteric motions and pathways of information flow in the *Geo*Rec lobe with solution NMR experiments. Such information transfer is critical to the crosstalk between Rec and HNH in several Cas9s. Despite NMR advancements in perdeuteration,[40] transverse relaxation-optimized spectroscopy (TROSY),[41] and sparse isotopic labeling,[42] per-residue dynamics underlying allosteric signaling in large multi-domain proteins such as *Geo*Cas9 (~126 kDa) have remained challenging to characterize. A novel cryo-EM structure of *Geo*Cas9[26] will facilitate the merging of future NMR studies with molecular dynamics simulations that we have previously used to report on RNP dynamics and atomic level networks of communication. The identification of additional (or synergistic) allosteric hotspots within *Geo*Rec using an integrated workflow will help to further resolve the balance between structural stability and flexibility in *Geo*Cas9, leading to new insight into targeted manipulation of RNA affinity and enhanced variants.

Materials and Methods

Expression and purification of *Geo*Rec1, *Geo*Rec2, *Geo*Rec, and *Geo*Cas9

The Rec1 (residues 90-225) and Rec2 (residues 245-456) subdomains, as well as the entire Rec lobe (residues 90-456) of *G. stearothermophilus* Cas9 were engineered into a pET28a vector with a N-terminal His_6-tag and a TEV protease cleavage site. The K267E and R332A mutations were separately introduced into the *Geo*Rec2 plasmid. Plasmids were transformed into BL21 (DE3) cells (New England Biolabs). Protein samples for CD spectroscopy, MST, and functional assays were grown in Lysogeny Broth (LB, Fisher), while isotopically labeled samples for NMR were grown in M9 minimal media (deuterated for *Geo*Rec2 and *Geo*Rec) containing $CaCl_2$, $MgSO_4$, MEM vitamins, and 1.0 g/L ^{15}N ammonium chloride and 2.0 g/L ^{13}C glucose (Cambridge Isotope Laboratories), as the sole nitrogen and carbon sources, respectively. Cells were induced with 1 mM IPTG after reaching an OD_{600} of 0.8−1.0 and grown for 4 hours at 37 °C post induction. The cells were harvested by centrifugation, resuspended in a buffer of 50 mM Tris-HCl, 250 mM NaCl, 5 mM imidazole, and 1 mM PMSF at pH 7.4, lysed by ultrasonication, and purified by Ni−NTA affinity chromatography. Following TEV proteolysis of the terminal His-tag, the samples were further purified on a Superdex75 size exclusion column. NMR samples were dialyzed into a buffer containing 20 mM NaPi, 80 mM KCl, 1 mM DTT, and 1 mM EDTA at pH 7.4.

The full-length *Geo*Cas9 plasmid was acquired from Addgene (#87700), expressed in LB media and purified as described above. The K267E, R332A, and K267E/R332A variants were introduced into full-length *Geo*Cas9 by modifying the original plasmid acquired from Addgene.

NMR spectroscopy

Backbone resonance assignments of *Geo*Rec1 and *Geo*Rec2 were carried out on a Bruker Avance NEO 600 MHz spectrometer at 25 °C. The following triple resonance experiments were

collected for each sample: ^{1}H-^{15}N TROSY-HSQC, HNCA, HN(CO)CA, HN(CA)CB, HN(COCA)CB, HN(CA)CO and HNCO. All spectra were processed in NMRPipe[43] and analyzed in Sparky[44]. Three-dimensional correlations and assignments were made in CARA[45] and *Geo*Rec1 and *Geo*Rec2 backbone assignments were deposited in the BMRB under accession numbers 52363 and 51197, respectively.

NMR spin relaxation experiments were carried out at 600 and 850 MHz on Bruker Avance NEO and Avance III HD spectrometers, respectively. CPMG experiments were adapted from the report of Palmer and coworkers[46] with a constant relaxation period of 20 ms and ν_{CPMG} values of 0, 25, 50 (×2), 75, 100, 150, 250, 500 (×2), 750, 800 (×2), 900, and 1000 Hz. Exchange parameters were obtained from global fits of the data carried out with RELAX[47] using the R2eff, NoRex, and CR72 models, as well as in-house fitting in GraphPad Prism via:

$$R_{2,eff} = R_2^0 + R_{ex} * (1 - \left(\frac{2 * x * \tan h \left({k_{ex}}/{\tau_{cp}{}^2} \right)}{k_{ex}} \right))$$

where R_{ex} is

$$R_{ex} = \frac{p_a p_b \Delta\omega^2}{k_{ex}}$$

Error was determined with replicate measurements.

Longitudinal and transverse relaxation rates were measured with relaxation times of 20 (x2), 60 (x2), 100, 200, 600(x2), 800, and 1200 ms for T_1 and 16.96, 33.92(x2), 50.88(x2), 67.84(x2), 84.8, and 101.76(x2) ms for T_2. Peak intensities were quantified in Sparky and the resulting decay profiles were analyzed in Sparky with errors determined from the fitted parameters. Steady-state ^{1}H-[^{15}N] NOE were measured with a 6 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 NMR data in RELAX[47] with fully automated protocols.

NMR nucleic acid titrations were performed on a Bruker Avance NEO 600 MHz spectrometer at 25 °C by collecting a series of ^{1}H-^{15}N TROSY HSQC spectra with increasing gRNA concentration. The ^{1}H and ^{15}N carrier frequencies were set to the water resonance and 120 ppm, respectively. Samples of WT, K267E, and R332A *Geo*Rec were titrated with gRNA until no further spectral perturbations were detected. NMR chemical shift perturbations were calculated as:

$$\Delta\delta = \sqrt{(\Delta\delta_{HN}^2 + \Delta\delta_{NH}^2/25)/2}$$

Microscale thermophoresis (MST)

MST experiments were performed on a Monolith X instrument (NanoTemper Technologies), quantifying WT, K267E, R332A, and K267E/R332A *Geo*Rec binding to a 39-nt Cy5-labeled gRNA at a concentration of 20 nM in a buffer containing 20 mM sodium phosphate, 150 mM KCl, 5 mM $MgCl_2$, and 0.1% Triton X-100 at pH 7.6. The *Geo*Rec proteins were serially diluted from a 200 μM stock into 16 microcentrifuge tubes and combined in a 1:1 molar ratio with serially diluted gRNA from a 40 nM stock. After incubation for 5 minutes at 37 °C in the dark, each sample was loaded into a capillary for measurement. K_d values for the various complexes were calculated using the MO Control software (NanoTemper Technologies).

Circular dichroism (CD) spectroscopy

All *Geo*Cas9 and *Geo*Rec proteins were buffer exchanged into a 20 mM sodium phosphate buffer at pH 7.5, diluted to 1 μM, and loaded into a 2 mm quartz cuvette (JASCO instruments). A CD spectrum was first measured between 200 - 250 nm, after which the sample was progressively heated from 20 – 90 °C in 1.0 °C increments while ellipticity was monitored at 222 and 208 nm. Phosphate buffer baseline spectra were subtracted from the sample measurements. Prior to CD

measurements, *Geo*Cas9-RNP was formed by incubating 3 μM *Geo*Cas9 with gRNA at a 1:1.5 molar ratio at 37 °C for 10 minutes.

X-ray crystallography

*Geo*Rec protein purified as described above was crystallized by sitting drop vapor diffusion at room temperature by mixing 1.0 μL of 15 mg/mL *Geo*Rec in a buffer of 20 mM HEPES and 100 mM KCl at pH 7.5 with 2.0 μL of crystallizing condition: 0.15 M calcium chloride, 15 % polyethylene glycol 6000, 0.1 M HEPES at pH 7.0. Crystals were cryoprotected in crystallizing condition supplemented with 30% ethylene glycol. Diffraction images were collected at the NSLS-II AMX beamline at Brookhaven National Laboratory under cryogenic conditions. Images were processed using XDS[48] and Aimless in CCP4.[49] Chain A of the *N. meningitidis* Cas9 X-ray structure (residues 249-445 only, PDB ID: 6JDQ) was used for molecular replacement with Phaser followed by AutoBuild in Phenix.[50] Electron density was only observed for the *Geo*Rec2 subdomain. The *Geo*Rec2 structure was finalized through manual building in Coot[51] and refinement in Phenix.

DNA cleavage assays

*Geo*Cas9 gRNA templates containing 21-nt, 23-nt, or 25-nt spacers targeting the mouse *Tnnt2* gene locus were introduced into EcoRI and BamHI sites in pUC57 (Genscript). The plasmid was transformed into BL21(DE3) cells (New England BioLabs) and subsequent restriction digest of the plasmid DNA was carried out using the BamHI restriction enzyme (New England BioLabs) according to the manufacturer's instructions. Linearized plasmid DNA was immediately purified using the DNA Clean and Concentrator-5 kit (Zymo Research) according to the manufacturer's instructions. RNA transcription was performed *in vitro* with the HiScribe T7 High Yield RNA Synthesis Kit (New England BioLabs). DNA substrates containing the target cleavage site (479

base pairs, **Table S2, S3**) were produced by polymerase chain reaction (PCR) using mouse genomic DNA as a template and primer pairs 5'CAAAGAGCTCCTCGTCCAGT3' and 5' ATGGACTCCAGGACCCAAGA3' followed by a column purification using the NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel). For the *in vitro* cleavage assay, RNP formation was achieved by incubating 3 μM *Geo*Cas9 (WT, K267E, R332A, or K267E/R332A variant) and 3 μM gRNA at either 37 °C, 60 °C, 75 °C, or 85 °C for 30 minutes in a reaction buffer of 20 mM Tris, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, and 5% glycerol at pH 7.5. The 10 μL cleavage reactions were set up by mixing RNP at varying concentrations with 149 nanograms of PCR products on ice followed by incubation at 37 °C for 30 minutes. The reaction was quenched by adding 1 μL of proteinase K (20 mg/mL) and subsequent incubation at 56 °C for 10 minutes. Two microliters of 6x DNA loading buffer were added to each reaction and 10 μL of the mixture per lane was loaded onto an 1% agarose gel. DNA band intensity measurements on a gel image were carried out with ImageJ.

For *in vitro* off-target cleavage assays, RNP formation was achieved by incubating 10 μM *Geo*Cas9 (WT, K267E, R332A, or K267E/R332A variant) and 10 μM gRNA at 37 °C for 30 minutes in the reaction buffer described above. The 10 μL cleavage reactions were set up by mixing 1 μM RNP with 150 nanograms of PCR products (off-target DNA sequences listed in **Table S2**) on ice followed by incubation at 37 °C for varying time points. The reaction was quenched and visualized as described above. WT and HiFi *Sp*Cas9 control proteins were purchased from Integrated DNA Technologies (IDT, cat. No. 108158 and No. 108160, respectively), as was the associated *Sp*Cas9 gRNA, Alt-R™ CRISPR-Cas9 gRNA, with an RNA spacer sequence complementing 5'- TGGACAGAGCCTTCTTCTTC-3'. The on-target and off-target DNA sequences used for the *Sp*Cas9 *in vitro* cleavage assay can be found in **Table S3**.

Author Contributions

HBB and **ALK** produced *Geo*Rec1, *Geo*Rec2, *Geo*Rec, and *Geo*Cas9 proteins and gRNA, conducted the NMR and biophysical experiments, analyzed the data, and wrote the original draft of the manuscript. **AMD** solved the X-ray crystal structure of *Geo*Rec2. **ZF** and **JL** conducted *Geo*Cas9 functional assays and analyzed the data. **GJ** supervised collection of X-ray crystallographic data. **GPL** conceived the study, supervised collection of NMR spectroscopic data, obtained funding, and wrote the original draft. The final manuscript was written and edited with contributions from all authors.

Acknowledgments

This work was supported by NIH grant R01 GM 136815 and NSF grant MCB 2143760 (to GPL). This research used the AMX beamline of the National Synchrotron Light Source II, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-SC0012704. The Center for BioMolecular Structure (CBMS) is primarily supported by the NIH (NIGMS) through a Center Core P30 Grant (P30GM133893), and by the DOE Office of Biological and Environmental Research (KP1607011).

References

1 Mir, A., Edraki, A., Lee, J. & Sontheimer, E. J. Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. *ACS Chem Biol* **13**, 357-365, doi:10.1021/acschembio.7b00855 (2018).

2 Belato, H. B. *et al.* Structural and dynamic insights into the HNH nuclease of divergent Cas9 species. *J Struct Biol* **214**, 107814, doi:10.1016/j.jsb.2021.107814 (2021).

3 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

4 Harrington, L. B. *et al.* A thermostable Cas9 with increased lifetime in human plasma. *Nat Commun* **8**, 1424, doi:10.1038/s41467-017-01408-4 (2017).

5 East, K. W. *et al.* Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J Am Chem Soc*, doi:10.1021/jacs.9b10521 (2020).

6 Wu, Z., Yang, H. & Colosi, P. Effect of genome size on AAV vector packaging. *Mol Ther* **18**, 80-86, doi:10.1038/mt.2009.255 (2010).
7 Skeens, E. *et al.* High-fidelity, hyper-accurate, and evolved mutants rewire atomic-level communication in CRISPR-Cas9. *Sci Adv* **10**, eadl1045, doi:10.1126/sciadv.adl1045 (2024).
8 Mukrasch, M. D. *et al.* Highly Populated Turn Conformations in Natively Unfolded Tau Protein Identified from Residual Dipolar Couplings and Molecular Simulation. *Journal of the American Chemical Society* **129**, 5235-5243, doi:10.1021/ja0690159 (2007).
9 Frederick, K. K., Marlow, M. S., Valentine, K. G. & Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **448**, 325-329, doi:10.1038/nature05959 (2007).
10 Salmon, L. *et al.* Multi-timescale conformational dynamics of the SH3 domain of CD2-associated protein using NMR spectroscopy and accelerated molecular dynamics. *Angew Chem Int Ed Engl* **51**, 6103-6106, doi:10.1002/anie.201202026 (2012).
11 Guzman, I., Ghaemi, Z., Baranger, A., Luthey-Schulten, Z. & Gruebele, M. Native Conformational Dynamics of the Spliceosomal U1A Protein. *The Journal of Physical Chemistry B* **119**, 3651-3661, doi:10.1021/jp511760m (2015).
12 Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278, doi:10.1016/j.cell.2014.05.010 (2014).
13 Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997, doi:10.1126/science.1247997 (2014).
14 Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573, doi:10.1038/nature13579 (2014).
15 Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096, doi:10.1126/science.1258096 (2014).
16 Casini, A. *et al.* A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol* **36**, 265-271, doi:10.1038/nbt.4066 (2018).
17 Vakulskas, C. A. *et al.* A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med* **24**, 1216-1224, doi:10.1038/s41591-018-0137-0 (2018).
18 Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407-410, doi:10.1038/nature24268 (2017).
19 Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci Adv* **3**, eaao0027, doi:10.1126/sciadv.aao0027 (2017).
20 Palermo, G. *et al.* Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. *Q Rev Biophys* **51**, doi:10.1017/s0033583518000070 (2018).
21 Ricci, C. G. *et al.* Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent Sci* **5**, 651-662, doi:10.1021/acscentsci.9b00020 (2019).
22 Nierzwicki, L. *et al.* Enhanced specificity mutations perturb allosteric signaling in CRISPR-Cas9. *Elife* **10**, doi:10.7554/eLife.73601 (2021).
23 Nierzwicki, Ł. *et al.* Principles of target DNA cleavage and the role of Mg2+ in the catalysis of CRISPR–Cas9. *Nature Catalysis* **5**, 912-922, doi:10.1038/s41929-022-00848-6 (2022).

24 Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477-1481, doi:10.1126/science.aab1452 (2015).
25 Palermo, G. *et al.* Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. *J Am Chem Soc* **139**, 16028-16031, doi:10.1021/jacs.7b05313 (2017).
26 Eggers, A. R. *et al.* Rapid DNA unwinding accelerates genome editing by engineered CRISPR-Cas9. *bioRxiv*, 2023.2012.2014.571777, doi:10.1101/2023.12.14.571777 (2023).
27 Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature* **578**, 229-236, doi:10.1038/s41586-020-1978-5 (2020).
28 Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-88, doi:10.1126/science.aad5227 (2016).
29 Belato, H. B. *et al.* Disruption of electrostatic contacts in the HNH nuclease from a thermophilic Cas9 rewires allosteric motions and enhances high-temperature DNA cleavage. *The Journal of Chemical Physics* **157**, 225103, doi:10.1063/5.0128815 (2022).
30 Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110-113, doi:10.1038/nature15544 (2015).
31 Lee, C. M., Cradick, T. J. & Bao, G. The Neisseria meningitidis CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells. *Mol Ther* **24**, 645-654, doi:10.1038/mt.2016.8 (2016).
32 Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proc Natl Acad Sci U S A* **110**, 15644-15649, doi:10.1073/pnas.1313587110 (2013).
33 Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495, doi:10.1038/nature16526 (2016).
34 Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63, doi:10.1038/nature26155 (2018).
35 Lee, J. K. *et al.* Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun* **9**, 3048, doi:10.1038/s41467-018-05477-x (2018).
36 Ma, E., Harrington, L. B., O'Connell, M. R., Zhou, K. & Doudna, J. A. Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol Cell* **60**, 398-407, doi:10.1016/j.molcel.2015.10.030 (2015).
37 Nguyen, V. *et al.* Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science* **355**, 289-294, doi:10.1126/science.aah3717 (2017).
38 Saavedra, H. G., Wrabl, J. O., Anderson, J. A., Li, J. & Hilser, V. J. Dynamic allostery can drive cold adaptation in enzymes. *Nature* **558**, 324-328, doi:10.1038/s41586-018-0183-2 (2018).
39 Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* **15**, 169-182, doi:10.1038/nrmicro.2016.184 (2017).
40 Venters, R. A., Farmer, B. T., 2nd, Fierke, C. A. & Spicer, L. D. Characterizing the use of perdeuteration in NMR studies of large proteins: 13C, 15N and 1H assignments of human carbonic anhydrase II. *J Mol Biol* **264**, 1101-1116, doi:10.1006/jmbi.1996.0699 (1996).
41 Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**, 12366-12371, doi:10.1073/pnas.94.23.12366 (1997).

42 Tugarinov, V., Kanelis, V. & Kay, L. E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc* **1**, 749-754, doi:10.1038/nprot.2006.101 (2006).
43 Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293, doi:10.1007/bf00197809 (1995).
44 Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325-1327, doi:10.1093/bioinformatics/btu830 (2015).
45 Keller, R. L. J. *Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment* Doctoral Thesis thesis, ETH Zürich, (2005).
46 Loria, J. P., Rance, M. & Palmer, A. G. A Relaxation-Compensated Carr−Purcell−Meiboom−Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *Journal of the American Chemical Society* **121**, 2331-2332, doi:10.1021/ja983961a (1999).
47 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 µs motion of proteins. *J Biomol NMR* **50**, 147-155, doi:10.1007/s10858-011-9509-1 (2011).
48 Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, doi:10.1107/s0907444909047337 (2010).
49 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242, doi:10.1107/s0907444910045749 (2011).
50 Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877, doi:10.1107/s2059798319011471 (2019).
51 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/s0907444910007493 (2010).