BIOPHYSICS

High-fidelity, hyper-accurate, and evolved mutants rewire atomic-level communication in CRISPR-Cas9

Erin Skeens[1]†, Souvik Sinha[2]†, Mohd Ahsan[2], Alexandra M. D'Ordine[1], Gerwald Jogl[1], Giulia Palermo[2,3]*, George P. Lisi[1]*

The high-fidelity (HF1), hyper-accurate (Hypa), and evolved (Evo) variants of the CRISPR-associated protein 9 (Cas9) endonuclease are critical tools to mitigate off-target effects in the application of CRISPR-Cas9 technology. The mechanisms by which mutations in recognition subdomain 3 (Rec3) mediate specificity in these variants are poorly understood. Here, solution nuclear magnetic resonance and molecular dynamics simulations establish the structural and dynamic effects of high-specificity mutations in Rec3, and how they propagate the allosteric signal of Cas9. We reveal conserved structural changes and dynamic differences at regions of Rec3 that interface with the RNA:DNA hybrid, transducing chemical signals from Rec3 to the catalytic His-Asn-His (HNH) domain. The variants remodel the communication sourcing from the Rec3 α helix 37, previously shown to sense target DNA complementarity, either directly or allosterically. This mechanism increases communication between the DNA mismatch recognition helix and the HNH active site, shedding light on the structure and dynamics underlying Cas9 specificity and providing insight for future engineering principles.

INTRODUCTION

CRISPR-associated protein 9 (CRISPR-Cas9) is a revolutionary genome editing system widely adapted for bioengineering purposes (*1*). Its use as a therapeutic for human diseases, often regarded as the pinnacle of Cas9 application, is hampered by the occurrence of off-target effects (*2*). To overcome this limitation, directed evolution and extensive engineering of the Cas9 enzyme have been performed, leading to a set of promising variants that exhibit enhanced specificity toward target DNA (*3–5*). Many of these variants include mutations distal from the catalytic sites, though the mechanisms that confer specificity from long-range mutational effects remain unclear. Elucidation of these specificity-enhancing mechanisms is critical for the rational design and further improvement of Cas9 variants that mitigate off-target cleavage in biomedical applications.

The *Streptococcus pyogenes* Cas9 (SpCas9), the most broadly used Cas9 enzyme, has a multi-domain architecture consisting of two nuclease domains, a His-Asn-His endonuclease (called HNH) and a Holliday junction resolvase C (RuvC), a protospacer adjacent motif (PAM)-interacting (PI) domain, and a recognition lobe that mediates nucleic acid binding through three distinct subdomains, Rec1-3 (Fig. 1A) (*6*, *7*). Upon recognition of the short PAM sequence, Cas9 locally unwinds the double-stranded DNA, and a Cas9-bound guide RNA (gRNA) forms an RNA:DNA hybrid with the target DNA strand (TS) (*8*). Coordinated cleavage of the TS and non-target strand then occurs via the HNH and RuvC nucleases, respectively. Biophysical studies and molecular dynamics (MD) simulations have indicated that the molecular function of Cas9 is governed by an intricate allosteric response driven by intrinsic dynamics, controlling the activation of the catalytic HNH domain (*9*, *10*). The flexibility of HNH ensures DNA cleavage at the proper location (*11*, *12*), but its conformational dynamics were shown to be dependent on the motions of the spatially distant Rec3 domain (*3*, *9*, *13*, *14*), which was therefore regarded as an "allosteric effector" of HNH function (*3*). Nevertheless, the molecular details of this Rec3-mediated activation are unknown.

Efforts to engineer SpCas9 variants for enhanced specificity revealed that Rec3 is an important driver of off-target recognition. Many notable variants house the majority of their specificity-conferring mutations within the Rec3 domain, including the "high fidelity" Cas9-HF1 (*4*), the "hyper-accurate" HypaCas9 (*3*), and the "evolved" evoCas9 (*5*) variants, suggesting that Rec3 plays a critical role in proofreading the DNA. However, the mechanisms by which mutations in Rec3 mediate the specificity of Cas9, especially considering their distance from the catalytic sites, are poorly understood. Single-molecule experiments have revealed that the HF1 and Hypa mutations alter the conformational equilibrium that controls Cas9 activation (*3*) and that high-specificity mutations in Rec3 affect DNA unwinding and the RNA:DNA heteroduplexation process (*15–17*). Further atomic-level description of the mutation-induced Cas9 structure and dynamics at the level of Rec3 could contribute to discerning the specificity-enhancing molecular mechanism and the allosteric role of Rec3.
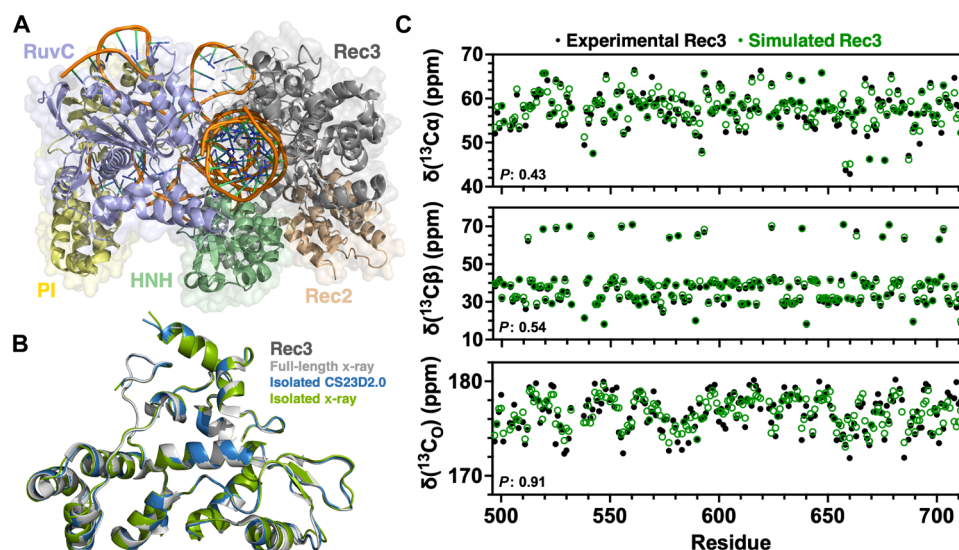
To understand the allosteric contributions of Rec3 to Cas9 specificity and function, we combined experimental and computational techniques to comprehensively characterize the structural and dynamic effects of high-specificity mutations on the Rec3 domain and more broadly, Cas9. We engineered a construct of the isolated Rec3 domain to experimentally probe its biophysical properties and the effect of the HF1, Hypa, and Evo mutations with atomic resolution through solution nuclear magnetic resonance (NMR). In parallel, multi-microsecond MD simulations of the full-length CRISPR-Cas9 system and its variants were used to describe how the allosteric signal propagates from Rec3 through the full complex. Our findings reveal conserved structural changes and altered dynamics within the high-specificity Rec3 variants, which rewire allosteric signaling in CRISPR-Cas9 and increase the communication between the DNA recognition region and the HNH catalytic core.

Fig. 1. The isolated Rec3 domain structure mirrors Rec3 from full-length Cas9. (A) Domain arrangement of the *Streptococcus pyogenes* Cas9, highlighting the RuvC (blue), PAM-interacting (PI; yellow), HNH (green), Rec3 (gray), and Rec2 (tan) domains, with the RNA and DNA displayed in orange. **(B)** The overlaid structures of Rec3 from the full-length Cas9 x-ray crystal structure [gray; Protein Data Bank (PDB): 4UN3] (7), our nuclear magnetic resonance (NMR)–derived CS23D2.0 model (blue), and the isolated Rec3 x-ray crystal structure (green; PDB: 8SCA), solved here at 1.67 Å, reveal remarkable structural consistency [i.e., Cα mean square deviation (RMSD) of 0.649 Å between Rec3 in PDB: 8SCA and in PDB: 4UN3; Cα RMSD of 0.598 Å between Rec3 in PDB: 8SCA and in the NMR-predicted model]. **(C)** Experimental (black dots) and simulated (green dots) chemical shifts of $^{13}C_\alpha$ (top), $^{13}C_\beta$ (center), and $^{13}C_O$ (bottom) are plotted for residues in the Rec3 domain. For each atom type, the estimated *P* values are in the 95% confidence interval, confirming the significance of the overlap between experimental and simulated chemical shift measurements.

RESULTS

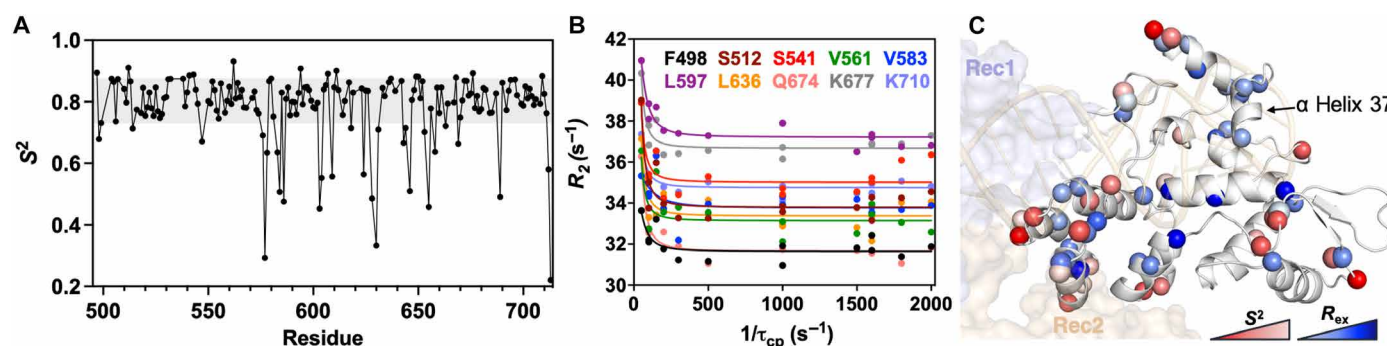### Structure and dynamics of the allosteric Rec3 domain

To probe the contribution of Rec3 to Cas9 function, x-ray crystallography and solution NMR were used to characterize the structure and dynamics of the isolated Rec3 domain, while MD simulations of the full-length CRISPR-Cas9 system were used to study Rec3 within the larger complex. The combination of these approaches can comprehensively describe allosteric mechanisms in large biomolecules (*18–23*). Solution NMR can assess the domain-specific dynamics with experimental accuracy, while MD simulations can interpret the dynamics in the context of the full-length assembly to understand the allosteric cross-talk and track the signal transmission.

For experimental studies of the Rec3 domain, an isolated construct of Rec3 (residues 497 to 713) was engineered. A structural model of the isolated Rec3 domain was predicted from the assigned ^1H-^{15}N heteronuclear single-quantum coherence (HSQC) NMR spectrum and carbon (Cα, Cβ, and C_O) chemical shifts using the CS23D2.0 server. Figure 1B shows an overlay of the CS23D2.0 model of the isolated Rec3 domain (blue) with Rec3 from the full-length Cas9 crystal structure [gray; Protein Data Bank (PDB): 4UN3]. The NMR-derived CS23D2.0 model is consistent with Rec3 from the Cas9 crystal structure. Secondary structure analysis of the isolated Rec3 domain from Cα/Cβ chemical shifts and circular dichroism (CD) spectroscopy supports a predominantly alpha helical structure (fig. S1, A and B), providing further indication that the engineered construct maintains the expected fold of the Rec3 domain within the full-length Cas9 protein.

We then determined the structure of the isolated Rec3 domain via x-ray crystallography (PDB: 8SCA at 1.67-Å resolution; Fig. 1B), which also aligns well with that of Rec3 from the Cas9 crystal structure and the NMR-predicted model, with a Cα root mean square deviation of 0.649 and 0.598 Å, respectively. There are no major

structural differences observed and minor differences are localized to flexible loops. This further confirms the strong NMR-predicted agreement between the isolated Rec3 and the domain within the full-length system. Multi-microsecond simulations were then performed on a full-length Cas9 complex, where an overall ensemble of ~16 μs (arising from four replicas of ~4 μs each) was used to compute Cα, Cβ, and C_O chemical shifts of Rec3 with SHIFTX2 (*24*) for comparison to those measured experimentally by NMR. We detected excellent agreement between the experimental and simulated values (Fig. 1C), which further supports that the isolated Rec3 construct and resulting NMR spectra are valuable for studying the structural and dynamic determinants of Rec3-mediated allostery and specificity (fig. S2). In turn, MD simulations of the full-length Cas9 properly represent the NMR data on the isolated Rec3, which offers the opportunity to study how allosteric information propagates from Rec3 to the rest of the system.

NMR relaxation experiments were conducted to probe the dynamics of Rec3 on timescales relevant to the transmission of allosteric signals. Fast molecular motions on the picosecond-nanosecond timescale were assessed via order parameters (S^2), calculated from the longitudinal (R_1) and transverse (R_2) relaxation rates and ^1H-[^{15}N] nuclear Overhauser effect (NOE; Fig. 2, A and C, and fig. S3) (*18*, *25*, *26*). For quantitation of slower dynamic processes on the microsecond-millisecond timescale, which are classically associated with the transmission of allosteric effects in many enzymes (*27*, *28*), including the HNH nuclease of Cas9 (*10*), Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments were performed (Fig. 2, B and C, and table S1). Residues in Rec3 with motions on the picosecond-nanosecond timescale are observed in intrinsically dynamic loops and on the face of the domain that does not interact with the RNA:DNA hybrid. Conversely, sites that exhibit microsecond-millisecond motions are generally localized to the RNA:DNA hybrid interface, with exchange

Fig. 2. Flexibility of the wild-type (WT) Rec3 domain. (**A**) Order parameters (S^2) of WT Rec3 reveal picosecond-nanosecond timescale motions. The gray shaded area denotes ±1.5σ from the 10% trimmed mean of the data, where order parameters below this area represent residues with picosecond-nanosecond motions. (**B**) A selection of residues in Rec3 exhibiting microsecond-millisecond timescale flexibility via representative Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion profiles. (**C**) Residues displaying picosecond-nanosecond motions via S^2 and CPMG-detected microsecond-millisecond motions are mapped onto the Rec3 structure, shown to display the face of the domain that interacts with the RNA:DNA hybrid. Red spheres represent picosecond-nanosecond motions (S^2), with a gradient denoting low S^2 values as dark red, signifying greater flexibility. Blue spheres denote microsecond-millisecond motions (CPMG), with a blue gradient that becomes darker as the exchange contribution to relaxation (R_{ex}) increases. Rec1, Rec2, and the RNA:DNA hybrid are shown for reference.

contributions to relaxation (R_{ex}) > 4, suggesting that the intrinsic flexibility of the Rec3 domain on the microsecond-millisecond timescale is important for facilitating its interaction with the RNA:DNA hybrid.

Evidence for a universal structural change in high-specificity Rec3

To probe whether structural and dynamic changes contribute to the enhanced specificity of the Cas9-HF1, HypaCas9, and evoCas9 variants, we introduced the specificity-conferring mutations into the isolated Rec3 domain to generate HF1 Rec3 (N497A/R661A/Q695A), Hypa Rec3 (N692A/M694A/Q695A/H698A), and Evo Rec3 (Y515N/K526E/R661Q). A ^{1}H-^{15}N HSQC spectral overlay of the Rec3 variants with wild-type (WT) Rec3 shows widespread chemical shift perturbations, indicative of local structural changes induced by the mutations (fig. S4), and line-broadened resonances, suggestive of a dynamic change. Analysis of the chemical shift perturbations, deemed statistically significant when 1.5σ above the 10% trimmed mean of all datasets, reveals a highly consistent pattern exhibited by the Rec3 variants (Fig. 3A). Three notable regions spanning residues 510 to 555, 576 to 595, and 678 to 698 are significantly perturbed and exhibit line broadening in all Rec3 variants, despite the disparate locations of the mutations within each variant. Residues exhibiting significant chemical shift perturbations in any single variant, any two variants, and all three variants are mapped onto the Rec3 structure in Fig. 3B, denoted by yellow, orange, and red spheres, respectively, to highlight residues that are most frequently perturbed by the mutations. The perturbations are largely present on the face of the Rec3 structure that interfaces with the RNA:DNA hybrid at the PAM distal ends. Nonspecific perturbations observed in only one of the three variants (Fig. 3B, yellow spheres) occur predominantly around the sites of mutation. Consistent patterns of chemical shift perturbation in all three variants face the RNA:DNA heteroduplex and are localized around a loop region in proximity to the PAM-distal end of the hybrid and distant from the neighboring domains. Such a structural change may alter points of contact between Rec3 and the RNA:DNA hybrid. Furthermore, this region is in close proximity to the α helix at residues 692 to 699 [α37 from PDB: 4UN3 (7)] that was previously shown to insert within the RNA:DNA hybrid upon off-target DNA binding, suggesting a possible proofreading
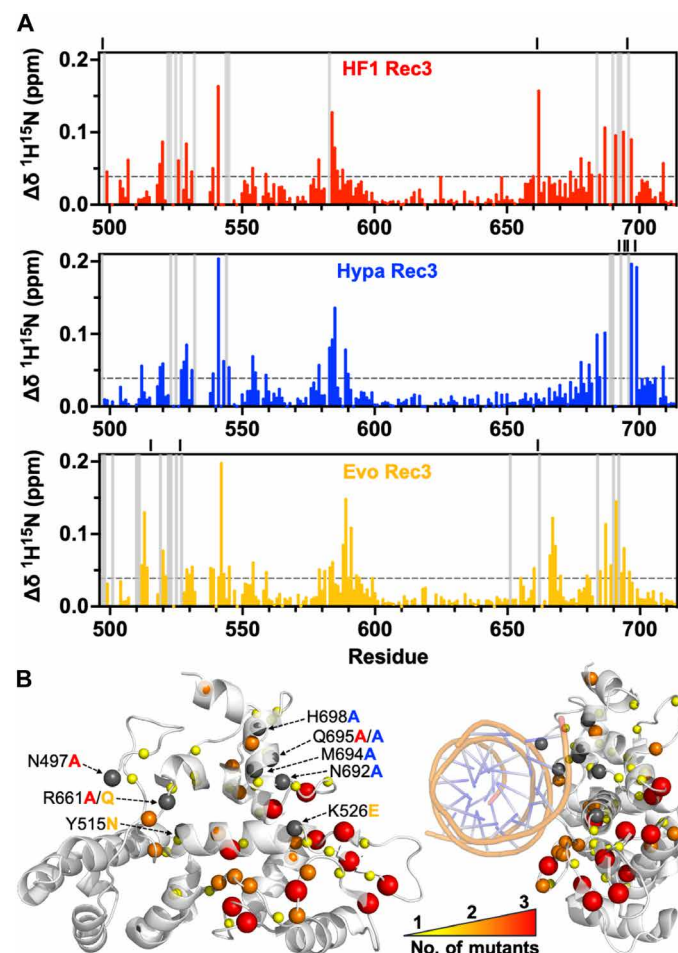
role in mismatch recognition (29). Altered hybrid contacts and perturbation of α37 may be critical for enhancing the specificity of these variants.

For a more comprehensive analysis of the structural changes induced by specificity-enhancing mutations in Rec3, the carbon chemical shifts for HF1 Rec3, Hypa Rec3, and Evo Rec3 were also determined by NMR. Predictions of the secondary structure via Cα/Cβ chemical shifts show a general conservation of secondary structural features among the variants, which is further supported by the strong overlap of CD spectra for WT Rec3 and the variants (fig. S1, A and B). ^{1}H-^{13}Cα/β/O spectral overlays of the variants with WT Rec3 reveal global spectral changes for each carbon position, indicating that the backbone and side chains are affected by specificity-enhancing mutations (fig. S5). These structural perturbations are consistent with thermal unfolding experiments that show modest changes to the structural stability of the variants (fig. S1C). Analysis of ^{13}Cα/β/O chemical shift perturbations shows similar trends across the Rec3 variants, with the strongest effects occurring in the loop region proximal to the PAM-distal end of the RNA:DNA hybrid, consistent with previously discussed ^{1}H-^{15}N chemical shift perturbations (Fig. 3 and fig. S6). Collectively, these findings suggest that the high-specificity mutations induce a universal Rec3 structural change at the PAM-distal ends of the RNA:DNA hybrid, which may promote the specificity-enhancing properties of the Cas9 variants.

Perturbation of the Rec3 dynamic landscape by high-specificity variants

Analysis of Rec3 dynamics in the high-specificity variants reveals that the variants exhibit differential flexibility on the fast and slow timescales, with an observed shift toward slower timescale motions (Fig. 4A and figs. S7 to S11). Decreased molecular motion on the picosecond-nanosecond timescale via S^2 is observed in all three variants, relative to WT Rec3 (Fig. 4B and fig. S10). CPMG relaxation dispersion experiments also reveal a change in quantitative microsecond-millisecond motions in the Rec3 variants (Fig. 4C, fig. S11, and tables S1 to S4). HF1 Rec3 exhibits fewer residues (−7 residues), Evo Rec3 exhibits slightly more residues (+4 residues), and Hypa Rec3 exhibits considerably more residues (+20 residues) with microsecond-millisecond flexibility than the WT Rec3. Regardless of the number of residues

Fig. 3. Chemical shift perturbations induced by high-specificity mutations reveal a structural signature in Rec3. (**A**) $^1H^{15}N$ combined chemical shift perturbations (Δδ) for HF1 Rec3 (red), Hypa Rec3 (blue), and Evo Rec3 (yellow). Gray vertical bars represent line-broadened resonances and gray dashed lines denote 1.5σ above the 10% trimmed mean of all datasets. Sites of mutation are indicated by black lines above each plot. (**B**) Significant $^1H^{-15}N$ chemical shift perturbations (+1.5σ) are mapped onto the Rec3 structure, with spheres denoting significant shifts observed in any one mutant (yellow), any two mutants (orange), and all three mutants (red). Sites of mutation are indicated by gray spheres and the mutated residues are labeled according to the color scheme in (A). The structure on the right shows the RNA:DNA hybrid for positional reference to the observed perturbations.
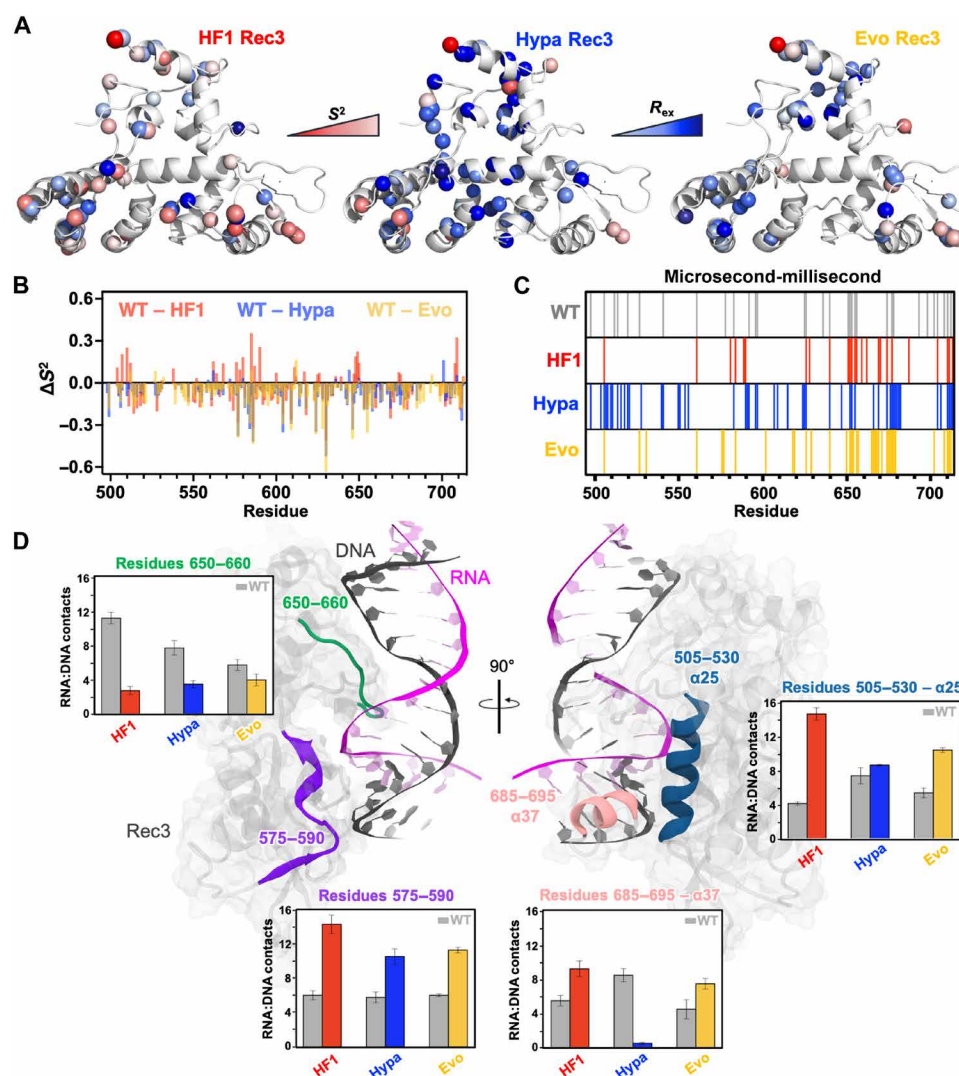
exhibiting microsecond-millisecond motions, ~50 to 60% of the sites identified in each variant are not found to undergo similar motions in WT Rec3, suggesting an overall rewiring of intradomain flexibility due to the specificity-enhancing mutations. In addition to a redistribution of slow timescale motions in Rec3, the residues that exhibit these motions in the variants have, on average, larger amplitudes to the CPMG profiles, reflecting a greater exchange contribution to relaxation (R_{ex}) than observed in WT Rec3 (Fig. 4A and tables S1 to S4). HF1 Rec3 and Evo Rec3 display clusters of heightened flexibility at the RNA:DNA hybrid interface, while Hypa Rec3 enhances microsecond-millisecond motion across the entire domain, including regions interfacing with the RNA:DNA hybrid. These data reveal that high-specificity mutations perturb the Rec3 dynamic profile, particularly at the regions interfacing the RNA:DNA hybrid, which could remodel allosteric signaling within Cas9 through the interacting heteroduplex.

Altered interactions between Rec3 and the RNA:DNA hybrid

Given the structural perturbations observed experimentally, and the dynamic changes at the region comprising the interface of Rec3 and the RNA:DNA hybrid, we sought to assess whether the high-specificity mutations alter the interactions of the Rec3 domain with the hybrid through MD simulations of the full-length CRISPR-Cas9 system and its mutants, considering the PDB: 4UN3 structure (7). The difference in the Rec3–RNA:DNA hybrid contact stability between WT and the variants was estimated from the simulated ensemble (i.e., ~16 μs for each system). A comparison of the total number of Rec3-hybrid contacts that are at least 10% more stable in either WT (gray) or the variants (red, blue, and yellow) shows noticeable mutation-induced changes (Fig. 4D). Specifically, the Rec3 regions comprising residues 505 to 530 (largely containing α helix 25, α25), at residues 512 to 525 from PDB: 4UN3 (7) and the 575 to 590 loop increase their contact stability with the RNA:DNA hybrid in all variants compared to the WT, with the α25 interactions remaining least perturbed by the HypaCas9 mutations. These regions also exhibit significant chemical shift perturbations (Fig. 3A). At the level of residues 685 to 695, which comprise part of α37 (residues 692 to 699), Cas9-HF1 and evoCas9 display improved interactions with the hybrid with respect to WT Rec3, while a significant reduction is observed in HypaCas9, due to the nonpolar alanine mutations exclusively present in this region (Fig. 4D). Notably, α37 was shown to be involved in mismatch recognition, inserting within the RNA:DNA hybrid upon off-target DNA binding (29). Resonances in the α37 region are also substantially line broadened and lose CPMG-detected microsecond-millisecond motions (for residues 693 and 698), suggesting altered dynamics (Figs. 3A and 4C). In light of this evidence, the different interactions of the α37 region observed here for the three variants suggest a critical proofreading role. Last, at odds with the regions noted above, residues 650 to 660 show a consistent loss of stable contacts with the hybrid in the variants relative to the WT. This region does not exhibit NMR chemical shift perturbations (Fig. 3A). These results highlight important hotspots at the Rec3-hybrid interface, where rearranged molecular interactions with the key α37 might create the foundation for off-target detection.

Propagation of mutation effects in Rec3 to affect HNH dynamics

Biochemical and single-molecule experiments indicated that Rec3 could act as an "allosteric effector" of HNH dynamics (3, 9, 13). As the high flexibility of HNH ensures DNA cleavage at the proper location, alterations in the motions of HNH due to distal mutations in Rec3 could be critical for the specificity-conferring mechanisms of the variants studied here. We, therefore, sought to investigate whether the rewired intradomain flexibility of Rec3 affects the HNH domain. From the simulated ensembles of full-length Cas9 and its variants, we estimated differences in the HNH conformational dynamics by computing the Jensen-Shannon distance (JSD), which measures how similar two distributions are, ranging from 0 (similar distribution) to 1 (dissimilar distribution) (30). The JSD was computed between the distributions of the intra-backbone distances ($JSD_{BB-dist}$) and side-chain torsions (JSD_{SC-tor}) of the HNH domain in the WT and variants (Fig. 5A). We observe that the HNH dynamics in evoCas9 are comparable to the WT for each estimated feature. On the other hand, in Cas9-HF1 and HypaCas9, the HNH dynamics are substantially reduced in terms of BB
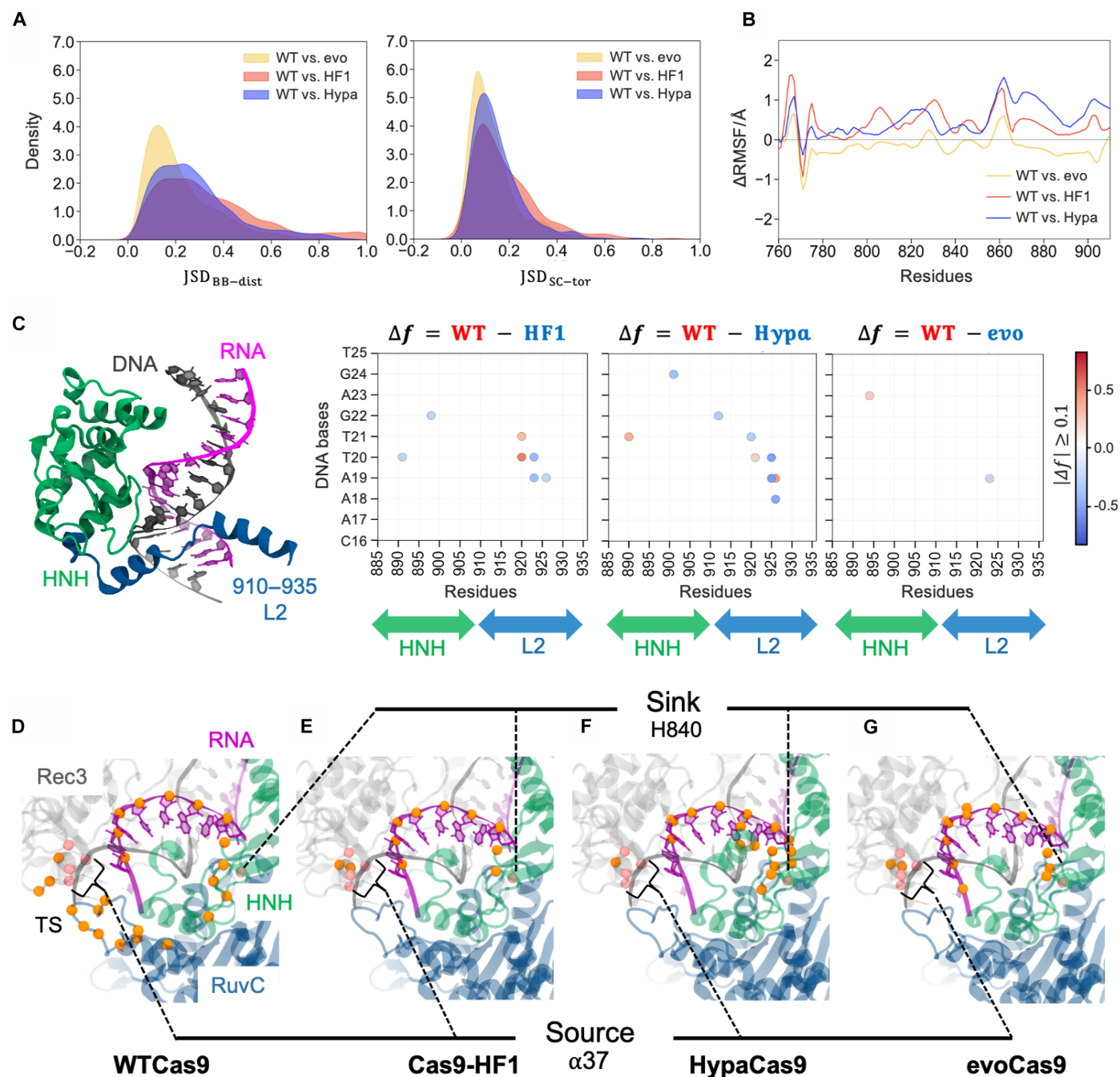
Fig. 4. High-specificity mutations alter the dynamic profile of Rec3 and its interactions with the RNA:DNA hybrid. (**A**) Residues exhibiting picosecond-nanosecond motions via S^2 and microsecond-millisecond motions via CPMG relaxation dispersion are mapped onto the Rec3 structure for HF1 Rec3 (left), Hypa Rec3 (center), and Evo Rec3 (right). Fast motions were determined via S^2 values 1.5σ below the 10% trimmed mean of all datasets. Red spheres represent picosecond-nanosecond motions (S^2), with a gradient denoting the lowest S^2 values as dark red, signifying greater picosecond-nanosecond flexibility. Blue spheres denote microsecond-millisecond motions (CPMG), with a blue gradient that becomes darker as the exchange contribution to relaxation (R_{ex}) increases. (**B**) Changes in S^2 (ΔS^2) between WT Rec3 and HF1 Rec3 (red), Hypa Rec3 (blue), and Evo Rec3 (yellow). Residues with positive ΔS^2 values are more flexible on the picosecond-nanosecond timescale in the variant, while residues with negative ΔS^2 values are more flexible in WT. (**C**) Residues exhibiting microsecond-millisecond motions via CPMG relaxation dispersion are indicated by a colored bar for a per-residue comparison between the WT Rec3 (gray), HF1 Rec3 (red), Hypa Rec3 (blue), and Evo Rec3 (yellow). (**D**) Differential contact analysis between the RNA:DNA hybrid and WT Rec3 (gray), HF1 Rec3 (red), Hypa Rec3 (blue), and Evo Rec3 (yellow). The number of stable contacts between Rec3 and the RNA:DNA hybrid is quantified in distinct regions of the protein, highlighted on the structure in blue (505 to 530), purple (575 to 590), green (650 to 660), and pink (685 to 695).

distances and SC torsions. Calculation of the differential root mean square fluctuations (ΔRMSF, in angstrom) of the Cα atoms between the WT and variants (Fig. 5B) further shows that HNH flexibility is reduced by the HF1 and Hypa mutations but remains comparable to WT in evoCas9. Principal components analysis also shows that Evo HNH samples a conformational space more similar to the WT HNH than to HF1 and Hypa HNH (fig. S12).

We next analyzed the interactions between the heteroduplex and the HNH domain in the three variants, thereby gaining insight into whether the RNA:DNA hybrid, interposing between Rec3 and HNH, could transduce the flexibility of Rec3 to the HNH domain. The gain

and loss of pairwise interactions between the RNA:DNA hybrid bases and the HNH residues were estimated as the number of contacts between pairs that are at least 10% more stable in the WT or the variants (details in the Supplementary Materials). Interactions between the DNA target strand bases and residues 910 to 935 of HNH, which comprise the L2 loop, are increased in Cas9-HF1 (Fig. 5C). HypaCas9 also displays an increase in DNA-L2 interactions compared to the WT, while these interactions are least perturbed in evoCas9. The L2 loop of HNH was previously shown to exert a critical role in mismatch recognition in WTCas9, as it tightly binds mismatched DNA bases (*29*). This reduces the conformational

Fig. 5. High-specificity mutations alter the HNH dynamics and reroute the allosteric signal in Cas9. (**A**) Distribution of JSD, comparing intra-backbone distances ($JSD_{BB-dist}$) and side-chain torsions (JSD_{SC-tor}) of the HNH domain in WTCas9 versus Cas9-HF1 (red), WTCas9 versus HypaCas9 (blue), and WTCas9 versus evoCas9 (yellow). (**B**) Differential root mean square fluctuations (ΔRMSF, in angstrom) of the HNH Cα atoms, computed between the WTCas9 and the HF1 (red), Hypa (blue), and Evo (yellow) variants. (**C**) Interactions between the RNA:DNA hybrid and the HNH domain. Differential contact maps report residue pairs that gain contact stability in either the WTCas9 (red) or the HF1, Hypa, and Evo variants (blue). Residues 910 to 935 of HNH, which comprise the L2 loop, are indicated on the x axis. The differential contact stability (Δf) is computed as the difference in the frequency f of formed contacts in the WTCas9 and its variants and is plotted as $|Δf| ≥ 0.1$, to account for contacts that are more stable in one system (e.g., WTCas9) by more than 10% with respect to the other (e.g., variant). (**D** to **G**) Signaling pathways connecting α helix 37 in Rec3 (i.e., source) to the HNH catalytic residue H840 (i.e., sink) in the WTCas9 (D), and the Cas9-HF1 (E), HypaCas9 (F), and evoCas9 (G) variants. Spheres of different colors are used to indicate the allosteric source/sink (pink) and the path nodes (orange). Analysis was performed on an aggregate sampling of ~16 μs (i.e., four MD replicas of ~4 μs each) for each system. Details are in Materials and Methods and in fig. S13.

mobility of HNH required for its transition to the active state, thus exerting a regulatory mechanism. In Cas9-HF1 and HypaCas9, the increase in interactions between the DNA and L2 is associated with altered dynamics of HNH (from the JSD plots). On the other hand, in evoCas9, where these interactions are less affected, the HNH dynamics are comparable to those of WT Cas9. This observation suggests that the RNA:DNA hybrid transduces the effect of the Rec3 mutations to HNH through its interaction with the L2 loop.

Rewired allosteric pathways between Rec3 and HNH

The perturbations in the Rec3 domain and associated changes to HNH dynamics observed via MD simulations hint at rewired allosteric pathways between the two domains, induced by the mutations. To gain insight into this communication, we performed shortest path calculations between the Rec3 mutation sites and the HNH catalytic core based on a dynamic network model from the simulated ensemble (see Materials and Methods). For each pair of "source" (Rec3 mutation sites) and "sink" (HNH catalytic residue H840), the optimal path (i.e., the shortest) as well as the top five sub-optimal paths were computed. Then, the distribution of all path lengths (i.e., the number of edges connecting the source and sink) corresponding to the optimal and sub-optimal paths was estimated, alongside the occurrence of residues in the paths to check for convergence in the communication pathways (figs. S13 to S15).

We observe an interesting trend in the communication pathways sourcing from residues located within the critical α37. In WT Cas9, α37 communicates with the HNH catalytic core via two main routes passing through the gRNA and the RuvC residues (Fig. 5D and fig. S13). In the Cas9-HF1, paths sourcing from α37 pass almost exclusively through the gRNA (Fig. 5E), thereby favoring the signal transduction. In this system, the communication routes that source from residue 695, which is within α37, also display a significant reduction in the path lengths with respect to the WT, while paths involving the residues 497 and 661 are less affected (fig. S14A). In HypaCas9, where all Rec3 mutations are clustered in α37 (N692A, M694A, Q695A, and H698A), the RuvC-mediated path observed in the WT is abolished, increasing the occurrence of the RNA nucleotides in the Rec3-HNH communication (Fig. 5F). Path lengths from this region to the HNH core are reduced relative to the WT, and communication routes also converge noticeably (fig. S13). We observe that common mutation sites at the level of α37 display similar communication patterns in the HF1 and Hypa variants, resulting in the replacement of the RuvC-medited path (observed in the WT) with gRNA nucleotides. We, therefore, sought to analyze the routes connecting α37 to the HNH core in the evoCas9 variant, which does not hold mutations in this α helix. The Evo mutations also abolish the RuvC-mediated pathway, rerouting the communication to the gRNA, and altering the WT Rec3-HNH communication similar to the HF1 and Hypa mutations located within α37 (Fig. 5G). On the other hand, pathways sourcing from the Evo mutation sites are preserved when moving from the WT to evoCas9, and path lengths are not significantly affected (fig. S15).

The three variants thereby alter the communication between α37 in Rec3, implicated in DNA mismatch recognition (29), and the HNH catalytic core. While HypaCas9 exerts this effect directly through mutations located within this key α helix, Cas9-HF1 uses the local Q695A mutation, alongwith distal mutations, to reroute the signal. evoCas9 most notably leverages purely distal mutations to allosterically affect the communication paths between the α37

and the HNH core. In this respect, it is notable that Cas9-HF1 and evoCas9 share a mutation of residue 661, distal to the key α helix. An in-depth analysis of the paths sourcing from the α37 also shows that the path lengths are reduced in all variants with respect to the WT (fig. S13), suggesting that the three variants enhance communication from the DNA mismatch recognition α helix 37 to the HNH active site, transferring the information of mismatch tolerance to the active site.

DISCUSSION

We combined solution NMR with molecular simulations to decipher the dynamics and allostery of three high-specificity variants of the Cas9 enzyme—Cas9-HF1, HypaCas9, and evoCas9—that contain multiple mutations within the Rec3 domain. Solution NMR was used to measure the structure and dynamics of the Rec3 domain with experimental accuracy, and MD simulations provided an all-atom description of the allosteric phenomenon, and how it propagates from Rec3 within the full-length complex. We detect conserved structural features and dynamic differences in the HF1, Hypa, and Evo variants, which rewire the Rec3 domain flexibility and remodel the allosteric signaling of Cas9. We found a consistent structural change in the Rec3 domain induced by the high-specificity mutations (Fig. 3A), with NMR chemical shift perturbations observed on the face of Rec3 that interfaces with the RNA:DNA hybrid at the PAM distal ends. Solution NMR also revealed altered molecular motions that perturb the Rec3 dynamic profile, particularly at regions interfacing with the heteroduplex. Relaxation dispersion experiments revealed microsecond-millisecond dynamics across the Hypa Rec3 domain, including at the RNA:DNA hybrid interface (Fig. 4, A and C), while HF1 and Evo Rec3 also show clusters of heightened flexibility near the same interface. These structural and dynamic changes suggest that high-specificity mutations could affect the interactions of the Rec3 domain with the heteroduplex. This is also supported by single-molecule data, showing that the Rec3 mutations interfere with the RNA:DNA heteroduplexation process (17) and with DNA unwinding/rewinding (15, 16). The Rec3-hybrid contacts were thereby obtained from multi-microsecond MD simulations of the full-length CRISPR-Cas9 systems (Fig. 4D). The three variants exhibited similar trends in altered interactions with the RNA:DNA hybrid at the level of the 575 to 590 and 650 to 660 loops, and the α25 helix. Most notably, they differentially interact with the heteroduplex at the level of α37 (Fig. 4B), which was shown to "sense" DNA mismatches within the RNA:DNA hybrid and is critically involved in mismatch recognition (29). In HypaCas9, which holds all of its mutations within α37, interactions with the hybrid are diminished relative to the WTCas9. Conversely, increased interactions between α37 and the hybrid are detected in the HF1 and Evo variants.

Critically, the high-specificity mutations in Rec3 also allosterically affect the dynamics of the catalytic HNH domain. Analysis of the interactions between HNH and the RNA:DNA heteroduplex reveals important differences at the level of the L2 loop (Fig. 5B), which was shown to tightly bind mismatched DNA bases in WT Cas9, reducing the conformational mobility of HNH as a regulatory mechanism. In HypaCas9, where the key α37 in Rec3 detaches from the hybrid, L2 in HNH tightly binds the hybrid. In Cas9-HF1 and evoCas9, where α37 has stronger interactions with the hybrid, the DNA-L2 interactions are less affected than observed for HypaCas9,

though still increased relative to WT Cas9. Consistent with the increased L2-hybrid interactions, the three variants affect the dynamics of the catalytic HNH domain (Fig. 5, A and B), which exhibits reduced flexibility from the WT in Cas9-HF1 and HypaCas9, while remaining comparable in evoCas9. This aligns with single-molecule studies of HF1 and Hypa, showing that distal mutations in Rec3 alter the conformational equilibrium of HNH (3), and is interesting in light of the well-characterized functional implications of high-specificity Cas9 variants (3–5, 31–36). Several studies have used high-throughput in vivo methods to measure on- and off-target cleavage to broadly assess the impacts of the high-specificity Cas9 variants (35, 36). Although it has been shown that Cas9 exhibits variability in its catalytic activity and target specificity dependent on the genomic loci targeted and assay methodology, the general trends supported by these high-throughput results reflect a trade-off between activity and specificity, where the increased specificity conferred by the Cas9 variants is paralleled by an overall decrease in catalytic activity. This is in line with stopped-flow kinetic experiments, which found that the slower rates of cleavage exhibited by the high-specificity variants promote more facile dissociation, rather than cleavage, of off-target DNA (31). Cas9-HF1, HypaCas9, and evoCas9 consistently rank among the most specific Cas9 variants, which is especially notable given that most, if not all, of their specificity-conferring mutations are contained in the Rec3 domain. Schmid-Burgk *et al.* (35) found that Cas9-HF1, HypaCas9, and evoCas9 exhibit >97% target specificity and <20 to 75% catalytic activity, compared to ~70% specificity and ~100% activity for WT Cas9 (fig. S16). Although the variants exhibit comparable specificity for off-target detection, Cas9-HF1 and HypaCas9 have significantly higher on-target activity than evoCas9. As Cas9 activity can be a direct consequence of HNH dynamics (37), the allosterically induced alterations to the motions of HNH, observed in the HF1 and Hypa variants in this study and through single-molecule experiments (3), could be a crucial component to maintaining their on-target activity while also conferring high specificity.

We also note that solution NMR detected conserved structural features in the HF1, Hypa, and Evo Rec3 domains, irrespective of the locations of their high-specificity mutations (Fig. 3). These structural changes, mainly found in regions of Rec3 interfacing with the RNA:DNA hybrid, could constitute a signature for the comparable specificity of the three variants (35). Alongside these conserved structural changes, dynamic differences in the three variants involve two critical elements of mismatch tolerance—α37 in Rec3 and L2 in HNH—that affect important contacts with the RNA:DNA hybrid and modulate Rec3 and HNH dynamics (Figs. 4 and 5) and likely contribute to the observed variability in catalytic activity among WT Cas9 and the three variants. Shortest-path analysis reveals that the variants rewire communication between α37 of Rec3 and the HNH catalytic core in a consistent manner (Fig. 5, D to G). The variants reroute the communication through the gRNA, reducing the path lengths and increasing the communication efficiency with respect to WTCas9. HypaCas9, which holds its mutations within α37, directly alters the signal transfer, while Cas9-HF1 uses the α37 Q695A mutation and its other distal mutations, and evoCas9 fully leverages mutations that are not located within α37, to reroute the signal from this critical α helix to the HNH core via protein dynamics. These data are consistent with our experimental findings, where the disparate mutations of the variants all induce structural perturbations and dynamic line broadening at α37 (Fig. 3), either directly or allosterically, while shifting the flexibility of the domain toward slower molecular motions distinct from WT Rec3. These data suggest an allosteric mechanism in which the three high-specificity Cas9 variants increase communication between α37 and the HNH active site. This outcome relies on subtle structural and dynamic perturbations to Rec3 that alter its contacts with the RNA:DNA hybrid, which, in turn, affects the hybrid contacts and dynamics of the catalytic HNH domain. It has previously been shown that specificity-conferring mutations in HNH itself similarly suppress HNH dynamics and Cas9 catalytic activity (15, 38, 39). Together, these findings suggest that direct or allosteric perturbation to the intrinsic motions of HNH contributes to the diminished catalytic rates of high-specificity Cas9 variants, which may temporally promote dissociation of off-target DNA following mismatch recognition by α37.

In summary, solution NMR and MD simulations reveal that the Cas9-HF1, HypaCas9, and evoCas9 high-specificity variants alter the structure and dynamics of the Rec3 domain and remodel the allosteric signaling within Cas9. High-specificity mutations induce conserved structural changes at regions of Rec3 that interface with the RNA:DNA hybrid, and subtle dynamic variations that remodel the communication from the DNA recognition region to the HNH catalytic core. In these variants, the Rec3 domain senses the RNA:DNA hybrid, while its interaction transduces the allosteric signal from Rec3 to HNH. These interactions are associated with altered dynamics of the HNH domain, which is more pronounced in Cas9-HF1 and HypaCas9, in agreement with single-molecule data. The three variants consistently reroute the communication between the α helix 37 in Rec3, previously shown to sense target DNA complementarity (29), and the HNH catalytic core. While HypaCas9 exerts this effect directly through mutations within α37, Cas9-HF1 uses the local Q695A mutation, alongside its other mutations. Most notably, evoCas9 fully leverages distal mutations to allosterically reroute the signal. These findings highlight a mechanism in which the three disparate Cas9 variants increase the communication from the DNA mismatch recognition helix to the HNH core through structural and motional perturbations, transferring the information of mismatch tolerance to the nuclease active site.

MATERIALS AND METHODS
Protein expression and purification
The WT Rec3 domain (residues 497 to 713) of the *S. pyogenes* Cas9 was engineered into a pET-28a(+) vector with an N-terminal His$_6$-tag and a Tobacco Etch Virus (TEV) protease cleavage site. High-specificity Rec3 variants, including HF1 Rec3 (N497A, R661A, and Q695A), Hypa Rec3 (N692A, M694A, and Q695A), and Evo Rec3 (Y515N, K526E, and R661Q), were generated via mutagenesis. The WT and high-specificity Rec3 plasmids were transformed into BL21-Gold (DE3) competent cells (Agilent) and expressed in lysogeny broth (LB) for x-ray crystallography and biochemical studies, and deuterated M9 minimal media, isotopically enriched with $^{15}NH_4Cl$ and $^{13}C_6H_{12}O_6$ (Cambridge Isotope Laboratories) as the sole nitrogen and carbon sources and supplemented with $MgSO_4$, $CaCl_2$, and minimum essential medium (MEM) vitamins, for NMR experiments. Cells were grown at 37°C to an OD$_{600}$ of 0.6 to 0.8 (LB) or 0.8 to 1.0 (deuterated M9), induced with 1.0 mM isopropyl-β-D-thiogalactopyranoside, incubated for 16 to 18 hours at 20°C, and then harvested by centrifugation.

During purification of WT Rec3 and the Rec3 variants, cells were resuspended in a buffer of 20 mM sodium phosphate, 300 mM sodium chloride, and 5 mM imidazole at pH 8.0 with a mini EDTA-free protease inhibitor cocktail tablet (Sigma-Aldrich) and lysed by sonication. Following the removal of cell debris by centrifugation, Rec3 was purified from the supernatant by Ni-NTA affinity chromatography. The column was washed with lysis buffer and the protein was eluted with a buffer of 20 mM sodium phosphate, 300 mM sodium chloride, and 250 mM imidazole at pH 8.0. The N-terminal His-tag was cleaved by TEV protease for 4 hours at room temperature while being dialyzed into lysis buffer and subsequently removed by Ni-NTA chromatography. Rec3 was further purified by size exclusion chromatography via a HiLoad 26/600 Superdex 200 column (GE) in a buffer of 20 mM sodium phosphate, 300 mM sodium chloride, and 2 mM EDTA at pH 8.0. Fractions containing purified Rec3 were then dialyzed into a buffer containing 40 mM sodium phosphate, 50 mM potassium chloride, 1 mM EDTA, and 6% D_2O at pH 7.5, and concentrated to 0.65 mM for NMR experiments.

NMR spectroscopy

NMR experiments were performed on Bruker Avance NEO 600 MHz and Bruker Avance III HD 850-MHz spectrometers at 25°C. NMR data were processed using NMRPipe (40) and analyzed in Sparky (41) with the help of in-house scripts. Backbone resonances for WT Rec3 were previously assigned (BMRB entry 50389) (42) and transferred to the HSQC spectra of the Rec3 variants. $^1H^{15}N$ combined chemical shift perturbations were determined from $^1H^{15}N$ TROSY HSQC (43) spectra by $\Delta\delta = |\delta_{WT} - \delta_{variant}|$, where $\delta = \sqrt{\frac{1}{2}[(\delta_{HN}^2) + (\delta_N^2/25)]}$ (44). Carbon chemical shifts ($^{13}C\alpha/\beta/_O$) were obtained from HNCA, HNCACB, and HNCO experiments on each variant and ^{13}C chemical shift perturbations were determined by $\Delta\delta = |\delta_{WT} - \delta_{variant}|$, where δ is the ^{13}C chemical shift value in parts per million.

TROSY-based spin relaxation experiments were performed with the ^1H and ^{15}N carriers set to the water resonance and 120 pm, respectively. Longitudinal relaxation rates (R_1) were measured with T_1 delays of 0, 20 (×2), 60 (×2), 100, 200, 600 (×2), 800, and 1200 ms. Transverse relaxation rates (R_2) were measured with T_2 delays of 0, 16.9, 33.9 (×2), 67.8, 136 (×2), and 203 (×2) ms. The recycle delay in these experiments was 2.5 s (45) and these data were collected in a temperature-compensated manner. Longitudinal and transverse relaxation rates were extracted by nonlinear least squares fitting the peak heights to a single exponential decay using in-house software. Uncertainties in these rates were determined from replicate spectra with duplicate relaxation delays. Steady-state ^1H-[^{15}N] NOE was obtained by interleaving pulse sequences with and without proton saturation and calculated from the ratio of peak heights (45). For the calculation of order parameters, model-free analysis was carried out using the Lipari-Szabo formalism in RELAX (46) with fully automated protocols (47). CPMG NMR experiments were adapted from the report of Palmer and coworkers (48) and performed at 25°C with a constant relaxation period of 40 ms, a 2.0-s recycle delay, and ν_{CPMG} values of 25, 50 (×2), 75, 100, 150, 250, 500 (×2), 750, 800 (×2), 900, and 1000 Hz. Relaxation dispersion profiles were generated by plotting R_2 versus $1/\tau_{cp}$ and exchange parameters were obtained from fits of these data using NESSY (49). Profiles were fit to three models, with the best model selected using Akaike information criteria with second-order correction.

Model 1: No exchange

$$R_2^{eff} = R_2^0 \qquad (1)$$

Model 2: Two-state, fast exchange [Meiboom equation (50)]

$$R_2^{eff} = R_2^0 + \frac{\Phi}{k_{ex}}\left[1 - \frac{4\nu_{CPMG}}{k_{ex}}\tan h\left(\frac{k_{ex}}{4\nu_{CPMG}}\right)\right] \qquad (2)$$

Model 3: Two-state, slow exchange [Richard-Carver equation (51)]

$$R_2^{eff} = R_2^0 + \frac{k_{ex}}{2} - \nu_{CPMG}\cos h^{-1}[D_+\cos h(\eta_+) - D_-\cos(\eta_-)] \quad (3)$$

CD spectroscopy

Rec3 samples (10 µM) were dialyzed into a buffer containing 40 mM sodium phosphate and 1 mM EDTA at pH 7.5 for CD experiments. CD spectra and thermal unfolding experiments were collected on a JASCO J-815 spectropolarimeter equipped with a variable temperature Peltier device and using a 2-mm quartz cuvette. CD spectra were collected at 25°C, and denaturation curves were recorded at 208 nm over a temperature range of 20° to 80°C. T_m values were determined via nonlinear curve fitting in GraphPad Prism.

X-ray crystallography

WT Rec3 was crystallized at room temperature by sitting drop vapor diffusion. Rec3 protein at 12 mg/ml in 20 mM tris (pH 7.5), 50 mM potassium chloride, and 1 mM EDTA was mixed with crystallizing condition 0.1 M bis-tris (pH 5.5), 0.2 M sodium chloride, 25% PEG3350 at a 2:1 ratio of protein-to-crystallizing condition. Crystals were cryo-protected in 30% ethylene glycol diluted in crystallizing conditions. Diffraction images were collected at the NSLS-II AMX beamline at Brookhaven National Laboratory. Images were processed using XDS (52) and Aimless (53) in CCP4 and the structure was solved by molecular replacement with Phaser in Phenix (54). The region of the full-length SpCas9 crystal structure corresponding to the Rec3 domain (residues 501 to 710, PDB ID: 4UN3) (7) was used as the search model for molecular replacement. The Rec3 structure was finalized by iterative rounds of manual building in Coot (55) and refinement with Phenix.

MD simulations

MD simulations were performed on the full-length CRISPR-Cas9 systems, based on the x-ray structure of the *S. pyogenes* Cas9 (PDB: 4UN3, at 2.59-Å resolution) (7). This structure captures the HNH domain in an inactivated state, a so-called "conformational checkpoint" between DNA binding and cleavage (9), in which the RNA:DNA complementarity is recognized before the HNH domain assumes an activated conformation. Single-molecule experiments reported alterations of the HNH dynamics in this "conformational checkpoint" in the high-specificity Cas9 variants (3), which motivated the use of this structure for MD simulations in the present study. Four model systems were considered: WT Cas9, Cas9-HF1, HypaCas9, and evoCas9. Each system was solvated in a ~145 × ~110 × ~147 Å periodic box of ~220,000 total atoms. In all systems, counterions were added to provide physiological ionic strength.

The systems were simulated using the AMBER ff99SBnmr2 force field for the protein (56), which improves the consistency of the backbone conformational ensemble with NMR experiments, also used in our previous NMR/MD study (38). Nucleic acids were described using the ff99bsc1 corrections for DNA (57) and the χOL3 corrections for RNA (58, 59). The TIP3P model was used for water (60). An integration time step of 2 fs was used. All bond

lengths involving hydrogen atoms were constrained using the SHAKE algorithm. Temperature control (300 K) was performed via Langevin dynamics (61), with a collision frequency γ = 1/ps. Pressure control was achieved through a Berendsen barostat (62), at a reference pressure of 1 atm and with a relaxation time of 2 ps. The systems were subjected to energy minimization to relax water molecules and counterions, keeping the protein, RNA, and DNA fixed with harmonic positional restraints of 300 kcal/mol Å^2. The systems were heated from 0 to 100 K in the canonical NVT ensemble, by running two simulations of 5 ps each, imposing positional restraints of 100 kcal/mol Å^2 on the abovementioned elements of the systems. The temperature was further increased to 200 K in ~100 ps of MD in the isothermal-isobaric NPT ensemble, reducing the restraint to 25 kcal/mol Å^2. Then, all restraints were released and the temperature was raised to 300 K in a single NPT run of 500 ps. Last, ~10 ns of NPT runs allowed the density of the system to stabilize around 1.01 g cm^{-3}. Production runs were carried out in the NVT ensemble reaching ~4 μs for each system, and in four replicates, collecting ~16 μs per system (for a total of ~64 μs of simulation time). This multi-microsecond simulation length with replicates was motivated by our previous work (10, 38, 63), showing that it provides a solid statistical ensemble for the analysis of allosteric mechanisms (described below). All simulations were conducted using the GPU-empowered version of AMBER 20 (64). Analysis of the results was performed on the aggregate ensemble (i.e., ~16 μs per system), after discarding the first ~200 ns of MD from each trajectory, to enable proper equilibration and fair comparison.

Analysis of Jensen-Shannon distances

To characterize the difference in the conformational dynamics of the HNH domain, we analyzed the distributions of all intra-backbone distances (BB − dist) and side-chain torsions (SC − tor) in the investigated systems. To compare the distributions of the abovementioned features between any two of our systems, we computed the JSD (JSD or D_{JS}), a symmetrized version of the Kullback-Leibler divergence (D_{KL}) (30). The JSD ranges from 0 to 1, where 0 corresponds to two identical distributions and 1 corresponds to a pair of separated distributions. For two distributions P_i and P_j, and considering a feature x_f from two different ensembles i and j

$$D_{JS} = \frac{1}{4}\sqrt{D_{KL}[P_i \| M] + D_{KL}[P_j \| M]} \qquad (4)$$

where $M = \frac{1}{2}(P_i + P_j)$. The Kullback-Leibler divergence, D_{KL}, corresponds to two distributions P_i and P_j is of the following form

$$D_{KL}[P_i \| P_j] = \sum P_j \log \frac{P_i}{P_j} \qquad (5)$$

JSD values were computed using the Python ENSemble Analysis (PENSA) open-source library (65). Kernel density estimations of the JSD values were plotted to describe the JSD range and compare the systems.

Shortest-path calculation

The allosteric pathways for information transfer were characterized through dynamic network analysis (66), a well-established method for the study of allosteric effects in proteins and nucleic acids (67–69). Through this analysis, the CRISPR-Cas9 system and its HF1, Hypa, and Evo variants are described as graphs of nodes and edges,

where nodes represent the amino acid (Cα atoms) and the nucleotides (P atoms, N1 in purines, and N9 in pyrimidines) (70) and edges denote the connection between them. To account for the information exchange between amino acids and nucleobases, the length of the edges connecting nodes was related to their motion correlations. A generalized correlation (GC) method (71) was used, which quantifies the system's correlations based on Mutual Information, describing the correlation between residue pairs independently on their relative orientation and capturing nonlinear contributions to correlations (details on GC analysis are reported in the Supplementary Materials).

On the basis of GC analysis, the weight (w_{ij}) of the edges connecting nodes i and j was computed as

$$w_{ij} = -\log GC_{ij} \qquad (6)$$

placing strongly correlated nodes close to each other in the dynamical network (i.e., displaying shorter edge lengths). To determine which nodes are in "effective contact," two nodes were considered connected if any heavy atom of the two residues is within 5 Å of each other (i.e., distance cutoff) for at least 70% of the simulation time (i.e., frame cutoff). This threshold was shown to properly optimize the network structure in our previous studies of CRISPR-Cas9 (10, 38, 63). The resulting "weighted graph" defines a dynamical network, used for shortest-path calculation.

Shortest-path calculations were carried out by computing the optimal (i.e., the shortest) and top five sub-optimal pathways between distally separated sites. Although the optimal path corresponds to the most likely mode of communication between nodes, sub-optimal paths can also be crucial for the transmission of allosteric communication by providing alternative routes (66). Hence, in addition to the optimal path, we also considered the top five sub-optimal pathways.

Well-established algorithms were used for shortest-path analysis. The Floyd-Warshall algorithm (72) was used to compute the optimal paths between the network nodes. This algorithm first creates a matrix of dimension $n \times n$, where n is the number of nodes. The elements of the matrix are the distance between all vertices in the network. The algorithm then recursively checks, for each pair of nodes (e.g., i and j), if there's any intermediate node (e.g., k) such that the sum of distances between $i − k$ and $k − j$ is less than the current distance of $i − j$. For each pair of nodes, the algorithm reiterates its checks as many times as the number of nodes in the graph. The resulting final matrix contains the shortest distance between all the vertices in the graph. The five sub-optimal paths were computed in rank from the shortest to the longest, using Yen's algorithm (73), which computes single-source K-shortest loop-less paths (i.e., without repeated nodes) for a graph with nonnegative edge weights.

For each system, the shortest pathways were computed between each mutation site in Rec3 (defined as source nodes) and the HNH catalytic residue H840 (defined as sink node). In Cas9-HF1, the shortest pathways were computed by sourcing from the N497A, R661A, and Q695A. In evoCas9, shortest-path calculations were sourced from the M495V, Y515N, K526E, and R661Q mutations. In HypaCas9, where all Rec3 mutations are clustered in the α helix 37 (N692A, M694A, Q695A, and H698A), the distribution of the path lengths was plotted by combining all the pathways connecting the N692A, M694A, Q695A, and H698A mutation sites to the H840 catalytic site. For comparison, shortest-path calculations were also performed in the WT Cas9, defining the same source and sink nodes of the variants. Given the importance of the α helix 37 for

mismatch recognition (*29*), we also performed shortest-path calculations sourcing from the HypaCas9 mutations, which are located within this key α helix, in the Cas9-HF1 and evoCas9 variants, as well as in the WT Cas9. To compare the systems, the distribution of the path lengths was plotted by combining all the pathways connecting all HypaCas9 mutation sites to H840.

All shortest-path calculations were performed over ~16 μs of aggregate sampling for each system (arising from four simulation replicates). To compare the lengths of the paths among the systems, we plotted the kernel density estimates as a function of the path lengths (i.e., the number of edges connecting the source and sink) for each of the source-sink pairs. To check for the convergence of communications in the systems, we estimated the occurrence of residues falling in the optimal and top five sub-optimal pathways. For each system, the aggregate ensemble arising from four simulation replicates (i.e., ~16 μs) was sectioned into three sample pools, which were used to compute the average occurrences and estimate the associated SEM. Networks of all the systems were built using the Dynetan Python library (*66*). Path-based analyses were performed using the NetworkX Python library (*74*) and our in-house Python scripts (available at: https://github.com/palermolab).

Supplementary Materials

This PDF file includes:
Supplementary Text
Figs. S1 to S16
Tables S1 to S4

REFERENCES AND NOTES

1. J. A. Doudna, The promise and challenge of therapeutic genome editing. *Nature* **578**, 229–236 (2020).
2. Y. Fu, J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung, J. D. Sander, High-frequency Off-target mutagenesis induced by CRISPR-cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).
3. J. S. Chen, Y. S. Dagdas, B. P. Kleinstiver, M. M. Welch, A. A. Sousa, L. B. Harrington, S. H. Sternberg, J. K. Joung, A. Yildiz, J. A. Doudna, Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
4. B. P. Kleinstiver, V. Pattanayak, M. S. Prew, S. Q. Tsai, N. T. Nguyen, Z. Zheng, J. K. Joung, High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
5. A. Casini, M. Olivieri, G. Petris, C. Montagna, G. Reginato, G. Maule, F. Lorenzin, D. Prandi, A. Romanel, F. Demichelis, A. Inga, A. Cereseto, A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.* **36**, 265–271 (2018).
6. F. Jiang, J. A. Doudna, The structural biology of CRISPR-Cas systems. *Curr. Opin. Struct. Biol.* **30**, 100–111 (2015).
7. C. Anders, O. Niewoehner, A. Duerst, M. Jinek, Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
8. S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
9. Y. S. Dagdas, J. S. Chen, S. H. Sternberg, J. A. Doudna, A. Yildiz, A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.* **3**, eaao0027-eaao0027 (2017).
10. K. W. East, J. C. Newton, U. N. Morzan, Y. B. Narkhede, A. Acharya, E. Skeens, G. Jogl, V. S. Batista, G. Palermo, G. P. Lisi, Allosteric motions of the CRISPR-Cas9 HNH nuclease probed by NMR and molecular dynamics. *J. Am. Chem. Soc.* **142**, 1348–1358 (2020).
11. F. Jiang, D. W. Taylor, J. S. Chen, J. E. Kornfeld, K. Zhou, A. J. Thompson, E. Nogales, J. A. Doudna, Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
12. G. Palermo, Y. Miao, R. C. Walker, M. Jinek, J. A. McCammon, Striking plasticity of CRISPR-Cas9 and key role of non-target DNA, as revealed by molecular simulations. *ACS Cent. Sci.* **2**, 756–763 (2016).
13. M. Yang, S. Peng, R. Sun, J. Lin, N. Wang, C. Chen, The conformational dynamics of Cas9 governing DNA cleavage are revealed by single-molecule FRET. *Cell Rep.* **22**, 372–382 (2018).
14. V. S. De Paula, A. Dubey, H. Arthanari, N. G. Sgourakis, A slow-exchange conformational switch regulates off-target cleavage by high-fidelity Cas9. bioRxiv 2020.12.06.413757 (2020). https://doi.org/10.1101/2020.12.06.413757.
15. D. Singh, Y. Wang, J. Mallon, O. Yang, J. Fei, A. Poddar, D. Ceylan, S. Bailey, T. Ha, Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat. Struct. Mol. Biol.* **25**, 347–354 (2018).
16. I. C. Okafor, D. Singh, Y. Wang, M. Jung, H. Wang, J. Mallon, S. Bailey, J. K. Lee, T. Ha, Single molecule analysis of effects of non-canonical guide RNAs and specificity-enhancing mutations on Cas9-induced DNA unwinding. *Nucleic Acids Res.* **47**, 11880–11888 (2019).
17. S. Y. Bak, Y. Jung, J. Park, K. Sung, H.-K. Jang, S. Bae, S. K. Kim, Quantitative assessment of engineered Cas9 variants for target specificity enhancement by single-molecule reaction pathway analysis. *Nucleic Acids Res.* **49**, 11312–11322 (2021).
18. A. J. Wand, Bringing disorder and dynamics in protein allostery into focus. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4278–4280 (2017).
19. S.-R. Tzeng, C. G. Kalodimos, Protein dynamics and allostery: An NMR view. *Curr. Opin. Struct. Biol.* **21**, 62–67 (2011).
20. S. J. Wodak, E. Paci, N. V. Dokholyan, I. N. Berezovsky, A. Horovitz, J. Li, V. J. Hilser, I. Bahar, J. Karanicolas, G. Stock, P. Hamm, R. H. Stote, J. Eberhardt, Y. Chebaro, A. Dejaegere, M. Cecchini, J.-P. Changeux, P. G. Bolhuis, J. Vreede, P. Faccioli, S. Orioli, R. Ravasio, L. Yan, C. Brito, M. Wyart, P. Gkeka, I. Rivalta, G. Palermo, J. A. McCammon, J. Panecka-Hofman, R. C. Wade, A. Di Pizio, M. Y. Niv, R. Nussinov, C.-J. Tsai, H. Jang, D. Padhorny, D. Kozakov, T. McLeish, Allostery in its many disguises: From theory to applications. *Structure* **27**, 566–578 (2019).
21. K. W. East, E. Skeens, J. Y. Cui, H. B. Belato, B. Mitchell, R. Hsu, V. S. Batista, G. Palermo, G. P. Lisi, NMR and computational methods for molecular resolution of allosteric pathways in enzyme complexes. *Biophys. Rev.* **12**, 155–174 (2020).
22. J. Liu, R. Nussinov, Allostery: An overview of its history, concepts, methods, and applications. *PLoS Comput. Biol.* **12**, e1004966–e1004966 (2016).
23. J. Guo, H.-X. Zhou, Protein allostery and conformational dynamics. *Chem. Rev.* **116**, 6503–6515 (2016).
24. B. Han, Y. Liu, S. W. Ginzinger, D. S. Wishart, SHIFTX2: Significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50**, 43–57 (2011).
25. N. Popovych, S. Sun, R. H. Ebright, C. G. Kalodimos, Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* **13**, 831–838 (2006).
26. A. J. Wand, The dark energy of proteins comes to light: Conformational entropy and its role in protein function revealed by NMR relaxation. *Curr. Opin. Struct. Biol.* **23**, 75–81 (2013).
27. G. P. Lisi, J. P. Loria, Using NMR spectroscopy to elucidate the role of molecular motions in enzyme function. *Prog. Nucl. Magn. Reson. Spectrosc.* **92-93**, 1–17 (2016).
28. G. P. Lisi, J. P. Loria, Solution NMR spectroscopy for the study of enzyme allostery. *Chem. Rev.* **116**, 6323–6369 (2016).
29. C. G. Ricci, J. S. Chen, Y. Miao, M. Jinek, J. A. Doudna, J. A. McCammon, G. Palermo, Deciphering off-target effects in CRISPR-Cas9 through accelerated molecular dynamics. *ACS Cent. Sci.* **5**, 651–662 (2019).
30. K. Lindorff-Larsen, J. Ferkinghoff-Borg, Similarity measures for protein ensembles. *PLOS ONE* **4**, e4203 (2009).
31. M. S. Liu, S. Gong, H. H. Yu, K. Jung, K. A. Johnson, D. W. Taylor, Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat. Commun.* **11**, 3576 (2020).
32. J. K. Lee, E. Jeong, J. Lee, M. Jung, E. Shin, Y. H. Kim, K. Lee, I. Jung, D. Kim, S. Kim, J. S. Kim, Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* **9**, 3048 (2018).
33. C. A. Vakulskas, D. P. Dever, G. R. Rettig, R. Turk, A. M. Jacobi, M. A. Collingwood, N. M. Bode, M. S. McNeill, S. Yan, J. Camarena, C. M. Lee, S. H. Park, V. Wiebking, R. O. Bak, N. Gomez-Ospina, M. Pavel-Dinu, W. Sun, G. Bao, M. H. Porteus, M. A. Behlke, 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 (2018).
34. D. Cerchione, K. Loveluck, E. L. Tillotson, F. Harbinski, J. DaSilva, C. P. Kelley, E. Keston-Smith, C. A. Fernandez, V. E. Myer, H. Jayaram, B. E. Steinberg, SMOOT libraries and phage-induced directed evolution of Cas9 to engineer reduced off-target activity. *PLOS ONE* **15**, e0231716 (2020).
35. J. L. Schmid-Burgk, L. Gao, D. Li, Z. Gardner, J. Strecker, B. Lash, F. Zhang, Highly parallel profiling of Cas9 variant specificity. *Mol. Cell* **78**, 794–800.e8 (2020).
36. N. Kim, H. K. Kim, S. Lee, J. H. Seo, J. W. Choi, J. Park, S. Min, S. Yoon, S. R. Cho, H. H. Kim, Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nat. Biotechnol.* **38**, 1328–1336 (2020).
37. S. H. Sternberg, B. LaFrance, M. Kaplan, J. A. Doudna, Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110–113 (2015).
38. L. Nierzwicki, K. W. East, U. N. Morzan, P. R. Arantes, V. S. Batista, G. P. Lisi, G. Palermo, Enhanced specificity mutations perturb allosteric signaling in CRISPR-Cas9. *eLife* **10**, e73601 (2021).

39. J. Wang, E. Skeens, P. R. Arantes, F. Maschietto, B. Allen, G. W. Kyro, G. P. Lisi, G. Palermo, V. S. Batista, Structural basis for reduced dynamics of three engineered HNH endonuclease lys-to-ala mutants for the clustered regularly interspaced short palindromic repeat (CRISPR)-associated 9 (CRISPR/Cas9) enzyme. *Biochemistry* **61**, 785–794 (2022).

40. F. Delaglio, S. Grzesiek, G. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

41. W. Lee, M. Tonelli, J. L. Markley, NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

42. E. Skeens, K. W. East, G. P. Lisi, (1)H, (13)C, (15) N backbone resonance assignment of the recognition lobe subdomain 3 (Rec3) from Streptococcus pyogenes CRISPR-Cas9. *Biomol. NMR Assign.* **15**, 25–28 (2021).

43. K. Pervushin, R. Riek, G. Wider, K. Wüthrich, 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 (1997).

44. S. Grzesiek, S. J. Stahl, P. T. Wingfield, A. Bax, The cd4 determinant for downregulation by HIV-1 nef directly binds to nef. mapping of the nef binding surface by NMR. *Biochemistry* **35**, 10256–10261 (1996).

45. G. Zhu, Y. Xia, L. K. Nicholson, K. H. Sze, Protein dynamics measurements by TROSY-based NMR experiments. *J. Magn. Reson.* **143**, 423–426 (2000).

46. M. Bieri, E. J. d'Auvergne, P. R. Gooley, 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 (2011).

47. G. Lipari, A. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559 (1982).

48. J. P. Loria, M. Rance, A. G. Palmer, A relaxation-compensated carr−purcell−meiboom−gill sequence for characterizing chemical exchange by NMR spectroscopy. *J. Am. Chem. Soc.* **121**, 2331–2332 (1999).

49. M. Bieri, P. R. Gooley, Automated NMR relaxation dispersion data analysis using NESSY. *BMC Bioinformatics* **12**, 421–421 (2011).

50. Z. Luz, S. Meiboom, Nuclear magnetic resonance study of the protolysis of trimethylammonium ion in aqueous solution—order of the reaction with respect to solvent. *J. Chem. Phys.* **39**, 366–370 (1963).

51. J. P. Carver, R. E. Richards, A general two-site solution for the chemical exchange produced dependence of T2 upon the carr-Purcell pulse separation. *J. Magn. Reson.* **6**, 89–105 (1972).

52. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).

53. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).

54. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).

55. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

56. L. Yu, D.-W. Li, R. Brüschweiler, Balanced amino-acid-specific molecular dynamics force field for the realistic simulation of both folded and disordered proteins. *J. Chem. Theory Comput* **16**, 1311–1318 (2020).

57. I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case, M. Orozco, Parmbsc1: A refined force field for DNA simulations. *Nat. Methods* **13**, 55–58 (2016).

58. P. Banáš, D. Hollas, M. Zgarbová, P. Jurečka, M. Orozco, T. E. Cheatham III, J. Šponer, M. Otyepka, Performance of molecular mechanics force fields for RNA simulations: Stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.* **6**, 3836–3849 (2010).

59. M. Zgarbová, M. Otyepka, J. Sponer, A. Mládek, P. Banáš, T. E. Cheatham III, P. Jurečka, Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **7**, 2886–2902 (2011).

60. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

61. P. Turq, F. Lantelme, H. L. Friedman, Brownian dynamics: Its application to ionic solutions. *J. Chem. Phys.* **66**, 3039–3044 (1977).

62. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

63. G. Palermo, C. G. Ricci, A. Fernando, R. Basak, M. Jinek, I. Rivalta, V. S. Batista, J. A. McCammon, Protospacer adjacent motif-induced allostery activates CRISPR-Cas9. *J. Am. Chem. Soc.* **139**, 16028–16031 (2017).

64. D. A. Case, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R.E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. S. P. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, L. Wilson, R. M. Wolf, X. Wu, Y. Xiong, Y. Xue, D. M. York, P. A. Kollman, AMBER 2020. University of California, San Francisco (2020).

65. M. Vögele, N. J. Thomson, S. T. Truong, J. McAvity, U. Zachariae, R. O. Dror, Systematic analysis of biomolecular conformational ensembles with PENSA. arXiv: 2212.02714 (2022). https://doi.org/10.48550/arXiv.2212.02714.

66. A. Sethi, J. Eargle, A. A. Black, Z. Luthey-Schulten, Dynamical networks in tRNA: Protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6620–6625 (2009).

67. P. R. Arantes, A. C. Patel, G. Palermo, Emerging methods and applications to decrypt allostery in proteins and nucleic acids. *J. Mol. Biol.* **434**, 167518–167518 (2022).

68. U. Doshi, M. J. Holliday, E. Z. Eisenmesser, D. Hamelberg, Dynamical network of residue-residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4735–4740 (2016).

69. T. Dodd, M. Botto, F. Paul, R. Fernandez-Leiro, M. H. Lamers, I. Ivanov, Polymerization and Editing modes of a High-fidelity DNA polymerase are linked by a Well-defined path. *Nat. Commun.* **11**, 5379–5379 (2020).

70. M. C. R. Melo, R. C. Bernardi, C. de la Fuente-Nunez, Z. Luthey-Schulten, Generalized correlation-based dynamical network analysis: A new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. *J. Chem. Phys.* **153**, 134104 (2020).

71. O. F. Lange, H. Grubmüller, Generalized correlation for biomolecular dynamics. *Proteins* **62**, 1053–1061 (2005).

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

73. J. Y. Yen, Finding the K shortest loopless paths in a network. *Manag. Sci.* **17**, 712–716 (1971).

74. A. A. Hagberg, D. A. Schult, P. J. Swart, Exploring network structure, dynamics, and function using NetworkX, in *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, G. Varoquaux, T. Vaught, J. Millman, Eds. (Los Alamos National Laboratory, 2008), pp. 11–15.