Nucleic Acids Research, 2023, 1–15
https://doi.org/10.1093/nar/gkad1127
RNA and RNA-protein complexes

Unveiling the RNA-mediated allosteric activation discloses functional hotspots in CRISPR-Cas13a

Souvik Sinha [1], Adrian M. Molina Vargas [2,3,4], Pablo R. Arantes [1,†], Amun Patel [1,†], Mitchell R. O'Connell [2,3,*] and Giulia Palermo [1,5,*]

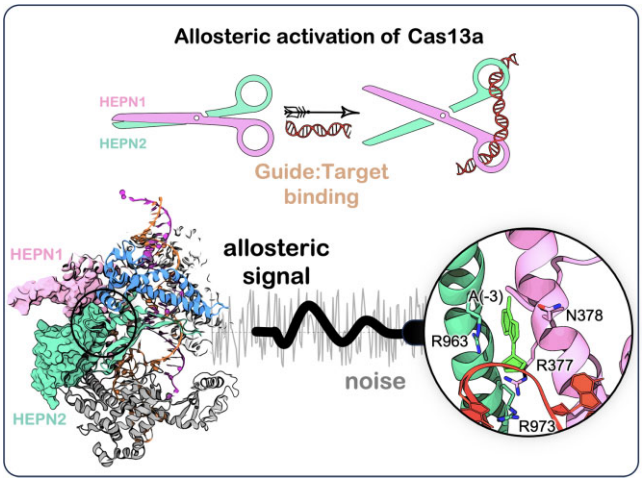
[1]Department of Bioengineering, University of California Riverside, Riverside, CA , 92521, USA
[2]Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA
[3]Center for RNA Biology, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA
[4]Department of Biomedical Genetics, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA
[5]Department of Chemistry, University of California Riverside, Riverside, CA, 92521, USA

[*]To whom correspondence should be addressed. Tel: +1 951 827 4303; Email: giulia.palermo@ucr.edu
Correspondence may also be addressed to Mitchell R. O'Connell. Tel: +1 585 276 3191; Email: mitchell_oconnell@urmc.rochester.edu
[†]The third and fourth authors should be regarded as as equally contributing authors.

Abstract

Cas13a is a recent addition to the CRISPR-Cas toolkit that exclusively targets RNA, which makes it a promising tool for RNA detection. It utilizes a CRISPR RNA (crRNA) to target RNA sequences and trigger a composite active site formed by two 'Higher Eukaryotes and Prokaryotes Nucleotide' (HEPN) domains, cleaving any solvent-exposed RNA. In this system, an intriguing form of allosteric communication controls the RNA cleavage activity, yet its molecular details are unknown. Here, multiple-microsecond molecular dynamics simulations are combined with graph theory to decipher this intricate activation mechanism. We show that the binding of a target RNA acts as an allosteric effector, by amplifying the communication signals over the dynamical noise through interactions of the crRNA at the buried HEPN1-2 interface. By introducing a novel Signal-to-Noise Ratio (SNR) of communication efficiency, we reveal critical allosteric residues—R377, N378, and R973—that rearrange their interactions upon target RNA binding. Alanine mutation of these residues is shown to select target RNA over an extended complementary sequence beyond guide-target duplex for RNA cleavage, establishing the functional significance of these hotspots. Collectively our findings offer a fundamental understanding of the Cas13a mechanism of action and pave new avenues for the development of highly selective RNA-based cleavage and detection tools.

Graphical abstract



Introduction

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and their associated (Cas) proteins are RNA-guided prokaryotic adaptive immune systems that protect bacteria against invading genetic elements (1). The transformative use of CRISPR-Cas9 for genome editing (2) led to the 'boom' of the 'CRISPR field' and the discovery of novel CRISPR-Cas systems that remarkably expand applications in genome editing and beyond (3–5). Cas13a (formerly known as C2c2) is a recently discovered CRISPR-associated protein that tar-

Received: July 27, 2023. **Revised:** October 25, 2023. **Editorial Decision:** October 31, 2023. **Accepted:** November 9, 2023

gets RNA (6,7), a property that is powerful for RNA detection, regulation, and imaging (8–13). The Cas13a effector protein complexes with a CRISPR RNA (crRNA) sequence, which is used as a guide to form the RNA-targeting interference complex (6,7,14). The latter can cleave single-stranded RNA (ssRNA) sequences using two Higher Eukaryotes and Prokaryotes Nucleotide (HEPN) catalytic domains (7,15), which are found in RNA targeting enzymes (16). Upon activation of the HEPN domains, Cas13a degrades its target RNAs through *cis* cleavage and other solvent-exposed ssRNAs through a non-specific *trans* cleavage activity (6,7). This is a potent 'collateral damage' that has enabled the development of ultrasensitive RNA detection tools like SHERLOCK (13), SPRINT (12), direct SARS-CoV-2 RNA detection assays (17) and many others.

The biophysical function of the Cas13a effector is characterized by intricate allosteric signalling, whose molecular details are highly unclear (18). Structures of Cas13a from *Leptotrichia buccalis* (Lbu) reveal a bilobed architecture, comprising a 'REC' α-helical lobe that recognizes the crRNA and a 'NUC' lobe consisting of the catalytic HEPN1-2 domains and a helical Linker (Figure 1) (19). Biochemical studies have shown that the binding of a complementary target RNA (tgRNA) to the REC lobe allosterically activates the HEPN domains (18), which are spatially distant, to form a composite active site for RNA cleavage (7,15). However, the mechanism of this activation and information transfer from the tgRNA binding site to the HEPN1-2 catalytic cleft is poorly understood.

Interestingly, the crRNA spacer—i.e. the 28 nucleotides (nt.) sequence that guides binding of the tgRNA—was shown to hold a critical role in the onset of the allosteric response (18). The spacer contains a 'seed' region (nt. 9–14), where perfect base pairing is required for tgRNA binding, and a 'switch' region (nt. 5–8, Figure 1D), whose binding of the tgRNA has been suggested to induce the activation of the catalytic cleft (18). Nevertheless, the 'seed' and 'switch' regions are approximately located 34 Å and 41 Å from the catalytic core respectively, puzzling the allosteric response of these regions toward activation.

Functional studies, both *in-vitro* and *in-vivo*, have also shown that extending the duplex complementarity at the 5′ flank of the crRNA spacer, impacts the RNA degrading ability of Cas13a (20). Such extended pairing beyond the guide-target duplex is called tag (segment from the crRNA)—anti-tag (segment from the tgRNA) pairing (blue segment in Figure 1D). This dependence on the complementarity length was proposed as a strategy to discriminate between self- and non-self-targets, which is key for Cas13-based technologies (21). This suggests the allosteric response is regulated by the degree of complementarity between the guide crRNA (also referred to as a guide-RNA) and tgRNA.

In light of this knowledge, understanding the allosteric regulation of Cas13a activation and the role of RNA is crucial to improving the RNA targeting specificity of Cas13a, and engineering its controlled function to discriminate non-self-targets.

Here, we used extensive molecular dynamics (MD) simulations, reaching a conformational sampling of ∼170 μs, and conducted thorough graph theory analysis, providing a comprehensive RNA-mediated allosteric mechanism of the Cas13a protein. Most significantly, our computational investigation discloses critical residues whose mutation in alanine is shown experimentally to select tgRNA over an extended tag–anti-tag complementarity, for RNA cleavages. By introducing a novel graph theory-based analysis—signal-to-noise ratio (*SNR*) of communication efficiency – we establish the critical hotspots for Cas13a activation, showing that our computational approach can guide the development of more selective RNA cleavage and detection strategies. These findings offer a fundamental understanding of the Cas13a mechanism of action and pave new avenues for the development of more selective RNA-targeting CRISPR-Cas13a systems.

Materials and methods

Structural models

Molecular simulations were based on the structure of the *Leptotrichia buccalis* (Lbu) Cas13a bound to a crRNA (PDB: 5XWY, at 3.2 Å resolution (15)), obtained via cryo-EM, and on the structure of the LbuCas13a in complex with a crRNA and tgRNA (PDB: 5XWP), obtained by single-wavelength anomalous diffraction at 3.08 Å resolution (15). The Lbu-Cas13a bound to an extended tag–anti-tag RNA (atgRNA) was built by including a longer duplex with eight base-pair extended atgRNA, obtained from the cryo-EM structure of the *Leptotrichia shahii* Cas13a bound to atgRNA (PDB: 7DMQ, at 3.06 Å resolution) (20). In all systems, we reinstated the catalytic H1053 and R1048 in the HEPN domains, which were mutated in alanine in the experimental structures (15). The protonation states of histidine residues have been computed using the H++ software (22) reporting singly protonated neutral states (protonated on the ε position). This allows histidine residues to act as a base within the HEPN1-2 cleft, as recently shown for a type III CRISPR-Cas system holding the HEPN1-2 RNase activity (23). The systems were solvated, leading to simulation cells of ∼138 × 97 × 130 $Å^3$, and neutralized by the addition of an adequate number of Na^+ ions.

Molecular dynamics simulations

Molecular dynamics (MD) simulations were performed using a simulation protocol tailored for protein/nucleic acid complexes (detailed in Supplementary Text). We employed the Amber ff19SB (24) force field, including the χOL3 corrections for RNA (25,26). The TIP3P model was used for explicit water molecules (27). Production runs were carried out in the NVT ensemble, using an integration time step of 2 fs. For each system, i.e. Cas13a bound to a crRNA (crRNA-Cas13a), in complex with a crRNA and tgRNA (tgRNA-Cas13a) and bound to an extended tag–anti-tag RNA (atgRNA-Cas13a), we performed ∼5 μs of MD simulations in three replicates. We also carried out a ∼5 μs long simulation of a tgRNA-bound complex substituting the A(-3) base of the crRNA with cytosine C(-3). Subsequently, we considered four variants (R377A, N378A, R963A, and R973A) of Cas13a bound to a tgRNA, including an atgRNA, similar to the wild-type (WT) Cas13a complexes. These systems were also simulated for ∼5 μs in three replicates. Overall, we accumulated ∼170 μs of total sampling. All simulations were performed using the GPU-empowered version of the AMBER 20 code (28). Analyses were performed over the aggregated multi-μs sampling collected for each complex (i.e. ∼15 μs). This was motivated by our previous studies of allostery in the CRISPR-Cas systems (29–33), showing that an aggregated multi-μs sampling offers

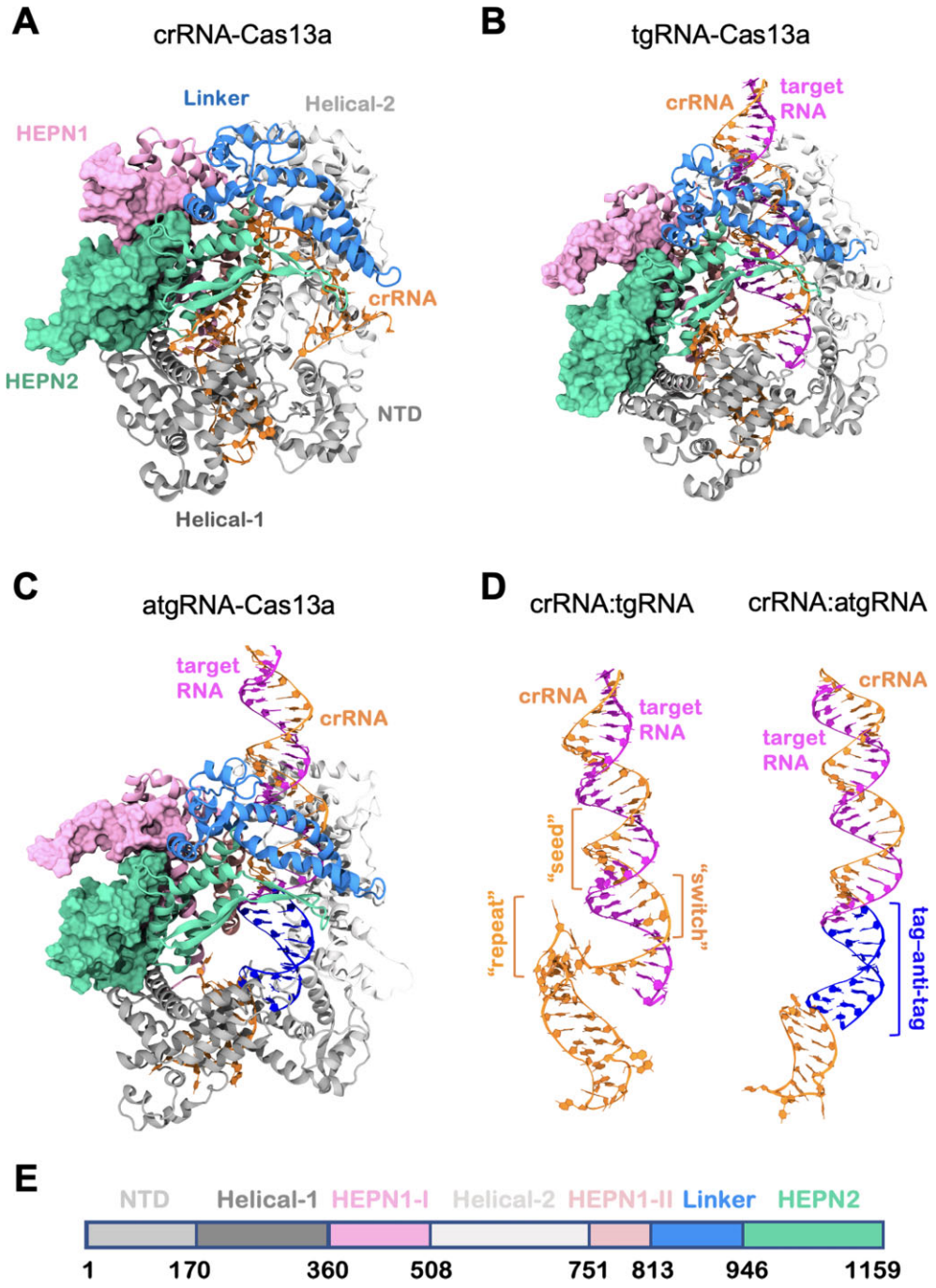


Figure 1. Overview of the RNA-bound Cas13a complexes from *Leptotrichia buccalis*. (**A–C**) Structures of the *Leptotrichia buccalis* Cas13a bound to a guide CRISPR RNA (crRNA, A), a target RNA (tgRNA, B), and a tag–anti-tag RNA (atgRNA, C) (19). The Cas13a protein is shown as cartoons, using a molecular surface representation for the HEPN1 (mauve) and HEPN2 (green) catalytic domains. (**D**) Structure of the RNA in the tgRNA- (left) and atgRNA- (right) bound systems. The crRNA (orange) forms a duplex with the tgRNA (magenta), whose complementarity is extended at the 5′ flank of the crRNA through a tag–anti-tag pairing (blue). The 'seed' (nucleotides, nt. 9–14) and 'switch' (nt. 5–6) regions of the crRNA, as well as the crRNA repeat region (nt. (A(-5)–(C-1)), are also indicated. (**E**) Protein sequence and domain colour code.

a robust ensemble for the purposes of our analysis (described below and in Supplementary Text).

Analysis of Jensen–Shannon distances

To characterize the difference in the conformational dynamics of the HEPN domains, we analysed the distributions of all intra-backbone dihedrals (BB torsions) and backbone C_α distances (BB distances) in the investigated systems. To compare the distributions of the abovementioned features between any two of our systems, we computed the Jensen-Shannon Distance (*JSD* or D_{JS}), a symmetrized version of Kullback-Leibler divergence (D_{KL}).(34) 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,

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

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:

$$D_{KL}\left[P_i||\ P_j\right] = \sum P_j \log \frac{P_i}{P_j} \quad (2)$$

JSD values were computed using the Python ENSemble Analysis (PENSA) open-source library (35). Kernel density estimations of the *JSD* values were plotted to describe the *JSD* range and compare the systems.

Inter-domain correlations

To evaluate the inter-dependent coupling between the Cas13a domains and the nucleic acids, we computed inter-domain generalized correlation scores (36). For each protein residue i, a correlation score (Cs) parameter can be computed as:

$$Cs = \sum_{j\neq i}^{N} GC_{ij} \quad (3)$$

which sums the atom-based generalized correlations, GC_{ij}, established by residue i with the residues j, based on Shannon's entropy estimation of mutual information (details in Supplementary Text) (37). Cs are a measure of the number and the intensity of the GC_{ij} coefficients displayed by each residue pair. To detail inter-domain correlations, the Cs were accumulated and normalized. First, the Cs were calculated for each residue i belonging to a specific protein domain (e.g. HEPN1(I)), with the residues j belonging to another protein domain of interest (e.g. HEPN2). Then, the Cs were accumulated over all residues j of each specific domain and normalized by the number of coupling residues. This resulted in a set of inter-domain Cs, ranging from 0 (not-correlated) to 1 (correlated), measuring the strength of the overall correlation that each domain establishes with the others.

Dynamic network and signal-to-noise ratio

To characterize the allosteric pathways of communication, network analysis was applied (38). In dynamical networks, Cα atoms of proteins and backbone P atoms of nucleotides, as well as N1 atoms in purines, and N9 in pyrimidines, are represented as nodes, connected by edges weighted by the generalized correlations GC_{ij} according to:

$$w_{ij} = -\log\left(GC_{ij}\right). \quad (4)$$

Details on network analysis are in Supplementary Text. From the dynamical network, we estimated the efficiency of crosstalk between the crRNA spacer regions (i.e. 'seed' (nt. 9–14); 'switch' (nt. 5–8); as well as nt. 1–4; nt. 15–18 and nt. 19–22) and the catalytic residues (R472, H477, R1048, H1053) by introducing a novel Signal-to-Noise Ratio (*SNR*) measure. *SNR* measures the preference of communication between predefined distant sites—i.e. the signal—over the remaining pathways in the network—i.e. the noise, estimating how allosteric pathways stand out (i.e. are favourable) over the entire communication network.

For the *SNR* calculation, we first computed the optimal (i.e. the shortest) and top five sub-optimal pathways (with longer lengths, ranked in comparison to the optimal path length) between all crRNA bases and the Cas13a residues, using well-established algorithms (*vide infra*). Then, the cumulative betweennesses of each pathway (S_k) was calculated as the sum of the betweennesses of all the edges in that specific pathway:

$$S_k = \sum_{i=1}^{n-1} b_i \quad (5)$$

where b_i is the edge betweenness (i.e. the number of shortest pathways that cross the edge, measuring the 'traffic' passing through them) between node i and $i+1$, and n is the number of edges in the k^{th} pathway. The distribution of S_k between the crRNA bases and all protein residues was defined as the noise, whereas the distribution of S_k between the crRNA nucleotide regions of interest (e.g. 'seed') and the HEPN1-2 catalytic residues were considered as signals. Since S_k depends on the number of edges present in the path, to ensure consistency, we characterized the *SNR* on the basis of shorter (edge count: 6–8), medium (edge count: 9–11), and longer (edge count: 12–14) paths (details in Supplementary Text). Notably, while the optimal path corresponds to the most likely mode of communication, suboptimal paths can also be crucial routes for communication transfer (31,38). Hence, in addition to the optimal path, we also considered the top five sub-optimal pathways for our *SNR* analysis. Well-established algorithms were employed for shortest-path analysis. The Floyd–Warshall algorithm was utilized to compute the optimal paths between the network nodes. The five sub-optimal paths were computed in rank from the shortest to the longest, using Yen's algorithm, which computes single-source K-shortest loop-less paths (i.e. without repeated nodes) for a graph with non-negative edge weights (39). To identify residues important for allosteric communication, we computed the occurrence of each residue appearing in at least one of the pathways (i.e. optimal and sub-optimal). This analysis also reports on the conservation of allosteric pathways, as pathways characterized by a lesser number of residues with higher occurrence are likely to be more conserved than those exhibiting a greater number of residues with lower occurrence.

Finally, the *SNR* corresponding to signals from each crRNA region (i.e. 'seed' (nt. 9–14); 'switch' (nt. 5–8); as well as nt 1–4; nt. 15–18 and nt. 19–22) to the HEPN1-2 catalytic core residues (R472, H477, R1048, H1053) was computed as:

$$SNR = \frac{E\left[S\right]/Var\left(S\right)}{E\left[N\right]/Var\left(N\right)} \quad (6)$$

where $E(S)$ and $Var(S)$ correspond to the expectation and variance of the signal distribution respectively; and $E(N)/Var(N)$ are the expectation/variance of the noise distribution. To provide the significance of the signal over the noise, we used a general approach based on p-value calculation. Our goal was to test the hypothesis that the signal is an outlier of the noise distribution. We can construct a best-fit probability distribution based on the noise data by treating our variable (the sum of betweennesses) as stochastic. Treating the signal as a sample from this population, we can assess the rarity of that sample's mean when randomly collecting samples of the same size. This rarity is defined as the p-value of the sample. The p-value is then computed using the Z-score:

$$Z_{signal} = \frac{E\left[S\right] - E\left[N\right]}{\sigma\left[N\right]/\sqrt{n}} \quad (7)$$

where $\sigma[N]$ corresponds to the standard deviation of the noise, and n is the number of signals. Here, we found that the

best-fit distributions for the noise follow a log-normal distribution. Hence, we took the logarithm of the signals and noise to transform the data points into a normal distribution. From this transformed distribution, we obtained the mean and standard deviations for the calculation of the Z_{signal}. All networks were built using the Dynetan Python library (38). Path-based analyses were performed using NetworkX Python library (40) and our in-house Python scripts.

Cas13a protein expression and purification

For expression of wild-type LbuCas13a we used Addgene Plasmid #83482 (7). LbuCas13a mutants were cloned from the WT vector via site-directed mutagenesis using the primers indicated in Supplementary Table S1.

All constructs were purified as previously described (7,41), with some modifications. Briefly, expression vectors were transformed into Rosetta2 DE3 grown in LB media supplemented with 0.5% w/v glucose at 37°C. Protein expression was induced at mid-log phase (OD_{600} ∼0.6) with 0.5 mM IPTG, followed by incubation at 16°C overnight. Cell pellets were resuspended in lysis buffer (50 mM HEPES [pH 7.0], 1 M NaCl, 5 mM imidazole, 5% (v/v) glycerol, 1 mM DTT, 0.5 mM PMSF, EDTA-free protease inhibitor [Roche]), lysed by sonication, and clarified by centrifugation at 15,000g. Soluble His_6-MBP-TEV-Cas13a was isolated over metal ion affinity chromatography, and in order to cleave off the His_6-MBP tag, the protein-containing eluate was incubated with TEV protease at 4°C overnight while dialyzing into ion exchange buffer (50 mM HEPES [pH 7.0], 250 mM NaCl, 5% (v/v) glycerol, 1 mM DTT). The cleaved protein was loaded onto a HiTrap SP column (GE Healthcare) and eluted over a linear KCl (0.25–1 M) gradient. Fractions containing LbuCas13a were pooled, concentrated, and further purified via size-exclusion chromatography on an S200 column (GE Healthcare) in gel filtration buffer (20 mM HEPES [pH 7.0], 200 mM KCl, 5% glycerol (v/v), 1 mM DTT), snap-frozen in liquid N_2 and were subsequently stored at −80°C. Protein purity was assessed by loading ∼2 ug of total protein in a 4–12% Bis–Tris gel and staining with Coomassie Blue dye.

Fluorescent ssRNA nuclease assays

Mature crRNAs and ssRNA targets were commercially synthesized (IDT; Supplementary Table S2). Cas13 trans-cleavage nuclease activity assays were performed in 10 mM HEPES pH 7.0, 50 mM KCl, 5 mM MgCl2, and 5% glycerol. Briefly, 100 nM LbuCas13a:crRNA complexes were assembled in for 30 min at 37°C. 100 nM of Rnase Alert reporter (IDT) and various final concentrations of ssRNA-target were added to initiate the reaction. These reactions were incubated in a fluorescence plate reader (Tecan Spark) for up to 60 min at 37°C with fluorescence measurements taken every 5 min (λ_{ex}: 485 nm; λ_{em}: 535 nm). Time-course and end-point values at 1 hour were background-subtracted, normalized, and analysed with their associated standard errors using GraphPad Prism9.

Results

Conformational changes of the HEPN1-2 domains

MD simulations were carried out on three RNA-bound complexes of the LbuCas13a: the binary complex bound to a guide crRNA only (crRNA-Cas13a), and the ternary complexes in which Cas13a binds to a 28 nt. tgRNA matching the crRNA (tgRNA-Cas13a), and a 36 nt. tag–anti-tag RNA (atgRNA-Cas13a, Figure 1). In these systems, we reinstated the catalytic H1053 and R1048 in the HEPN domains, which were mutated to alanine in the structures to prevent RNA cleavage (19). An aggregate sampling of ∼15 μs was collected for each system, from three simulation replicas of ∼5 μs each.

To characterize the difference in the HEPN1-2 conformational dynamics among the systems studied, we first analysed the distribution of the Jensen-Shannon Distances (*JSD*) (34), measuring the similarity of two distributions, ranging from 0 (similar) to 1 (dissimilar). The *JSD* distributions were computed to compare all intra-backbone distances (BB distances) and backbone torsions (BB torsions) of the HEPN domains in the investigated systems (see *Material and Methods*). We observe that, while overall the *JSD* distributions are similar for the overall HEPN1-2 domains (Supplementary Figure S1), the backbone dynamics of the catalytic cleft (residues 470–480, 1045–1055) clearly display a separation between the crRNA- and tgRNA-Cas13a systems (Figure 2A, B, red distribution). A separation in the dynamics of the catalytic cleft is also observed when comparing the crRNA- and atgRNA-bound Cas13a (blue distribution), while the catalytic cleft dynamics in the atgRNA-Cas13a are least separated from those of the tgRNA-Cas13a (grey distribution). This observation is complemented by the Solvent Accessible Surface Area (SASA), showing that the HEPN1-2 catalytic cleft is significantly more accessible to the solvent in the tgRNA-bound system than in the crRNA-Cas13a binary system (Figure 2C), also deviating from the experimental structures (19). Interestingly, time-evolution plots of the SASA across our simulation replicas show that in the tgRNA-bound system, the catalytic pocket more frequently transitions from open-to-close conformation with respect to the crRNA-Cas13a, which displays a relatively closed pocket (Supplementary Figure S2). This indicates that a closed catalytic core in the crRNA-bound complex opens upon tgRNA binding (Figure 2D). Interestingly, in the presence of an atgRNA, the catalytic cleft samples both open and closed conformations, also accessing states where the pocket is even less accessible than in the crRNA-Cas13a. This is also notable from the time-evolution plots, reporting a higher amplitude and frequency of the open-to-close transition in the atgRNA-Cas13a compared to the tgRNA-bound system (Supplementary Figure S2), overall indicating a higher plasticity of the pocket.

These observations indicate that when Cas13a is bound to the crRNA alone, the HEPN1-2 catalytic cleft assumes a closed conformation. On the other hand, tgRNA binding results in an opening of the cleft. Moreover, the plasticity of the HEPN1-2 cleft appears to be a key distinguishing factor between the tgRNA- and atgRNA-Cas13a systems. In the atgRNA-Cas13a complex, the catalytic cleft is more flexible in comparison to the tgRNA-bound Cas13a, suggesting that also the length of base-pair complementarity impacts the HEPN dynamics.

Target RNA shifts the dynamics of Cas13a

To understand how the binding of the tgRNA, in the presence and absence of an extended tag–anti-tag pair (atgRNA), impacts the dynamics of the spatially distant HEPN1-2 domains, it is imperative to measure the dynamic correlations among these spatially distant sites (42). We employed a generalized correlation (GC) analysis, relying on Shannon's entropy-based

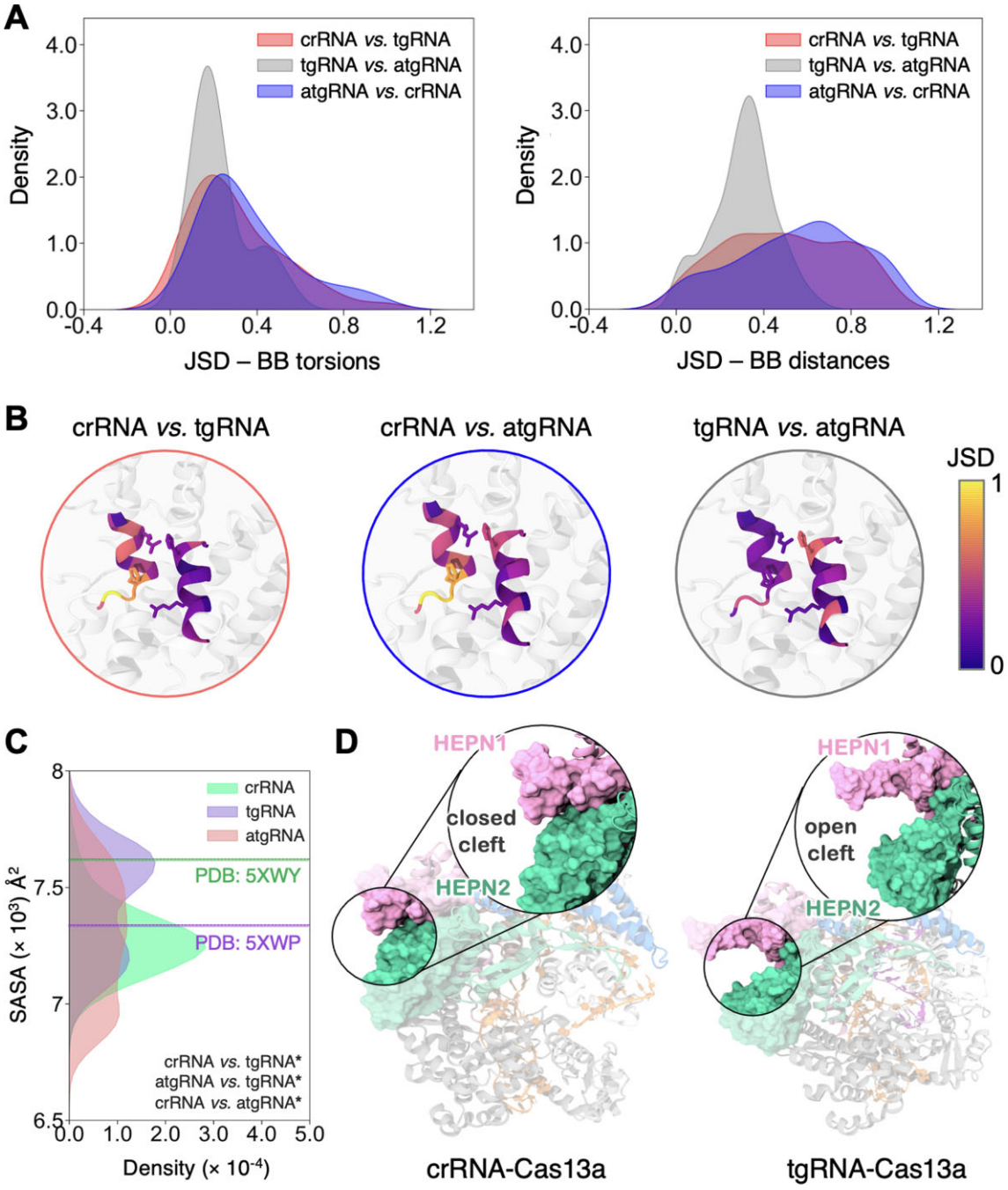


Figure 2. Structural dynamics of the HEPN1 and HEPN2 catalytic domains. (**A**) Distribution of Jensen-Shannon Distances (*JSD*), comparing intra-backbone distances (BB distances) and backbone torsions (BB torsions) of the HEPN1-2 catalytic core helices (residues 470–480 and 1045–1055) in the crRNA *vs.* tgRNA (red), tgRNA *vs.* atgRNA (gray) and atgRNA *vs.* crRNA (blue) bound systems. (**B**) Projection of the *JSD* BB torsions on the three-dimensional structure of the HEPN1-2 catalytic core, color-coded according to the scale on the right. (**C**) Distribution of the Solvent Accessible Surface Area (SASA) in the crRNA- (green), tgRNA- (violet), and atgRNA- (magenta) bound Cas13a. Vertical lines indicate the experimental values in the crRNA- (PDB: 5XWY) and tgRNA- (PDB: 5XWP) bound proteins. The statistical significance among the distributions was calculated using a two-tailed unpaired *t*-test (*P*-value reference: not significant, ns $P > 0.05$, * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$). (**D**) Representative snapshots of Cas13a bound to crRNA (top) and tgRNA (bottom), displaying a closed (crRNA-bound) and open (tgRNA-bound) conformation of the HEPN1-2 catalytic cleft.

estimation of mutual information (37), to describe the linear and non-linear couplings of amino acids and nucleobases. Analysis of the per-domain GC reveals that tgRNA binding increases the overall inter-domain correlations in Cas13a, including the catalytic HEPN1-2 domains, with respect to the crRNA- and atgRNA-bound complexes (Figures 3A, S3). This observation reflects the biophysical basis of protein allostery (43,44). Accordingly, the binding of an allosteric effector—in the present case the tgRNA that activates Cas13a—shifts the protein conformational landscape (32). This results in an overall increase of coupled motions, which mediate the communication between distant sites. Correlations between the crRNA and the Cas13a domains are found to improve in the ternary complexes, compared to the crRNA–Cas13a. Dynamical

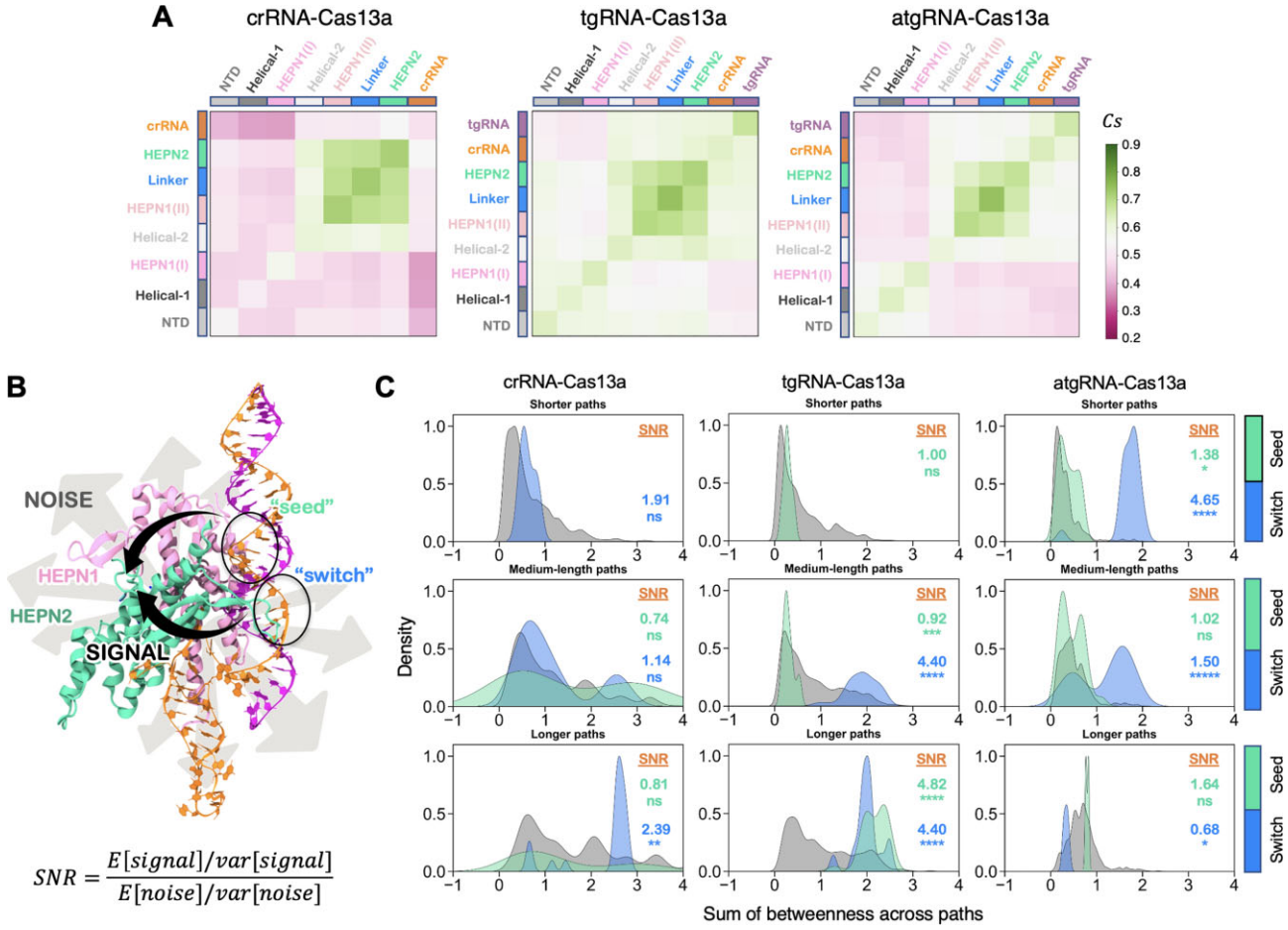


Figure 3. Coupled dynamics and signal-to-noise ratio of communication efficiency. (**A**) Generalized inter-domain correlation maps for the crRNA-, tgRNA- and atgRNA-bound Cas13a complexes. A pink–to–green colour code is used to describe low-to-high correlations. (**B**) Schematic representation of the allosteric signals from the crRNA 'seed' (nt. 9–14) and 'switch' (nt. 5–8) regions to the HEPN1-2 catalytic core residues (R472, H1053, H477, R1048) over the noise, computed as the communication between all pairs of crRNA bases and Cas13a residues. Two black arrows are used to indicate the signal standing up over the noise, depicted using grey arrows. The signal-to-noise Ratio (*SNR*) is computed as the mean-variance ratio ($E[]/var[]$) of the signal over the noise (details in the Material and methods). (**C**) Distribution of the signals from the crRNA 'seed' (green) and 'switch' (blue) regions to the catalytic core residues, plotted on the background of noise (grey), across short (6–8 edge counts), medium (9–11), and long paths (12–14). The statistical significance of the signals over the noise was computed using *Z*-score statistics with a two-tailed hypothesis (*P*-value reference: not significant, ns $P > 0.01$, * $P \leq 0.01$, ** $P \leq 0.001$, *** $P \leq 0.0001$, **** $P \leq 0.00001$). Values of *SNR* are also reported for each sourcing region.

differences are further captured by Principal Component Analysis (PCA), which separates the binary crRNA–Cas13a complex from the ternary complexes along the first principal component (PC1), while distinguishing the tgRNA- and atgRNA-bound complexes along PC2 (Supplementary Figure S4). This observation is in line with the domain-wise PCs, which indicate that the main differentiation between the tgRNA- and atgRNA-bound Cas13a systems lies in the PC2 of the NUC domains, particularly the Linker and the HEPN1-2 domains.

Signal-to-noise ratio of communication efficiency

To understand how the observed dynamical differences impact allosteric signalling, we performed graph theory-derived network analysis (38). This approach is suited for the characterization of allosteric mechanisms, as shown in a number of studies (38,45–48), including those performed by our research group (29–33).

We intended to estimate the communication efficiency between the crRNA spacer region and the catalytic cleft in the systems. In this respect, traditional shortest-path measurements are useful to find the most likely communication pathways between sites, but they do not report how the identified pathways stand out (i.e. are favourable) over the entire communication network (38). Assessing this favourability is crucial since the dominant allosteric pathways are particularly effective in transmitting communication between distantly coupled subunits. Hence, we introduced a novel signal-to-noise ratio (*SNR*) estimate of the information transfer, measuring the preference of communication between predefined distant sites – i.e. the signal – over the remaining pathways of comparable length in the network—i.e. the noise. High *SNR* values indicate the preference of the network to communicate through the signal over other noisy routes (see Materials and methods).

First, we computed the optimal (i.e. the shortest) and top five sub-optimal pathways between all crRNA bases and the Cas13a residues, obtaining a distribution of communication efficiency in terms of the sum of edge betweennesses (i.e. the number of shortest pathways that cross the edge, measuring the 'traffic' passing through them). This provided a

comprehensive overview of all communications, constituting the crosstalk noise between crRNA and protein. Then, we computed the optimal and sub-optimal pathways communicating two regions of the crRNA spacer (i.e. 'seed' and 'switch'), which have been reported to be critical for Cas13a activation) (18), with the HEPN1-2 catalytic residues (R472, H477, R1048, H1053). This represents the signal of our interest. To ensure consistency in the scale of comparison, we characterized the *SNR* across short (6–8 edge counts), medium (9–11), and long paths (12–14), based on the number of edges communicating the crRNA regions with the catalytic residues (details in Supplementary Text, Supplementary Figure S5). The obtained *SNR* indicates the extent to which the signal deviates from the distribution of noise, thus reflecting the prevalence of the signal over the noise (Figure 3B). Little–to–no overlap of the signal with the noise (i.e. high *SNR*) indicates the prevalence of the allosteric signal.

In the crRNA-Cas13a, broad noise distributions overlap with the signals irrespective of the path lengths, dampening the *SNR* compared to the tgRNA-Cas13a. In the tgRNA-Cas13a, amplification of signals from the 'switch' regions for medium–to–long path lengths indicate that tgRNA binding improves the crosstalk efficiency between the crRNA spacer and the HEPN1-2 catalytic residues. The 'seed' region also amplifies the signal across longer paths. Signals sourcing from other regions of the crRNA spacer were also computed in the tgRNA-bound complex (Supplementary Figure S6), reporting a lower *SNR* with respect to the 'switch' and 'seed' regions. This observation agrees with previous biochemical data (18), suggesting that the complementarity at the 'switch' region could trigger the allosteric activation of HEPN1-2, with mismatches in this region preventing LbuCas13a activation. In the atgRNA-Cas13a, a modest *SNR* is detected over medium and long path lengths, while the communication increases over shorter paths for signals sourcing from the 'switch' region.

Overall, the communication is remarkably efficient in the tgRNA-bound Cas13a, in line with the experimental suggestion that tgRNA binding allosterically triggers activation of the HEPN1-2 domains (18). As noted above, allosteric phenomena are commonly associated with a shift in dynamics, resulting in the efficient transfer of signals. Our findings thereby suggest that efficient signalling is associated with a shift in correlations that prioritize the communication signals over communication noise. Therefore, shifted correlations result in noise-free communication pathways for allosteric signals.

Key residues for allosteric coupling

The *SNR* revealed that the communication signal from the crRNA spacer to the HEPN1-2 catalytic cleft is remarkably efficient in the tgRNA-Cas13a. To further understand the signal transduction mechanism in this system, we computed the signalling pathways (i.e. the optimal and top five suboptimal paths) connecting the 'switch' and 'seed' regions of the crRNA to the catalytic core residues (Figure 4A).

The pathways connecting the 'switch' region to the catalytic residues exclusively follow a route that directly connects the crRNA bases of the repeat region (A(-5)–C(-1)) to the catalytic core through the HEPN1(I)-2 interface (Figures 4B, S7A). On the other hand, the 'seed' region communicates with the catalytic core through multiple routes, involving the Linker-HEPN2 interface, the HEPN1(II) residues, as well as the HEPN1(I)-2 interfacial residues through the crRNA repeat, similar to the communication observed for the 'switch'.

Hence, the pathways connecting the 'switch' nucleotides to the catalytic residues display a lower number of residues with increased occurrences, tracing a more efficient communication path, compared to the pathways connecting the 'seed'. This observation further affirms the critical role of the 'switch' in the allosteric activation of HEPN1-2 (18). Extending our analysis to 50 suboptimal pathways reported a similar trend (Supplementary Figure S7B), where a direct route consistently connects the 'switch' region to the HEPN1(I)-2 catalytic core through the crRNA repeat bases A(-4) and A(-3). This evidence pinpoints a pivotal role of the crRNA repeat region in signal transmission.

To better understand our observations, we analysed the interactions between the crRNA repeat region (A(-5)–(C-1)) and the proximal HEPN1(I)-2 interface (residues: 371–383; 963–975). Notably, this interface region is located distally with respect to the catalytic cleft (Figure 4A). We observed that in the tgRNA-bound system, the A(-3) base of the crRNA repeat region penetrates the interface (Figure 4C), hampering interactions between HEPN1(I) and HEPN2. This is confirmed by the histogram of differential contact stability ($\Delta f_{crRNA-tgRNA}$, Supplementary Figure S8A), showing that, in the crRNA-Cas13a, the flipped-out A(-3) base causes increased interactions at the HEPN1(I)-2 interface, with respect to the tgRNA-bound system. In the atgRNA-Cas13a, the crRNA repeat region is sequestered due to the extended tag–anti-tag complementarity, leading to stable interfacial contacts compared to the tgRNA-Cas13a (Supplementary Figure S8B).

To detail the interactions of the A(-3) base in the tgRNA-bound Cas13a, we conducted an in-depth contact analysis at the HEPN1(I)-2 interface. A Sankey plot was used to report the frequency, f, of stable contacts between residues of the HEPN1(I) and HEPN2 domains, and the crRNA, forming for $\geq$10% of the simulation time ($f \geq 0.1$) in the tgRNA-Cas13a (Figure 4D, details in Supplementary Text). In this plot, residues are connected through edges, whose width is proportional to f. We observe that A(-3) substantially interacts with polar/positively charged residues, mainly R377, N378, and R963. Analysis of the differential contact stability was also carried out to characterize interactions that gain stability in tgRNA-Cas13a, compared to the crRNA-Cas13a (details in Supplementary Text).

As expected, a loss of stable contacts at the HEPN1(I)-2 interface is observed in the tgRNA-Cas13a, compared to the crRNA-bound crRNA system (Supplementary Figure S9A). Upon tgRNA binding, R377 and N378 of HEPN1(I) gain interactions with A(-3); and R963 and R973 of HEPN2 increase their contacts with the neighbouring bases, compared to the crRNA-Cas13a (Supplementary Figure S9B). To test whether this observation is sequence dependent, we substituted the A(-3) base with a smaller cytosine in the tgRNA-Cas13a and carried out an additional ~5 μs-long simulation. In this system, the C(-3) base stably locates at the HEPN1(I)-2 interface and interacts with R377, N378, and R963 (Figure 4E). Taken together, these observations suggest that these rearranged interactions involving charged/polar residues between HEPN1(I)-2 could be critical for allosteric signalling from the crRNA spacer to the catalytic core.

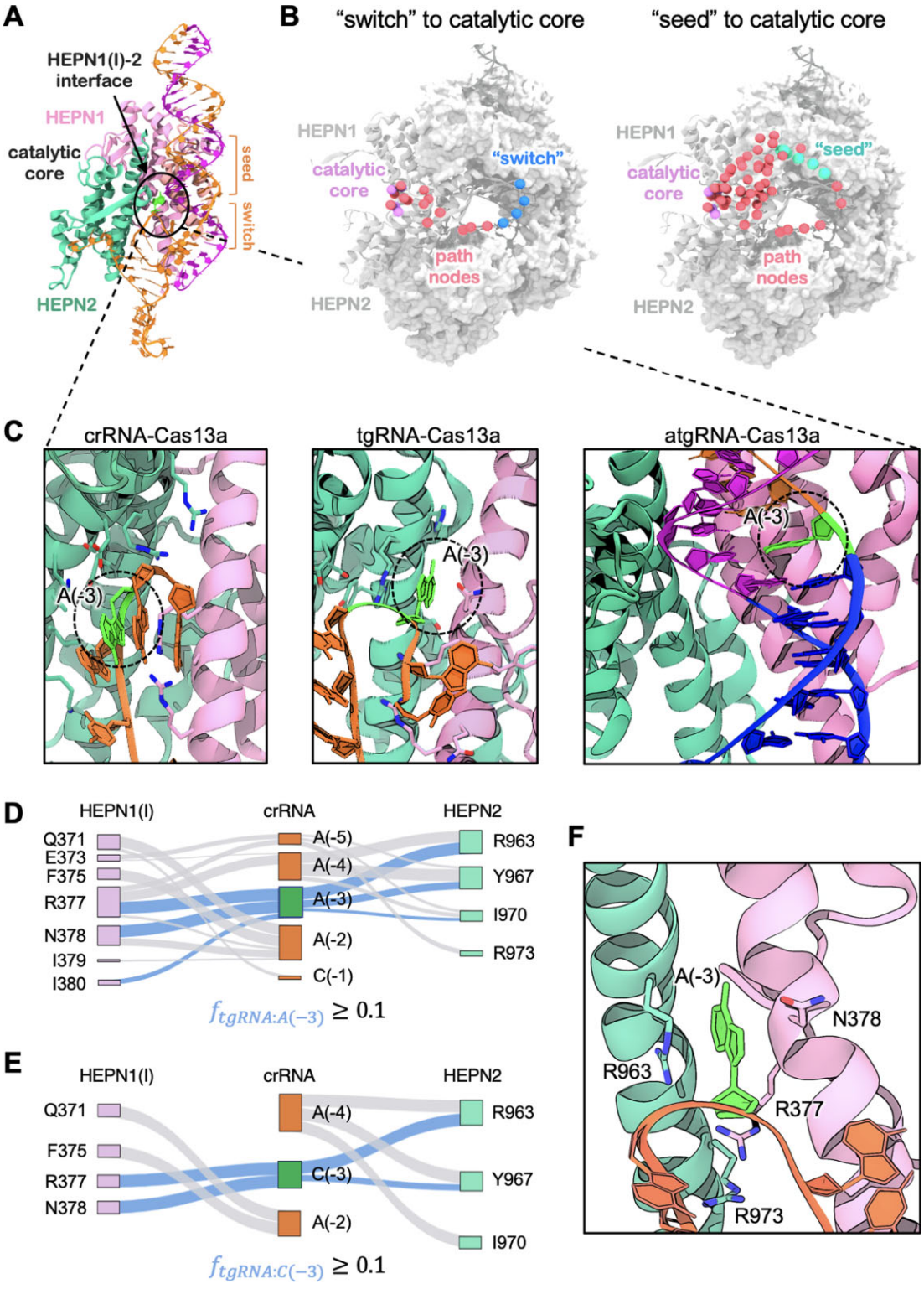


Figure 4. Allosteric signalling from the crRNA 'switch' and 'seed' regions in the tgRNA-Cas13a. (**A**) Overview of the crRNA 'seed' and 'switch' regions binding the HEPN1**(I)**-2 interface, and location with respect to the catalytic core. (**B**) Signalling pathways connecting the crRNA 'switch' (left) and 'seed' (right) regions to the HEPN1-2 catalytic residues (R472, H1053, H477, R1048). Residues occurring in the top five optimal pathways of communication are plotted on the three-dimensional structure of CRISPR-Cas13a (grey). Residues are plotted using spheres of different colours according to the region of interest: 'switch' (blue), 'seed' (green), catalytic residues (pink), and remaining path residues (red). (**C**) Conformation of the A(-3) base (green) of the crRNA repeat region, in proximity to the HEPN1**(I)**-2 interface of the three RNA-bound complexes. (D, E) Sankey plots reporting the frequency, f, of formed contacts between residues of the HEPN1**(I)** and HEPN2 domains, and the crRNA, forming for $\geq$ 10% of the simulation time ($f \geq 0.1$) in the tgRNA-Cas13a **(D)**, and in the system substituting the A(-3) base with cytosine **(E)**. Residue pairs of HEPN1(I) (left), HEPN2 (right), and the crRNA bases (centre) are connected by edges whose width is proportional to f. Contact edges that involve A/C(-3) are shown in blue, to highlight them with respect to the remaining interactions (grey). (**F**) Close-up view of the HEPN1(I)-2 interface in the tgRNA-bound Cas13a. Four polar/positively charged residues (R377, N378, R963, and R973; cyan) have been mutated in alanine to explore their impact on Cas13a activity.

Role of HEPN1-2 interfacial interactions

To experimentally verify our observations, we generated and purified the wild-type (WT) Cas13a and four variants, mutating charged/polar residues to alanine at the HEPN1(I)-2 interface (i.e. R377A, N378A, R963A, and R973A, Figure 4F). We designed and generated tgRNAs and crRNAs containing the spacer used by Liu *et al.*(19) (PDB: 5XWP, Figure 5A) and an anti-tag containing target RNA (atgRNA), holding an extended 8-nt. sequence with complementarity to the crRNA direct repeat (Figure 5B). We performed fluorescent RNA transcleavage assays with WT LbuCa13a and the variants bound to our tgRNA and atgRNA sequences (see *Material and Methods*). In the WT LbuCas13a, there is robust activation of the nuclease activity with a tgRNA, while the presence of an atgRNA results in a small decrease in the apparent cleavage rates over time (Figure 5C), albeit the amount of end-point cleavage product was only slightly reduced relative to the tgRNA (Figure 5D). In our LbuCas13a variants, N378A and R973A maintained a robust activity for the tgRNA, the R377A variant suffered a reduction in cleavage efficiency and R963A was not active at all (Figure 5D). In the presence of an atgRNA, none of the variants displayed any significant nuclease activation, suggesting that these variants are more sensitive to anti-tag containing RNAs than the WT LbuCas13a (Figure 5C, D). To further validate these observations, we obtained crRNA-tgRNA (Figure 5E) and crRNA-atgRNA pairs (Figure 5F) corresponding to the sequences used by Wang *et al.* (PDB:7DMQ) (20). Our cleavage assays showed a similar activation pattern to our data above, where our LbuCas13a variants show increased inhibition in the presence of an atgRNA (Figure 5G, H).

Taken together, these experimental data, using a 28 nt. spacer similar to the simulations, reveal that the WT LbuCas13a maintains its activity irrespective of tgRNA or atgRNA binding. On the other hand, our variants still cleave RNA with tgRNA bound, while hampered nuclease activation in the presence of an extended tag–anti-tag complementarity is observed.

Cas13a variants reorganize the allosteric communication network

To evaluate how the signalling transfer is affected by the mutations we made, we collected a ∼15 μs ensemble for each of our four Cas13a variants bound to a tgRNA and atgRNA and compared them with the corresponding WT Cas13a complexes. Analysis of the *JSD* and SASA distributions reveal that the conformational dynamics of the HEPN1-2 catalytic cleft are perturbed by single mutations at the HEPN1(I)-2 interface compared to the WT Cas13a (Supplementary Figure S10). We then analysed the signalling transfer by comparing the maximal *SNR* detected in the variants with that of the WT Cas13a, considering all signals sourcing from the critical 'switch' region, and establishing a consistent scale for comparison (Figures 6A, S11–12). The highest *SNR* across different path lengths indicates the most favoured communication route in the system, irrespective of path lengths. We thereby computed the ratio between the maximal *SNR* in the variants and in the WT Cas13a ($SNR_{ratio} = SNR_{max-variant}/SNR_{max-wt}$). This comparison indicates whether point mutations at the HEPN1(I)-2 interface impact the strength of the communication between the crRNA 'switch' and the catalytic residues compared to the WT.

We observe that in the tgRNA-bound systems, the R377A, N378A, and R973A variants maintain a high SNR_{ratio}, as evidence of efficient communication compared to the WT Cas13a (Figure 6A). On the other hand, R963A reduces the signal with respect to the WT, indicating altered communication. Upon atgRNA binding, the SNR_{ratio} reaches 40–60% of perturbations in all variants with respect to the WT, with R973A maintaining a SNR_{ratio} approximately within 30% of the WT. This suggests that, in the atgRNA-bound variants, the signalling from the 'switch' to the HEPN1-2 catalytic core is reduced. This is in line with the experimental activity, showing that none of the variants displays a significant nuclease activation in the presence of an atgRNA (Figure 5C, D).

To further understand the signalling transfer in the tgRNA-bound variants, and the observation of a reduced SNR_{ratio} in the inactive R963A mutant, we analysed the interactions of the crRNA repeat bases and the proximal HEPN1(I)-2 interface. In the WT tgRNA-Cas13a, the A(-3) base forms stable contacts with R377, N378, and R963, along with several other HEPN2 residues (Figure 4D). These interactions are preserved in the tgRNA-bound mutants except, as expected, for the mutated residue (Supplementary Figure S13). Nevertheless, in the tgRNA-bound R963A mutant, the A(-3) base is more flexible and is frequently extruded from the HEPN1(I)-2 interface (Figure 6B). The A(-3) conformations are monitored on a polar plot reporting the distance d, which describes the displacement of A(-3) with respect to the Cα atom at position 963, and the dihedral angle θ, reporting the rotation of the A(-3) purine base with respect to the crRNA backbone (Figure 6C, S14). The polar plot evidences the higher flexibility of A(-3) in the R963A mutant, compared to the remaining systems, and its extrusion from the HEPN1(I)-2 interface. This observation can be ascribed to the loss of interaction between the R963 guanidinium side chain and the crRNA phosphate backbone that, on the other hand, is maintained in the other systems (Supplementary Figure S15). We recall that the R963A substitution hampers the tgRNA-Cas13a activity (Figure 5). This suggests that the positioning of the A(-3) base, and the dynamics of the crRNA repeat region at the HEPN1(I)-2 interface, critically affects the transmission of the signal from the 'switch' to the catalytic core, as evidenced by lower *SNR* with respect to the WT.

To characterize the allosteric communication in our atgRNA-bound variants, and how it compares to the tgRNA-Cas13a, we inspected the pathways communicating the 'switch' to the HEPN1-2 catalytic core. In the atgRNA-bound WT Cas13a, allosteric pathways mainly involve the HEPN1(I), HEPN1(II), and HEPN2 interfaces and the tgRNA (Figures 6D, S16), at odds with the direct routes passing through the crRNA repeat and HEPN1(I)-2 observed in the tgRNA-Cas13a (Figure 4B and S6). When introducing our point mutations in the tgRNA-bound Cas13a, the major routes of communication are similar to those of the WT system, with the addition of a few residues from HEPN1(II) (Supplementary Figure S17A). In the atgRNA-bound variants, allosteric pathways are more sensitive to perturbations, compared to the tgRNA-bound variants. In the presence of an atgRNA, the variants increase the number of crRNA bases involved in the communication with respect to WT atgRNA-Cas13a, with the N378A and R963A variants also losing the tgRNA communication channel (Supplementary Figure S17B). An analysis of the interactions at the major interfaces along the allosteric pathways (i.e. HEPN1(I), HEPN1(II) and

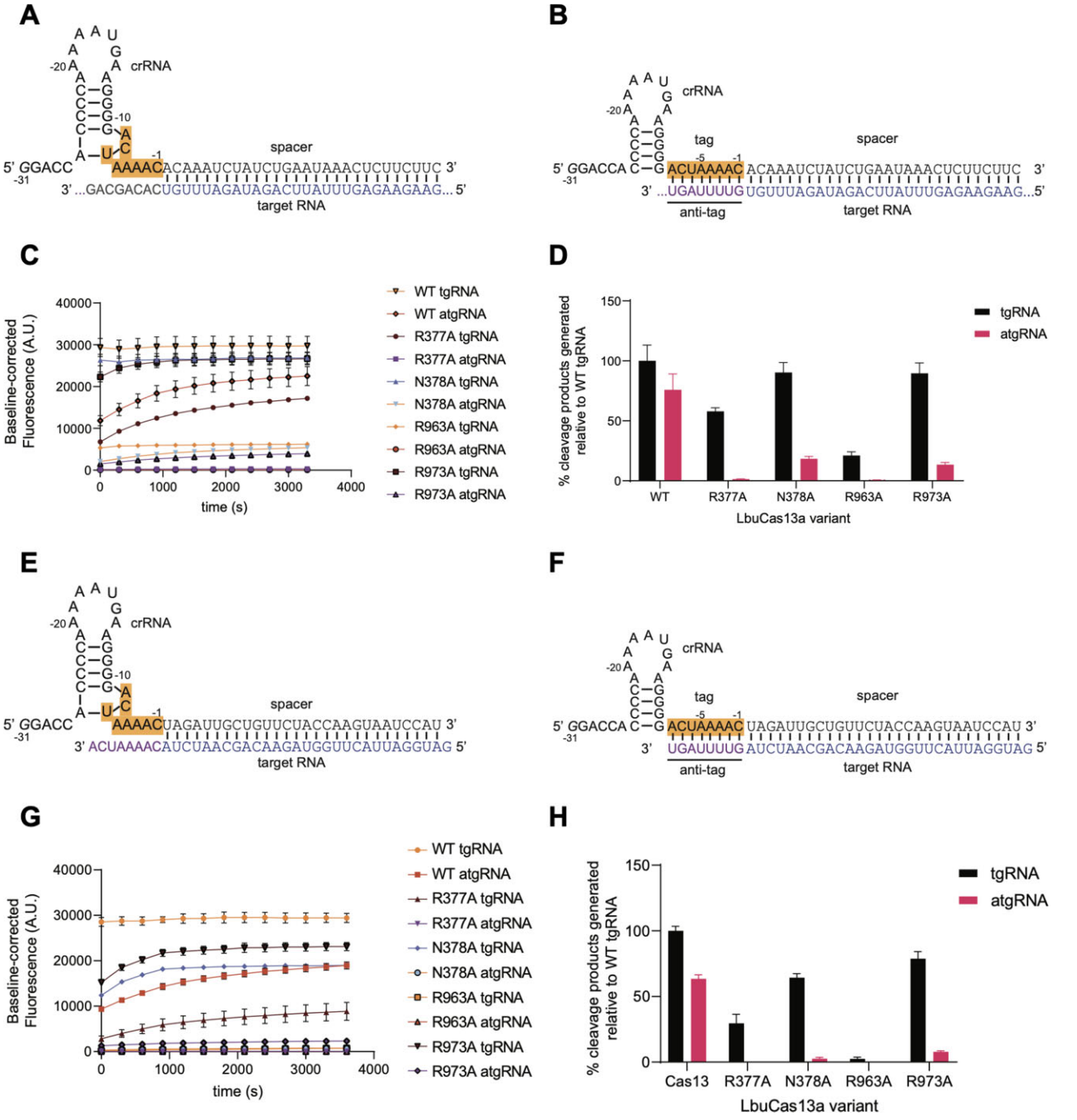


Figure 5. LbuCas13a variants are more sensitive to inhibition by anti-tag-containing RNAs. (A, B) crRNA-target RNA (tgRNA) pair (**A**) and crRNA-anti-tag RNA (atgRNA) pair (**B**), designed based on Liu *et al.* (19). (**C**) One-hour time course of background-corrected fluorescence measurements from RNA cleavage experiments by LbuCas13a WT and variants initiated by the addition of 100 pM tgRNA or atgRNA shown in (A) and (B). (**D**) Data from (C) normalized as percent cleavage product generated relative to WT LbuCas13a with tgRNA. (E, F) crRNA-target RNA (tgRNA) pair (**E**) and crRNA-anti-tag RNA (atgRNA) pair (**F**), designed based on Wang *et al.* (20). (**G**) One-hour time course of background-corrected fluorescence measurements from RNA cleavage experiments by LbuCas13a WT and variants initiated by the addition of 100 pM tgRNA or atgRNA shown in (**I**) and (F). (**H**) Data from (G) normalized as percent cleavage product generated relative to WT LbuCas13a with tgRNA.

HEPN2, Figure 6E) also shows that the mutations result in a reduction of stable contacts at the HEPN1(I)-HEPN2 interface with respect to the WT Cas13a, while gained back at the HEPN1(I)-HEPN1(II) interface. These rearrangements at the critical HEPN1(I)-(II)-HEPN2 interfaces are the basis of the altered signalling pathways observed in the atgRNA-bound variants (Supplementary Figure S17).

In summary, analysis of the allosteric communication in our mutants reveals that the inactive mutants reduce the signal with respect to the WT. This altered communication can be attributed to the dynamics of the crRNA repeat region in the tgRNA-bound systems, and to an overall perturbation of the crosstalk pathways in the presence of an atgRNA. All in all, alterations of the protein-RNA interactions at the

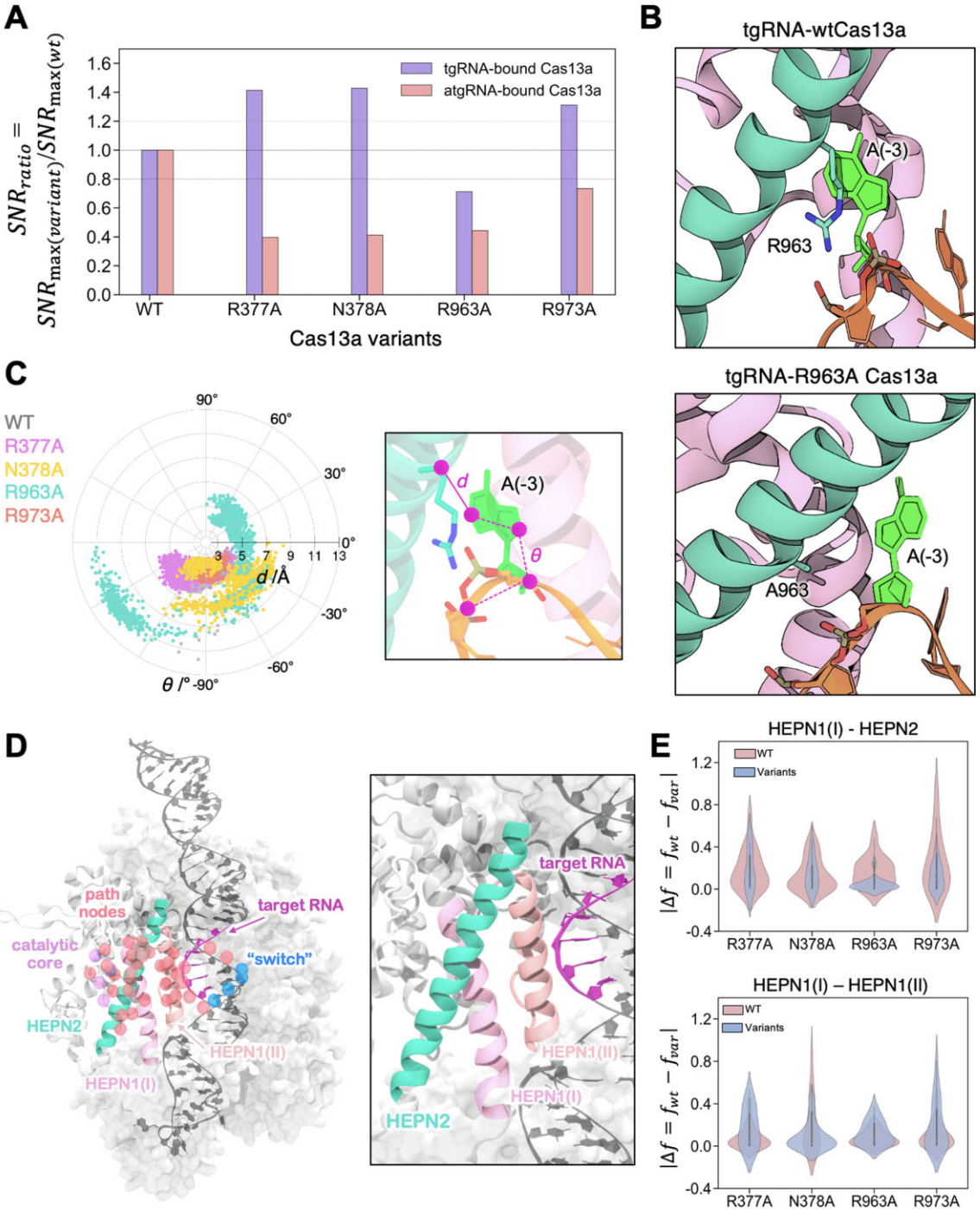


Figure 6. Mutation-induced reorganized communication network. (**A**) Ratio between the maximal Signal-to-Noise Ratio (SNR_{max}) in our Cas13a variants and the WT Cas13a ($SNR_{ratio} = SNR_{max-variant}/SNR_{max-wt}$), computed for the tgRNA- (violet) and atgRNA- (mauve) bound systems. (**B**) Close-up view of the HEPN1(I)-HEPN2 interface in the WT tgRNA-Cas13a (top) and its R963A mutant (bottom), showing the extrusion of A(-3) from the protein in the R963A mutant. (**C**) Polar plot of (*i*) the distance *d* between the Cα atom at position 963 and the centre of mass of the N1-C6 ring (polar coordinate, in Å), and (*ii*) the dihedral angle θ between the C3'@A(-4), C3'@A(-3), C8@A(-4) and C2@A(-4) atoms (angular coordinate, in degrees), computed from the simulated ensembles of the WT tgRNA-bound Cas13a and its mutants. The *d* distance and θ angle are shown on the right. (**D**) Signalling pathways connecting the crRNA 'switch' region to the HEPN1-2 catalytic residues (R472, H1053, H477, R1048) in the atgRNA-Cas13a. A close-up view of the HEPN1(I), HEPN1(II) and HEPN2 interfaces involved in the allosteric pathways is reported on the right. (**E**) Distributions of differential contact stability at the HEPN1(I)-HEPN2 (top) and HEPN1(I)-HEPN1(II) (bottom) interfaces, between the atgRNA-Cas13a and its variants ($|\Delta f = f_{wt} - f_{var}|$).

HEPN1(I)-2 interface in the mutants with respect to the WT, point for the allosteric regulation of the activity.

Discussion

Here, extensive MD simulations were combined with graph theory and experimental assays to characterize the allosteric activation mechanism in Cas13a, a newly emerged CRISPR-Cas system, which is being developed as a powerful tool for RNA cleavage, detection and imaging.

Multiple-μs simulations reveal that the binding of a target RNA (tgRNA) at the recognition lobe impacts the dynamics of the spatially distant HEPN1-2 catalytic core, resulting in altered dynamics with respect to the crRNA-bound form (Figure 2A-B), and in the opening of the catalytic cleft (Figure 2C). This is in line with biochemical data, suggesting that tgRNA binding allosterically activates HEPN1-2 to form a composite active site (15,18). In the presence of an extended tag–anti-tag pairing (atgRNA), the HEPN1-2 catalytic cleft increases its flexibility, compared to both the crRNA- and tgRNA-Cas13a, indicating that the length of the complementarity also affects the HEPN1-2 dynamics.

Analysis of mutually coupled motions shows that tgRNA binding induces a shift in the system's dynamics (Figure 3A), evidenced by increased inter-domain correlations. Coupled dynamics are also found in the presence of an atgRNA, mainly involving the crRNA and the Cas13a domains. This 'shift in dynamics' is typical evidence of an allosteric response (32,43), reinforcing the notion of the tgRNA as an effector of Cas13a's allosteric function.

To better understand how the observed differences in the systems' dynamics impact the flux of information from the sites of tgRNA binding to the catalytic core, we estimated the crosstalk efficiency across the studied systems. We analysed how the communication signal from the crRNA (site of tgRNA binding) to the HEPN1-2 cleft emerges over the remaining pathways (i.e. the noise) using a signal-to-noise ratio (*SNR*) measure. We found that, upon tgRNA binding, the signal from the 'switch' region of the crRNA amplifies over the noise, providing evidence of efficient crosstalk (Figure 3B). The communication sourcing from the 'switch' is also maintained upon atgRNA binding. These observations underscore the critical role of this region in the allosteric signalling between the sites of effector binding and RNA cleavage. We recall that, in line with our observations, biochemical data have noted the 'switch' region to be crucial in triggering the allosteric activation of HEPN1-2 (18).

Analysis of the allosteric pathways also shows that the 'switch' region efficiently communicates with the HEPN1-2 catalytic core through the crRNA bases of the repeat region (i.e. A(-5)-C(-1), Figure 4B). This pinpoints a cardinal role for the crRNA repeat region, which prompted us to analyse its interactions with Cas13a. We examined the interactions with the proximal HEPN1(I)-2 interface, notably located distally with respect to the catalytic cleft (Figure 4C). We observe that in the tgRNA-Cas13a, the A(-3) base of the crRNA repeat penetrates the HEPN1(I)-2 interface, decreasing inter-domain interactions compared to the crRNA-Cas13a where A(-3) is extruded. The extended tag–anti-tag complementarity also sequesters A(-3) from the HEPN1(I)-2 interface. The interaction network further reveals several polar/charged residues critical for HEPN(I)-2 binding, that exhibit increased interactions with A(-3) upon tgRNA biding (Figure 4D-E). In detail, N378, R963A, R973 and R377A rearrange their interactions at the HEPN(I)-2 interface upon tgRNA binding, which could be critical in mediating the allosteric information transfer process.

To experimentally test the role of these polar/charged residues in the allosteric activation of Cas13a, we performed mutagenesis and RNA cleavage experiments. Alanine substitution of N378, R377, and R973 maintains a robust activity for the tgRNA, while R963A is not active (Figure 5). In the presence of an atgRNA, none of the variants displays significant nuclease activation. This concludes that our N378A, R377A, and R973A variants can discriminate tgRNA for cleavage over atgRNA. Our computational analysis and the experimental assays thereby disclose polar/charged residues that are pivotal for the allosteric signalling and whose mutation in alanine can push the preference for tgRNA-activated cleavage vs. atgRNA-activated cleavage.

To understand the mechanistic function of our mutants, we performed additional multi-μs long simulations of our variants. Analysis of the allosteric signal sourcing from the 'switch' region to the HEPN1-2 catalytic cleft, shows that our variants mostly perturb the *SNR* upon atgRNA binding (with respect to the WT atgRNA-Cas13a, Figure 6A), while the tgRNA-bound variants preserve efficient signalling compared to the WT tgRNA-Cas13a. Interestingly, the tgRNA-bound R963A Cas13a, which is experimentally shown to be inactive (Figure 5D), reduces the signal with respect to the WT, indicating altered communication. It is also notable that R963 is among the highest-occurring residues in the allosteric routes of the WT Cas13a (Supplementary Figure S7), while it does not appear in the allosteric pathways of the tgRNA-bound R963A mutant (Supplementary Figure S17A). Hence, mutating a residue belonging to the WT allosteric pathway has a significant impact on the activity, while other residues not included in the path show a less dramatic impact, in particular when the allosteric effector is bound (tgRNA). This further supports that the allosteric pathway connecting the sites of tgRNA binding with the catalytic cleft controls the activity in the WT Cas13a. In the R963A-tgRNA system, the A(-3) base is frequently extruded from the HEPN1(I)-2 interface (Figure 6C), suggesting that the positioning of the A(-3) base (and the dynamics of the crRNA repeat region) critically affects the transmission of the signal from the 'switch' to the catalytic core, as evidenced by lower *SNR* with respect to the WT.

Analysis of the allosteric pathways also shows that our variants preserve the communication connecting the 'switch' to the catalytic cleft in the tgRNA-bound system (Supplementary Figure S17A). On the other hand, in the presence of a tag–anti-tag, our mutations perturb and reduce the conservation of the signalling pathways observed in the WT atgRNA-Cas13a (Supplementary Figure S17B). This observed perturbation of the allosteric crosstalk pathways, observed in the atgRNA-bound variants, along with altered *SNR* compared to the WT, is qualitatively consistent with the experimentally observed inactivity of the atgRNA-bound variants (Figure 5). Overall, the inactive mutants (i.e. the atgRNA-bound variants and the tgRNA-bound R963A) reduce the signal with respect to the WT. This reduced signalling can be ascribed to the dynamics of the crRNA repeat region in the tgRNA-bound systems, and to an overall perturbation of the crosstalk pathways in the presence of an atgRNA.

Conclusions

In summary, our study characterizes the allosteric activation mechanism of Cas13a and discloses critical point mutations able to promote tgRNA-mediated over an extended tag-anti-tag-mediated complementarity, for RNA cleavage activation. We show that the binding of a tgRNA acts as an allosteric effector of the spatially distant HEPN1-2 catalytic cleft, by amplifying the allosteric signals that connect the sites of tgRNA binding with the HEPN1-2 catalytic site. By introducing a novel graph theory-based analysis—signal-to-noise ratio (*SNR*) of communication efficiency—we show that the allosteric signal stands out over the dynamical noise when passing through the crRNA repeat region. Critical residues at this interface (R377, N378, and R973) rearrange their interactions upon tgRNA binding and are experimentally shown to select tgRNA, over an extended atgRNA, for RNA cleavages. Considering this selectivity, we speculate here that alanine mutation of R377, N378, and R973 could improve (or alter) the selectivity of Cas13a. This hypothesis is confirmed in our companion paper by Molina-Vargas *et al.* (49), showing that our computational approach can guide the development of more selective RNA cleavage and detection strategies. Taken together, our findings offer a fundamental understanding of the CRISPR-Cas13a mechanistic function, and pave the way for harnessing innovative computational approaches for engineering innovative RNA-based cleavage and detection tools.

Data availability

All in-house Python scripts are available in Figshare at https://doi.org/10.6084/m9.figshare.24498937. Further data are available from the corresponding authors upon reasonable request.

Supplementary Data

Supplementary Data are available at NAR Online.

Acknowledgements

Author contributions: Souvik Sinha: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing-original draft. Adrian M. Molina Vargas: Data curation, Investigation, Formal Analysis, Writing—review & editing. Pablo R. Arantes: Formal Analysis, Methodology, Validation, Visualization, Writing—review & editing. Amun Patel: Formal Analysis, Methodology, Validation, Visualization, Writing—review & editing. Mitchell R. O'Connell: Supervision, Project administration, Funding acquisition, Writing—review & editing. Giulia Palermo: Conceptualization, Supervision, Project administration, Funding acquisition, Writing—original draft.

Funding

National Institutes of Health [R01GM141329 to G.P., R35GM133462 to M.R.O.]; National Science Foundation [CHE-2144823 to G.P.]. Funding for open access charge: National Institutes of Health; part of this work used Expanse at the San Diego Supercomputer Center through allocation MCB160059 and Bridges2 at the Pittsburgh Supercomputer Center through allocation BIO230007 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation [2 138 259, 2 138 286, 2 138 307, 2 137 603, 2138296]; computer time was also provided by the National Energy Research Scientific Computing Center (NERSC) [M3807].

Conflict of interest statement

S.S., A.M.M.V., P.R.A., A.P, M.R.O. and G.P. are co-inventors on patent applications filed by the University of Rochester and University of California, Riverside relating to work in this manuscript. M.R.O. is an inventor of patent applications related to CRISPR-Cas systems and uses thereof. M.R.O. is a member of the scientific advisory boards for Dahlia Biosciences and Locana Bio, and an equity holder in Dahlia Biosciences and LocanaBio.

References

1. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
2. Doudna,J.A. (2020) The promise and challenge of therapeutic genome editing. *Nature*, **578**, 229–236.
3. Knott,G.J. and Doudna,J.A. (2018) CRISPR-Cas guides the future of genetic engineering. *Science*, **361**, 866–869.
4. Anzalone,A.V., Koblan,L.W. and Liu,D.R. (2020) Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.*, **38**, 824–844.
5. Zhang,F. (2019) Development of CRISPR-Cas systems for genome editing and beyond. *Q. Rev. Biophys.*, **52**, e6.
6. Abudayyeh,O.O., Gootenberg,J.S., Konermann,S., Joung,J., Slaymaker,I.M., Cox,D.B.T., Shmakov,S., Makarova,K.S., Semenova,E., Minakhin,L., *et al.* (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, **353**, 6299.
7. East-Seletsky,A., O'Connell,M.R., Knight,S.C., Burstein,D., Cate,J.H.D., Tjian,R. and Doudna,J.A. (2016) Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*, **538**, 270–273.
8. Cox,D.B.T., Gootenberg,J.S., Abudayyeh,O.O., Franklin,B., Kellner,M.J., Joung,J. and Zhang,F. (2017) RNA editing with CRISPR-Cas13. *Science*, **358**, 1019–1027.
9. Ackerman,C.M., Myhrvold,C., Thakku,S.G., Freije,C.A., Metsky,H.C., Yang,D.K., Ye,S.H., Boehm,C.K., Kosoko-Thoroddsen,T.-S.F., Kehe,J., *et al.* (2020) Massively multiplexed nucleic acid detection with Cas13. *Nature*, **582**, 277–282.
10. Gootenberg,J.S., Abudayyeh,O.O., Kellner,M.J., Joung,J., Collins,J.J. and Zhang,F. (2018) Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a and Csm6. *Science*, **360**, 439–444.
11. Gootenberg,J.S., Abudayyeh,O.O., Lee,J.W., Essletzbichler,P., Dy,A.J., Joung,J., Verdine,V., Donghia,N., Daringer,N.M., Freije,C.A., *et al.* (2017) Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*, **356**, 438–442.
12. Iwasaki,R.S. and Batey,R.T. (2020) SPRINT: a Cas13a-based platform for detection of small molecules. *Nucleic Acids Res.*, **48**, e101.
13. Kellner,M.J., Koob,J.G., Gootenberg,J.S., Abudayyeh,O.O. and Zhang,F. (2019) SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nat. Protoc.*, **14**, 2986–3012.
14. Knott,G.J., East-Seletsky,A., Cofsky,J.C., Holton,J.M., Charles,E., O'Connell,M.R. and Doudna,J.A. (2017) Guide-bound structures

of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme. *Nat. Struct. Mol. Biol.*, **24**, 825–833.
15. Liu,L., Li,X., Wang,J., Wang,M., Chen,P., Yin,M., Li,J., Sheng,G. and Wang,Y. (2017) Two distant catalytic sites are responsible for C2c2 RNase activities. *Cell*, **168**, 121–134.
16. O'Connell,M.R. (2019) Molecular mechanisms of RNA targeting by Cas13-containing type VI CRISPR–Cas systems. *J. Mol. Biol.*, **431**, 66–87.
17. Fozouni,P., Son,S., Díaz de León Derby,M., Knott,G.J., Gray,C.N., D'Ambrosio,M.V., Zhao,C., Switz,N.A., Kumar,G.R., Stephens,S.I., *et al.* (2021) Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy. *Cell*, **184**, 323–333.
18. Tambe,A., East-Seletsky,A., Knott,G.J., Doudna,J.A. and O'Connell,M.R. (2018) RNA binding and HEPN-nuclease activation are decoupled in CRISPR-Cas13a. *Cell Rep.*, **24**, 1025–1036.
19. Liu,L., Li,X., Ma,J., Li,Z., You,L., Wang,J., Wang,M., Zhang,X. and Wang,Y. (2017) The molecular architecture for RNA-guided RNA cleavage by Cas13a. *Cell*, **170**, 714–726.
20. Wang,B., Zhang,T., Yin,J., Yu,Y., Xu,W., Ding,J., Patel,D.J. and Yang,H. (2021) Structural basis for self-cleavage prevention by tag:anti-tag pairing complementarity in type VI Cas13 CRISPR systems. *Mol. Cell*, **81**, 1100–1115.
21. Meeske,A.J. and Marraffini,L.A. (2018) RNA guide complementarity prevents self-targeting in type VI CRISPR systems. *Mol. Cell*, **71**, 791–801.
22. Anandakrishnan,R., Aguilar,B. and Onufriev,A.V. (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.*, **40**, W537–W541.
23. Garcia-Doval,C., Schwede,F., Berk,C., Rostøl,J.T., Niewoehner,O., Tejero,O., Hall,J., Marraffini,L.A. and Jinek,M. (2020) Activation and self-inactivation mechanisms of the cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6. *Nat. Commun.*, **11**, 1596.
24. Tian,C., Kasavajhala,K., Belfon,K.A.A., Raguette,L., Huang,H., Migues,A.N., Bickel,J., Wang,Y., Pincay,J., Wu,Q., *et al.* (2020ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.*, **16**, 528–552.
25. Banas,P., Hollas,D., Zgarbova,M., Jurecka,P., Orozco,M., Cheatham,T.E. 3rd, Sponer,J. and Otyepka,M. (2010) Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins. *J. Chem. Theor. Comput.*, **6**, 3836–3849.
26. Zgarbova,M., Otyepka,M., Sponer,J., Mladek,A., Banas,P., Cheatham,T.E. and Jurecka,P. (2011) 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.
27. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
28. Case,D.A., Aktulga,H.M., Belfon,K., Ben-Shalom,I.Y., Brozell,S.R., Cerutti,D.S.T.E., Cheatham,I., Cruzeiro,V.W.D., Darden,T.A., Duke,R.E., *et al.* (2020) In: *AMBER 2020*. Univ. California, San Francisco.
29. Nierzwicki,L., East,K.W., Morzan,U.N., Arantes,P.R., Batista,V.S., Lisi,G.P. and Palermo,G. (2021) Enhanced specificity mutations perturb allosteric signaling in CRISPR-Cas9. *eLife*, **10**, e73601.
30. Arantes,P.R., Patel,A.C. and Palermo,G. (2022) Emerging methods and applications to decrypt allostery in proteins and nucleic acids. *J. Mol. Biol.*, **434**, 167518.
31. East,K.W., Newton,J.C., Morzan,U.N., Narkhede,Y.B., Acharya,A., Skeens,E., Jogl,G., Batista,V.S., Palermo,G. and Lisi,G.P. (2020) Allosteric motions of the CRISPR–Cas9 HNH nuclease probed by NMR and molecular dynamics. *J. Am. Chem. Soc.*, **142**, 1348–1358.
32. Palermo,G., Ricci,C.G., Fernando,A., Basak,R., Jinek,M., Rivalta,I., Batista,V.S. and McCammon,J.A. (2017) Protospacer adjacent motif-induced allostery activates CRISPR-Cas9. *J. Am. Chem. Soc.*, **139**, 16028–16031.
33. Saha,A., Arantes,P.R., Hsu,R.V., Narkhede,Y.B., Jinek,M. and Palermo,G. (2020) Molecular dynamics reveals a DNA-induced dynamic switch triggering activation of CRISPR-Cas12a. *J. Chem. Inf. Model.*, **60**, 6427–6437.
34. Lindorff-Larsen,K. and Ferkinghoff-Borg,J. (2009) Similarity Measures for Protein Ensembles. *PLoS One*, **4**, e4203.
35. Vögele,M., Thomson,N.J., Truong,S.T., McAvity,J., Zachariae,U. and Dror,R.O. (2022) Systematic analysis of biomolecular conformational ensembles with PENSA. arXiv doi: https://arxiv.org/abs/2212.02714, 06 December 2022, preprint: not peer reviewed.
36. Palermo,G., Miao,Y., Walker,R.C., Jinek,M. and McCammon,J.A. (2016) Striking plasticity of CRISPR-Cas9 and key role of non-target DNA, as revealed by molecular simulations. *ACS Cent. Sci.*, **2**, 756–763.
37. Lange,O.F. and Grubmüller,H. (2006) Generalized correlation for biomolecular dynamics. *Proteins Struct. Funct. Genet.*, **62**, 1053–1061.
38. Sethi,A., Eargle,J., Black,A.A. and Luthey-Schulten,Z. (2009) Dynamical networks in tRNA: protein complexes. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 6620–6625.
39. Yen,J.Y. (1971) Finding the *K* shortest loopless paths in a network. *Manage. Sci.*, **17**, 712–716.
40. Hagberg,A., Schult,D. and Swart,P. (2008) Exploring network structure, dynamics, and fusing NetworkX. In: Varoquaux,G., Vaught,T and Millman,J (eds.) *Proceedings of the 7th Python in science conference*. SciPy, pp. 11–15.
41. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
42. East-Seletsky,A., O'Connell,M.R., Burstein,D., Knott,G.J. and Doudna,J.A. (2017) RNA targeting by functionally orthogonal type VI-A CRISPR-Cas enzymes. *Mol. Cell*, **66**, 373–383.
43. Dokholyan,N.V. (2016) Controlling allosteric networks in proteins. *Chem. Rev.*, **116**, 6463–6487.
44. Guo,J. and Zhou,H.X. (2016) Protein allostery and conformational dynamics. *Chem. Rev.*, **116**, 6503–6515.
45. Liu,J. and Nussinov,R. (2016) Allostery: an overview of its history, concepts, methods, and applications. *PLoS Comput. Biol.*, **12**, e1004966.
46. Melo,M.C.R., Bernardi,R.C., de la Fuente-Nunez,C. and Luthey-Schulten,Z. (2020) Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. *J. Chem. Phys.*, **153**, 134104.
47. Dodd,T., Botto,M., Paul,F., Fernandez-Leiro,R., Lamers,M.H. and Ivanov,I. (2020) Polymerization and editing modes of a high-fidelity DNA polymerase are linked by a well-defined path. *Nat. Commun.*, **11**, 5379.
48. Doshi,U., Holliday,M.J., Hamelberg,D. and Eisenmesser,E.Z. (2016) 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.
49. Molina Vargas,A. M., Sinha,S., Osborn,R., Arantes,R.R., Patel,A., Dewhurst,S., Hardy,S.J., Cameron,A., Palermo,G. and O'Connell,M.R. (2023) New design strategies for ultra-specific CRISPR-Cas13a-based RNA detection with single-nucleotide mismatch sensitivity. *Nucleic Acids Res.*, https://doi.org/10.1093/nar/gkad1132.

Received: July 27, 2023. **Revised:** October 25, 2023. **Editorial Decision:** October 31, 2023. **Accepted:** November 9, 2023