

Intersegmental transfers drive target search in an RNA-targeting CRISPR system

Ofer Kimchi^{1,2,*†}, Benjamin B. Larsen^{3,*}, Emily Gibson³, Aartjan J. W. te Velthuis³, and Cameron Myhrvold^{3,4,5,6,†}

¹Department of Mathematics, Courant Institute School of Mathematics, Computing, and Data Science, New York University, New York, NY 10012, USA

²Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey, 08544, USA

³Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

⁴Omenn-Darling Bioengineering Institute, Princeton University, Princeton, NJ 08544, USA

⁵Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA

⁶Department of Chemistry, Princeton University, Princeton, NJ 08544, USA

*These authors contributed equally

†Address correspondence to: C.M. (cmmyhrvol@princeton.edu) or O.K. (ok2294@nyu.edu)

ABSTRACT

Sequence-specific RNA-binding proteins (RBPs) must efficiently locate their targets among a multitude of cellular RNAs. Cas13, an RNA-guided CRISPR protein, represents an ideal model system in which to study this search process. Cas13 combats bacteriophage infection by cleaving RNA nonspecifically upon binding of its crRNA to the target RNA sequence; thus, Cas13's search for its RNA target comes with a time constraint determined by the rate of phage multiplication. The mechanism by which Cas13 locates its target within this critical window remains unknown. Here, we investigate Cas13's mechanism of target search through integration of biophysical modeling, activity assays, and biochemical characterization. We show that Cas13 employs facilitated diffusion to accelerate its search, and find that Cas13's search time when targeting RNAs of different lengths cannot be explained by 1D sliding, the search mechanism used by many DNA-binding proteins. We propose that Cas13 primarily searches for its RNA target by intersegmental transfers (ITs), non-specifically binding the RNA at two locations and directly switching between them without fully dissociating from the RNA. We develop a biophysical model for ITs in an RNA context that we subsequently validate experimentally. Furthermore, we demonstrate that ITs can differentially accelerate the search process for a broad class of RNA-binding proteins, as opposed to their DNA-binding counterparts, due to RNA's short persistence length and the heterogeneity of RNA lengths in the cell. Our results illuminate how Cas13 achieves rapid target recognition in a complex RNA environment, and implicate ITs as a potentially widespread solution to the RNA search problem.

INTRODUCTION

Sequence-specific nucleic acid binding proteins must efficiently search vast physical and sequence space to locate their targets on a biologically relevant timescale. The search processes of RNA-binding proteins (RBPs) in particular present a unique challenge compared to their DNA-binding counterparts. First, the scale of the search is substantial: for example, an *E. coli* cell contains hundreds of thousands of unique RNA molecules (Fig. 1A) including an estimated ~10 times more non-ribosomal RNA nucleotides than DNA base pairs^{1–4}. Moreover, these RNA molecules have a large length distribution, ranging from under 50 nt for the shortest sRNAs to well over 4 kb for the longest mRNA species, in addition to wide variation in features such as GC content and secondary structure⁵.

CRISPR-Cas13 represents an effective model system for studying protein-RNA interactions, including the RNA search process. Cas13 is an RNA-guided, RNA-targeting CRISPR nuclease which has been widely adopted for technological applications including viral diagnostics, RNA knockdown, and RNA imaging^{6–10}. Upon binding its RNA target, Cas13 nonspecifically cleaves other RNAs in solution, enabling a clear readout of Cas13 activity in different environments through the addition of short quenched fluorescent RNA reporters. Cas13's canonical function in bacteria is to provide immunity from bacteriophage infection by sequence-specifically binding to phage mRNAs and subsequently unleashing nonspecific RNA cleavage, which induces cell dormancy and terminates phage infection prior to cell lysis^{11,12}. In the absence of phage defenses such as Cas13, a lytic infection leads to infection-competent progeny within as few as 30 minutes for many phages, at which point it is likely too late to initiate an effective immune response^{13,14}. Cas13 must locate its target RNA quickly enough to mount a successful response during that window, and thereby prevent a runaway phage infection and collapse of the bacterial population (Fig. 1B).

To speed their search beyond the rate set by 3D diffusion, nucleic acid binding proteins typically employ a process termed facilitated diffusion: diffusing in 3D until binding non-specifically to a section of DNA or RNA, then searching locally on the molecule, before dissociating and repeating the process^{15–17}. Several methods for local search have been hypothesized based on studies of DNA-binding proteins: 1D sliding, in which the protein binds nonspecifically to DNA and moves randomly along the coil before dissociating; intracoil diffusion (also sometimes referred to as “hopping” and “jumping”) in which the protein dissociates from the DNA and diffuses in 3D within the coil before reassociating; and intersegmental transfers (ITs), in which the protein utilizes multiple non-specific binding patches to directly transfer between different DNA segments without fully dissociating (Fig. 1C)^{17–19}. The presence or absence of multiple nonspecific binding patches is the key difference between intracoil diffusion and ITs, as the former involves multiple rounds of dissociation and reassociation of the protein from the polymer, while in the latter scheme the protein remains (non-specifically) bound to the polymer throughout the search process. As a result, intracoil diffusion is almost entirely driven by the polymer itself, and is largely invariant to properties of the protein. On the other hand, ITs are driven by both polymer properties and protein properties, and thus can be evolutionarily tuned in different ways by different proteins, just as the distance of 1D sliding can be tuned evolutionarily^{20,21}.

Facilitated diffusion has been observed in numerous DNA-binding proteins, including CRISPR-Cas9 and Cas12 whose local search is dominated by 1D sliding^{21–26}. Though far less studied than their DNA counterparts, the search processes of certain RBPs have also been investigated. Recent work on the RBPs Hfq, RNase E, and Argonautes have argued that they perform 1D sliding as part of their target search^{27,28}. Another study on Argonautes argued that intracoil diffusion plays a dominant role²⁹. To our knowledge, in none of these studies of RNA-

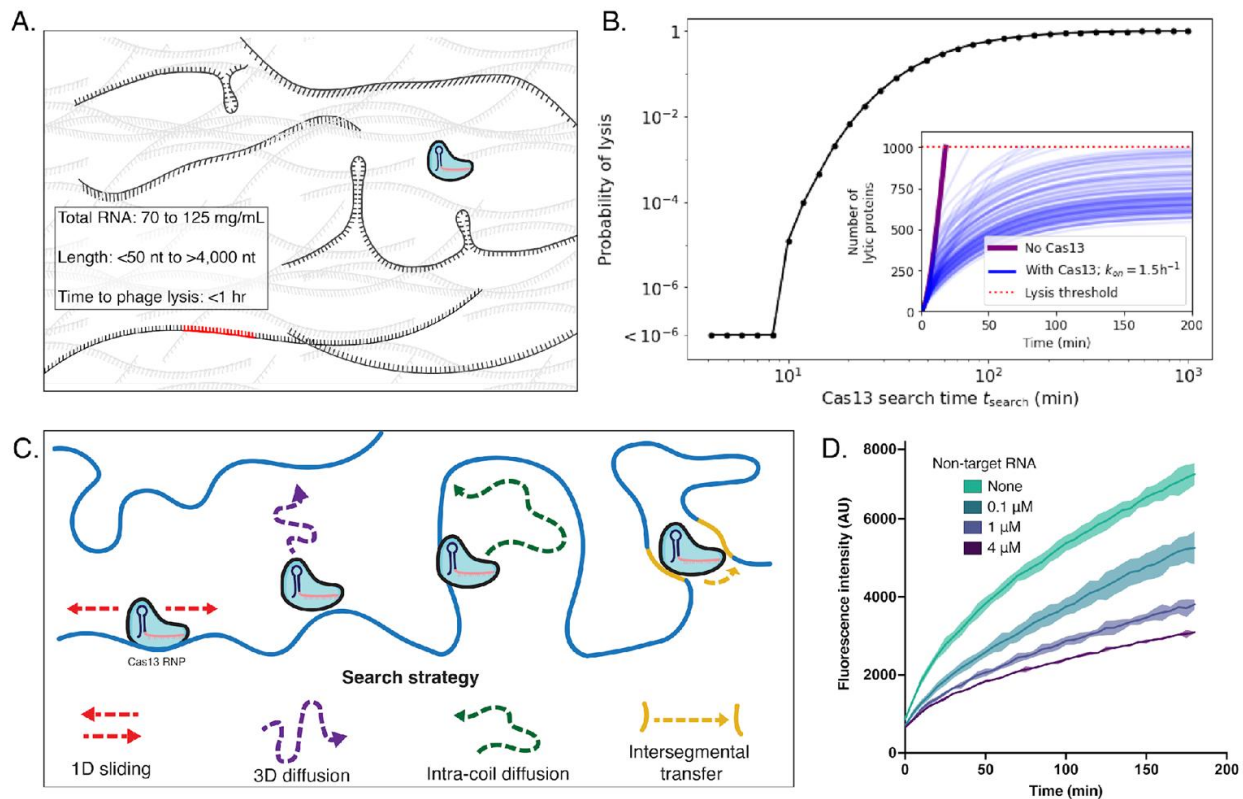


Figure 1: Cas13 efficiently searches an expansive RNA sequence space. **A.** Cas13 must search through abundant cellular RNAs of varying lengths and secondary structures to locate a 20-28 nt protospacer sequence on a relatively short timescale. **B.** Inset: Stochastic trajectories of the accumulation of lytic proteins for a given system with $k_{\text{on}}=(t_{\text{search}})^{-1}=1.5/\text{h}$. Main figure: Fraction of stochastic trajectories resulting in cell lysis, as a function of Cas13 search time t_{search} . Model is described in Section S5. **C.** Visual summary of four previously described search mechanisms that a sequence-specific RNA- or DNA- binding protein may utilize to find its binding partner. **D.** Fluorescence kinetic curves show that addition of increasing amounts of single-stranded non-target RNA leads to slower Cas13 activation *in vitro*. Target RNA concentration: 0.17 nM. Error of $n = 2$ technical replicates is shown by shading.

binding proteins were intersegmental transfers explicitly considered as a potential search strategy, nor have quantitative arguments been made for the circumstances in which different facilitated diffusion strategies may be preferred for an RNA search.

Here, we explore how sequence-specific RBPs can rapidly find their targets within the massive and heterogeneous landscape of RNA molecules in the cell. We use Cas13 as a model RBP for our study since, beyond its applications in basic research and biotechnology, its highly active *trans*-cleavage activity provides a simple fluorescent readout for target binding and efficient multiplexing of experimental conditions^{30,31}. Coupling Cas13 binding and activity assays together with biophysical modeling, we demonstrate that Cas13 utilizes facilitated diffusion to accelerate its search beyond the limit imposed by 3D diffusion. Furthermore, we show that Cas13 uses ITs as the dominant search modality, and argue that ITs can be more efficient than other facilitated diffusion strategies for a broad class of RNA-binding proteins, especially those targeting long RNAs ($\geq 1,000$ nucleotides in length).

RESULTS

Non-target RNAs slow Cas13's search

A major search obstacle faced by RNA-binding proteins, in comparison to DNA-binding proteins, is the quantity and heterogeneity of RNA molecules in solution. We first sought to establish whether the presence of RNAs lacking the 28-nt target or “protospacer” sequence (referred to hereon as non-target RNAs) substantially affects Cas13 search kinetics. To quantify our expectation for this effect, we consider the rate of target binding, k_{on} , for a searching protein which is initially not bound to the target RNA. This rate is equal to the inverse of the average search time τ_{search} , and has previously been found by several groups^{17,20,32,33} (also, see Supp. Section S1) to scale with system properties as

$$(k_{on})^{-1} = \tau_{search} = \tau_{3D} + \frac{L_t}{n}(\tau_{3D} + \tau_t^{local}) \quad (1)$$

where τ_{3D} is the time it takes the protein to bind non-specifically to some section of the target RNA, L_t is the number of possible binding sites in the target RNA (one of which is the protospacer, i.e. the target site), n is the number of sites typically searched upon each non-specific binding event (and is >1 for facilitated diffusion) and τ_t^{local} is the time taken to search those sites. In a solution with only target RNAs (i.e. RNAs containing the protospacer), τ_{3D} is given by the inverse of the Smoluchowski rate $k_t^{3D} = 4\pi D_t r_t c_t$, where D_t is the sum of the diffusion coefficients of the protein and target, r_t is the radius of gyration of the target RNA, and c_t is the target RNA concentration. However, in the presence of non-target RNAs, this rate is modified to be

$$\tau_{3D} = (k_t^{3D})^{-1} (1 + \sum_i k_i^{3D} \tau_i^{local}) = \frac{1}{4\pi D_t r_t c_t} + \sum_i \frac{D_i r_i c_i}{D_t r_t c_t} \tau_i^{local} \quad (2)$$

where terms with the i subscript denote the i^{th} species of non-target RNA in solution. Eq. (2) quantifies how the presence of non-target RNAs will slow down the search process.

To measure this effect experimentally in our model system, we utilized a Cas13 activity assay in which Cas13's activation by target binding leads to the trans-cleavage of a quenched fluorescent reporter molecule. The rate of the resulting fluorescence increase is referred to as “activity” and labeled ξ . As fluorescence assays using *in vitro* transcribed RNA tend to exhibit minimal variation (coefficient of variation $<5\%$), we used two technical replicates for most activity assays throughout^{34–36}. We measured LwaCas13a activation kinetics in the presence of increasing concentrations of non-target RNA, while keeping the target RNA concentration constant. We designed both target and non-target RNAs to be single-stranded to control for structure effects (see Methods). We find that increasing concentrations of non-target RNA reduce Cas13 activity (Fig. 1D). Since no element of the target RNA was changed, this establishes that Cas13 activity can be used as a proxy for the search time kinetics modeled by Eq. (1); although the precise relationship between the two is not established here, we make use of the fact that $\xi(k_{on})$ is a monotonically increasing function throughout this work. An alternative hypothesis that non-target RNA predominantly affects Cas13's *trans*-cleavage rather than its search time is not supported by follow-up experiments (see Supplement, Section S2). Notably, Cas13 is still activated to a significant degree even when the target RNA is outnumbered by non-target RNAs by a factor of over 20,000.

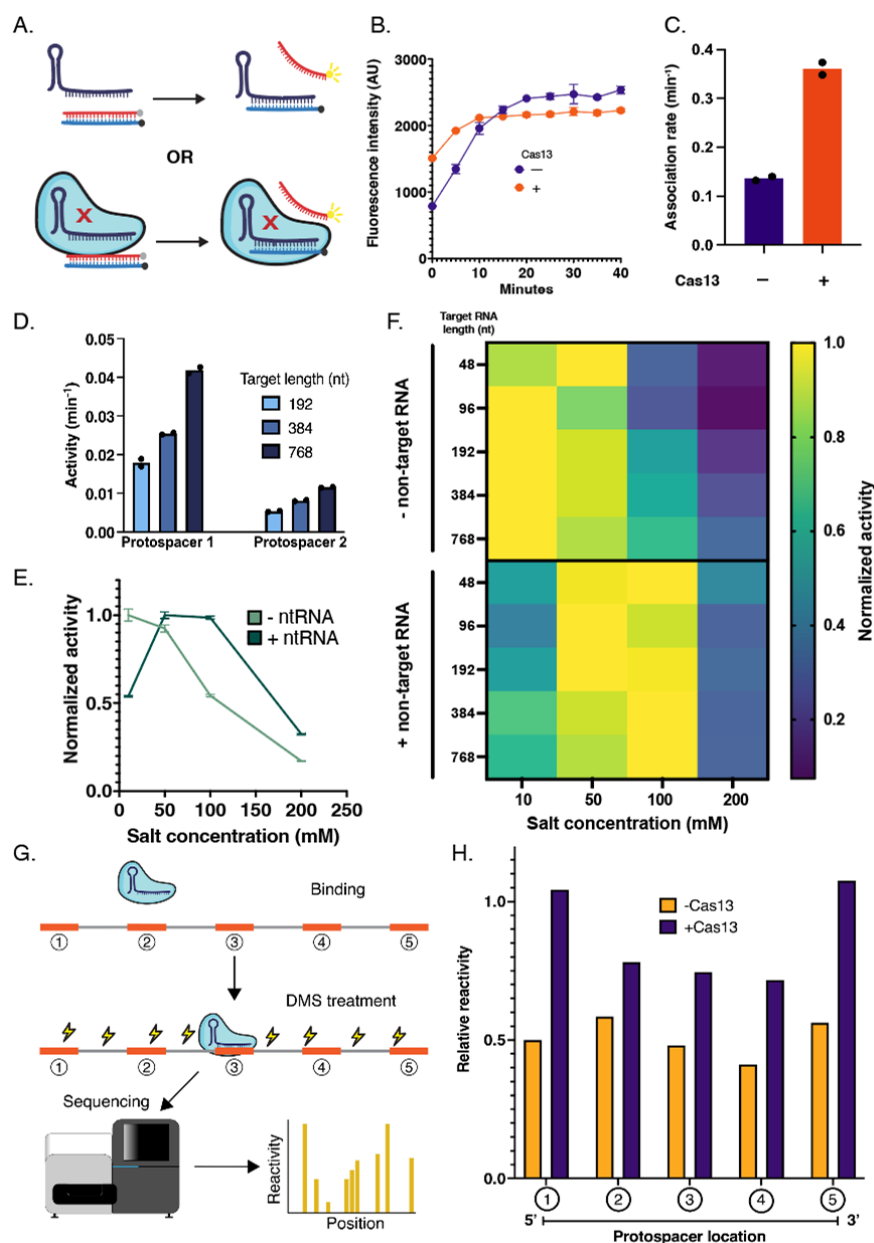


Figure 2: Cas13 locates its RNA target by facilitated diffusion. **A.** Schematic of an *in vitro* assay which directly reports kinetics of target RNA binding with or without the presence of catalytically inactive (dead) Cas13 via a strand-displacement readout; red X indicates dead Cas13. **B.** Fluorescence kinetic curves reveal faster target binding in the presence of dead Cas13 compared to the complementary crRNA alone. Error bars represent the standard deviation of two technical replicates. **C.** Activity quantification (pseudo-first order rate constant) of the kinetic curves shown in B. **D.** Activity quantification showing an increase in Cas13 activity when targeting RNAs of increasing length, for two unique protospacer sequences. **E.** Cas13 activity profile across different salt concentrations in an *in vitro* assay in the presence or absence of RNA for a 192 nt target RNA. Error bars represent the standard deviation of two technical replicates. **F.** Heatmap showing Cas13 activity across varying salt concentrations for target RNAs of different lengths in the presence or absence of RNA. Y axis markings indicate target RNA length. **G.** Schematic of a DMS Cas13 footprinting assay using a target RNA containing five identical protospacer sequences. In brief, the binding of the crRNA to the target RNA protects the bound region from DMS-induced mutagenesis, which can be detected with a sequencing readout and used to construct a crRNA occupancy profile. **H.** Relative DMS reactivity of the five different protospacer sequences in the presence or absence of Cas13, highlighting Cas13-mediated binding bias toward the center of the molecule.

Cas13 employs facilitated diffusion in its search

We next sought to establish whether Cas13 uses facilitated diffusion in its search process. We reasoned that if Cas13 uses facilitated diffusion, the Cas13-crRNA complex may locate its target RNA faster than the crRNA alone, which must search solely through 3D diffusion. We designed a strand displacement assay to directly report target binding independent of Cas13 activity (Fig. 2A). We found that fluorescence kinetics were ~3x faster in the presence of catalytically inactive Cas13 (dCas13) compared to the crRNA alone, despite Cas13's larger size and slower diffusion (Fig. 2B-C)³⁷.

To further probe whether Cas13 uses facilitated diffusion to speed up its search, we examined the scaling behavior of the search rate k_{on} as a function of target length, L_t . Changing L_t affects k_{on} in two different ways: first, directly (Eq. 1) as longer targets have more non-target sites which Cas13 must search before finding the protospacer; second, indirectly (Eq. 2) as longer targets are physically larger, leading to faster non-specific binding to the molecule (i.e. increasing k_t^{3D}). For a pure 3D search, the former effect is always dominant, such that increasing the target length will slow down the search (see Supp. Section S2). However, that is not necessarily the case for a search using facilitated diffusion. For small enough target RNAs, a protein employing facilitated diffusion may display an increase in k_{on} as L_t is increased¹⁸. We find that Cas13 activity (and therefore also k_{on}) increases with L_t for target RNAs between 192 and 768 nt. This scaling behavior was consistent across two different protospacer sequences, despite variations in absolute activity between them (Fig. 2D, Supp. Fig. 1).

Another hallmark of facilitated diffusion concerns the effect of ionic conditions on the search time. Lower salt concentrations are expected to increase the dwell time of a (positively-charged) protein on a (negatively-charged) polymer such as RNA. Therefore, in the absence of facilitated diffusion, i.e. for $n=1$, decreasing the salt concentration will simply increase τ_t^{local} , and thereby decrease k_{on} . However, a search employing facilitated diffusion can exhibit a regime in which decreasing the salt concentration will increase k_{on} due to an increase in n . We find that Cas13 activity monotonically increases with decreased [NaCl] under the conditions tested (Fig. 2E; light green). To control for potential effects of salt on the equilibrium Cas13 activity, we tested the effects of changing [NaCl] in the presence of non-target RNA. With non-target RNA present, low values of [NaCl] are expected to lead Cas13 to dwell longer on non-target RNAs, increasing τ_t^{local} , and thereby decreasing k_{on} . In agreement with this prediction, we find a non-monotonic dependence of Cas13 activity on [NaCl] in the presence of non-target RNAs (Fig. 2E; dark green). Across different target RNA lengths, the maximum Cas13 activity was consistently found at a higher salt concentration in the presence of non-target RNAs than in their absence (Fig. 2F, Supp. Fig 1).

Finally, we explored the effects of different protospacer placements within the target molecule. While a search using pure 3D diffusion will tend to find any protospacer location with equal probability, a process employing local search with correlated motion between sites searched will preferably find protospacers located nearer the center of the RNA³⁸. To observe potential preferences in Cas13's target binding across the length of an RNA molecule, we created a novel methylation-based footprinting assay based on dimethyl sulfate mutational profiling with sequencing (DMS-MaPSeq) using dCas13^{39,40}. We designed a 254 nt-long unstructured target RNA containing five identical protospacer sequences equally spaced throughout the molecule (Methods). The dCas13-crRNA complex or crRNA alone was then allowed to bind to the target RNA. The RNA was then treated with DMS, a strong methylating agent which acts on unpaired adenine and guanine bases, resulting in a methylation profile corresponding to the unbound regions of RNA and leaving behind unmethylated "footprints" where dCas13-crRNA or crRNA alone was bound (Fig. 2G). A highly processive reverse transcriptase converts these methylation marks to mutations in the resulting cDNA, which can then be analyzed by next-generation sequencing to establish a binding profile. We found that

while the crRNA alone has a uniform binding profile across the five protospacer sequences, the dCas13-crRNA complex is disfavored from binding to the two protospacers at the 5' and 3' termini of the RNA as evidenced by the elevated rate of DMS reactivity within these protospacers (Fig. 2H, Supp. Fig. 2). Taken together, the dependence of Cas13 activity on target length, effects of ionic concentration, and observed edge effects provide strong evidence of facilitated diffusion by Cas13 in its target search process.

Scaling of search time with target length can distinguish between search strategies

We next asked which facilitated diffusion strategy Cas13 uses in its search. As the search process for several other RBPs has been found to involve 1D sliding^{27,41}, and given the prevalence of the 1D sliding strategy in DNA-binding proteins including other CRISPR proteins^{22–26}, we initially hypothesized that Cas13 would employ 1D sliding in its search process. A sliding based search has the property that $\tau_{\text{local}} = n^2/2D_{1D}$ (with D_{1D} a 1-dimensional diffusion constant) is independent of L_t , while $\tau_{3D} \sim (L_t)^{-\nu}$, where $\nu < 1$ is the Flory exponent ($\nu = 1/2$ for a random walk) (see Section S2 for a more detailed calculation). The overall target acquisition rate for a sliding-based search is then given by

$$(k_{on})^{-1} \approx L_t^{-\nu}(n + L_t) f + \frac{n}{2D_{1D}} L_t \quad (1D \text{ sliding}) \quad (3)$$

where f is independent of L_t . Therefore, in the absence of non-target RNAs, the overall relationship of $\xi(L_t)$ is expected to be a non-monotonic concave function, as has been noted in other work¹⁸.

An alternative search strategy would be intracoil diffusion. Intracoil diffusion is a qualitatively different form of facilitated diffusion because it is unclear to what extent the protein itself can be evolutionarily tuned to modulate its behavior. Prior work has shown that intracoil diffusion plays a dominant role in Argonautes' search (though we note the authors of this paper used the phrase “intersegmental transfer” to refer to this process; this should not be confused with intersegmental transfers as they are defined in our work and elsewhere)²⁹. Intracoil diffusion is typically separated into two qualitative regimes: short “hops” of distance $< b$ wherein the protein binds to a spot correlated with its previous binding position, and longer “jumps” of distance $\gg b$ with no such correlation^{32,42}. Given the short persistence length of RNA, we neglect hops in this work. As derived by Halford and Marko with simple scaling arguments, for intracoil diffusion, $n \sim (L_t)^{1-\nu}$ ²⁰. Assuming that the time spent diffusing intramolecularly is negligible, and assuming a uniform dwell time τ_d for each non-specific binding event, a search dominated by intracoil diffusion has:

$$(k_{on})^{-1} \approx \frac{1}{4\pi D_t a c_t} (1 + \tau_d \sum_i k_i^{3D} n_i) + L_t \tau_d \quad (\text{intracoil diffusion}) \quad (4)$$

where $a = 0.33$ nm is the length of each site of the RNA (see derivation in Supp. Section S2). In other words, intracoil diffusion enables the search process to largely behave as if the target site is on its own, separate, RNA strand. In the absence of non-target RNA, Eq. 4 implies that a search dominated by intracoil diffusion would exhibit an association rate that decreases monotonically with L_t .

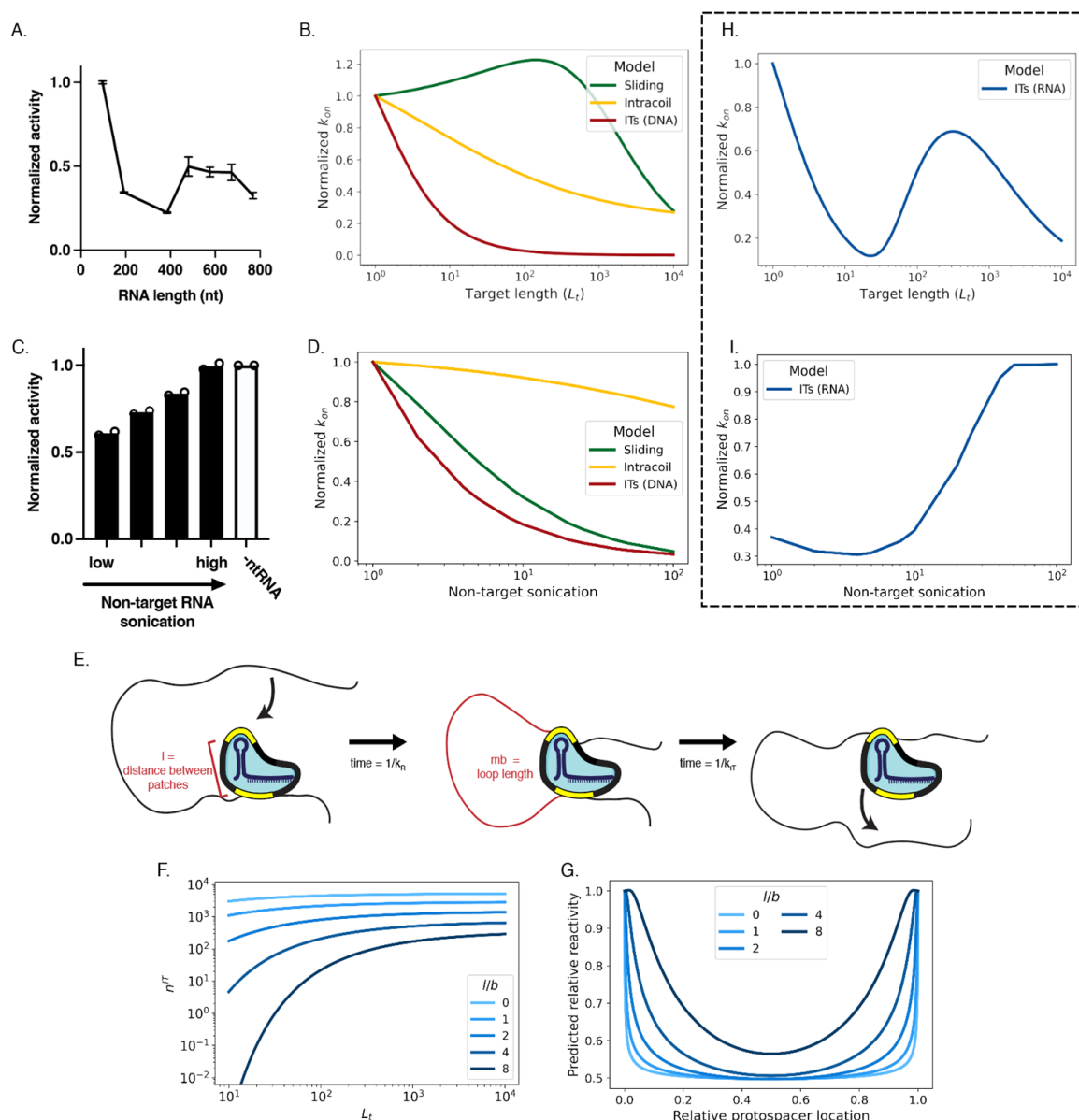


Figure 3: Observed effects of RNA length on Cas13 activation are consistent with intersegmental transfer, but not with other models. **A.** Cas13 activity profile when targeting RNAs of varying lengths containing a protospacer in the middle of the molecule. Error bars represent the standard error of two separate experiments with two technical replicates each. **B.** Modeling predictions of the rate of target binding as a function of RNA target length for different search strategies. **C.** Cas13 activity as a function of average non-target RNA length, generated by spiking target RNA into solutions containing a constant concentration of non-target RNAs that were subjected to different amounts of sonication. **D.** Modeling predictions of the rate of target binding as a function of the degree of non-target RNA sonication for different search strategies. **E.** Schematic illustrating the intersegmental transfer process for Cas13. **F.** The predicted number of ITs performed is plotted as a function of RNA length; colors correspond to the distance between binding . **G.** An IT-based search model predicts edge effects in binding, where sites near the middle of the RNA are searched more than sites near the edges. **H.** The predicted effect of target length on the rate of target binding for an IT-based search. **I.** The predicted rate of target binding as a function of the degree of non-target sonication for an IT-based search.

Neither 1D sliding nor intracoil diffusion are the dominant search strategy used by Cas13

Eqs. 3 and 4 demonstrate that studying the qualitative properties of $\xi(L_t)$ could enable us to distinguish between different search strategies. For example, finding that Cas13 activity scales with target length as a non-monotonic concave function would rule out intracoil diffusion as the dominant search mechanism. We therefore designed three sets of unstructured target RNAs with lengths ranging from 48 to 768 nt, containing the same protospacer sequence but different flanking sequence contexts. For all target sequences, measured Cas13 activity in the absence of non-target RNA was non-monotonic and convex—initially high at very short target lengths before a decrease and subsequent increase in activity as L_t was increased (Fig. 3A, Supp. Fig 3). This qualitative behavior cannot be explained by either Eq. 3 (1D sliding; green curve in Fig. 3B) or Eq. 4 (intracoil diffusion; yellow curve in Fig. 3B).

To further explore the role of RNA length on Cas13 activity, we took ~1000 nt-long non-target RNAs subjected to different amounts of sonication. We then spiked in a constant concentration of target RNA, thereby creating samples with constant values of L_t and constant amounts of non-target RNA nucleotides, but with differing length distributions for the non-target RNA (Fig. 3C, Supp. Fig. 3). The most highly sonicated sample contained a large number of shorter non-target RNA molecules and the less-sonicated samples contained fewer and longer non-target RNA molecules. Following a similar analysis to Eq. 3 (see Supp. Section S2), a search dominated by 1D sliding would exhibit faster association for lower amounts of non-target RNA sonication, as only the number of non-target RNAs, and not their lengths, affect the search time (green curve in Fig. 3E). Similarly, Eq. 4 indicates that the behavior of a search dominated by intracoil diffusion is expected to be largely invariant to sonication for small amounts of sonication, and exhibit similar qualitative behavior as that predicted for sliding for large amounts of sonication (yellow curve in Fig. 3D). However, we once again see behavior inconsistent with both of these predictions: higher Cas13 activity for *higher* amounts of non-target RNA sonication (Fig. 3C).

A model for intersegmental transfers (ITs)

Having found that Cas13's observed behavior is incompatible with both a search dominated by 1D sliding and one dominated by intracoil diffusion, we consider a third possibility, that of intersegmental transfers (ITs). ITs can be performed when the protein has more than one patch capable of associating non-specifically with the RNA. A segment of the RNA non-specifically associates to the protein at one of these patches, and while that segment remains bound, stochastic thermal motion brings a second region of RNA to bind transiently to the second patch. If the first RNA segment dissociates before the second, the protein has effectively transferred between regions of RNA which can be quite distant along the RNA contour, but which were transiently near each other in 3D space (Fig. 3E). As an analogy, consider the protein as a passenger at a railway transit station; it can switch between different train lines (different RNA segments) without leaving the station.

The generic nature of these non-specific RNA-binding patches on the protein is a key point. In prior work, ITs have typically been considered for multimeric proteins, which have multiple sites able to bind *specifically* to the target⁴³. However, this need not be the case. In order to perform ITs, an RBP need only be capable of associating non-specifically with two RNA segments simultaneously. We refer to the locations for non-specific association as “binding patches” on the protein, but these may be as simple as a patch of positively-charged residues on the protein surface, capable of non-specifically attracting negatively-charged RNA.

Existing models for ITs on the DNA cannot explain the behaviors observed in the experiments shown in Fig. 3 (red curves in Fig. 3B and 3D); however, these existing models are not necessarily appropriate for a search on the RNA. For example, prior work has indicated that

ITs can particularly speed up the search on the DNA when the protein rarely or never dissociates from the polymer^{44,45}; such a search would be highly inefficient in the context of RNA given the prevalence of non-target RNAs. Importantly, though, prior models have not accounted for the relationship between the rate of performing ITs and the overall dissociation rate of the protein from the polymer, nor have they considered the effects of a short polymer persistence length compared to protein size, as is the case for RNA^{17,44–47}. Therefore, motivated by our findings that Cas13 uses a facilitated diffusion strategy that is neither sliding nor intracoil diffusion, we decided to revisit this problem and develop a model for an IT-based RNA search.

Short ITs on RNA are exponentially suppressed

The first question we ask is how quickly each IT can be performed. ITs are mediated by the dynamic fluctuations of the RNA, which have previously been found to be well described by the Zimm model⁴⁸. If a section of RNA is bound to one of the non-specific binding patches, the rate at which a region of RNA a contour length of m Kuhn lengths away binds to the other binding patch, located a distance l away in 3D space, scales approximately as

$$k_R(m) \sim \frac{k_B T}{\eta} b^{-3} m^{-3/2} e^{-3l^2/2mb^2} \quad (5)$$

where b is the Kuhn length of the RNA (~ 2 nm⁴⁹; equal to twice the persistence length), k_B is Boltzmann's constant, T is temperature in Kelvin, η is the viscosity of solution, and for clarity we have omitted a numerical prefactor as well as a term logarithmic in the reaction radius^{50,51}. The heavy-tailed scaling of $k_R \sim m^{-3/2}$ in Eq. 5 is a key feature of ITs: they are superdiffusive, and enable the protein to search far from the site of initial binding, in contrast to diffusive 1D sliding which is better suited to short-distance searches. Specifically, ITs are an example of a Lévy flight, and the mean step size of an IT scales with the total length of the RNA. Because the Kuhn length of DNA is ~ 100 nm⁵² and protein diameters are typically much smaller, $l \ll b$ when considering a DNA search and the exponential term has been neglected in prior models of ITs for DNA-binding proteins. However, this exponential is non-negligible for RNA-binding proteins, and can have a substantial impact on the search, as we will show.

k_R gives the rate at which RNA strands encounter the protein, but the rate at which the protein actually performs an IT is given by the dissociation rate of each binding patch: ITs are performed when one binding patch dissociates and rebinds before the other dissociates. We therefore argue that the rate of performing ITs and the overall dissociation rate of the protein from the polymer must be coupled; this coupling is the second major deviation of our work from prior models of ITs. Here, we assume for simplicity that the dissociation rates of the two binding patches are equal and given by k_{IT} . We note that k_{IT} could naturally be tuned by evolution.

An IT-based search can explain the RNA-length-dependence of Cas13 activity

ITs are only useful if k_{IT} is small enough that the RNA is mostly able to re-equilibrate between each IT. (We verify the self-consistency of this assumption in Supp. Section S4). Under this assumption, we can sum over m to get the total rate of the free binding patch encountering any stretch of RNA along the chain, which will be used to calculate the probability that the dissociation of a binding patch results in an IT rather than in a complete dissociation of the protein. This rate depends on the number of sites upstream and downstream of where the protein is bound, which we denote by L_u and L_d , respectively:

$$k_R^{tot} = \sum_m k_R(m) = \alpha (S_q(L_u) + S_q(L_d)) \quad (6)$$

where $\alpha \sim k_B T / \eta b^3$ has units of inverse time, $q = 3l^2/2b^2$, and we define

$$S_q(x) = \sum_{m=1}^x m^{-3/2} e^{-q/m} \approx \sqrt{\pi} q^{-1/2} - 2x^{-1/2} \quad (7)$$

where the approximation is valid in the regime $1 < q < x$ (see derivation and approximations for other parameter regimes in Supp. Section S3; see also Supp. Fig. 4). Eqs. 6-7 reveal a key feature of ITs for RBPs, which does not apply to DNA-binding proteins. Recognizing that the total number of ITs taken before dissociating, n^{IT} , is approximately given by k_R^{avg}/k_{IT} where k_R^{avg} is the average of k_R^{tot} taken over all positions (see Supp. Section S4), these equations indicate that the number of ITs performed by an RBP is dependent on the length of the RNA to which it is bound: the RBP will perform more ITs on longer RNAs, and fewer ITs on shorter RNAs (Fig. 3F). Or, put another way, an RBP using ITs will dissociate more slowly from longer RNAs, and more quickly from shorter RNAs. As shown in the figure, this effect is particularly apparent for large values of l .

Moreover, Eqs. 6-7 demonstrate that an RBP performing ITs will experience “edge effects” in qualitative agreement with Cas13’s behavior. When the protein is bound at a certain position, the probability of it performing a successful IT from that position rather than dissociating completely from the RNA is given by $(1 + k_{IT}/k_R^{tot})^{-1}$. As k_R^{tot} is a concave function of position along the RNA, the likelihood of the protein binding to the center of the RNA is higher than the likelihood it binds to the edge of the RNA (Fig. 3G), in agreement with our findings in Fig. 2H. As shown in Fig. 3G, the width of the “edge” grows with the distance l between the binding patches on the protein (see Supp. Section S4).

To explore the ability of our RNA-IT model to explain the dependence of Cas13 activity on RNA length, which could not be recapitulated by models of sliding or intracoil diffusion, we consider the dependence of k_{on} on L_t . We plot this dependence in Fig. 3H, for $l/b=8$. We find that unlike the corresponding plot for a sliding-based search (Fig. 3B) which is concave for all L_t , an IT-based search yields a convex dependence of k_{on} on L_t for small L_t (Fig. 3H). This prediction shows good agreement with our experimental findings for Cas13 (Fig. 3A). This convexity is present only when $q \geq 1$ (i.e. $l \geq b$): in this case, for small L_t , the RNA is too short for ITs to transport the protein to an arbitrary segment of the RNA, reducing the efficacy of ITs for short targets.

Similarly, ITs can also explain the effect of non-target RNA sonication on Cas13. The key feature is that when performed on RNAs, ITs lead the protein to dissociate faster from shorter RNA molecules. When Cas13 binds a short RNA at one non-specific binding patch, the probability of a segment of the RNA reaching the second binding patch is small, leading Cas13 to dissociate without performing many ITs. For this reason, higher amounts of sonication of non-target RNAs leads to higher Cas13 activity, as Cas13 wastes less time searching shorter non-target RNAs (Fig. 3I). This prediction is also in good agreement with our experimental results, which displayed a surprising increase in Cas13 activity with greater sonication of non-target RNAs (Fig. 3C). Thus, Fig. 3 demonstrates that Cas13’s responses to changes in the lengths of both target and non-target RNAs can be explained by a target search dominated by ITs, but not by a search dominated by other modes of facilitated diffusion.

Further experiments confirm that Cas13 employs ITs in its search process

In addition to enabling the protein to transfer between different regions of the same RNA molecule, ITs can also enable transfers between different (nearby) molecules. As searches using 1D sliding do not involve such intermolecular transfers, the ability of the protein to transfer intermolecularly can be used to distinguish searches dominated by sliding from those dominated

by ITs. We therefore designed two assays to measure Cas13's ability to transfer between different RNA molecules.

The first such assay we designed takes its inspiration from Gowers and Halford's study of the restriction enzyme EcoRV⁵³. We tested Cas13's activity in three distinct scenarios (Fig. 4A): (1) with the protospacer in the middle of a long RNA; (2) with the protospacer in the middle of a short RNA, in the presence of a long non-target RNA; (3) with the protospacer in the middle of a short RNA which has been hybridized to a long non-target RNA. The essential test is in scenario (3), as a search using ITs is expected to find the target at different rates for scenarios 2 & 3, while a search using sliding is expected to find the target in scenarios 2 & 3 at similar rates. The reason for this is that ITs enable the protein to transfer between different nearby RNA chains, while a search using sliding is restricted to the RNA chain to which it initially bound non-specifically. In other words, ensuring the long and short RNAs are spatially co-localized is expected to significantly accelerate a search dominated by ITs, but not a search dominated by sliding. In agreement with our hypothesis that Cas13 uses ITs in its search, we find similar rates of target acquisition for scenarios 1 & 3, and a substantially slower rate for scenario 2 (Fig. 4B, Supp. Fig. 1).

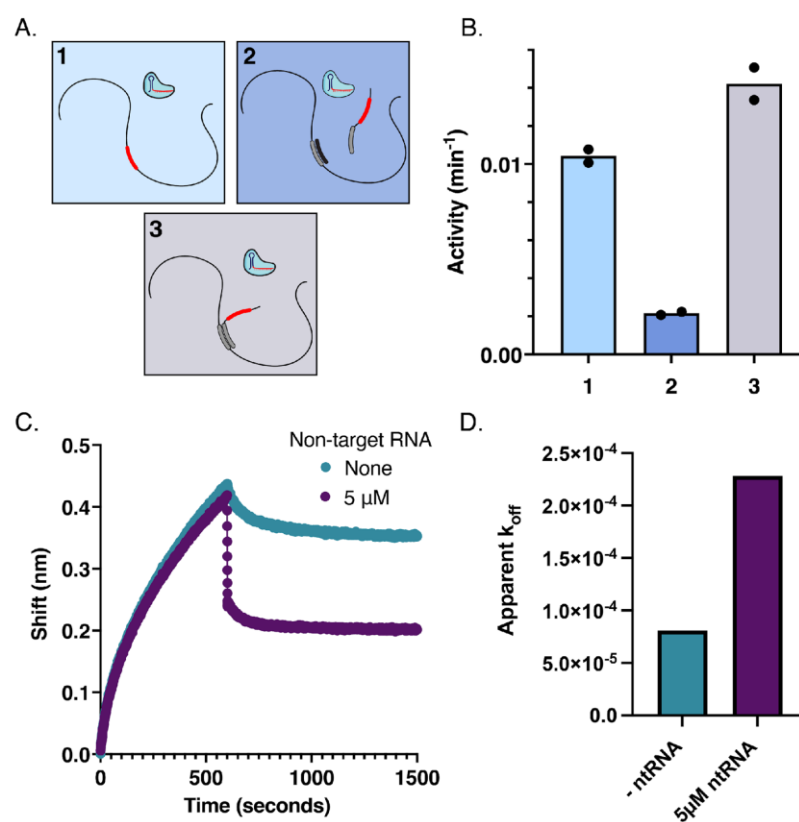


Figure 4: Cas13 uses intersegmental transfers during its search process. **A.** Schematic of an assay to differentiate 1D sliding and intersegmental transfer. In scenario 1, Cas13 targets a protospacer in the middle of a 768 nt target RNA. In scenario 2, Cas13 targets a protospacer on a 71 nt target RNA in the presence of a 768 nt RNA. In scenario 3, the 71 nt target RNA is annealed to the middle of the 768 nt RNA via a complementary region downstream of the protospacer sequence. **B.** Quantification of Cas13 activity for the three scenarios illustrated in F. **C.** Biolayer interferometry data showing that Cas13 dissociates from its target RNA more quickly in the presence of RNA. **D.** Quantification of apparent dissociation rate constants (k_{off}) from C.

Motivated by the same principles, we designed an additional assay, in which we measured the dissociation rate of Cas13 from the target RNA as a function of the concentration of non-target RNAs in solution. A protein performing a sliding-based search would have the same dissociation rate, regardless of how much non-target RNA was in solution; however, a protein whose searches are dominated by ITs would dissociate much faster in the presence of a high concentration of non-target RNAs. This is because ITs are realized by many microscopic dissociation and reassociation events performed by each binding patch; if a non-target RNA is transiently located near the protein, one of these reassociation events can result in the protein transferring to the non-target RNA. Using biolayer interferometry to directly study binding and dissociation rates, we found that Cas13 binds strongly to its target RNA and, in the absence of non-target RNA, dissociates slowly. However, when non-target RNA is added to the dissociation buffer, the rate of dissociation is greatly increased (Fig. 4C,D). Both this and the previous assay indicate that the mechanism of facilitated diffusion employed by Cas13 is better explained by ITs than by 1D sliding.

Search time analysis indicates an optimal rate of performing ITs

To understand the evolutionary forces acting on proteins performing ITs, we sought to model how the optimal search would depend on various system parameters. First, we consider a system with no non-target RNAs, using Eq. 1 with $n=n^{IT}+1$ and $\tau_t^{local}=1/k^{off}$, where $k^{off}=k_{IT}/(n^{IT}+1)$ is the dissociation rate of the protein (see Supp. Section S4). Note that these are not independent, but are both functions of L_t and of k^{IT} . In Fig. 5A, we plot the target acquisition rate k_{on} as a function of k_{IT} for different values of k_R^{avg} ; Fig. 5B shows a similar plot for different concentrations of target RNA. Although models with independent k^{off} and n^{IT} can naturally lead to search times being optimized by the protein never dissociating from the target⁴⁴, Fig. 5A-B demonstrate that coupling the two leads the search time to be optimized at a finite value of k_{IT} . This optimal value is given by

$$k_{IT}^* = \frac{k_R^{avg}}{\sqrt{k_R^{avg} \tau_{3D} - 1}} \quad (\text{no non-target RNAs}) \quad (8)$$

Larger values of k_{IT} are suboptimal as they lead to faster dissociation and therefore fewer sites searched per round of binding, while smaller values of k_{IT} are suboptimal as they lead each IT to take longer. In the regime where $\tau_{3D} < (k_R^{avg})^{-1}$, k_{IT}^* is negative and ITs slow down the search for any finite value of k_{IT} . However, given that k_R^{avg} is of order inverse nanoseconds, this regime is not physically relevant for RNA. We note that the presence of non-target RNAs increases the optimal value of k_{IT} compared to Eq. 8, since time that the protein spends searching non-target RNAs before dissociating from them is wasted (Fig. 5C).

The corresponding optimal search time is given by

$$(k_{on}^*)^{-1} = \tau_{3D} + \frac{L_t}{k_{IT}^*} \left(2 + \frac{k_{IT}^*}{k_R^{avg}} \right) \quad (\text{no non-target RNAs}) \quad (9)$$

Recognizing that L_t/k_{IT} represents the total time the protein spends performing ITs, and that $k_R^{avg} \gg k_{IT}^*$, this indicates that, aside from the first 3D search to find the target, the protein optimally spends nearly the same amount of time searching locally through ITs as searching globally through 3D diffusion. Similar results have previously been found for the optimal search using 1D sliding³² suggesting that equipartitioning local and global search times may be a feature of facilitated diffusion more broadly.

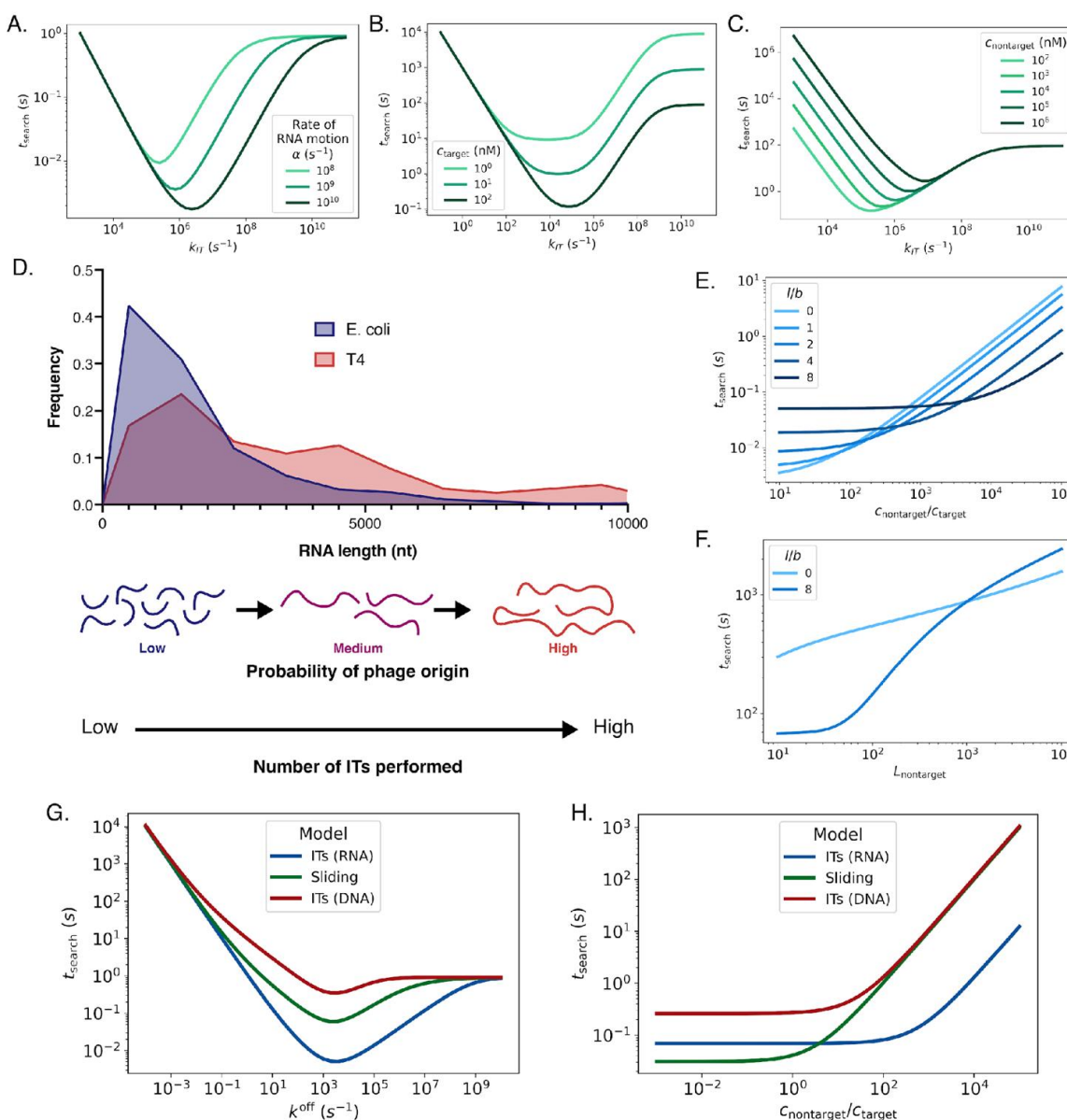


Figure 5: Modeling reveals general advantages of intersegmental transfer in RNA-based search processes. **A.** Curves showing the dependence of search time on the rate of intersegmental transfer, for different rates of overall RNA binding. **B.** Modeling results as in A, but for different concentrations of target RNA in the solution. **C.** Modeling results as in A, but for different concentrations of non-target RNA in the solution. **D.** Histogram of the known RNA length distributions for *E. coli* and bacteriophage T4. **E.** Curves showing the dependence of search time on the ratio of non-target to target RNA in the system, for different values for the distance between binding domains. **F.** Modeling results as in E, but for different non-target RNA lengths. **G.** Direct comparison of search time for three different modes; ITs searching DNA, sliding, and ITs searching RNA. **H.** Modeling the search time for the same three search strategies as a function of the ratio of non-target to target RNA.

Benefits of ITs for a search on RNA

Having found that Cas13 employs ITs in its search, we asked whether such a mechanism may benefit RNA-based search processes more generally. First, Eq. 5 indicates that ITs may be substantially faster on RNA than on DNA, because of the dependence of k_R on the persistence length. For single-stranded RNA as we consider here, $b \approx 2$ nm, while for double-stranded DNA $b \approx 100$ nm^{49,52}. For RNA, $k_R \sim b^{-3}$ as the internal polymer motion is Zimm-like (Eq. 5); for DNA, the motion is more appropriately described by the Rouse model⁴⁸ and the scaling is moderated somewhat⁵⁴ to $k_R \sim b^{-2}$. Therefore, put together, ITs can be performed 3-5 orders of magnitude faster on RNA than on DNA.

In addition to its effect on the overall search rate, the short persistence length of RNA has other qualitative effects on an IT-based search. In particular, for DNA, an IT-based search will always have $l \ll b$ (i.e. nearby binding patches, relative to the persistence length of DNA), while for RNA, $l \gtrsim b$ can be relevant. Indeed, the qualitative features of Cas13's search time scaling with target RNA length (Fig. 3) can only be understood in the context of large values of l/b (i.e. a large separation between the binding patches). On its surface, this appears suboptimal, as increased values of l lead to fewer ITs performed on each RNA (Fig. 3F) and ultimately to slower searches in the absence of non-target RNA (Fig. 5E; left side of the plot). The key to making sense of this finding is recognizing that an optimal search process must not only find the target molecule quickly, but avoid non-target RNAs as much as possible. When the solution contains many short non-target RNAs, large values of l enable the protein to dissociate quickly from these RNAs. As described in Eq. 7, an IT-based search enables the protein to dissociate faster from short RNAs than long ones. As seen in Fig. 3F, especially for $l \gg b$, the dissociation rate from short RNAs can be orders of magnitude faster than the corresponding rate for long RNAs.

To explore this idea and its potential advantages, we used publicly available datasets to analyze the lengths of all known transcriptional units in two well-characterized systems, *E. coli* and bacteriophage T4^{55,56}. We found that phage RNAs tend to be longer, on average, than bacterial RNAs, and thus infer that a mechanism which prioritizes thoroughly searching long RNAs while avoiding shorter ones will be highly efficient for a phage RNA-targeting system like Cas13, or other RNA-targeting CRISPR systems (Fig. 5D). In this sense, we propose ITs may be acting as a kind of length filter, in order to selectively enrich for longer RNAs likely to be of bacteriophage origin during the search process and thus help narrow the search space.

In Fig. 5E, we consider a solution consisting of long target RNAs ($L_t = 10^3$) and short non-target RNAs ($L_n = 40$). We plot the search time as a function of the relative concentration of the non-target RNA for proteins with different amounts of separation between the binding patches (i.e. different values of l/b). While for small amounts of non-target RNA, the search is much faster for proteins with nearby binding patches (i.e. smaller values of l/b), for high enough non-target RNA concentrations, the search is optimized by proteins with large separations between the binding patches. The reason for this concerns the length-dependent dissociation rate of an RBP performing ITs: even though a protein whose binding patches are farther apart will dissociate faster from all RNAs, it will especially dissociate faster from short RNAs.

Similarly, in Fig. 5F, we plot the search time as a function of the length of non-target RNAs for given values of the non-target RNA concentration and of l/b . We find that even though large values of l/b slow down the rate at which the protein can perform each IT, they enable the protein to avoid short non-target RNAs, and thus can speed up a search in which short non-target RNAs are present at high quantities.

Finally, to compare between sliding and ITs more directly for an RNA-based search, in Fig. 5G we plot the search time $t_{\text{search}} = (k_{\text{on}})^{-1}$ for ITs (blue), as well as for a sliding-based search (green) as a function of the dissociation rate of the protein from the RNA, k^{off} . For very large

values of k^{off} , the two searches are equivalent to a pure 3D search, and for very small values both are inefficient as the protein gets stuck searching the non-target RNAs. For intermediate dissociation rates, ITs are substantially faster than sliding, given realistic 1D diffusion coefficients of $D_{1D}=10^6$ nm²/s (at $D_{1D}=10^9$ nm²/s, sliding and ITs are similarly fast). ITs are particularly faster in the presence of non-target RNAs; in their absence, sliding can be faster than ITs (Fig. 5H). The speed of ITs is dependent on the short persistence length of RNA; even in the absence of non-target molecules, our model finds ITs to be a relatively inefficient method of searching DNA (Fig. 5G-H, red).

DISCUSSION

In this study, we used a feedback loop of biophysical modeling and experimental validation to study how RBPs can quickly find their RNA targets among the vast number of cellular RNAs. While our modeling work is able to consider generic RBPs, our experimental studies focused on a model RBP, Cas13. We established an experimental system that uses Cas13 binding and activation kinetics as a proxy for RNA search speed. We then showed that Cas13 uses facilitated diffusion in its target search process and considered several potential mechanistic models. Multiple orthogonal lines of evidence, including the effect of RNA lengths on Cas13 search time and Cas13's ability to transfer between different RNA molecules without dissociating, indicate that Cas13 uses a search dominated by intersegmental transfers.

We find that the search process of RBPs differs from that of DNA-binding proteins in three primary ways, all of which lead ITs to be particularly efficient for RNA searches. First, the short persistence length of RNA compared to DNA means that ITs can occur $\sim 10^4$ times faster on RNA than on DNA. Second, the secondary structures formed by RNA can create obstacles that impede 1D diffusion²⁷ which can be circumvented by ITs (Fig. 4A-B). Third, cellular RNAs are heterogeneous in length; a protein which spends equal time on each local search would search short RNAs many times over before ever fully searching a long RNA. While the increased radius of gyration of longer RNAs coupled with intracoil diffusion can partially address this inefficiency, ITs enable a search that is counterintuitively more efficient for long RNAs than for short ones. In particular, ITs enable the protein to naturally (and without any energy consumption) dissociate quickly from short RNAs and spend more time searching longer RNAs. This "length-filtering" effect is strengthened when the two binding patches on the protein are far apart, representing a tradeoff in the position of these patches: when the patches are closer together, the protein searches more sites on each RNA before dissociating; when they are farther apart, the protein searches fewer sites, but in a length-dependent manner, dissociating particularly quickly from shorter RNAs.

For this reason, our results indicate that ITs may be a generically efficient facilitated diffusion mechanism for a broad class of sequence-specific RNA-targeting proteins. Those proteins whose RNA targets are longer than the typical cellular RNA, for instance proteins targeting mRNAs, can utilize the length-filtering effect we have described by placing their binding patches farther apart (yielding qualitatively similar behaviors to our findings on Cas13), while those proteins whose RNA targets are shorter than the typical cellular RNA can place their patches closer together, eschewing length-filtration in favor of faster overall search. In contrast to previous work which considered ITs primarily for multimeric proteins with multiple recognition domains, we propose that ITs may be broadly applicable for proteins with only one recognition domain but multiple non-specific binding patches. While the non-specific binding patches are unable to distinguish the target sequence from non-target themselves (see Supp. Section S4), they facilitate faster binding between the specific binding domain and the target by enabling the protein to perform ITs. For example, Cas13 has only one RNA-recognition domain which can distinguish the specific target sequence from other RNAs, but positively charged patches on its surface can serve as separate locations for non-specific RNA association (Supp. Fig. 5). In fact,

RBP have been found to be highly enriched for such positively charged patches, and roughly 45% of eukaryotic RBPs contain multiple experimentally validated RNA-binding domains^{57,58,59}. The search behavior of several RBPs appears to also be consistent with ITs^{27–29,41}, though ITs have rarely been considered due to their relative inefficiency on DNA compared to sliding^{45,60}.

While we have treated them separately in this work, ITs have a close connection to intracoil diffusion. In the limit of fast polymer diffusion compared to protein diffusion, an intracoil diffusion process may be qualitatively similar to ITs with $q=0$. The model we use for intracoil diffusion in this work is appropriate in the opposite limit. Developing a comprehensive theory of the intracoil diffusion process across different regimes of polymer diffusion rates may be a fruitful subject for future work⁶¹. Intracoil diffusion, as well as 1D sliding, may indeed play a substantial role in Cas13's search process; our model does not rule out these mechanisms entirely but rather shows that these alone cannot explain Cas13's search behavior. Intracoil diffusion occurs as part of the 3D diffusion process, and it is possible that a mixture of longer-distance transfers and local 1D searches may lead to more efficient search. Future work will identify what role, if any, is played by other search modalities beyond ITs in Cas13's search process.

In an effort to isolate the RNA search process and minimize confounding factors, we performed all assays in highly controlled environments using carefully designed RNAs. Therefore, this work neglects many features of the cellular environment that may influence the search process *in vivo*, including different secondary structures, the effects of other RNA-binding proteins, and molecular crowding. These effects will be the subject of future work. Our IT model will also be expanded upon, to include effects such as asymmetry between binding patches, the presence of more than 2 patches, or extended binding regions.

Though we have framed the RNA search problem as it pertains to Cas13 operating in its native prokaryotic context, Cas13 has also been broadly utilized in eukaryotic cells for RNA knockdown, base editing, and RNA imaging. In these contexts, the search problem is even greater; eukaryotic cells are on average around 1,000 times larger by volume and contain hundreds of times as many RNAs spread across multiple cellular compartments^{4,62,63}. Thus, improved understanding of the RNA search process is critical for furthering the biotechnological applications of Cas13, as it may lead to improvements in RNA targeting efficiency via intelligent crRNA design, protein engineering, or other approaches.

ACKNOWLEDGMENTS

We thank Long Nguyen and Soonwoo Hong for protocols and technical assistance related to protein purification, and Owen Abad for helpful discussions and preliminary experimental planning. We thank Venu Vandavasi and the Princeton Biophysics Core Facility for access to and training on the Octet RED96e instrument. We thank Alexander Grosberg for discussions regarding the scaling of intracoil diffusion with target length. We thank Ned Wingreen and the members of the Myhrvold and te Velthuis labs for helpful discussions.

This work was supported by the Branco Weiss Fellowship–Society in Science, administered by the ETH Zürich (O.K.), the Peter B. Lewis '55 Lewis-Sigler Institute/Genomics Fund (O.K.), the National Institutes of Health (R01AI182281, C.M.; R01AI177487, A.J.W.t.V.; and DP2AI175474, A.J.W.t.V.; T32GM007388, B.B.L; T32GM148739, B.B.L and E.G.), and the Princeton Catalysis Initiative (C.M.). Research reported in this publication was supported by the National Center For Advancing Translational Sciences of the National Institutes of Health under Award Number UM1TR004789. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

METHODS

Target RNA design

To control for effects of RNA secondary structure, we designed target RNAs to include minimal structure, as described previously in Ref.³⁷. For most experiments, crRNA 4 and its corresponding protospacer were used, as that crRNA showed high activity in our previous studies, and was compiled from the ADAPT dataset which has a sequence composition representative of viral diversity³⁶. The data labeled “protospacer 2” in Fig. 2D represent results using crRNA 5 from the same paper; in other experiments, RNAs including the protospacer corresponding to crRNA 5 were used as non-target RNA. RNAs of different lengths were created by concatenating the protospacer with RNA sequences compiled out of a set of 64 16-nt-long DNA sequences with minimal internal structure from Shortreed *et al.*⁶⁴. The protospacer was placed in the middle of the sequence, aside from the experiments described in Fig. 2G-H which included 5 equally-spaced identical protospacers. NUPACK⁶⁵ was used to predict the structure of the resulting sequences and ensure minimal secondary structure. All sequences used are shown in Supp. Table 1.

RNA preparation

Templates for target RNAs for use in Cas13 assays were purchased from IDT (kinetics experiments and BLI) or Ansa Biotechnologies (DMS experiments) as DNA. Templates containing a double-stranded T7 promoter sequence were transcribed to RNA using homemade T7 polymerase mix. For a 20 μ L IVT reaction, 1.5 μ L enzyme master mix, 10 μ L 2 $\times$ IVT reaction mix, 1 μ L 0.1 M DTT, DNA template, and nuclease-free water were combined. The 2 $\times$ IVT reaction mix contained 20 mM rNTP mix, 80 mM Tris-HCl, 12 mM MgCl₂, 20 mM DTT, and 2 mM spermidine. Reactions were carried out at 37 °C for 16 hours followed by DNase I XT template digestion and subsequent purification using 1.8 $\times$ SPRI beads and the addition of 1.6 $\times$ isopropanol for RNAs of length < 100 nt, and were eluted into 20 μ L of nuclease-free water. RNAs were quantified using a NanoDrop One (Thermo Fisher Scientific) or Biotek Take3Trio (Agilent) and stored in nuclease-free water at -80 °C.

When pre-annealing RNAs or DNA-RNA hybrids, equimolar amounts of the two species (or a 2:1 ratio of DNA to RNA in certain cases) were mixed in a 60 mM KCl annealing buffer, then annealed in a thermocycler by heating to 85 °C and gradually cooling to 10 °C at 0.1 °C/second, before final cooling to 4 °C.

Cas13 activity assays

Plate reader detection assays were performed by mixing target RNA at a ratio of 10% v/v with 90% Cas13 detection mix. The detection mix consisted of 1 $\times$ SHINE buffer (20 mM HEPES pH 8.0, 60 mM KCl and 3.5% PEG-8000 in nuclease-free water), supplemented with 45 nM purified homemade LwaCas13a, 1 U per μ L murine RNase Inhibitor (New England Biolabs), 62.5 nM fluorescent reporter (/5FAM/rUrUrUrUrUrU/IABkFQ/; IDT), 22.5 nM processed crRNA (IDT) and 14 mM magnesium acetate. Reactions (15 μ L each) were then loaded in technical duplicate onto a 384 well clear-bottom Greiner plate. The plate was then loaded onto a Biotek Cytation5 or Synergy H1 plate reader, and fluorescence intensity was measured every 5 minutes for 3 hours at 37 °C.

Highly multiplexed chip-based assays were performed using a BioMark HD system with a Standard Biotools genotyping IFC (96.96 format). The assay mix (7% of reaction volume) contained 1 $\times$ Assay Loading Buffer (Standard Biotools), 293 nM LwaCas13a, and 1 $\times$ SHINE buffer (see above section). The sample mix (93% of reaction volume) contained 1.07 U/ μ L murine RNase Inhibitor (New England Biolabs), 15 mM MgOAc, 110 nM FAM reporter, 1 $\times$ ROX reference dye (Standard Biotools), 1 $\times$ GE buffer (Standard Biotools), 1 $\times$ SHINE buffer, and target RNA (variable concentration).

Sample volumes of 6 μ l and assay volumes of 6 μ l, in addition to control line fluid, were loaded onto the IFC chip (Standard Biotools). Chips were then placed into the IFC Controller HX and primed, loaded and mixed using the Prime and Load Mix 96.96 GE scripts (Standard Biotools). After mixing, chips (with two technical replicates per condition) were run on BioMark HD at 37 °C for 8 h, with measurements taken in the FAM and ROX channels every 5 min. Normalized and background-subtracted fluorescence for a given time point was calculated as (FAM – FAM background)/(ROX – ROX background).

For the strand displacement assay, reaction conditions were the same as standard Cas13 detection assays, with the following adjustments: the strand-displacement reporter concentration was set to 20 nM, the dCas13 concentration (in the +dCas13 condition) was set to 9 nM, and the crRNA concentration was set to 4.5 nM.

Protein purification

For LwaCas13a: a plasmid containing the human codon-optimized LwaCas13a sequence (Addgene ID: 90097) was transformed into Rosetta2(DE3) Singles™ Competent Cells according to the manufacturer's protocol. Single colonies were inoculated into 30 mL LB broth and incubated at 37 °C overnight. The following day, overnight cultures were inoculated into 2 L Terrific Broth and incubated at 37 °C with shaking at 250 rpm until reaching an OD of 0.6. Cultures were then placed on ice and induced with IPTG at a final concentration of 0.5 mM and cultured overnight at 18 °C with shaking.

The cells were then pelleted by centrifugation (4000 $\times$ g, 10 min) and resuspended in lysis buffer (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM imidazole, 1 mM DTT) supplemented with protease inhibitors (Complete Ultra EDTA-free tablets) and lysozyme (500 μ g). Cells were then lysed by sonication (10 s on/10 s off for 20 min with 50% power) and the lysate subsequently clarified by centrifugation (20,000 $\times$ g, 1 hr, 4 °C). The supernatant was passed through a 0.2 micron filter before being injected onto the Biorad NGS FPLC system connected to a Ni-NTA column and purified using a wash buffer (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM DTT) and an elution buffer (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 300 mM imidazole, 1 mM DTT). Eluted fractions were pooled and mixed with 300 μ L of 1 mg/mL SUMO protease then dialyzed against SUMO cleavage buffer (30 mM Tris HCl pH 8.0, 500 mM NaCl 1 mM DTT, 0.15% Igepal (NP-40)) overnight. The protein was concentrated to 50% of its original volume using an Amicon 50 kDa MWCO spin column then mixed at a 1:1 ratio with cation exchange buffer (20 mM HEPES pH 7.5, 250 mM NaCl, 1 mM DTT, 5% glycerol). The protein was then further purified by injecting it onto a 5 mL HiTrap SP HP cation exchange column and eluted over a salt gradient of 250 mM to 2 M NaCl in elution buffer (20 mM HEPES pH 7.5, 2 M NaCl, 1 mM DTT, 5% glycerol).

Protein purity was checked on a Coomassie-stained PAGE gel, after which fractions were pooled. If purity was not sufficient, the sample was subjected to an additional purification step consisting of size exclusion chromatography over a Superdex 200 Increase 10/300 GL column. The protein was then concentrated and exchanged into storage buffer (600 mM NaCl, 50 mM Tris-HCl pH 7.5, 5% glycerol, 2 mM DTT) again using a 50 kDa MWCO spin column and subsequently aliquoted and flash-frozen in liquid nitrogen. Aliquots were stored at -80 °C until use.

For T7 RNA polymerase, RNase inhibitor, and IPPase: The T7 RNAP variant (G47A+884G) coding sequence was cloned into a pET-derived N-terminal 6 $\times$ His-SUMO expression vector. The RNase inhibitor expression construct was generated in-house, and inorganic pyrophosphatase (iPPase) was expressed from a gift from Sebastian Maerkl and Takuya Ueda (Addgene plasmid #124137).

RNA Bacterial expression plasmids were transformed into Rosetta 2(DE3)pLysS Singles competent cells (Millipore Sigma, Cat. #71401). Individual colonies were picked into 50 mL Luria Broth (Fisher Scientific, Cat. #BP9723-2) containing the appropriate antibiotic(s) and grown

overnight at 37°C with shaking. Overnight cultures were used to inoculate 4 L Terrific Broth (RPI, Cat. #T15000-10000.0), and cultures were grown at 37 °C until the OD600 reached 0.6–0.8.

Cultures were cooled rapidly on ice for 10–15 min, induced with 1 mM IPTG (Gold Biotechnology, Cat. #I2481C100), and incubated overnight at 18°C. Cells were harvested by centrifugation at 4,000 ×g for 15 min. Cell pellets from 4 L culture were resuspended in 150 mL lysis buffer (50 mM Tris-HCl pH 7.9, 5% glycerol, 500 mM NaCl, 20 mM imidazole, and 1 mM TCEP). Lysozyme (0.5 mg/mL; Astatech, Cat. #AT25331 5G), DNase I (0.01 mg/mL; Worthington, Cat. #LS002145), and PMSF (0.5 mM; Tokyo Chemical Industry, Cat. #B1062) were added to the lysis buffer. Cells were lysed by sonication on ice. Lysates were clarified by centrifugation at 186,000 ×g for 30 min using a Beckman Coulter Type 45 Ti rotor, and the supernatant was passed through a 0.45 µm filter. Clarified lysate was loaded onto a prepacked Ni-NTA affinity column (EconoFit Nuvia IMAC column, Bio-Rad, Cat. #12009287) using a Bio-Rad NGC Quest Plus FPLC system. Bound protein was washed with lysis buffer and eluted with elution buffer (50 mM Tris-HCl pH 7.9, 5% glycerol, 500 mM NaCl, 400 mM imidazole, and 1 mM TCEP). Fractions containing the target protein were identified by SDS-PAGE and pooled.

For T7 RNAP, pooled Ni-NTA eluate was digested with SUMO protease to remove the N-terminal 6× His-SUMO tag. Briefly, eluate was combined with 2 mg SUMO protease in dialysis tubing (Ward's Science, Cat. #470163-414) and dialyzed overnight at 4°C with rotation against SUMO digestion buffer (30 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM DTT, and 0.15% Igepal/NP-40). Following digestion, the protein was buffer-exchanged into buffer A (50 mM Tris-HCl pH 7.9, 5% glycerol, 200 mM NaCl, and 1 mM TCEP), using 50 kDa MWCO centrifugal filters (Millipore Sigma, Cat. #UFC905024). The buffer-exchanged sample was applied to a 5 mL HiTrap Heparin HP column (Cytiva, Cat. #17040701), and T7 RNAP was eluted using a linear gradient from buffer A to buffer B (50 mM Tris-HCl pH 7.9, 5% glycerol, 2 M NaCl, and 1 mM TCEP). Fractions containing purified T7 RNAP were pooled.

Purified T7 RNAP, RNase inhibitor, and iPPase were buffer-exchanged and concentrated using 50 kDa MWCO centrifugal filters into storage buffer (50 mM Tris-HCl pH 7.9, 5% glycerol, 100 mM NaCl, 0.1% Triton X-100, 1 mM EDTA, and 1 mM DTT). Proteins were individually aliquoted and stored at 80 °C. In-house IVT enzyme master mix was prepared by combining purified T7 RNAP, RNase inhibitor, and iPPase at final concentrations of 2.67 µg/µL, 2.67 µg/µL, and 0.27 µg/µL, respectively, in a buffer containing 50 mM Tris-HCl pH 7.9, 100 mM NaCl, 50% glycerol, 0.1% Triton X-100, 1 mM EDTA, and 1 mM DTT. The enzyme master mix was then stored at -20 °C.

DMS footprinting assay

The Cas13 binary complex, or crRNA alone, was prepared at a final concentration of 1 µM in 1× binding buffer as described above (see “EMSA” section). This complex was then mixed with target RNA in 1× binding buffer at a ratio of 1:4 to ensure excess unbound target was present, and allowed to complex for 15 minutes at 37 °C. Each 10 µL sample was added to 35 µL of refolding buffer (0.38 M sodium cacodylate pH 7.2, 14 mM MgCl₂) then treated with 7 µL of DMS solution (15% v/v dimethyl sulfate, 85% v/v ethanol) or ethanol-only control at 37 °C. Reactions were quenched by the addition of 50 µL of beta-mercaptoethanol. Samples were cleaned up in two steps; first, they were subjected to column-based purification (Zymo RNA Clean and Concentrator 5), then purified by 5× SPRI beads. Samples were eluted into nuclease free water, quantified by NanoDrop One, and stored for downstream library preparation.

For each sample, 200 ng of RNA was dephosphorylated using Quick CIP (NEB) according to the manufacturer's instructions, before halting reactions by heat-inactivation at 85°C. Next, a single phosphate group was restored to the 5' ends of the entire sample using T4 polynucleotide kinase (NEB). The reaction was allowed to proceed at 37 °C overnight, before heat-inactivation at 65 °C for 20 minutes. Each sample was once again purified using 1× SPRI

beads and eluted into 10 μ L of nuclease-free water. The eluted RNA was then circularized using RNA Ligase I (NEB), with the addition of 20 units of murine RNase inhibitor. Reactions were allowed to proceed overnight at 16 °C, before heat-inactivation at 65 °C for 15 minutes. First-strand cDNA was synthesized from the circular RNA using UltraMarathonRT (RNAConnect), with the addition of High Boost. The resulting cDNA was then used as the template for PCR amplification using Q5 DNA polymerase (NEB) and primers containing 5' sequencing barcodes. The PCR products were purified and size-selected using SPRI beads, eluted into nuclease-free water, then pooled for next-generation sequencing.

DMS assay next-generation sequencing analysis

The reads were first demultiplexed using Barcode Splitter, and subsequently the adapters were trimmed using cutadapt. The reads were then split into the regions 5' and 3' to the ligation site in order to remove gaps and nontemplated bases inserted by T7 polymerase or uMRT. Cutadapt was used to remove barcodes (--cut) and to trim sequences to the desired length (--length). The two sections were then separately aligned to their respective sections of the template using Bowtie2 with settings --local, -N 1, -L 18. Reads with MAPQ (mapping quality) score <25 were discarded. A custom script was used to analyze the per-nucleotide mutation rate for each sample, using the rate of mutagenesis per A nucleotide (excluding primer-binding regions) to generate an overall mutation rate per protospacer.

Biolayer interferometry

Biotinylated DNA adapters were pre-annealed to target RNAs containing a 3' complementary region, and the duplexes were then immobilized on streptavidin biosensors (FortéBio). The sensors were then equilibrated in glycerol-free protein storage buffer (50 mM Tris-HCl pH 7.5, 600 mM NaCl, 2 mM DTT) before submerging into 100 nM Cas13-crRNA complex (prepared as previously described) and allowed to associate for 10 minutes at 37 °C. Sensors were then moved to fresh storage buffer either containing or lacking non-target RNA, and allowed to dissociate for 20 minutes at 37 °C. All data collection and processing were performed on an Octet RED96e biolayer interferometry system using Octet Data Analysis HT software version 11.1.3.50.

Activity fits and plotting

Activity scores were calculated by fitting fluorescence data to pseudo-first order kinetic curves (one-phase association, GraphPad Prism version 10.6), then defining “activity” as the rate constant (k, with units of inverse time). Separate curves were fit for each technical replicate and used to calculate descriptive statistics per condition. Curves and corresponding activity scores were plotted in Prism.

RNA length distribution analysis

Bacteriophage T4 transcription units were determined using promoter and terminator annotations from Table 1 of Ref. ⁵⁶. Transcript lengths were determined by calculating the distance between the genomic location of each promoter and the nearest terminator, accounting for the direction (+ strand or - strand) of the promoter and terminator. All promoters and terminators were assumed to be unidirectional based on their annotations in the table. *E. coli* transcription units were taken from Supplementary Table 12 of Ref. ⁵⁵. Transcription unit lengths were determined by subtracting the “Start” and “End” columns of the table.

REFERENCES

1. Cayley, S., Lewis, B. A., Guttman, H. J. & Record, M. T., Jr. Characterization of the cytoplasm of *Escherichia coli* K-12 as a function of external osmolarity. Implications for protein-DNA interactions in vivo. *J. Mol. Biol.* **222**, 281–300 (1991).
2. Elowitz, M. B., Surette, M. G., Wolf, P. E., Stock, J. B. & Leibler, S. Protein mobility in the cytoplasm of *Escherichia coli*. *J. Bacteriol.* **181**, 197–203 (1999).
3. Nakazato, H., Venkatesan, S. & Edmonds, M. Polyadenylic acid sequences in *E. coli* messenger RNA. *Nature* **256**, 144–146 (1975).
4. Neidhardt, F. C. Chemical Composition of *Escherichia coli*. in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (ed. John L. Ingraham, K. Brooks Low, Boris Magasanik, Moselio Schaechter, H. Edwin Umbarger) 3–6 (American Society for Microbiology, Washington, D.C., DC, 1987).
5. Barik, A. & Das, S. A comparative study of sequence- and structure-based features of small RNAs and other RNAs of bacteria. *RNA Biol.* **15**, 95–103 (2018).
6. East-Seletsky, A. *et al.* Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538**, 270–273 (2016).
7. Abudayyeh, O. O. *et al.* \RNA\ targeting with \CRISPR\-\Cas13}. *Nature* **550**, 280–284 (2017).
8. Gootenberg, J. S. *et al.* Nucleic acid detection with \CRISPR\-\Cas13a\-\C2c2}. *Science* **356**, 438–442 (2017).
9. Yang, L.-Z. *et al.* Dynamic imaging of RNA in living cells by CRISPR-Cas13 systems. *Mol. Cell* **76**, 981–997.e7 (2019).
10. Schertzer, M. D. *et al.* Cas13d-mediated isoform-specific RNA knockdown with a unified computational and experimental toolbox. *Nat. Commun.* **16**, 6948 (2025).
11. Meeske, A. J., Nakandakari-Higa, S. & Marraffini, L. A. Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature* **570**, 241–245 (2019).
12. Williams, M. C. *et al.* Restriction endonuclease cleavage of phage DNA enables resuscitation from Cas13-induced bacterial dormancy. *Nat. Microbiol.* **8**, 400–409 (2023).
13. Nguyen, H. M. & Kang, C. Lysis delay and burst shrinkage of coliphage T7 by deletion of terminator T ϕ reversed by deletion of early genes. *J. Virol.* **88**, 2107–2115 (2014).
14. Wang, I.-N. Lysis timing and bacteriophage fitness. *Genetics* **172**, 17–26 (2006).
15. G. Adam, M. D. Reduction of Dimensionality in Biological Diffusion Processes. in *Structural Chemistry and Molecular Biology* (ed. A. Rich, N. D.) (W.H. Freeman, San Francisco, CA, 1968).
16. Richter, P. H. & Eigen, M. Diffusion controlled reaction rates in spheroidal geometry. Application to repressor--operator association and membrane bound enzymes. *Biophys. Chem.* **2**, 255–263 (1974).
17. Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
18. von Hippel, P. H. & Berg, O. G. Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678 (1989).
19. Hippel, P. H. von, Revzin, A., Gross, C. A. & Wang, A. C. Interaction of lac repressor with non-specific DNA binding sites. in *Protein-Ligand Interactions* (eds. Sund, H. & Blauer, G.) 274–288 (De Gruyter, Berlin, Boston, 1975).
20. Halford, S. E. & Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052 (2004).
21. Hammar, P. *et al.* The lac repressor displays facilitated diffusion in living cells. *Science* **336**, 1595–1598 (2012).
22. Globyte, V., Lee, S. H., Bae, T., Kim, J.-S. & Joo, C. CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion. *EMBO J.* **38**, EMBJ201899466 (2019).

23. Singh, D. *et al.* Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). *Proc. Natl. Acad. Sci. U. S. A.* **115**, 5444–5449 (2018).
24. Aldag, P. *et al.* Dynamic interplay between target search and recognition for a Type I CRISPR-Cas system. *Nat. Commun.* **14**, 3654 (2023).
25. Knight, S. C. *et al.* Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **350**, 823–826 (2015).
26. Lee, J. & Jeong, C. Single-molecule perspectives of CRISPR/Cas systems: target search, recognition, and cleavage. *BMB Rep.* **58**, 8–16 (2025).
27. Richards, J. & Belasco, J. G. Obstacles to scanning by RNase E govern bacterial mRNA lifetimes by hindering access to distal cleavage sites. *Mol. Cell* **74**, 284–295.e5 (2019).
28. Chandradoss, S. D., Schirle, N. T., Szczepaniak, M., MacRae, I. J. & Joo, C. A dynamic search process underlies MicroRNA targeting. *Cell* **162**, 96–107 (2015).
29. Cui, T. J. *et al.* Argonaute bypasses cellular obstacles without hindrance during target search. *Nat. Commun.* **10**, 4390 (2019).
30. Ackerman, C. M. *et al.* Massively multiplexed nucleic acid detection with Cas13. *Nature* **582**, 277–282 (2020).
31. Welch, N. L. *et al.* Multiplexed CRISPR-based microfluidic platform for clinical testing of respiratory viruses and identification of SARS-CoV-2 variants. *Nat. Med.* **28**, 1083–1094 (2022).
32. Mirny, L. *et al.* How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A Math. Theor.* **42**, 434013 (2009).
33. Hu, T. & Shklovskii, B. I. How does a protein search for the specific site on DNA: The role of disorder. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **74**, 021903 (2006).
34. Molina Vargas, A. M. *et al.* New design strategies for ultra-specific CRISPR-Cas13a-based RNA detection with single-nucleotide mismatch sensitivity. *Nucleic Acids Res.* **52**, 921–939 (2024).
35. Bell, A. G. *et al.* A streamlined CRISPR-based test for tuberculosis detection directly from sputum. *Sci. Adv.* **11**, eadx2067 (2025).
36. Metsky, H. C. *et al.* Designing sensitive viral diagnostics with machine learning. *Nat. Biotechnol.* **40**, 1123–1131 (2022).
37. Larsen, B. B. *et al.* RNA structure modulates Cas13 activity and enables mismatch detection. *Nat. Biotechnol.* 1–9 (2025).
38. Coppey, M., Bénichou, O., Voituriez, R. & Moreau, M. Kinetics of target site localization of a protein on DNA: a stochastic approach. *Biophys. J.* **87**, 1640–1649 (2004).
39. Zubradt, M. *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* **14**, 75–82 (2017).
40. Tomezsko, P., Swaminathan, H. & Rouskin, S. DMS-MaPseq for genome-wide or targeted RNA structure probing in vitro and in vivo. *Methods Mol. Biol.* **2254**, 219–238 (2021).
41. Małecka, E. M. & Woodson, S. A. RNA compaction and iterative scanning for small RNA targets by the Hfq chaperone. *Nat. Commun.* **15**, 2069 (2024).
42. Loverdo, C. *et al.* Quantifying hopping and jumping in facilitated diffusion of DNA-binding proteins. *Phys. Rev. Lett.* **102**, 188101 (2009).
43. Shimamoto, N. One-dimensional diffusion of proteins along DNA. Its biological and chemical significance revealed by single-molecule measurements. *J. Biol. Chem.* **274**, 15293–15296 (1999).
44. Sheinman, M. & Kafri, Y. The effects of intersegmental transfers on target location by proteins. *Phys. Biol.* **6**, 016003 (2009).
45. Hu, T. & Shklovskii, B. I. How a protein searches for its specific site on DNA: the role of intersegment transfer. *arXiv [q-bio.BM]* (2007).
46. Lomholt, M. A., Ambjörnsson, T. & Metzler, R. Optimal target search on a fast-folding

- polymer chain with volume exchange. *Phys. Rev. Lett.* **95**, 260603 (2005).
47. de la Rosa, M. A. D., Koslover, E. F., Mulligan, P. J. & Spakowitz, A. J. Dynamic strategies for target-site search by DNA-binding proteins. *Biophys. J.* **98**, 2943–2953 (2010).
48. Shusterman, R., Alon, S., Gavrinov, T. & Krichevsky, O. Monomer dynamics in double- and single-stranded DNA polymers. *Phys. Rev. Lett.* **92**, 048303 (2004).
49. Chen, H. *et al.* Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 799–804 (2012).
50. Levernier, N. *et al.* Non-Markovian closure kinetics of flexible polymers with hydrodynamic interactions. *J. Chem. Phys.* **143**, 204108 (2015).
51. Doi, M. Diffusion-controlled reaction of polymers. *Chem. Phys.* **9**, 455–466 (1975).
52. Bustamante, C., Smith, S. B., Liphardt, J. & Smith, D. Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.* **10**, 279–285 (2000).
53. Gowers, D. M. & Halford, S. E. Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO J.* **22**, 1410–1418 (2003).
54. Toan, N. M., Morrison, G., Hyeon, C. & Thirumalai, D. Kinetics of loop formation in polymer chains. *J. Phys. Chem. B* **112**, 6094–6106 (2008).
55. Cho, B.-K. *et al.* The transcription unit architecture of the Escherichia coli genome. *Nat. Biotechnol.* **27**, 1043–1049 (2009).
56. Miller, E. S. *et al.* Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* **67**, 86–156, table of contents (2003).
57. Cook, K. B., Kazan, H., Zuberi, K., Morris, Q. & Hughes, T. R. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.* **39**, D301–8 (2011).
58. Stitzinger, S. H., Sohrabi-Jahromi, S. & Söding, J. Cooperativity boosts affinity and specificity of proteins with multiple RNA-binding domains. *NAR Genomics and Bioinformatics* **5**, (2023).
59. Lambert, S. A. *et al.* The human transcription factors. *Cell* **172**, 650–665 (2018).
60. Mahmutovic, A., Berg, O. G. & Elf, J. What matters for lac repressor search in vivo—sliding, hopping, intersegment transfer, crowding on DNA or recognition? *Nucleic Acids Res.* **43**, 3454–3464 (2015).
61. Berg, O. G. Diffusion-controlled protein-DNA association: influence of segmental diffusion of the DNA. *Biopolymers* **23**, 1869–1889 (1984).
62. Velculescu, V. E. *et al.* Analysis of human transcriptomes. *Nat. Genet.* **23**, 387–388 (1999).
63. Heim, N. A. *et al.* Hierarchical complexity and the size limits of life. *Proc. Biol. Sci.* **284**, 20171039 (2017).
64. Shortreed, M. R. *et al.* A thermodynamic approach to designing structure-free combinatorial DNA word sets. *Nucleic Acids Res.* **33**, 4965–4977 (2005).
65. Fornace, M. E. *et al.* NUPACK: Analysis and Design of Nucleic Acid Structures, Devices, and Systems. *ChemRxiv* (2022) doi:[10.26434/chemrxiv-2022-xv98l](https://doi.org/10.26434/chemrxiv-2022-xv98l).
66. Werner, A. Predicting translational diffusion of evolutionary conserved RNA structures by the nucleotide number. *Nucleic Acids Res.* **39**, e17 (2011).
67. Krouglova, T., Vercammen, J. & Engelborghs, Y. Correct diffusion coefficients of proteins in fluorescence correlation spectroscopy. Application to tubulin oligomers induced by Mg²⁺ and Paclitaxel. *Biophys. J.* **87**, 2635–2646 (2004).
68. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **38**, D750–3 (2010).
69. Lapham, J., Rife, J. P., Moore, P. B. & Crothers, D. M. Measurement of diffusion constants for nucleic acids by NMR. *J. Biomol. NMR* **10**, 255–262 (1997).
70. Hyeon, C., Dima, R. I. & Thirumalai, D. Size, shape, and flexibility of RNA structures. *arXiv [q-bio.BM]* (2006) doi:[10.48550/arXiv.q-bio/0609035](https://doi.org/10.48550/arXiv.q-bio/0609035).
71. Berg, O. G. & von Hippel, P. H. Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Biophys. Chem.* **14**, 131–160 (1985).

72. Fei, J. *et al.* Determination of in vivo target search kinetics of regulatory noncoding RNA. *Science* **347**, 1371–1374 (2015).