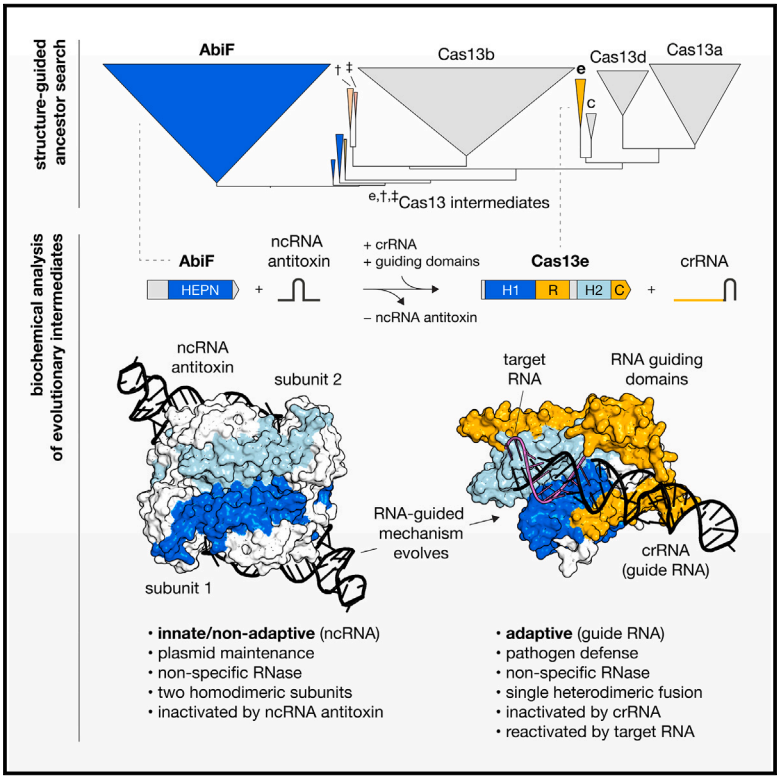
Cell

Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins

Graphical abstract



Highlights

- Hybrid structure-sequence search identifies Cas13s likely evolved from AbiF
- AbiF is a toxin-antitoxin (TA) system with a ncRNA antitoxin that inhibits the RNase toxin
- An intermediate Cas13e system possesses both CRISPR and TA biochemical properties
- Insertion of CRISPR RNA binding and REC domains led to complex RNA-guided CRISPR systems

Authors

Shai Zilberzwige-Tal, Han Altae-Tran, Soumya Kannan, ..., Rhiannon K. Macrae, Eugene V. Koonin, Feng Zhang

Correspondence

zhang@broadinstitute.org

In brief

Hybrid structure-sequence mining and biochemical/structural characterization trace the evolutionary origins of CRISPR-Cas13, elucidating how their RNA-guided mechanism likely evolved from an ancient scaffold comprising an RNase toxin and an ncRNA antitoxin.

Zilberzwige-Tal et al., 2025, Cell *188*, 1–16
April 3, 2025 © 2025 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2025.01.034

CellPress

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

Cell

CellPress
OPEN ACCESS

Article

Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins

Shai Zilberzwige-Tal,[1,2,3,4,5,9] Han Altae-Tran,[1,2,3,4,5,6,9] Soumya Kannan,[1,2,3,4,5] Max E. Wilkinson,[1,2,3,4,5] Samuel Chau-Duy-Tam Vo,[1,2,3,4,5] Daniel Strebinger,[1,2,3,4,5] KeHuan K. Edmonds,[1,2,3,4,5] Chun-Chen Jerry Yao,[1,2,3,4,5,7] Kepler S. Mears,[1,2,3,4,5] Sergey A. Shmakov,[8] Kira S. Makarova,[8] Rhiannon K. Macrae,[1,2,3,4,5] Eugene V. Koonin,[8] and Feng Zhang[1,2,3,4,5,10,*]

[1]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[2]McGovern Institute for Brain Research at MIT, Cambridge, MA 02139, USA
[3]Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[5]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[6]Institute for Protein Design, University of Washington, Seattle, WA 98195, USA
[7]Department of Molecular Cellular Biology, Harvard University, Cambridge, MA, USA
[8]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA
[9]These authors contributed equally
[10]Lead contact
*Correspondence: zhang@broadinstitute.org
https://doi.org/10.1016/j.cell.2025.01.034

SUMMARY

Despite ongoing efforts to study CRISPR systems, the evolutionary origins giving rise to reprogrammable RNA-guided mechanisms remain poorly understood. Here, we describe an integrated sequence/structure evolutionary tracing approach to identify the ancestors of the RNA-targeting CRISPR-Cas13 system. We find that Cas13 likely evolved from AbiF, which is encoded by an abortive infection-linked gene that is stably associated with a conserved non-coding RNA (ncRNA). We further characterize a miniature Cas13, classified here as Cas13e, which serves as an evolutionary intermediate between AbiF and other known Cas13s. Despite this relationship, we show that their functions substantially differ. Whereas Cas13e is an RNA-guided RNA-targeting system, AbiF is a toxin-antitoxin (TA) system with an RNA antitoxin. We solve the structure of AbiF using cryoelectron microscopy (cryo-EM), revealing basic structural alterations that set Cas13s apart from AbiF. Finally, we map the key structural changes that enabled a non-guided TA system to evolve into an RNA-guided CRISPR system.

INTRODUCTION

CRISPR-Cas systems are RNA-guided adaptive immune systems that guard bacteria and archaea against invading viruses and mobile genetic elements. Their diverse reprogrammable RNA-guided mechanisms have ushered in transformative methods for precisely editing mammalian genomes and transcriptomes.[1–5] Recently, the DNA targeting CRISPR-Cas9 and CRISPR-Cas12 systems were shown to have emerged from the transposon-associated RNA-guided nucleases OMEGA-IscB and OMEGA-TnpB.[6–9] However, the origin of the RNA-targeting CRISPR-Cas13 and, more generally, the origin of RNA-guided targeting mechanisms in CRISPR systems have remained elusive.

All known Cas13s contain two nuclease domains from the higher eukaryotes and prokaryotes nucleotide-binding (HEPN) superfamily,[10,11] comprising RNA nucleases found widely among prokaryotic defense systems and beyond.[12–15] Although previous genome mining work used the hallmark presence of two HEPN domains to successfully discover new Cas13s,[11,16,17] the high level of sequence divergence among the HEPN superfamily[11] made it challenging to identify related homologs and to reconstruct their evolutionary relationships. Indeed, the only potential remote homology reported for Cas13 is between Cas13b and the ApeA family of nucleases.[18] Recently, structure-based comparisons, such as distance-matrix aligment (DALI),[19–21] have been shown to outperform sequence-based comparisons in classifying proteins at the superfamily level.[22] The 3D structures of proteins, as predicted by deep learning methods such as AlphaFold2,[23] combined with algorithms for quantifying structural similarity, therefore offer the potential to address the limitations of sequence-only approaches in cases of high sequence divergence by allowing for more reliable remote homology searches and grouping of divergent domains into evolutionarily related families.[16,24,25]

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

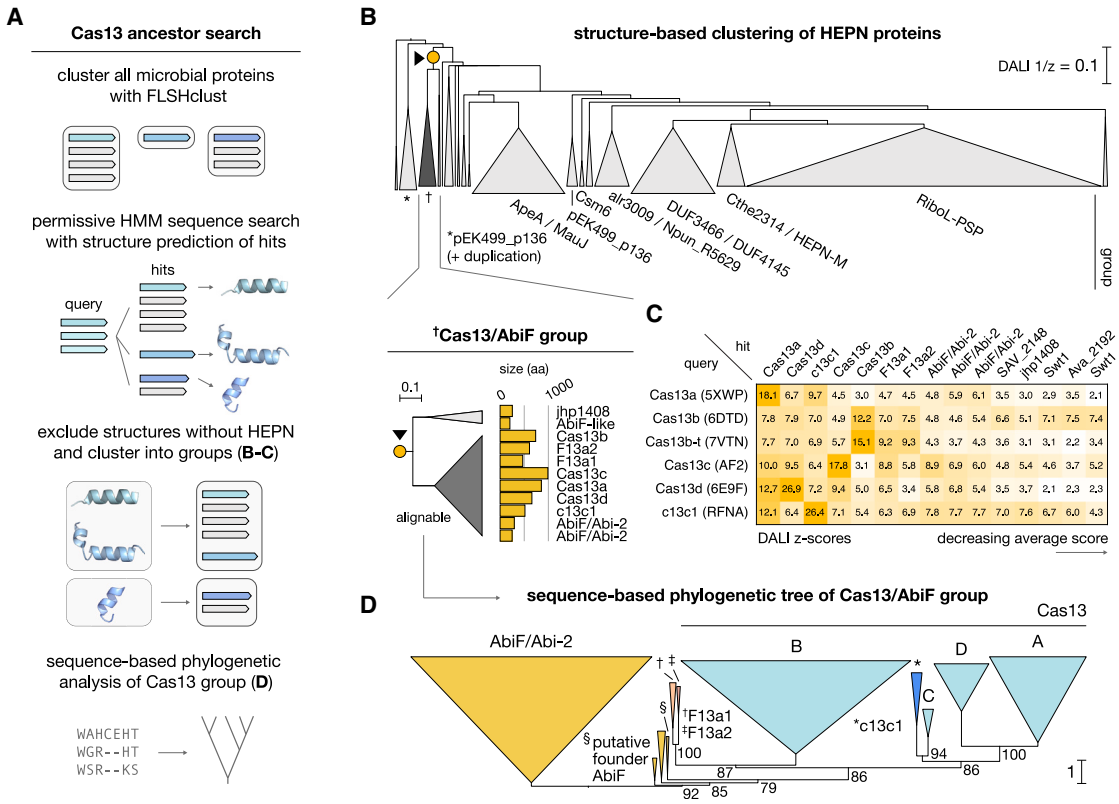


Figure 1. Identification of Cas13's ancestors

(A) Schematic of the search for Cas13 ancestors using a hybrid sequence/structure search.

(B) Top: an AF2 structure-based clustering dendrogram (not a phylogenetic tree) classifying Cas13-related HEPN domain-containing protein families into distinct groups. Bottom: Cas13-containing group with a sequence alignable subset (dark gray).

(C) Heatmap of structure alignment *Z* scores of comprehensive structure search proteins along with representative solved and predicted HEPN domain structures. Top 15 hits are shown, sorted by average *Z* score across all queries. PDB IDs or structure predictor (AF2 vs. RFNA[26]) are shown in parentheses.

(D) Maximum-likelihood sequence-based phylogenetic tree of the HEPN domains from proteins in the AbiF/Cas13 group alignable subset with midpoint rooting for visualization. Bootstrap values (out of 100) are shown.

See also Figures S1, S2, and S3.

Here, we describe a hybrid sequence and structure search to explore the evolutionary origins of Cas13 (Figure 1). Combined with a comprehensive RNA-protein association analysis (Figure 2), we discovered that Cas13 likely evolved from an abortive infection F (AbiF) protein associated with a non-coding RNA (ncRNA). Furthermore, characterization of the AbiF system revealed a distinct mechanism of action: unlike CRISPR systems, AbiF-ncRNA pairs function as type III toxin-antitoxin (TA) systems. A cryoelectron microscopy (cryo-EM) structure of the AbiF ribonucleoprotein (RNP) complex revealed that the ncRNA antitoxin is processed and then blocks the AbiF nuclease active site. Using sequence-based phylogenetic analysis, we traced the evolution from AbiF to Cas13. In the emerging evolutionary scenario, the AbiF gene first duplicated, producing a double-HEPN nuclease with additional domain insertions. The duplicated protein gradually lost its interaction with *cis*-encoded ncRNA antitoxin, retained by only a few systems. However, a new interaction with CRISPR RNA (crRNA) emerged, facilitated by a small set of helical insertions and a single Recognition (REC)-like insertion in the effector protein, which likely enable the recognition of the guide RNA-target RNA duplex. Finally, we bioinformatically investigated the diversity of Cas13 and AbiF-related proteins, uncovering a spectrum of defense-related systems linked to the AbiF toxin. Overall, this study elucidates how an RNA-guided mechanism evolved from a non-specific toxin RNA nuclease and broadens our understanding of the diverse molecular mechanisms underlying RNA-based immune systems in prokaryotes.

RESULTS

Tracing the ancestry of Cas13

To gain insight into the evolution of Cas13, we first examined a previously reported but uncharacterized miniature (~500 amino acid [aa]) Cas13-like, dual-HEPN domain-containing protein[27,28] (designated c13c1). A CRISPR association analysis identified a cluster of at least five c13c1[28] loci associated with CRISPR arrays, suggesting that c13c1 is a *bona fide* and potentially ancestral Cas13 effector. This prompted us to trace the lineage of Cas13s from within the greater HEPN superfamily. However,

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

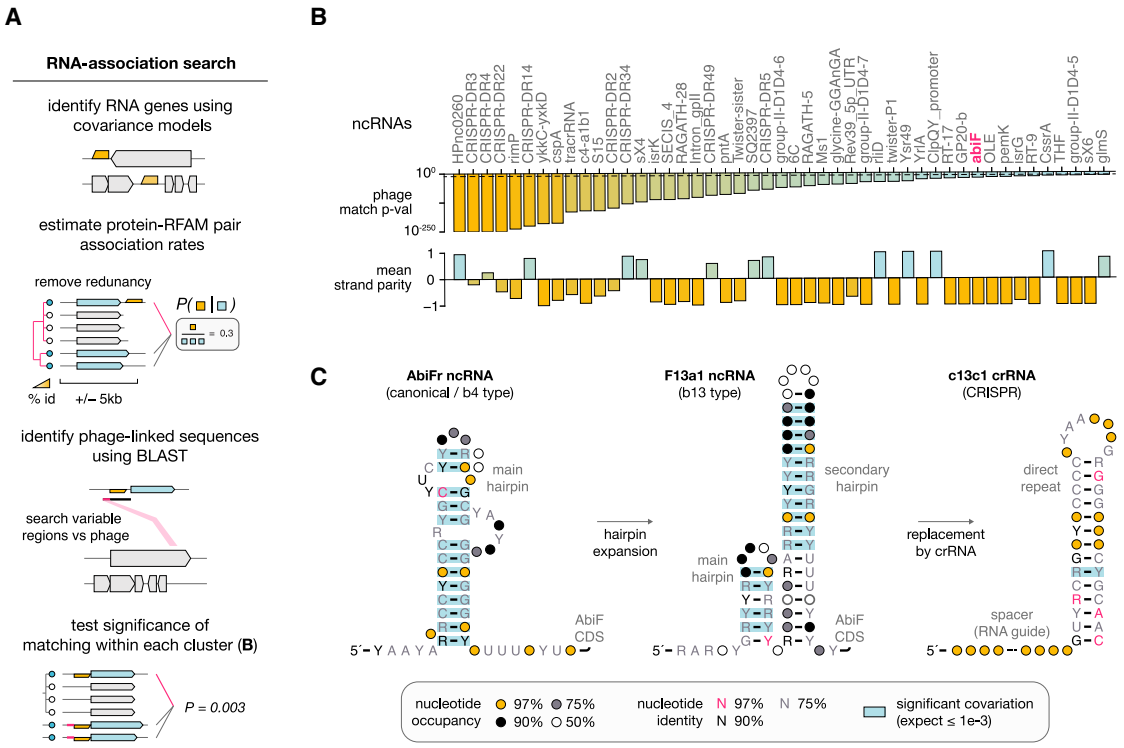


Figure 2. Phage-linked ncRNA search

(A) Schematic of the RNA association search for conserved prokaryotic protein-ncRNA systems that recombine with phage.

(B) Phage-matching RNA analysis of all RFAM RNA profile candidates that are stably associated with protein clusters and pass filters. Above, *p* values of phage matches, with dashed line indicating cutoff of 1e−12. Below, mean RNA strand preference (1 for matches to the [+] strand and −1 for matches to the (−) strand of phage-coding sequences).

(C) RNA covariance models for the canonical PbAbiF RNA motif (left), F13a1-associated ncRNA (middle), and the c13c1 CRISPR RNA (right). Significant covariation is shown with teal rectangles, indicating support for the existence of RNA structure. Major RNA changes under the inferred evolutionary scenario of AbiF/Cas13 evolution are shown using arrows.

See also Figures S2 and S3.

phylogenetic analysis of proteins with low sequence conservation can produce inaccurate reconstructions.[29] Given the extreme sequence and structural diversity of the HEPN domain superfamily,[12] we sought to delineate the specific group of HEPN protein families most closely related to Cas13 before conducting phylogenetic analyses. To this end, we designed a hybrid sequence-structure homolog search that starts with a permissive hidden Markov model (HMM) search against a clustered protein database, followed by structure-based filtering and grouping to identify HEPN families related to Cas13 (Figure 1A).

For the homolog search, we clustered 9 billion proteins from prokaryotic (meta)genome databases (retrieved in 2022)[30–32] down to 30% sequence identity by using fast locality-sensitive hashing-based clustering (FLSHclust).[28] We then performed a profile-profile search using extended HEPN domains from Cas13a/b/c/d and c13c1 as queries against the clusters, retaining clusters with a conserved HEPN catalytic motif indicative of active nuclease activity (Data S1).[33] The retained clusters were merged into superclusters using profile alignment, and representative protein structures for each protein supercluster were modeled using AlphaFold2.[23] The proteins were then hierarchically clustered by DALI structural similarity,[19] using the unweighted pair-group method of analysis (UPGMA) method (Figure 1B). From this, we identified at least eight major groups, each comprising one or more HEPN families (Figure 1B, top). One group consisted of CRISPR-linked Csm6 (see discussion) and another included Cas13, c13c1, and two sets of AbiF/Abi-2-like proteins (Figure 1B, bottom inset). Within the Cas13-containing group, we further identified two previously undescribed families, which we designated F13a1 (~450–500 aa) and F13a2 (~550–650 aa). Similarly to Cas13s, F13a1/2 contain two HEPN domains and are distantly related to both Cas13b's HEPN domains (up to 25% sequence similarity according to BLOSUM62) and AbiF's HEPN domain (also up to 25% sequence similarity). F13a1/2 share a conserved C-terminal domain (F13a-D1), and F13a2 contains an additional domain (F13a-D2) between the two HEPN domains (Figure S1).

Next, we compared the solved Cas13 structures with representative solved and predicted HEPN structures using DALI, revealing high structural similarity between Cas13s and AbiF (DALI *Z* score of up to 8.9) as well as between Cas13s and F13a1/2 (*Z* scores of up to 9.3) (Figure 1C). We additionally found that all HEPNs from Cas13s, as well as c13c1 and F13a1/2,

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

exhibit a circular permutation of the four main helices that define the HEPN domain, denoted H1–H4 in the order the helices appear in protein (Figure S2). This permutation is also present in AbiF-like HEPNs and Swt1 but not in the other examined HEPNs (Figure S2). AbiF and related HEPN domains contain a conserved, AbiF-specific subdomain consisting of AbiF-subdomain-I and AbiF-subdomain-II, which are split by the HEPN helix 1 and appear to interact across the dimer interface (Figures S1A and S2). Notably, c13c1 also contains a two-part subdomain with both structural and sequence similarity to the AbiF subdomain, whereas F13a1/2 appear to have small remnants of the AbiF subdomain (Figures S1 and S2; Data S1). Furthermore, c13c1 and the AbiF-like Ava_2192 encompass related active site rearrangements whereby the highly conserved glutamic acid is replaced by an aspartic acid on AbiF-subdomain-II (Figure S2). Moreover, AbiF, c13c1, and F13a1/2 share a structurally conserved positively charged groove potentially involved in nucleic acid binding (Figures S1A and S1B). Structural comparisons of individual HEPNs suggest that Cas13a/b/c/d, F13a1/2, and c13c1 derived both of their HEPNs from AbiF-like HEPNs (except perhaps in the limited context of Cas13b's [although not Cas13b-t's] HEPN2 domain) (Data S2.1). Paired with remote sequence similarity[34] (Data S2.2), these structural comparisons indicate that across all considered proteins from the HEPN superfamily, F13a1/2 and AbiF/Abi-2 are most closely related to Cas13a/c/d and c13c1 (Figures 1B, 1C, and S2; Table S1; Data S1).

We then conducted a phylogenetic analysis of the HEPN domains from Cas13a/b/c/d, c13a1, F13a1/2, and AbiF (Figures 1D and S3). First, a sequence alignment of the HEPN domain core (helices 1–4) from these HEPN domains was constructed using multiple aligment using fast Fourier transform (MAFFT).[35] We then used IQ-Tree2 to generate a maximum-likelihood phylogenetic tree, retaining the best tree out of five independent runs.[36] In the resulting tree, Cas13a and Cas13d are sister clades (bootstrap 100/100), recapitulating the previously reported similarities between these Cas13 subtypes.[11,37] Similarly, F13a1/2 are closely related (bootstrap 100/100). Overall, the tree topology is concordant with the similarities observed in sequence comparisons (Figure 1D; Data S2.2). A small set of AbiF-like proteins appear to be the evolutionary intermediate between the main AbiF clade and Cas13/F13a1/2/c13c1 (bootstrap 92/100). We denoted the clade of AbiFs most closely linked to Cas13s as the putative founder AbiF clade (Figure 1D). Regarding the early evolution of Cas13s from AbiF, we found that ancestral F13a1/2 appear to be unique intermediates between AbiF and Cas13b (bootstrap 87/100) that, together with Cas13b, form a distinct lineage apart from Cas13a/c/d and c13c1 (bootstrap 86/100). By contrast, c13c1 and Cas13c form sister clades (bootstrap 94/100) with a common ancestor that serves as an evolutionary intermediate between AbiF and Cas13a/d (bootstrap 86/100) (Figures 1D and S3). The placement of F13a1/2, c13c1, and Cas13c (in particular Cas13c-t) as evolutionary intermediates is consistent with their smaller sizes relative to Cas13a/b/d (Figure S3). Overall, the structural comparisons, sequence similarities, and phylogenetic analyses concordantly indicate that AbiF, F13a1/2, c13c1, and Cas13a/b/c/d likely share common ancestry, with the putative founder AbiF clade, c13c1, and F13a1/2 serving as key evolutionary intermediates in the evolution of Cas13s.

Identification of ncRNAs associated with AbiF, F13a1, and c13c1

The ancestors of Cas9 and Cas12, the OMEGA effectors IscB and TnpB, are associated with a ncRNA (ωRNA).[6] We sought to determine if the likely ancestors of Cas13, c13c1, AbiF, and F13a1/2 are also associated with ncRNAs, and if so, what their biological function is. To assess the potential association between known ncRNAs, as well as the broader landscape of prokaryotic ncRNAs, we conducted a comprehensive search for ncRNA-protein associations using the guilt-by-association approach (Figure 2A).[2,11,28,38–40] For this analysis, we tabulated genomic co-occurrences (within 10 kb) between protein clusters and known ncRNAs, finding 175,510 significant associations (Figure 2A; Table S2; STAR Methods).[41] Unexpectedly, we found a statistically significant association between AbiF and the Rfam (RNA family) abiF ncRNA, so named owing to the occasional presence of neighboring *abiF* genes in ∼25% of the loci reported previously (Table S3).[42] To avoid naming ambiguity, we hereinafter refer to the AbiF-associated ncRNA as AbiFr. For one AbiF protein cluster, 75% of the corresponding loci were associated with the *abiFr* ncRNA gene, indicating that some *AbiF* genes are stably associated with *AbiFr* ncRNA genes at high rates, suggesting a functional association between AbiF and AbiFr (Table S3). Our analysis further revealed that *abiFr* ncRNA genes, along with 46 other defense-enriched ncRNAs, recombine with phage genes with a strand preference, suggesting AbiFr may be involved in host-pathogen conflict using an RNA (strand-dependent) mechanism (Figure 2B).

We next developed a covariance-based structure prediction of the full *abiFr* ncRNA gene to better understand the 5′ phage-matching region, revealing conserved terminal extensions not found in the Rfam ncRNA model (Figure 2C).[43] We found that the conserved positions at the 5′ end of the locus likely correspond to a bacterial σ70 promoter (Data S2.3A), suggesting a short single-stranded portion of the 5′ hairpin extension may be expressed.[44] The 5′ upstream region of the *abiFr* gene that matches phage sequences occasionally corresponds to this promoter region (Data S2.3B and S2.3C), suggesting *abiFr* promoters may recombine with those from other genes, explaining the strand preference we found (Figure 2B). The 3′ end of the *abiFr* ncRNA gene is a poly U stretch that likely serves as a Rho-independent terminator signal. Using these signals, we predicted the full ncRNA gene and gene product (Figure 2C; Data S2.4A). A covariance model of the c13c1 CRISPR direct repeat (DR) showed that it likely adopts a stem loop without mismatches or bulges and differs substantially at the sequence level from the AbiFr ncRNA, suggesting the two ncRNAs may perform distinct functions (Figure 2C).

Given that only ∼19% of AbiF clusters contained an Rfam *abiFr* ncRNA gene nearby (<500 bp), we explored if *abiF* genes are associated with other diverse ncRNAs beyond Rfam *abiFr*. We identified conserved ncRNAs upstream and downstream of *abiF* genes along various branches in the dual-HEPN phylogenetic tree, resulting in 29 diverse AbiFr ncRNA covariance models (named b2, …, b29) (see STAR Methods). Our analysis

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

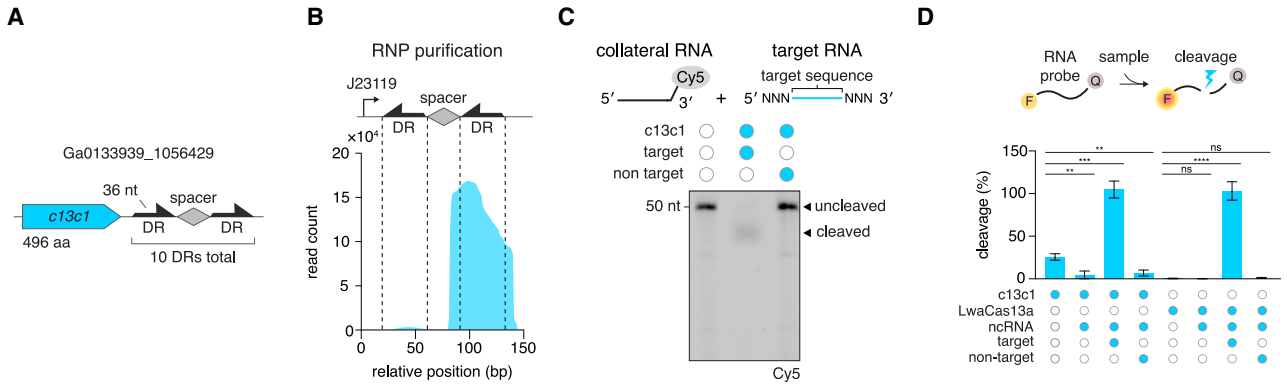


Figure 3. Characterization and analysis of c13c1

(A) Locus schematic of c131 locus (JGI: contig id Ga0133939_1056429) (not to scale).

(B) RNA sequencing reads from c13c1 (JGI: contig id Ga0133939_1056429) RNP pull-down aligned to the c13c1 locus.

(C) Collateral activity assay of c13c1. c13c1 was incubated with either a target or non-target RNA PFS library and a separate RNA reporter labeled with Cy5 at its 3′ end.

(D) Top: schematic of RNaseAlert reporting system. Bottom: collateral activity assay of c13c1 and LwaCas13a with or without crRNA and with or without target or non-target RNA PFS libraries. $^{**}p < 0.01$, $^{****}p < 0.0001$ (two-tailed Student's t test).

revealed that AbiF-like systems are broadly associated with ncRNAs, with 88% of AbiF clusters located in the vicinity of one of the 29 *abiFr* ncRNAs (Figures 2C and S3; Data S3). Beyond the canonical single-kinked hairpin structure, AbiFr ncRNAs can also contain two variably spaced hairpins (b29-1, b13, b16), large bulges (b23), or conserved bifurcated stem loops (b7).

We remapped these ncRNA families back onto the phylogenetic tree and found that while most F13a12 genes were not associated with any *abiFr* ncRNAs, one branch containing three F13a1 systems was associated with an upstream b13-like *abiFr* ncRNA gene corresponding to a ncRNA structure with two hairpins (Figure S3; Data S2.3D and S2.3E). This clade was additionally associated with two conserved downstream protein-coding genes, one of which encodes an nonsense mRNA decay 7 (NAM7)-like RNA helicase. A ncRNA in the vicinity of only some F13a1 genes suggests that F13a1 proteins may also utilize ncRNAs encoded in *trans*. Combined with our phylogenetic analysis, the presence of diverse associated ncRNAs for AbiF, F13a1, and c13c1 suggests that *cis*-encoded ncRNAs were a key feature throughout the evolution of Cas13s from AbiF.

Biochemical characterization and analysis of c13c1

In contrast to Cas13a/b/c/d proteins, c13c1 is substantially smaller (~450–550 aa) and lacks a large REC domain, which is known to be involved in crRNA binding (Figure S3; Data S1).[45] To test if c13c1 can bind a crRNA, we co-expressed c13c1 (JGI: contig id Ga0133939_1056429) with a DR-spacer-DR (Figure 3A). Small RNA sequencing (sRNA-seq) on the RNP pull-down revealed c13c1 associates with a mature crRNA consisting of a 14-nt spacer followed by a 36-nt DR and 5 additional downstream nucleotides (Figure 3B). We note that for sRNA-seq library preparation, a protocol without adaptor ligations enabled amplification and mapping to the *c13c1* ncRNA locus, suggesting terminal modifications may hinder ligation-based methods.

We next tested if c13c1 has RNA-guided RNA cleavage activity. Cas13s are known to recognize a 1- to 3-nt protospacer flanking sequence (PFS) and exhibit non-specific collateral RNase activity following target recognition.[24,46] We generated a library with one target sequence flanked by 3 random nucleotides. c13c1 RNPs were incubated with Cy5-labeled random RNA (collateral RNA) and either the target or non-target RNA libraries. Cleavage of collateral RNA was observed only in the presence of the target RNA (Figure 3C).[24,47] We quantified RNase activity upon target recognition using the RNaseAlert reporter assay in which c13c1 was incubated with either the target or non-target PFS library and a reporter single-stranded RNA (ssRNA). We observed robust, target-dependent collateral activity of c13c1 (Figure 3D). Unexpectedly, a fluorescence signal was observed when apo c13c1 was incubated with the reporter RNA (but it was not observed with the apo form of *Leptotrichia wadei* Cas13a). This suggests that c13c1 is a non-specific nuclease that is weakly autoinhibited on its own, strongly inhibited when bound to crRNA, and fully activated when bound to both crRNA and target RNAs. Based on these experimental results, along with the phylogenetic distinctness of c13c1 relative to Cas13a/b/c/d, we designate c13c1 as Cas13e to represent a separate subtype of type VI CRISPR-Cas systems in accordance with previous criteria for designating subtypes.[11]

AbiF is a toxin with an RNA antitoxin

Our sequence and structural comparisons suggested that the potential ancestor of Cas13 is AbiF, a protein linked to abortive infection through unknown mechanisms.[48–50] To better characterize AbiF, we performed sRNA-seq on *Paenibacillus donghaensis*, which encodes an AbiF and AbiFr ncRNA pair, and observed the expression of both components (Figure S4A). Heterologous expression in *E. coli* of *Prevotellaceae bacterium* AbiF (PbAbiF), a system in which the AbiFr ncRNA includes an upstream region matching to phage (E value < 1e−3) (Data S2.3B and S2.3C) inhibited bacterial growth in both liquid culture and

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

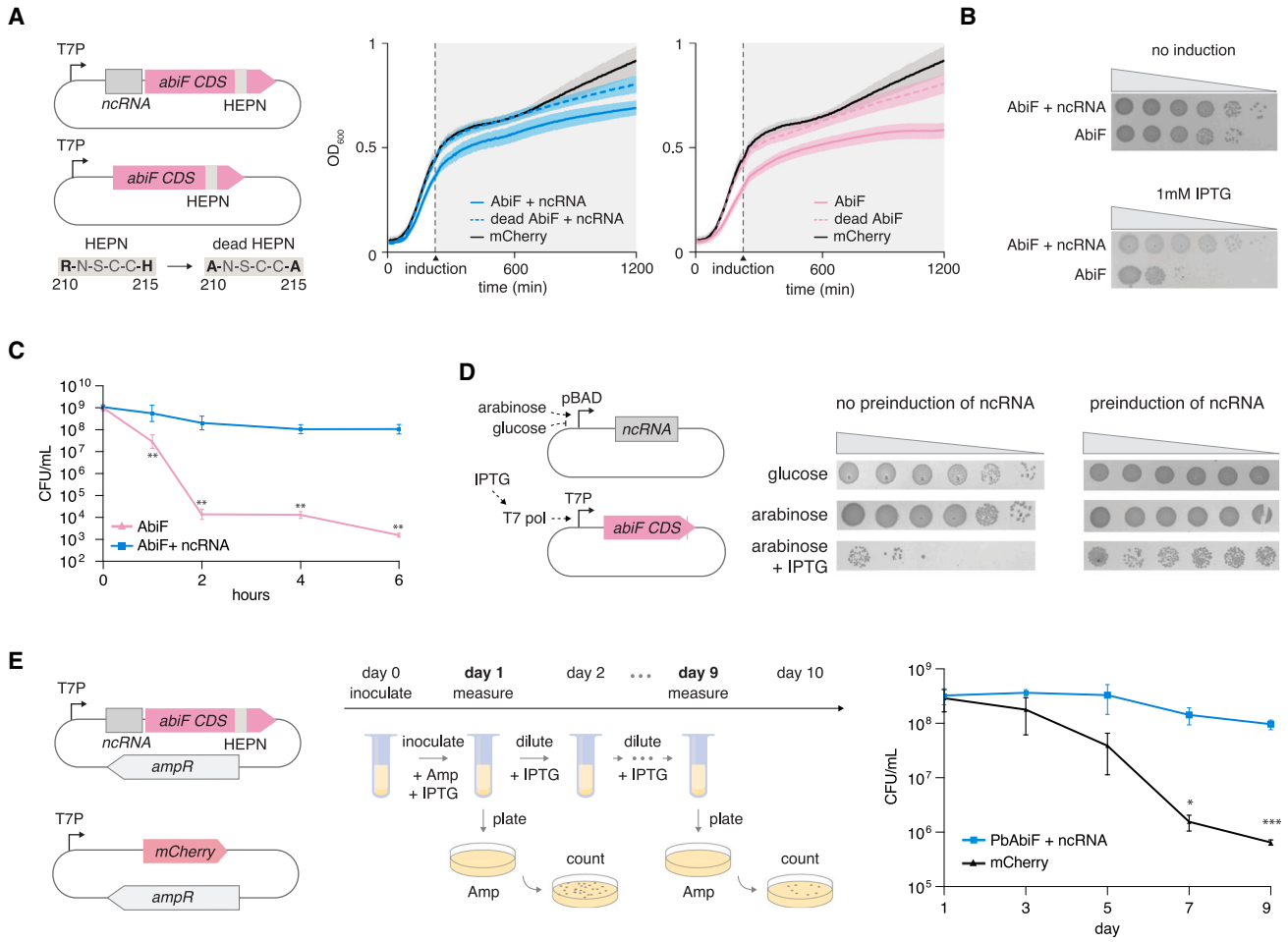


Figure 4. AbiF is a type III toxin with an RNA antitoxin

(A) Left: schematic representation of the *E. coli* expression vectors for PbAbiF and its ncRNA. Bottom left: point mutations (R210A and H215A) resulting in dead PbAbiF. Right: growth assays of PbAbiF and dead PbAbiF induction, with and without PbAbiFr ncRNA along with an mCherry control. All samples were generated in the same assay but plotted separately for clarity.

(B) Spotting assay for PbAbiF + ncRNA and PbAbiF without ncRNA, with and without IPTG induction using the plasmids described in (A).

(C) CFU count over time of *E. coli* expressing either PbAbiF + ncRNA or PbAbiF without ncRNA. The bacteria were grown in liquid culture until reaching a logarithmic phase, induced using IPTG, and then plated on induction plates at different time points. $^{**}p < 0.01$ (two-tailed Student's t test, $n = 3$).

(D) Left: schematic of the *E. coli* expression vectors for PbAbiF and its ncRNA used for the spotting assay. Right: spotting assay of the PbAbiF system with and without the pre-induction of ncRNA.

(E) Left: schematic of the *E. coli* expression vectors for the PbAbiF and PbAbiFr ncRNA, and mCherry used for the plasmid maintenance assay. Middle: schematic of plasmid maintenance assay. Right: CFU counts from Amp plates over time. $^{*}p < 0.05$, $^{***}p < 0.001$ (two-tailed Student's t test, $n = 3$).

See also Figure S4.

solid medium culture spotting assays (Figures 4A and 4B). The growth inhibition phenotype was more severe when the protein was expressed alone (Figures 4A and 4B). Similar results were also observed for an additional AbiF protein encoded by *E. coli* strain SMN197SH3 (EcAbiF) (Figures S4B–S4D). Expression of a HEPN-inactive variant (R210A+H215A) rescued the growth inhibition phenotype irrespective of ncRNA expression (Figure 4A).

To determine if the growth defects were bacteriostatic (inhibits growth) or bactericidal (leads to cell death), we assessed viable cell counts. When the PbAbiF native locus was induced, co-expression of the PbAbiF and its ncRNA exhibited bacteriostatic effects, while expression of PbAbiF without its ncRNA proved to be bactericidal (Figure 4C). We found that pre-induction of the associated ncRNA prior to AbiF expression completely abolished the toxic effect of PbAbiF (Figure 4D). HEPN-inactive PbAbiF showed no growth inhibition in liquid culture, but it caused a slight defect in the spotting assay (Figure S4E). This was rescued by pre-induction of the ncRNA, possibly reflecting the reduced solubility of the AbiF protein in the absence of its associated ncRNA (Figure S4F). A plasmid addiction assay showed that the AbiF system effectively promotes stable plasmid inheritance in *E. coli*, as has been shown for some other TA systems[51–53] (Figures 4E and S4G).

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

CellPress
OPEN ACCESS

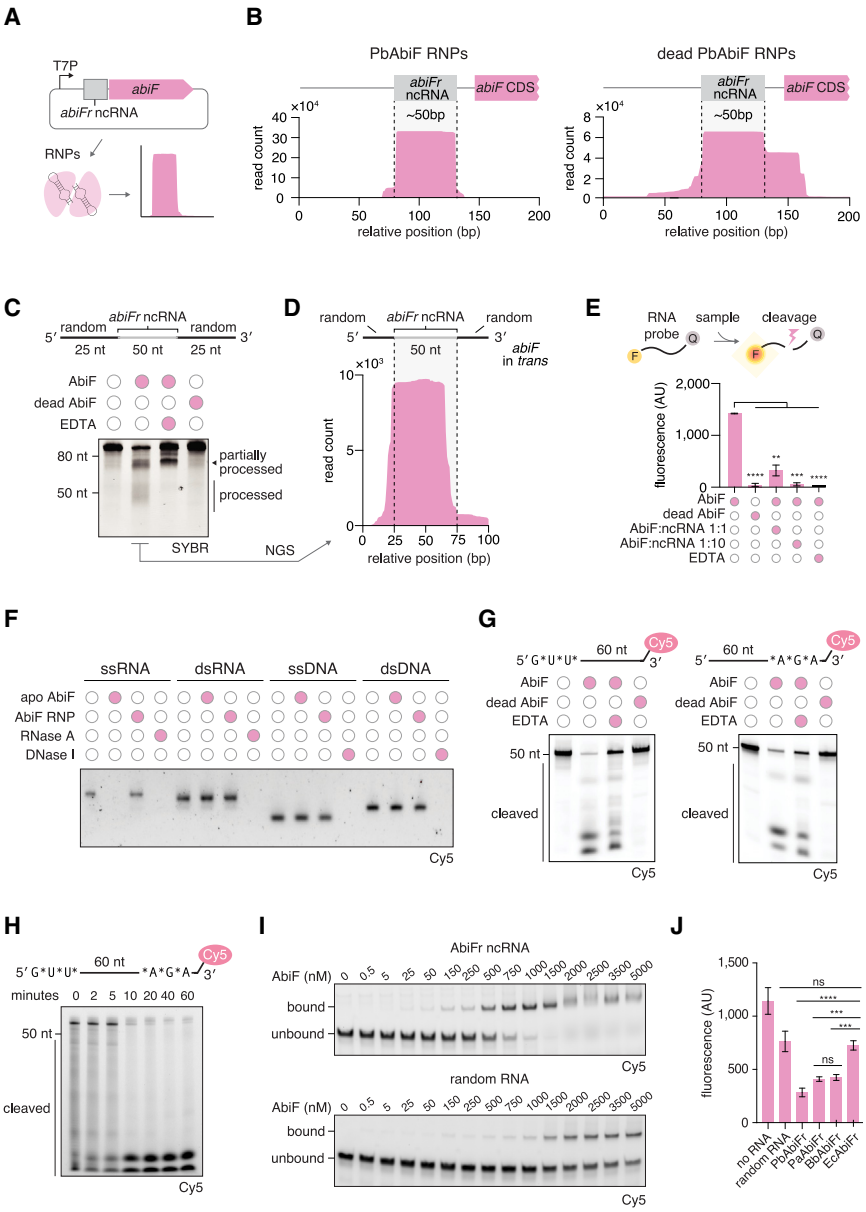


Figure 5. AbiF RNase activity is inhibited by the AbiFr ncRNA

(A) RNP pull-down and sequencing workflow. The plasmid was expressed in BL21(DE3), containing T7 RNA polymerase under IPTG induction.

(B) RNA sequencing reads from PbAbiF RNP pull-down (left) and catalytically dead PbAbiF RNP (right) pull-down aligned to the *PbAbiF* locus.

(C) AbiFr ncRNA processing by PbAbiF. Catalytically active and dead PbAbiF were incubated with ssRNA composed of the AbiFr ncRNA and 25 random nucleotides upstream and downstream of the ncRNA sequence. The RNA was visualized using a SYBR stain.

(D) RNA sequencing reads from the PbAbiFr ncRNA processing *in vitro* reaction aligned to the template RNA of the ncRNA processing reaction.

(E) Top: schematic of PbAbiF RNase activity assay. Bottom: activity assay of the PbAbiF with or without PbAbiFr ncRNA. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$ (two-tailed Student's t test, $n = 3$).

(F) Substrate specificity assay of PbAbiF. Apo PbAbiF or PbAbiF RNP were incubated with ssRNA, dsRNA, ssDNA, and dsDNA.

(G) PbAbiF endonuclease activity assay. Catalytically active PbAbiF and dead PbAbiF were incubated with a random ssRNA sequence harboring either three 5′-terminal (left) or three 3′-terminal (right) phosphorothioate bonds (asterisks), which were subsequently visualized using Cy5.

(H) Time course assay of PbAbiF endonuclease activity. A random ssRNA sequence harboring three 5′- and 3′-terminal phosphorothioate bonds (asterisks) was incubated with PbAbiF and visualized using Cy5.

(I) Left: EMSA of catalytically dead PbAbiF combined with Cy5-labeled AbiFr ncRNA (top) or Cy5-labeled random RNA (bottom).

(J) RNaseAlert activity assay for PbAbiF protein in the presence of apo AbiF (no additional RNA) or PbAbiF with one of the following: random RNA, PbAbiFr, PaAbiFr, BbAbiFr, or EcAbiFr. $*p < 0.05$, $**p < 0.01$ (two-tailed Student's t test, $n = 3$).

See also Figures S3, S4, and S5.

Based on these observations, we propose that AbiF is a type III TA system. Type III TA systems consist of two components: a ncRNA antitoxin, often featuring a pseudoknot structure, and a cognate protein toxin, usually an RNase.[15,54,55] These systems also have a distinct locus architecture, with the antitoxin gene located upstream of the toxin gene, as we observed in most of the AbiF loci. We additionally identified a poly-T tract at the 3′ end of the *abiFr* ncRNA gene, suggesting the presence of a Rho-independent transcriptional terminator, as is found in other type III TA systems.[56]

AbiF RNase activity is inhibited by the AbiFr ncRNA

We next confirmed that AbiF binds its cognate ncRNA by performing sRNA-seq following pull-down of an overexpressed PbAbiF or EcAbiF in *E. coli* (Figures 5A, 5B, and S4H). The bound RNA matched the predicted AbiFr ncRNA, and sequencing confirmed the presence of the 5′ and 3′ single-stranded extensions of the hairpin in both systems (Data S2.4C and S2.4D). In other type III TA systems, the maturation of the RNA antitoxin is catalyzed by the toxin itself,[54,55] and several Cas13s also process their own crRNA.[10] To investigate if AbiF similarly processes its ncRNA, we purified the HEPN-inactive PbAbiF RNP (R210A, H215A) or EcAbiF RNP (R213A, H218A) complexes and performed RNA-seq, finding longer extensions at both ends of the ncRNA, suggesting processing of the ncRNA by the HEPN domain of AbiF (Figures 5B and S4I). To test this activity biochemically, we designed a 100-nt RNA substrate containing the PbAbiFr ncRNA, flanked by 25-nt extensions at both ends. Incubation with active PbAbiF revealed a prominent 75-nt fragment, suggesting incomplete processing (Figures 5C

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

and S5A). RNA-seq analysis of the processed PbAbiF RNA from the *in vitro* reaction revealed AbiF-mediated cleavage at both AbiFr ends, with a distinct bias toward the 5′ flanking region based on read coverage (Figure 5D). These differences may indicate that the *in vitro* reaction setup might not accurately capture all the necessary conditions to achieve complete processing.

Given that many Cas13s exhibit non-specific collateral activity following target recognition, we tested if AbiF RNase activity extends to other RNA species, using the RNaseAlert reporter system (Figures 5E, S5B, and S5C). When PbAbiF was incubated with the reporter RNA alone, a fluorescence signal was detected. HEPN-inactive PbAbiF lacked detectable RNase activity, as did the catalytically active PbAbiF RNP complex (PbAbiF incubated with its cognate ncRNA), indicating that the ncRNA strongly inhibits AbiF. When PbAbiF was incubated with the reporter RNA and random RNA, a slight decrease in signal was detected, attributed to the overall increase in ssRNA concentration in the sample. Similar results were obtained with EcAbiF (Figure S5D). Furthermore, EDTA abolished the RNase activity of PbAbiF (Figures 5E, S5B, and S5C). We note, however, that processing of the PbAbiFr ncRNA is only partially inhibited by EDTA, although this assay was performed in the same buffer as the ncRNA gel cleavage assay (e.g., $MgCl_2$ was present in both conditions at the same concentration) (Figure 5C).

To assess the substrate specificity of AbiF, we incubated PbAbiF with ssRNA, double-stranded RNA (dsRNA), ssDNA, and dsDNA substrates. ssRNA was cleaved by PbAbiF, but no nuclease activity was detected for other substrates, as has been reported for other HEPN nucleases (Figure 5F).[14,57,58] PbAbiF RNPs showed no detectable cleavage activity for ssRNA or any other tested substrate, emphasizing the inhibition of PbAbiF by its cognate ncRNA. To evaluate the cleavage mode of AbiF, we employed a random RNA substrate harboring three 5′- or 3′-terminal phosphorothioate (PS) bonds and a 3′-terminal Cy5 fluorescent label. Following incubation with PbAbiF, the full-length, modified RNA levels decreased and small Cy5-labeled RNA fragments were detected, indicative of endonucleolytic activity, consistent with the properties of other type III toxin endonucleases[55,59] (Figure 5G). Conducting a time series cleavage assay with an RNA substrate harboring three PS bonds at both termini confirmed endonucleolytic cleavage by AbiF that occurs rapidly over 10 min (Figure 5H).

To further understand AbiF's binding preferences between AbiFr ncRNA and random RNA, we performed an electrophoretic mobility shift assay (EMSA). Using the AbiFr ncRNA sequence identified through sRNA-seq (Data S2.4C and S2.4D), we found that the dead AbiF protein binds to both its cognate ncRNA and the random RNA control with a higher affinity for its ncRNA (Figure 5I). We analyzed the binding curve resulting from the RNA bound percentage as a function of AbiF concentration (Figure S5E; Data S2.5). Due to the steepness of the binding curve for AbiF-AbiFr ncRNA binding, we fit each binding curve to the Hill equation and found PbAbiF-PbAbiFr ncRNA binding exhibited a dissociation constant (K_D) of 410 ± 28 nM with a Hill coefficient of 2.0 ± 0.2, indicating cooperative binding, whereas PbAbiF-random RNA binding exhibited a K_D of 4.92 ± 0.71 μM and Hill coefficient of 1.3 ± 0.25, indicating AbiF preferentially binds the AbiFr ncRNA over random RNA (Figure S5E). The association rate between AbiF and the AbiFr ncRNA may be bolstered *in vivo* by the co-transcription of these two components.

To characterize the specificity of the AbiF-ncRNA inhibition across distinct systems, we tested the RNase activity of PbAbiF in the presence of its cognate ncRNA (b4 type from the diverse ncRNA search) as well as three other ncRNAs from *Prevotellaceae copri* (PcAbiF, b4 type) and *Bacteroidaceae bacterium* (BbAbiF, b4 type), of the same branch as PbAbiF, and the distantly related EcAbiF (b18_U1 type) (Figure S3; Data S3). We observed that PbAbiF activity was inhibited by PcAbiF and BbAbiFr ncRNAs, suggesting robust AbiF binding to the cognate and related ncRNAs (Figure 5J). By contrast, the EcAbiFr ncRNA showed a minimal inhibitory effect, highlighting orthogonality across divergent evolutionary lineages and structurally divergent ncRNAs.

We then explored potential anti-AbiF mechanisms. Previous work compared sequences from *Lactococcus* phages sk1 and 712, identifying a minimal sequence shared between the two that includes DRs and a hairpin structure. Despite sequence similarity, only the sk1 sequence confers immunity against the *Lactococcus Lactis* AbiF-like system, suggesting that specific structural features within these DRs and the hairpin play a key role in inhibiting the AbiF-like activity.[50] Although containing bulges, these hairpins do not resemble the PbAbiFr ncRNA at the sequence level (Data S2.4B). Nevertheless, we found that the sk1 hairpin substantially inhibited PbAbiF activity, compared with random RNA (Figure S5F). This finding underscores the unique mechanism by which sk1 resists the *Lactococcus Lactis* AbiF-like system, potentially linked to these structural elements, and underscores the involvement of AbiF in the host-pathogen arms race.

Structure of the AbiF RNP complex

To characterize the molecular basis of the inhibition of AbiF by the ncRNA, we purified the PbAbiF RNP and used cryo-EM to resolve its structure at 3.6 Å resolution (Figures 6A, 6B, and S6A–S6E; Data S2.6). PbAbiF adopted a four-helix HEPN domain fold that is augmented by an N-terminal extension (AbiF-subdomain-I) and an insertion between HEPN helices H1 and H2 (AbiF-subdomain-II) (Figure 6C). Note HEPN helices are numbered H1–H4 by their sequential order in the HEPN domain. As described above, Cas13s contain a circular permutation of their HEPN domain relative to canonical HEPNs. For example, Cas13a's HEPN-1 domain has a circular permutation where "helix 1" of the canonical HEPN occurs at the C terminus of Cas13a as helix H4; this non-canonical rearrangement is also observed in PbAbiF, Cas13b/c/d, F13a1/2, and Cas13e AlphaFold2 structures, further supporting their shared ancestry (Figures 6D and S2). Like other HEPN-containing proteins,[14] the PbAbiF RNP complex forms a symmetric dimer (Figure 6B), with the HEPN domains of each monomer interacting face to face, bringing the active site residues of each monomer (H210 and R215) into close proximity around the 2-fold symmetry axis (Figures 6B and 6C).

Each PbAbiF monomer binds to a single AbiFr ncRNA molecule, forming a symmetric RNP with a 2:2 protein:RNA ratio, similar to other type III TA systems.[61,62] The 49 observed

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

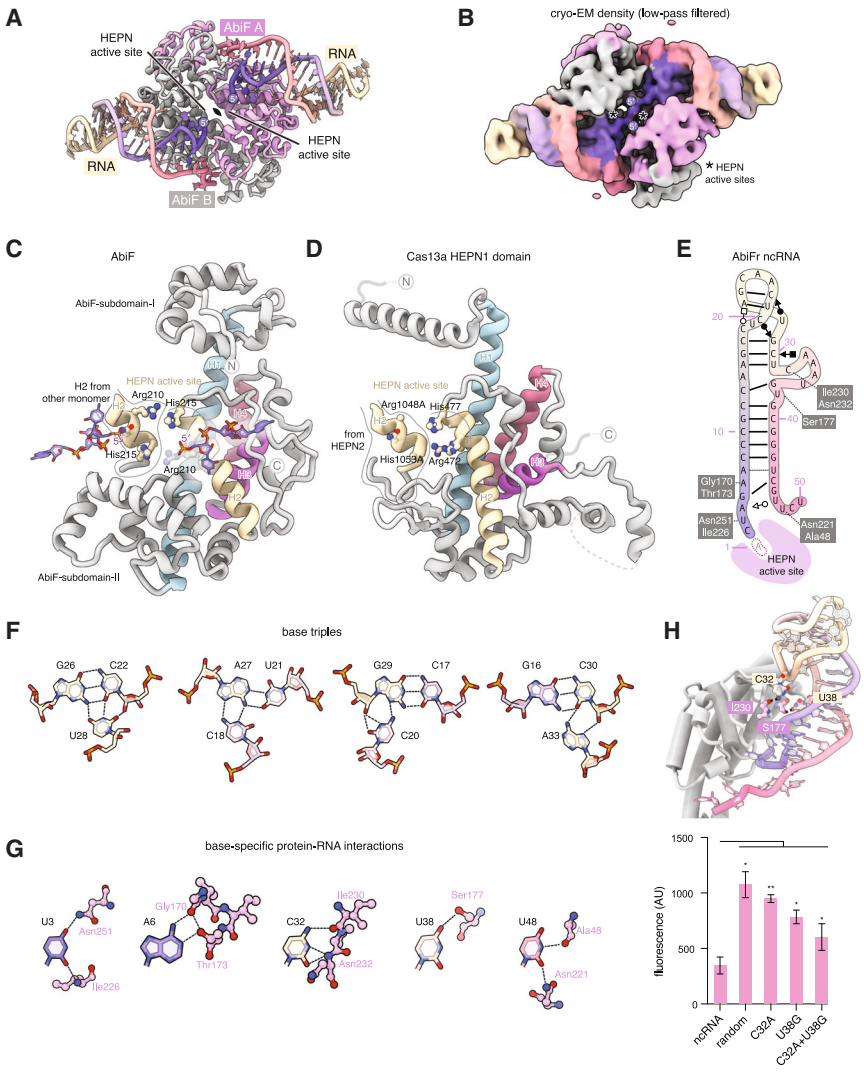


Figure 6. Structure of the AbiF RNP complex

(A) Cryo-EM reconstruction of the PbAbiF-ncRNA complex.

(B) Model of the PbAbiF-ncRNA complex. Each protein monomer is colored pink or white. The RNA is shaded 5′ to 3′ with a purple-yellow-pink gradient from. The view is aligned with the 2-fold symmetry axis (almond symbol). Inset: cryo-EM density for a putative nucleotide ligand and its coordinating residues.

(C) A highlighted single protein monomer in PbAbiF. The four core helices of a HEPN domain are colored, along with helix H2 from the other protein monomer. The three 5′-most bases visible in the ncRNA are shown as sticks.

(D) The same view as (C) but for Cas13a from *Lepotrichia buccalis* (PDB: 5XWP).[45]

(E) Secondary structure of the PbAbiFr ncRNA. Solid lines represent Watson-Crick base pairs. Non-Watson-Crick base pairs are shown[60] with squares representing Hoogsteen edges, triangles representing sugar edges, black fill representing pairing in *cis*, and white fill representing pairing in *trans*. Protein residue hydrogen bonding with bases are shown.

(F) Details of the four base triples in the PbAbiFr ncRNA.

(G) Details of base-specific interactions between PbAbiF protein and ncRNA.

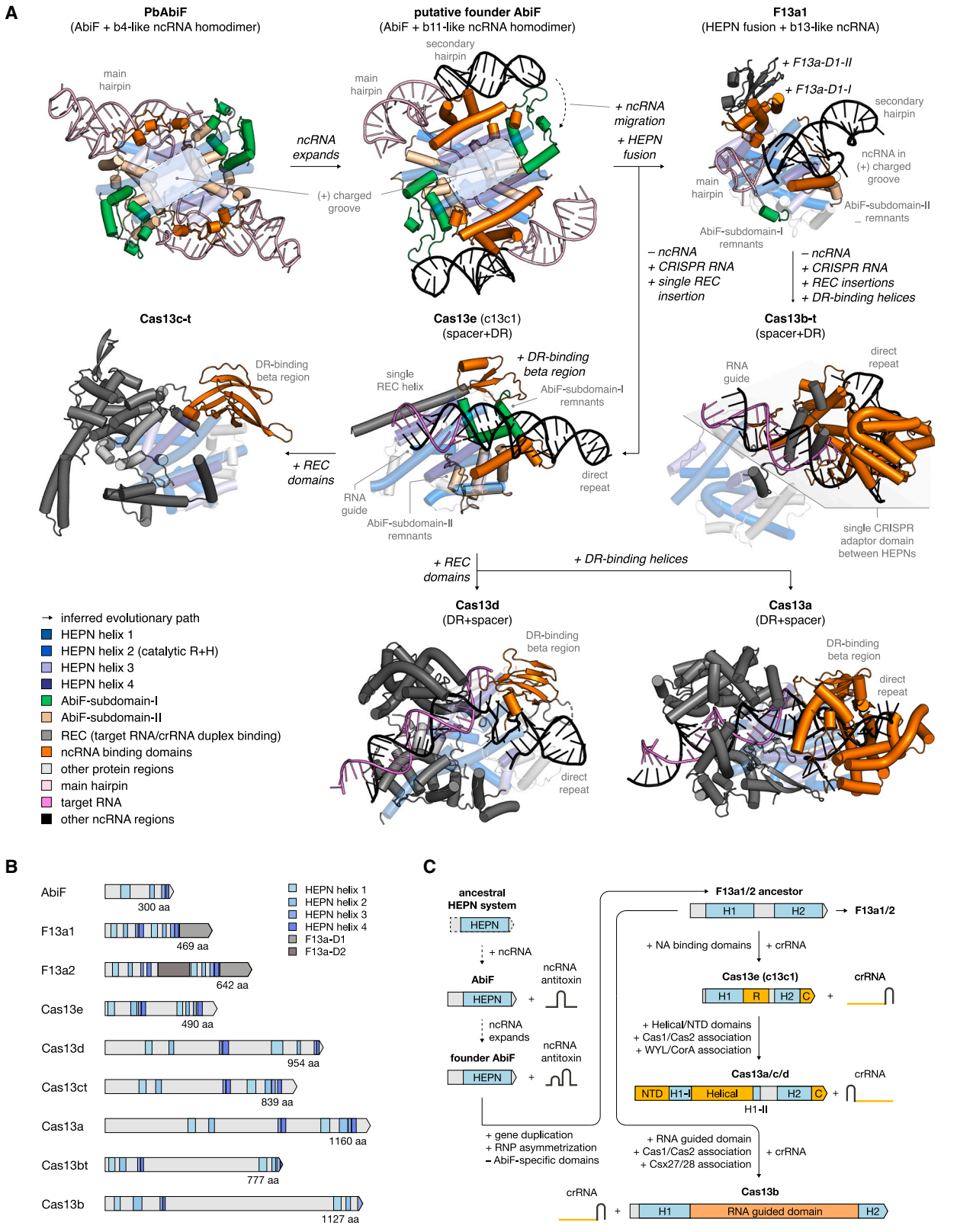
(H) Top: specific protein-RNA interactions between the PbAbiF protein and PbAbiFr ncRNA. Bottom: activity assay of PbAbiF in the presence of wild-type (WT) ncRNA, random RNA, and mutated ncRNA. $^{*}p < 0.05$, $^{**}p < 0.01$ (two-tailed Student's t test, $n = 3$). See also Figures S2 and S6.

nucleotides of the ncRNA in the cryo-EM density form two stacked helices (Figure 6E). The basal helix, between nucleotides 7 and 12 and 39 and 44, is GC-rich and lies over an AbiF-specific insertion between helices 1 and 3 of the HEPN domain. The distal helix, comprising nucleotides 15–31, forms an intricate hairpin with four base triple interactions that likely stabilize the AbiFr ncRNA structure (Figure 6F). A single-stranded, relatively conserved bulge between the two helices (nucleotides 32–38) forms base-specific interactions with the AbiF protein (Figure 6E). C32 forms hydrogen bonds with the backbone of residues 230 and 232, while U38 interacts with the Ser177 side chain (Figure 6G). To assess the functional significance of these interactions, we mutated the ncRNA motif at positions C32 and U38. We then assessed their effect on the ability of PbAbiF to cleave random RNA using the RNaseAlert assay. Both single mutants showed reduced inhibition of PbAbiF activity (Figure 6H). The double mutant also exhibited a significant reduction in the fluorescence signal, compared with a random RNA control, but it was less pronounced.

The structure also showed 5- and 6-nt ssRNA extensions of the ncRNA 5′ and 3′ of the basal hairpin, respectively (Figures 6A and 6E). These extensions snake over opposing surfaces of the AbiF protein, each making base-specific contacts with the protein. The most 3′ base observed was U50, but the structure does not rule out further bases in the sample. However, the 5′ end of the ncRNA points directly into PbAbiF active site, coming close to the 5′ end of its symmetry counterpart. A potential extra base is also present directly over the active site, but the proximity of this position to the symmetry axis of the cryo-EM density precludes detailed interpretation. This positioning suggests that the 5′ end of the ncRNA is processed by PbAbiF, and the RNP represents the post-cleavage state; alternatively, or additionally, the ncRNA 5′ end might directly block access to the active site and thus inhibit the cleavage activity of PbAbiF. Comparison with the PbAbiF-based ncRNA covariance model showed that these terminal cleavage sites are conserved among homologs (Data S2.4 and S3).

The cryo-EM reconstruction additionally contains an unassigned density within the AbiF subdomain that might represent a co-purified ligand (Figure 6A). This putative ligand is stacked at the dimer interface between Tyr43 of one monomer and

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034



(legend on next page)

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

Trp161 of the other monomer. The positively charged side chains Arg24 and His61 further coordinate the ligand. Based on these characteristics, the ligand could be tentatively identified as a bound nucleotide, potentially AMP, but the resolution was not high enough for definitive assignment. We hypothesized that this bound ligand might contribute to the activity of the PbAbiF protein. Accordingly, we screened the activity of apo-form PbAbiF in the presence of 20 different nucleotides/nucleosides. Activity was inhibited in the presence of nucleoside diphosphates, monophosphates, and cyclic nucleotides (Figure S6F), suggesting that the bound ligand allosterically regulates enzymatic function.

Evolution of the RNA-guided nuclease activity of Cas13

Finally, we mapped protein structures and other features onto the evolutionary scenario inferred from the sequence-based maximum-likelihood phylogenetic analysis to infer a detailed account of how the RNA-guided nuclease function of Cas13 might have emerged (Figures 1D and S3). For Cas13e, F13a1, and putative founder AbiF, RNPs were modeled with AlphaFold3,[63] achieving predictions with minimum protein/ncRNA interface errors of 4.4, 3.7, and 3.1 Å, respectively. For Cas13c-t, we used AlphaFold2 without RNA because the crRNA orientation is unknown.

Under the inferred evolutionary scenario, the AbiFr ncRNA apparently expanded, yielding two hairpins (main and secondary hairpins), resulting in a ncRNA with a large stem loop found in the putative founder AbiF clade (Figure 7A). Despite this expansion, the ncRNA's main hairpin still encompassed a 5′ extension that apparently blocks the HEPN active site, suggesting the ncRNA still functioned as an antitoxin (Figure 7A). The AbiF protein then likely underwent gene duplication and fusion, resulting in a dual-HEPN protein interacting with a b13-like AbiFr ncRNA. The binding site for this ncRNA was predicted to migrate into the positively charged groove (Data S2.7), which apparently co-evolved with the insertion of additional RNA binding helices alongside reductions to the AbiF subdomain (Figure 7A; Data S2.7). This transition was likely paired with the insertion of positively charged, beta-rich domains (F13a-D1-I and F13a-D1-II), which further expanded the positively charged groove, resulting in the emergence of extant F13a1/2 systems. The ancestral F13a1/2 systems subsequently lost their genomic association with the ncRNA antitoxin, potentially enabling them to interact with other RNAs such as crRNAs. During the subsequent evolution toward Cas13e, the ancestral F13a-D1-I domain likely persisted but rapidly evolved, allowing the protein to bind the crRNA stem loop. Additional helical insertions between the HEPN-1 helix 4 and the AbiF-subdomain-I remnant from HEPN-2 appear to have further facilitated binding to the crRNA stem loop (Figure 7). However, binding the DR hairpin alone was likely insufficient for full RNA-guided conformational activation of the HEPN domains. The crucial evolutionary step was likely the insertion of a highly charged REC helix, recognizing the matched guide-target RNA duplex and thus facilitating specific target recognition (Figure 7). Together, these evolutionary events appear to have given rise to an ancestral form of RNA-guided Cas13 similar to Cas13e.

Subsequent evolution of more complex forms, such as Cas13a, c, and d, involved additional structural adaptations. The Cas13e-like ancestor acquired extensive REC insertions, resulting in the increased guide length seen in Cas13a and b and likely enhancing mismatch detection (Figure 7). Notably, evolution of Cas13b appears distinct, possibly stemming from an ancestral system predating F13a1/2. In contrast to Cas13a, c, and d, both Cas13b and Cas13b-t contain unique insertions located almost exclusively between the two HEPN domains, facilitating binding to both the crRNA hairpin and the guide RNA/target RNA duplex (Figure 7). This divergence between Cas13b and Cas13a, c, and d is underscored by the distinct spacer configuration of Cas13b crRNAs (spacer + DR) and the lack of a structurally analogous DR-binding beta region (found in Cas13a, c, d, and e). In terms of the catalytic activity, the HEPN domain active sites of all Cas13s face away from the target RNA, suggesting that the collateral activity persisted as a core function starting from ancestral AbiFs. This scenario is supported by our observation that Cas13e has basal non-specific RNase activity that is inactivated by crRNA (Figure 3D), analogous to AbiF and its ncRNA (Figure 5E), indicating that Cas13e exhibits characteristics of both TA and CRISPR systems where the crRNA serves as both a guide and a target-dependent antitoxin. As Cas13e evolved into the larger Cas13a, the apo-form non-specific RNase activity appears to have been lost, marking the complete transition from a TA system to a specialized CRISPR system.

Notwithstanding the complexity of advanced RNA-guided systems such as Cas13a, b, c, and d, our findings suggest a relatively straightforward path of evolution of RNA-guided RNA-targeting nucleases from non-RNA-guided nucleases containing the same catalytic domain. The transition from a protein inherently capable of RNA recognition, such as AbiF, to a rudimentary RNA-guided system, such as Cas13e, apparently was enabled by a minimal set of insertions, one for binding the crRNA DR and another for binding the guide RNA-target RNA duplex.

DISCUSSION

Here, we showed that among all identified Cas13-like HEPN domain-containing proteins, AbiF and F13a1/2 are the most

Figure 7. Evolution of the RNA-guided nuclease activity of Cas13

(A) Structural comparison highlighting evolutionary domain acquisitions and modifications. Arrows represent likely evolutionary steps, with modern ancestors shown via structures. Italicized labels indicate key changes from previous step in evolution. Some domains are translucent for clarity. Structure sources: PbAbiF (PDB: 8VZ6), Cas13a (PDB: 5XWP[45]), Cas13b-t (PDB: 7VTN[64]), Cas13d (PDB: 6E9F[65]), Cas13e (AF3[63]), F13a1 (AF3), and putative founder AbiF (AF3), Cas13c-t (AF2).

(B) Protein domain architectures for the Cas13/AbiF/Abi-2 clade, highlighting the location of the four helical regions of the HEPN domain, with HEPN helix 2 containing the catalytic motif.

(C) Simplified evolutionary schematic detailing the evolution of Cas13 from AbiF.

See also Figure S3.

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

similar to Cas13 at both the structural and sequence levels and are likely the ancestral nucleases from which Cas13s evolved. Although other HEPN-containing nucleases, such as the CRISPR-associated Csm6 or the defense-associated DndB, could be appealing candidates for Cas13 ancestors from a functional perspective, they are notably less similar to Cas13s. Following our hybrid structure/sequence-based search for Cas13 homologs, we used state-of-the-art sequence-based phylogenetic analyses to trace the evolution of Cas13 from AbiF, the likely ancestral form containing a single HEPN nuclease and associating with diverse AbiFr ncRNAs. We structurally and biochemically characterized AbiF and showed that it is a toxin nuclease controlled by an ncRNA antitoxin. Additionally, we characterized a distinct subtype of Cas13, designated Cas13e, a compact protein containing two HEPN domains, which is a likely evolutionary intermediate between AbiF and large Cas13s containing REC/helical insertions. Lastly, we computationally explored the diversity of AbiF and AbiF-like HEPN-containing proteins.

The HEPN-mediated toxicity of AbiF and its inhibition by AbiFr ncRNA confirm that AbiF is a type III TA system. We demonstrated that, like other type III toxins, AbiF catalyzes the maturation of its antitoxin ncRNA and that the PbAbiFr ncRNA sterically masks the catalytic site of PbAbiF and inhibits the RNase. The PbAbiFr antitoxin ncRNA adopts a predominantly double-stranded structure with numerous RNA base triples (hydrogen bonding involving three bases) when it binds to the AbiF protein, offering a unique structural scaffold that potentially explains the specificity between AbiFs and their cognate antitoxins that we observed. As observed both in our solved cryo-EM model of PbAbiF and the AF3 model of the putative founder AbiF clade, the ncRNA is also rigidly oriented relative to the protein and contains a 5′ ssRNA extension that is precisely positioned in the HEPN active site, likely explaining the precise 5′ processing. However, the specific mechanism by which PbAbiF processes the 3′ end of the ncRNA remains unclear. It is possible that, like Cas13a,[47,66] AbiF may contain a secondary, potentially metal-independent, nuclease active site that helps processing the ncRNA, which could explain the cleavage effects that are sometimes observed when EDTA is present. Further investigation is needed to elucidate the precise molecular events facilitating PbAbiF's processing of the 3′ end of the antitoxin ncRNA.

Like other type III TA systems, PbAbiF forms oligomeric assemblies that interact with the AbiFr antitoxin ncRNA.[67] In the case of type III TA systems with resolved structures, such as ToxIN, each pseudoknotted ToxI RNA is bound in a head-to-tail orientation to two ToxN monomers, obstructing their active sites.[68] Despite observing a comparable inhibitory mechanism involving the occupation of active sites by the ncRNA antitoxin in PbAbiF, it is noteworthy that the ncRNA does not adopt a head-to-tail configuration but rather binds to a single monomer.

We further found that other systems, such as phage anti-defense modules, might co-opt or interfere with AbiF and/or AbiFr ncRNAs. For example, the sk1 and 712 RNA hairpins found in *Lactococcus lactis* phages might mimic the function of AbiF-ncRNA antitoxins to evade AbiF's RNase activity. This form of mimicry-based evasion appears in other phage anti-defense systems, such as Racrs, which are RNA-based CRISPR mimics used by phage to evade bacterial CRISPR systems.[28,69] AbiFs appear to have been directly incorporated into other biological pathways involved in defense and cellular signaling. For example, *abiF* genes can be found in bacteriophage exclusion (BREX) gene cassettes, embedded in type I restriction systems and even fused to Cas3 in some type I CRISPR systems, suggesting a potential role for AbiF in safeguarding or regulating defense systems (Figure S7). We additionally found *abiF*s associated with cyclic oligonucleotide-based anti-phage signaling system (CBASS)-likenucleotidyltransferases (NTases) and trypsin proteases where the antitoxin appears to have been lost, potentially replaced by the accompanying gene (Figure S7). Moreover, while AbiFs do not appear to be RNA-guided, the unique predicted placement of the ncRNA for F13a1 (Figure 7) suggests that F13a1 and/or F13a2 may function as unique RNA-guided systems not associated with CRISPR.

The diversity of the AbiF and other HEPN-containing nucleases underscores the complex backdrop of abortive infection-linked systems from which RNA-guided defense functions evolved. Unlike the case of the transposon-linked OMEGA-TnpB and OMEGA-IscB,[6,7] which already possessed the RNA-guided mechanism and likely evolved into type V and type II CRISPR-Cas systems on numerous occasions, the origin of Cas13 from AbiF appears to have been a unique or at best a rare event. Conceivably, this is the case because AbiF, although interacting with a ncRNA, does not employ it as a guide, making the transition to an RNA-guided CRISPR effector far more challenging. Through extensive sequence mining, deep learning-based structure prediction, structure/sequence comparisons, and sequence-based phylogenetic analysis, we reconstructed the likely path of evolution from AbiF—a small, non-specific toxin RNA nuclease—to Cas13—a large, elaborate, RNA-guided and RNA-targeting CRISPR effector. The Cas13 subtype that appears to be an evolutionary intermediate between AbiF and Cas13a/c/d, Cas13e, is the smallest known RNA-guided nuclease in the AbiF/Cas13 family. Taken together, the results of this work offer a detailed account of how a reprogrammable RNA-guided CRISPR mechanism likely evolved from a non-RNA-guided system, specifically a protein toxin-ncRNA antitoxin pair. More broadly, this study sheds light on the paths by which innate immune systems can evolve into adaptive ones in the everlasting arms race between hosts and pathogens.

Limitations of the study

Only protein families for which a confident AlphaFold2 prediction could be obtained were included in the structural UPGMA analysis; however, only a small minority did not pass (66/456, 14%), consisting mainly of incomplete proteins. Phylogenetic analyses rely on specific statistical assumptions and the quality of available data. As both genomic datasets and computational methods continue to improve, the accuracy and resolution of phylogenetic reconstructions may be further refined. Our RNP structure models for Cas13e, F13a1, and putative founder AbiF may have errors in the placement of the RNA, as they are predicted structures; however, the AlphaFold3 model is highly confident in these predictions given that the model is well calibrated to predict its own error on new unseen examples.[63] We further identified a large number of AbiF-related systems. We focused here on a limited set of AbiF-like systems, but owing

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

to the diversity of AbiF-related systems, there were many systems that we were unable to characterize.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Feng Zhang (zhang@broadinstitute.org).

Materials availability

Plasmids generated in this study are available from Addgene. All other reagents are available upon request.

Data and code availability

The cryo-EM map has been deposited in the Electron Microscopy Data Bank with accession code EMD: 43682. The coordinates of the atomic model have been deposited in the Protein Data Bank under accession code PDB: 8VZ6. All primers and genes used in this study are reported in Table S4. Analysis code and additional genomic loci used in this study are available through Zenodo.[70]

ACKNOWLEDGMENTS

S.Z.-T. is supported by The Gruss Lipper Postdoctoral Research Fellowship. H.A.-T. is supported by the Life Science Research Foundation. M.E.W. and K.K.E. are supported by the Helen Hay Whitney Foundation. F.Z. is supported by an NIH grant (2R01HG009761-05); the Howard Hughes Medical Institute; K. Lisa Yang and Hock E. Tan Molecular Therapeutics Center at MIT; Broad Institute Programmable Therapeutics Gift Donors; The Pershing Square Foundation, William Ackman, and Neri Oxman; James and Patricia Poitras; and BT Charitable Foundation. K.S. Makarova and E.V.K. are supported by the Intramural Research Program of the National Library of Medicine (National Institutes of Health).

AUTHOR CONTRIBUTIONS

S.Z.-T., H.A.-T., and F.Z. conceived the project. S.Z.-T., S.K., M.E.W., S.C.-D.-T.V., D.S., K.K.E., and K.S. Mears designed and performed the experiments with input from H.A.-T. and F.Z. H.A.-T. designed and performed computational analyses with feedback from S.Z.-T., S.K., C.-C.J.Y., K.S. Makarova, S.A.S., E.V.K., and F.Z. F.Z. supervised the research with support from R.K.M. S.Z.-T., H.A.-T., S.K., M.E.W., R.K.M., K.S. Makarova, E.V.K., and F.Z. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

F.Z. is a scientific advisor and cofounder of Editas Medicine, Beam Therapeutics, Pairwise Plants, Arbor Biotechnologies, Aera Therapeutics, and Moonwalk Biosciences. F.Z. is a scientific advisor for Octant.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Bacterial strains
- METHOD DETAILS
  - Plasmid construction
  - Purification of recombinant proteins
  - Small RNA sequencing
  - Electrophoretic mobility shift assay (EMSA)
  - Hill equation fit for Kd estimation
  - Liquid growth assay
  - Spotting assays
  - Viable count assay
  - Plasmid maintenance assay
  - *In vitro* AbiFr ncRNA processing
  - Cleavage of ssRNA, dsRNA, ssDNA, and dsDNA
  - AbiF activity assays
  - Cryo-EM
  - Cas13e targeted activity
  - Cas13e collateral activity assay
  - Endonuclease activity assay
  - Permissive sequence-based Cas13 homolog search
  - Profile-Profile UPGMA dendrogram
  - Structure prediction and structure-based HEPN dendogram / Structure-based Cas13 ancestor search
  - Phage-linked RNA association analysis
  - Structure predictions for structure comparison
  - Sequence-based phylogenetic analysis of Cas13/AbiF/F13a1/Fa13a2
  - Generation of PbAbiFr ncRNA profile
  - Generation of diverse AbiFr ncRNA profiles
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2025.01.034.

Received: February 12, 2024
Revised: December 12, 2024
Accepted: January 24, 2025
Published: February 18, 2025

REFERENCES

1. 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. https://doi.org/10.1126/science.aaq0179.
2. Kannan, S., Altae-Tran, H., Jin, X., Madigan, V.J., Oshiro, R., Makarova, K.S., Koonin, E.V., and Zhang, F. (2022). Compact RNA editors with small Cas13 proteins. Nat. Biotechnol. *40*, 194–197. https://doi.org/10.1038/s41587-021-01030-2.
3. 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. https://doi.org/10.1126/science.aaq0180.
4. 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.e3. https://doi.org/10.1016/j.molcel.2017.04.008.
5. Smargon, A.A., Cox, D.B.T., Pyzocha, N.K., Zheng, K., Slaymaker, I.M., Gootenberg, J.S., Abudayyeh, O.A., Essletzbichler, P., Shmakov, S., Makarova, K.S., et al. (2017). Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. Mol. Cell *65*, 618–630.e7. https://doi.org/10.1016/j.molcel.2016.12.023.
6. Altae-Tran, H., Kannan, S., Demircioglu, F.E., Oshiro, R., Nety, S.P., McKay, L.J., Dlakić, M., Inskeep, W.P., Makarova, K.S., Macrae, R.K., et al. (2021). The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. Science *374*, 57–65. https://doi.org/10.1126/science.abj6856.
7. Altae-Tran, H., Shmakov, S.A., Makarova, K.S., Wolf, Y.I., Kannan, S., Zhang, F., and Koonin, E.V. (2023). Diversity, evolution, and classification of the RNA-guided nucleases TnpB and Cas12. Proc. Natl. Acad. Sci. USA *120*, e2308224120. https://doi.org/10.1073/pnas.2308224120.
8. Kapitonov, V.V., Makarova, K.S., and Koonin, E.V. (2015). ISC, a Novel Group of Bacterial and Archaeal DNA Transposons That Encode Cas9

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

Homologs. J. Bacteriol. *198*, 797–807. https://doi.org/10.1128/JB.00783-15.

9. Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., et al. (2017). Diversity and evolution of class 2 CRISPR-Cas systems. Nat. Rev. Microbiol. *15*, 169–182. https://doi.org/10.1038/nrmicro.2016.184.
10. Zhang, B., Ye, Y., Ye, W., Perčulija, V., Jiang, H., Chen, Y., Li, Y., Chen, J., Lin, J., Wang, S., et al. (2019). Two HEPN domains dictate CRISPR RNA maturation and target cleavage in Cas13d. Nat. Commun. *10*, 2544. https://doi.org/10.1038/s41467-019-10507-3.
11. Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., et al. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. *18*, 67–83. https://doi.org/10.1038/s41579-019-0299-x.
12. Anantharaman, V., Makarova, K.S., Burroughs, A.M., Koonin, E.V., and Aravind, L. (2013). Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. Biol. Direct *8*, 15. https://doi.org/10.1186/1745-6150-8-15.
13. Koga, M., Otsuka, Y., Lemire, S., and Yonesaki, T. (2011). Escherichia coli rnlA and rnlB compose a novel toxin-antitoxin system. Genetics *187*, 123–130. https://doi.org/10.1534/genetics.110.121798.
14. Pillon, M.C., Gordon, J., Frazier, M.N., and Stanley, R.E. (2021). HEPN RNases – An Emerging Class of Functionally Distinct RNA Processing and Degradation Enzymes. Crit. Rev. Biochem. Mol. Biol. *56*, 88–108. https://doi.org/10.1080/10409238.2020.1856769.
15. Jurėnas, D., Fraikin, N., Goormaghtigh, F., and Van Melderen, L. (2022). Biology and evolution of bacterial toxin–antitoxin systems. Nat. Rev. Microbiol. *20*, 335–350. https://doi.org/10.1038/s41579-021-00661-1.
16. Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., et al. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol. Cell *60*, 385–397. https://doi.org/10.1016/j.molcel.2015.10.008.
17. Koonin, E.V., Makarova, K.S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. *37*, 67–78. https://doi.org/10.1016/j.mib.2017.05.008.
18. Koonin, E.V., and Makarova, K.S. (2019). Origins and evolution of CRISPR-Cas systems. Philos. Trans. R. Soc. Lond. B Biol. Sci. *374*, 20180087. https://doi.org/10.1098/rstb.2018.0087.
19. Holm, L. (2020). DALI and the persistence of protein shape. Protein Sci. *29*, 128–140. https://doi.org/10.1002/pro.3749.
20. Holm, L. (2022). Dali server: structural unification of protein families. Nucleic Acids Res. *50*, W210–W215. https://doi.org/10.1093/nar/gkac387.
21. Holm, L., Laiho, A., Törönen, P., and Salgado, M. (2023). DALI shines a light on remote homologs: one hundred discoveries. Protein Sci. *32*, e4519. https://doi.org/10.1002/pro.4519.
22. van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding, J., and Steinegger, M. (2024). Fast and accurate protein structure search with Foldseek. Nat. Biotechnol. *42*, 243–246. https://doi.org/10.1038/s41587-023-01773-0.
23. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
24. 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*, aaf5573. https://doi.org/10.1126/science.aaf5573.
25. Huang, J., Lin, Q., Fei, H., He, Z., Xu, H., Li, Y., Qu, K., Han, P., Gao, Q., Li, B., et al. (2023). Discovery of deaminase functions by structure-based protein clustering. Cell *186*, 3182–3195.e14. https://doi.org/10.1016/j.cell.2023.05.041.
26. Baek, M., McHugh, R., Anishchenko, I., Jiang, H., Baker, D., and DiMaio, F. (2024). Accurate prediction of protein-nucleic acid complexes using RoseTTAFoldNA. Nat. Methods *21*, 117–121. https://doi.org/10.1038/s41592-023-02086-5.
27. Shmakov, S.A., Utkina, I., Wolf, Y.I., Makarova, K.S., Severinov, K.V., and Koonin, E.V. (2020). CRISPR Arrays Away from Cas Genes. CRISPR J. *3*, 535–549. https://doi.org/10.1089/crispr.2020.0062.
28. Altae-Tran, H., Kannan, S., Suberski, A.J., Mears, K.S., Demircioglu, F.E., Moeller, L., Kocalar, S., Oshiro, R., Makarova, K.S., Macrae, R.K., et al. (2023). Uncovering the functional diversity of rare CRISPR-Cas systems with deep terascale clustering. Science *382*, eadi1910. https://doi.org/10.1126/science.adi1910.
29. Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G., and Baurain, D. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. *9*, e1000602. https://doi.org/10.1371/journal.pbio.1000602.
30. Grigoriev, I.V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R.A., et al. (2012). The genome portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res. *40*, D26–D32. https://doi.org/10.1093/nar/gkr947.
31. NCBI (2004). National Center for Biotechnology Information (Bethesda, MD). Available from: https://www.ncbi.nlm.nih.gov/genomes/
32. Richardson, L., Allen, B., Baldi, G., Beracochea, M., Bileschi, M.L., Burdett, T., Burgin, J., Caballero-Pérez, J., Cochrane, G., Colwell, L.J., et al. (2023). MGnify: the microbiome sequence data analysis resource in 2023. Nucleic Acids Res. *51*, D753–D759. https://doi.org/10.1093/nar/gkac1080.
33. Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2011). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat. Methods *9*, 173–175. https://doi.org/10.1038/nmeth.1818.
34. Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S.J., and Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics *20*, 473. https://doi.org/10.1186/s12859-019-3019-7.
35. Katoh, K., Misawa, K., Kuma, K.-I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066. https://doi.org/10.1093/nar/gkf436.
36. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.
37. Yan, W.X., Chong, S., Zhang, H., Makarova, K.S., Koonin, E.V., Cheng, D.R., and Scott, D.A. (2018). Cas13d is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein. Mol. Cell *70*, 327–339.e5. https://doi.org/10.1016/j.molcel.2018.02.028.
38. Altae-Tran, H., Gao, L., Strecker, J., Macrae, R.K., and Zhang, F. (2020). Computational Identification of Repeat-Containing Proteins and Systems. QRB Discov. *1*, e10. https://doi.org/10.1017/qrd.2020.14.
39. Shmakov, S.A., Makarova, K.S., Wolf, Y.I., Severinov, K.V., and Koonin, E.V. (2018). Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. Proc. Natl. Acad. Sci. USA *115*, E5307–E5316. https://doi.org/10.1073/pnas.1803440115.
40. Makarova, K.S., Wolf, Y.I., Snir, S., and Koonin, E.V. (2011). Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. J. Bacteriol. *193*, 6039–6056. https://doi.org/10.1128/JB.05535-11.

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

41. Kalvari, I., Nawrocki, E.P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., Griffiths-Jones, S., Toffano-Nioche, C., Gautheret, D., Weinberg, Z., et al. (2021). Rfam 14: expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Res. *49*, D192–D200. https://doi.org/10.1093/nar/gkaa1047.

42. Weinberg, Z., Lünse, C.E., Corbino, K.A., Ames, T.D., Nelson, J.W., Roth, A., Perkins, K.R., Sherlock, M.E., and Breaker, R.R. (2017). Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions. Nucleic Acids Res. *45*, 10811–10823. https://doi.org/10.1093/nar/gkx699.

43. Rivas, E., Clements, J., and Eddy, S.R. (2020). Estimating the power of sequence covariation for detecting conserved RNA structure. Bioinformatics *36*, 3072–3076. https://doi.org/10.1093/bioinformatics/btaa080.

44. Li, R.W. (2013). Metagenomics and Its Applications in Agriculture, Biomedicine and Environmental Studies (Nova Science Publishers).

45. 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.e10. https://doi.org/10.1016/j.cell.2017.06.050.

46. Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., Verdine, V., Cox, D.B.T., Kellner, M.J., Regev, A., et al. (2017). RNA targeting with CRISPR–Cas13. Nature *550*, 280–284. https://doi.org/10.1038/nature24049.

47. 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. https://doi.org/10.1038/nature19802.

48. McLandsborough, L.A., Sechaud, L., and McKay, L.L. (1998). Synergistic Effects of abiE or abiF from pNP40 When Cloned in Combination with abiD from pBF611. J. Dairy Sci. *81*, 362–368. https://doi.org/10.3168/jds.S0022-0302(98)75585-7.

49. Tangney, M., and Fitzgerald, G.F. (2002). Effectiveness of the lactococcal abortive infection systems AbiA, AbiE, AbiF and AbiG against P335 type phages. FEMS Microbiol. Lett. *210*, 67–72. https://doi.org/10.1111/j.1574-6968.2002.tb11161.x.

50. Rincé, A., Tangney, M., and Fitzgerald, G.F. (2000). Identification of a DNA region from lactococcal phage sk1 protecting phage 712 from the abortive infection mechanism AbiF. FEMS Microbiol. Lett. *182*, 185–191. https://doi.org/10.1111/j.1574-6968.2000.tb08894.x.

51. Wu, A.Y., Kamruzzaman, M., and Iredell, J.R. (2020). Specialised functions of two common plasmid mediated toxin-antitoxin systems, ccdAB and pemIK, in Enterobacteriaceae. PLoS One *15*, e0230652. https://doi.org/10.1371/journal.pone.0230652.

52. Díaz-Orejas, R., Espinosa, M., and Yeo, C.C. (2017). The Importance of the Expendable: Toxin–Antitoxin Genes in Plasmids and Chromosomes. Front. Microbiol. *8*, 1479. https://doi.org/10.3389/fmicb.2017.01479.

53. Kamruzzaman, M., Wu, A.Y., and Iredell, J.R. (2021). Biological Functions of Type II Toxin-Antitoxin Systems in Bacteria. Microorganisms *9*, 1276. https://doi.org/10.3390/microorganisms9061276.

54. Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S., and Salmond, G.P.C. (2009). The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair. Proc. Natl. Acad. Sci. USA *106*, 894–899. https://doi.org/10.1073/pnas.0808832106.

55. Samson, J.E., Spinelli, S., Cambillau, C., and Moineau, S. (2013). Structure and activity of AbiQ, a lactococcal endoribonuclease belonging to the type III toxin-antitoxin system. Mol. Microbiol. *87*, 756–768. https://doi.org/10.1111/mmi.12129.

56. Goeders, N., Chai, R., Chen, B., Day, A., and Salmond, G.P.C. (2016). Structure, evolution, and functions of bacterial Type III toxin-antitoxin systems. Toxins (Basel) *8*, 282. https://doi.org/10.3390/toxins8100282.

57. Songailiene, I., Juozapaitis, J., Tamulaitiene, G., Ruksenaite, A., Šulčius, S., Sasnauskas, G., Venclovas, Č., and Siksnys, V. (2020). HEPN-MNT Toxin-Antitoxin System: The HEPN Ribonuclease Is Neutralized by OligoAMPylation. Mol. Cell *80*, 955–970.e7. https://doi.org/10.1016/j.molcel.2020.11.034.

58. Garcia-Rodriguez, G., Charlier, D., Wilmaerts, D., Michiels, J., and Loris, R. (2021). Alternative dimerization is required for activity and inhibition of the HEPN ribonuclease RnlA. Nucleic Acids Res. *49*, 7164–7178. https://doi.org/10.1093/nar/gkab513.

59. Blower, T.R., Pei, X.Y., Short, F.L., Fineran, P.C., Humphreys, D.P., Luisi, B.F., and Salmond, G.P.C. (2011). A processed noncoding RNA regulates an altruistic bacterial antiviral system. Nat. Struct. Mol. Biol. *18*, 185–190. https://doi.org/10.1038/nsmb.1981.

60. Leontis, N.B., and Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. RNA *7*, 499–512. https://doi.org/10.1017/s1355838201002515.

61. Manikandan, P., Sandhya, S., Nadig, K., Paul, S., Srinivasan, N., Rothweiler, U., and Singh, M. (2022). Identification, functional characterization, assembly and structure of ToxIN type III toxin-antitoxin complex from E. coli. Nucleic Acids Res. *50*, 1687–1700. https://doi.org/10.1093/nar/gkab1264.

62. Rao, F., Short, F.L., Voss, J.E., Blower, T.R., Orme, A.L., Whittaker, T.E., Luisi, B.F., and Salmond, G.P.C. (2015). Co-evolution of quaternary organization and novel RNA tertiary interactions revealed in the crystal structure of a bacterial protein-RNA toxin-antitoxin system. Nucleic Acids Res. *43*, 9529–9540. https://doi.org/10.1093/nar/gkv868.

63. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w.

64. Nakagawa, R., Kannan, S., Altae-Tran, H., Takeda, S.N., Tomita, A., Hirano, H., Kusakizako, T., Nishizawa, T., Yamashita, K., Zhang, F., et al. (2022). Structure and engineering of the minimal type VI CRISPR-Cas13bt3. Mol. Cell *82*, 3178–3192.e5. https://doi.org/10.1016/j.molcel.2022.08.001.

65. Zhang, C., Konermann, S., Brideau, N.J., Lotfy, P., Wu, X., Novick, S.J., Strutzenberg, T., Griffin, P.R., Hsu, P.D., and Lyumkis, D. (2018). Structural Basis for the RNA-Guided Ribonuclease Activity of CRISPR-Cas13d. Cell *175*, 212–223.e17. https://doi.org/10.1016/j.cell.2018.09.001.

66. 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.e12. https://doi.org/10.1016/j.cell.2016.12.031.

67. Kang, S.-M., Kim, D.-H., Jin, C., and Lee, B.-J. (2018). A Systematic Overview of Type II and III Toxin-Antitoxin Systems with a Focus on Druggability. Toxins *10*, 515. https://doi.org/10.3390/toxins10120515.

68. Short, F.L., Pei, X.Y., Blower, T.R., Ong, S.-L., Fineran, P.C., Luisi, B.F., and Salmond, G.P.C. (2013). Selectivity and self-assembly in the control of a bacterial toxin by an antitoxic noncoding RNA pseudoknot. Proc. Natl. Acad. Sci. USA *110*, E241–E249. https://doi.org/10.1073/pnas.1216039110.

69. Camara-Wilpert, S., Mayo-Muñoz, D., Russel, J., Fagerlund, R.D., Madsen, J.S., Fineran, P.C., Sørensen, S.J., and Pinilla-Redondo, R. (2023). Bacteriophages suppress CRISPR-Cas immunity using RNA-based anti-CRISPRs. Nature *623*, 601–607. https://doi.org/10.1038/s41586-023-06612-5.

70. Code and Loci for "Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins" (Zenodo). https://doi.org/10.5281/zenodo.14600232.

71. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

72. Anaconda Inc. (2022). Distribution (Anaconda). https://www.anaconda.com/download.

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

73. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

74. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

75. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.

76. Illumina (2025). bcl2fastq and bcl2fastq2 Conversion Software. https://emea.support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html.

77. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

78. Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. *35*, 1026–1028. https://doi.org/10.1038/nbt.3988.

79. Rivas, E., Clements, J., and Eddy, S.R. (2017). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat. Methods *14*, 45–48. https://doi.org/10.1038/nmeth.4066.

80. Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*, W29–W37. https://doi.org/10.1093/nar/gkr367.

81. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421. https://doi.org/10.1186/1471-2105-10-421.

82. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589. https://doi.org/10.1038/nmeth.4285.

83. Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113. https://doi.org/10.1186/1471-2105-5-113.

84. Schrödinger, L.L.C. The PyMOL Molecular Graphics System (Version 2)

85. Fu, L., Cao, Y., Wu, J., Peng, Q., Nie, Q., and Xie, X. (2022). UFold: fast and accurate RNA secondary structure prediction with deep learning. Nucleic Acids Res. *50*, e14. https://doi.org/10.1093/nar/gkab1074.

86. Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*, 26. https://doi.org/10.1186/1748-7188-6-26.

87. Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics *29*, 2933–2935. https://doi.org/10.1093/bioinformatics/btt509.

88. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: the protein families database in 2021. Nucleic Acids Res. *49*, D412–D419. https://doi.org/10.1093/nar/gkaa913.

89. Kimanius, D., Dong, L., Sharov, G., Nakane, T., and Scheres, S.H.W. (2021). New tools for automated cryo-EM single-particle analysis in RELION-4.0. Biochem. J. *478*, 4169–4185. https://doi.org/10.1042/BCJ20210708.

90. Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A.J., and Berger, B. (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. Nat. Methods *16*, 1153–1160. https://doi.org/10.1038/s41592-019-0575-8.

91. Casañal, A., Lohkamp, B., and Emsley, P. (2020). Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. Protein Sci. *29*, 1069–1078. https://doi.org/10.1002/pro.3791.

92. Croll, T.I. (2018). ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. Acta Crystallogr. D Struct. Biol. *74*, 519–530. https://doi.org/10.1107/S2059798318002425.

93. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr. D Struct. Biol. *75*, 861–877. https://doi.org/10.1107/S2059798319011471.

94. Hyatt, D., Chen, G.-L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics *11*, 119. https://doi.org/10.1186/1471-2105-11-119.

95. Steinegger, M., and Söding, J. (2018). Clustering huge protein sequence sets in linear time. Nat. Commun. *9*, 2542. https://doi.org/10.1038/s41467-018-04964-5.

96. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

97. Shindyalov, I.N., and Bourne, P.E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng. *11*, 739–747. https://doi.org/10.1093/protein/11.9.739.

98. Katoh, K., and Toh, H. (2008). Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. BMC Bioinformatics *9*, 212. https://doi.org/10.1186/1471-2105-9-212.

99. Li, M., Gong, L., Cheng, F., Yu, H., Zhao, D., Wang, R., Wang, T., Zhang, S., Zhou, J., Shmakov, S.A., et al. (2021). Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems. Science *372*, eabe5601. https://doi.org/10.1126/science.abe5601.

100. Terashima, H., Koike, M., Kojima, S., and Homma, M. (2010). The flagellar basal body-associated protein FlgT is essential for a novel ring structure in the sodium-driven Vibrio motor. J. Bacteriol. *192*, 5609–5615. https://doi.org/10.1128/JB.00720-10.

101. Yamazaki, H., Kasai, S., Mimura, J., Ye, P., Inose-Maruyama, A., Tanji, K., Wakabayashi, K., Mizuno, S., Sugiyama, F., Takahashi, S., et al. (2020). Ribosome binding protein GCN1 regulates the cell cycle and cell proliferation and is essential for the embryonic development of mice. PLoS Genet. *16*, e1008693. https://doi.org/10.1371/journal.pgen.1008693.

102. Galperin, M.Y. (2023). All DACs in a Row: Domain Architectures of Bacterial and Archaeal Diadenylate Cyclases. J. Bacteriol. *205*, e0002323. https://doi.org/10.1128/jb.00023-23.

103. Gao, L.A., Wilkinson, M.E., Strecker, J., Makarova, K.S., Macrae, R.K., Koonin, E.V., and Zhang, F. (2022). Prokaryotic innate immunity through pattern recognition of conserved viral proteins. Science *377*, eabm4096. https://doi.org/10.1126/science.abm4096.

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E. coli Endura electrocompetent	Lucigen	60242
E. coli C43(DE3) electrocompetent cells	Biosearch Technologies	60345-1
Paenibacillus donghaensis	Belgian Coordinated Collections of Microorganisms	LMG 23780
Chemicals, peptides, and recombinant proteins		
glycoblue	Thermo Fisher Scientific	AM9515
cOmplete, EDTA-free	Roche	04693124001
Strep-Tactin Superflow Plus resin	Qiagen	30004
cOmplete ULTRA Tablets	Millipore Sigma	6538282001
RESOURCE S column	Cytiva	17-1180-01
Superdex 75 Increase 10/300 GL column	Cytiva	29148721
Ni Sepharose 6 Fast Flow	Cytiva	17531801
T4 Polynucleotide Kinase	New England Biolabs	M0201L
RNA 5′ polyphosphatase	Biosearch Technologies	RP8092H
pCp-Cy5	Jena Biosciences	NU-1706-CY5
T4 RNA ligase 1, ssRNA ligase (High Concentration)	New England Biolabs	M0437M
NEBuffer r3.1	New England Biolabs	B6003S
Proteinase K	New England Biolabs	P8107S
RNase A	New England Biolabs	T3018L
SUPERase In RNase inhibitor	Invitrogen	AM2694
LwaCas13a	New England Biolabs	Custom order
Critical commercial assays		
NucleoBond Xtra Midi kit	Machery-Nagel	740410.50
Direct-zol RNA miniprep plus	Zymo	R2073
Zymo RNA Clean & Concentrator-5	Zymo	R1013
NEBNext rRNA depletion kit (Bacteria)	New England Biolabs	E7850L
SMARTer smRNA-Seq Kit for Illumina	Takara Bio	635030
HiScribe T7 High Yield RNA Synthesis Kit	New England Biolabs	E2040S
RNaseAlert	Thermo Fisher Scientific	44-797-68
KAPA Library Quantification Kit for Illumina	Roche	KK4824
NEBNext Multiplex Small RNA Library Prep kit	New England Biolabs	E7330S
Deposited data		
PbAbiF RNP structure	This paper	PDB: 8VZ6, EMD: 43682
Oligonucleotides		
For all sequences, see Table S4	IDT, TWIST	N/A
Recombinant DNA		
For all sequences, see Table S4	IDT, TWIST	N/A
Software and algorithms		
ImageJ	Schneider et al.[71]	https://imagej.nih.gov/ij/
Anaconda	Anaconda Software Distribution. Anaconda Inc[72]	https://www.anaconda.com/download

(*Continued on next page*)

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BWA (Burrows-Wheeler Aligner)	Li and Durbin[73]	http://bio-bwa.sourceforge.net/
Samtools	Li et al.[74]	http://www.htslib.org/
Cutadapt	Martin[75]	https://cutadapt.readthedocs.io/
bcl2fastq	Illumina[76]	https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html.
Python v3.5	Python Software Foundation	https://www.python.org
SciPy	Virtanen et al.[77]	https://docs.scipy.org/
PySpark 2.4	Apache Foundation	https://spark.apache.org/docs/latest/api/python/index.html
DALI	Holm[19]	https://ekhidna2.biocenter.helsinki.fi/dali/
AlphaFold2	Jumper et al.[23]	https://deepmind.google/technologies/alphafold/
AlphaFold3	Abramson et al.[63]	https://alphafoldserver.com/
RFNA	Baek et al.[26]	https://github.com/uw-ipd/RoseTTAFold2NA
MMSeqs2	Steinegger and Söding[78]	https://github.com/soedinglab/MMseqs2
HH-suite	Steinegger et al.[34]	https://github.com/soedinglab/hh-suite
R-scape	Rivas et al.[43,79]	http://eddylab.org/R-scape/
FLSHclust	Altae-Tran et al[28]	Code in corresponding paper
BPROM	Li[44]	https://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb
HMMER	Finn et al.[80]	http://hmmer.org/
BLAST	Camacho et al.[81]	https://blast.ncbi.nlm.nih.gov/doc/blast-help/downloadblastdata.html
IQ-Tree2+ModelFinder	Kalyaanamoorthy et al.[36,82]	https://www.iqtree.org/
MAFFT	Katoh et al.[35]	https://mafft.cbrc.jp/alignment/server/index.html
MUSCLE	Edgar[83]	https://www.drive5.com/muscle/
Pymol	Schrödinger[84]	https://pymol.org/support.html
UFold	Fu et al.[85]	https://github.com/uci-cbcl/UFold
RNAfold	Lorenz et al.[86]	https://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi
Infernal	Nawrocki and Eddy[87]	http://eddylab.org/infernal/
Rfam	Kalvari et al.[41]	https://rfam.org/
PFAM	Mistry et al.[88]	https://pfam-docs.readthedocs.io/en/latest/summary.html
NCBI	NCBI[31]	https://www.ncbi.nlm.nih.gov/
JGI	Grigoriev et al.[30]	https://jgi.doe.gov/
EMBL	Richardson et al.[32]	https://www.ebi.ac.uk/
PDB	PDB	https://www.rcsb.org/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Bacterial strains

For protein purification, *E. coli* C43(DE3) from a single colony was grown in Terrific Broth media (TB) at 37°C until OD600 reached 0.6, and further grown at 16°C with IPTG supplementation for 16 hours for expression. For growth assays, *E. coli* C43(DE3) from a single colony was grown in TB at 37°C overnight and then plated on LB-agar plates supplemented with appropriate antibiotics and IPTG, glucose, or arabinose and induced at 37°C for 16 hours. For CFU assays *E. coli* C43(DE3) from a single colony were grown in TB at 37°C overnight, IPTG was then added, and culture samples were collected at different time points following induction. Culture for *P.donghaensis* followed the conditions suggested by the Belgian Coordinated Collections of Microorganisms (BCCM).

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

CellPress
OPEN ACCESS

METHOD DETAILS

Plasmid construction

For protein purification of Cas13e, a codon-optimized sequence of *cas13* (contig ID 37790, Joint Genome Institute (JGI)) was cloned into a modified pET45(+) backbone with an N-terminal 14xHIS-Twin-Strep-bdSUMO tag. The ncRNA was cloned into a separate pCOLADuet-1 vector downstream of the J23119 promoter. For the 3N degenerate library, the target sequence was ordered as an ssDNA oligo from IDT and amplified by PCR using NEBNext High-Fidelity 2X Master Mix (NEB) and the fragment was gel purified. Simultaneously, 4 ug of pUC19 vector was digested with SspI and PdmI (Thermo Fisher Scientific) and the resulting vector backbone was gel purified. The vector and insert were assembled using Gibson assembly with 2X Gibson Assembly Master Mix (NEB) in an 80 uL reaction with 800 ng of backbone vector and an 8x molar quantity of insert at 50 C for 1 hour. The assembled plasmid was purified by isopropanol precipitation by adding 80 uL of isopropanol, 1.6 uL of 5 M NaCl (final reaction concentration 50 mM) and 1 uL of glycoblue co-precipitant (Thermo Fisher Scientific) and incubating at room temperature for 15 min. The precipitated DNA was then centrifuged for 15 min at max speed at 4 C, then washed twice with freshly prepared ice cold 70% ethanol, air-dried at room temperature for 1 min and resuspended in 12 uL TE buffer at 55 C for 10 min. 4 uL of the resuspension was electroporated in 50 uL of Endura electrocompetent *E. coli* (Lucigen) as per the manufacturer's instructions and recovered in 2 mL recovery media at 37 C for 1 hour. The cells were then plated across 2 22.7 cm x 22.7 cm plates with the appropriate antibiotic resistance and grown for 12-16 h at 37 C. The resulting colonies were then scraped off the plate and midiprepped using a NucleoBond Xtra Midi kit (Machery-Nagel). For quality control, next-generation sequencing libraries were prepared from the resulting plasmids and sequenced on an Illumina MiSeq. For growth assays, CFU assays, and protein purification of AbiF RNPs the native locus of AbiF from its natural host was cloned into a modified pET45(+) backbone with C-terminal Strep-tag II using Gibson assembly (NEB). For apo AbiF, the RNA upstream of the AbiF coding sequence was removed using reverse PCR (NEB). For dead AbiF plasmids, point mutations in the HEPN catalytic motif (R210A, H215A for PbAbiF and R213A, H218A for EcAbiF) were introduced using PCR. For preinduction of RNA spotting assay, apo AbiF was cloned into a modified pCOLADuet backbone with N-terminal HISX14 Twin-Strep and bdSUMO tag using Gibson assembly (NEB). The AbiFr ncRNA was cloned into pBAD backbone using Gibson assembly.

Purification of recombinant proteins

AbiF RNP complex: Vectors were transformed into *E. coli* C43(DE3) electrocompetent cells (Biosearch Technologies). Cells were grown at 37°C in TB supplemented with the appropriate antibiotics until reaching OD_{600}=0.6, then shifted to 18°C and induced overnight at 18°C with 1 mM IPTG (Gold Biotechnology). Cells were then harvested by centrifugation for 10 minutes at 4°C at 5,000 rpm (Beckman Coulter Avanti J-E, rotor JLA-8.1000). The cell paste was resuspended in lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl, 5% glycerol, 5 mM $MgCl_2$, and 5 mM 2-mercaptoethanol) supplemented with protease inhibitors (Roche cOmplete, EDTA-free), and lysed by two passes with a high-pressure homogenizer (LM20 Microfluidizer, Microfluidics). The lysate was cleared by centrifugation, and the soluble fraction was gently mixed with Strep-Tactin Superflow Plus resin (Qiagen) at 4°C. The resin was first washed with lysis buffer, then briefly with buffer A (10 mM HEPES pH 7.9, 250 mM potassium acetate, 2.5 mM magnesium acetate, 5 mM 2-mercaptoethanol) on a gravity flow column. Bound RNP was then eluted in buffer A supplemented with 5 mM desthiobiotin (Sigma).

Apo AbiF: The RNP vectors were transformed into *E. coli* C43(DE3) electrocompetent cells (Biosearch Technologies). Cells were cultured at 37 °C until reaching OD_{600}=1.8, and protein expression was then induced by the addition of 1 mM IPTG (Gold Biotechnology). Cells were then allowed to grow further at 21°C overnight. All subsequent steps were carried out at 4°C. Cells were harvested by centrifugation for 10 minutes at 4°C at 5,000 rpm (Beckman Coulter Avanti J-E, rotor JLA-8.1000). The cell paste was resuspended in 5 volumes of lysis buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% [vol/vol] glycerol, 5 mM 2-Mercaptoethanol) containing cOmplete ULTRA Tablets (Millipore Sigma 6538282001) and disrupted using an LM20 microfluidizer (Microfluidics). Cell debris was removed by centrifugation for 1 hour at 4°C at 9,500 rpm (Beckman Coulter Avanti J-E, rotor JLA-10.500). Nucleic acids were precipitated by 1% [vol/vol] polyethylenimine (PEI), followed by centrifuging at 9,500 rpm at 4°C for 30 minutes. The pellet was washed with buffer A (20 mM Tris-HCl, pH 8.0, 10% [vol/vol] glycerol, 5 mM 2-Mercaptoethanol) followed by centrifuging at 9,500 rpm at 4°C for 30 minutes. The pellet was discarded and the proteins in the supernatant were loaded onto a RESOURCE S column (Cytiva) pre-equilibrated with buffer A. The column was washed with 10 column volumes of buffer A and then eluted with a linear gradient of 10 column volumes of buffer B (50 mM Tris-HCl, pH 8.0, 1 M NaCl, 10% [vol/vol] glycerol, 5 mM 2-Mercaptoethanol). Fractions containing apo PbAbiF were pooled, concentrated, and then loaded on a Superdex 75 Increase 10/300 GL column (Cytiva) pre-equilibrated in buffer C (10 mM HEPES pH 7.5, 250 mM Potassium acetate, 2.5 mM Magnesium acetate, 20% [vol/vol] glycerol, 5 mM 2-Mercaptoethanol). The purified proteins were concentrated, aliquoted, frozen in liquid nitrogen, and stored at -80°C.

The Cas13e RNP complex and apo Cas13e were purified similarly to the AbiF RNP complex with the following modifications: lysis buffer contained 50 mM Tris pH 8, 200 mM NaCl, 5% glycerol, 5 mM MgCl2, and 5 mM 2-mercaptoethanol. The soluble fraction was gently mixed with Ni Sepharose 6 Fast Flow (Cytiva) at 4°C. The resin was first washed with lysis buffer, then briefly with buffer A (50 mM Tris pH 8, 1 M NaCl, 5% glycerol, 5 mM MgCl2, and 5 mM 2-mercaptoethanol), and buffer B (50 mM Tris pH 8, 500 mM NaCl, 5% glycerol, 5 mM MgCl2, and 5 mM 2-mercaptoethanol) on a gravity flow column. Bound RNP and apo protein were cleaved using bdSENP1 protease.

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

Small RNA sequencing

For Cas13e RNPs, 200 μL of RNPs were mixed with 600 μL of TRI reagent (Zymo) and incubated for 5 minutes at room temperature. 120 μL chloroform (Sigma Aldrich) was then added, the sample was inverted gently to mix, and then incubated at room temperature for 3 minutes, followed by spinning at 12000xg at 4°C for 15 minutes. The aqueous phase was used as input for RNA extraction using a Direct-zol RNA miniprep plus kit including in-column DNase I treatment (Zymo). The purified RNA was then subject to treatment as per the manufacturer's instructions with 20 U of T4 PNK (NEB) for 6 hours at 37°C, then 20 U of RNA 5′ polyphosphatase (Biosearch Technologies) for 30 minutes at 37°C, with each enzymatic step followed by cleanup with a Zymo RNA Clean & Concentrator-5 kit as per the manufacturer's instructions. Following enzymatic treatments, purified RNA was subject to library preparation with a SMARTer smRNA-Seq Kit for Illumina (Takara) as per the manufacturer's instructions. The library was sequenced on an Illumina MiSeq with Read 1 80 cycles, Index 1 8 cycles, Index 2 8 cycles. Adapters were trimmed using CutAdapt[75] and mapped to loci of interest using BWA.

For AbiF RNPs, the above protocol was followed with the following modifications: 1 hour incubation with PNK and following enzymatic treatments, purified RNA was subject to library preparation with an NEBNext Multiplex Small RNA Library Prep kit (NEB) as per the manufacturer's instructions, with an extension time of 1 minute and 12 cycles in the final PCR. Amplified libraries were gel extracted, quantified by qPCR using a KAPA Library Quantification Kit for Illumina (Roche) on a CFX Opus 384 (BioRad), and sequenced on an Illumina MiSeq with Read 1 45 cycles, Read 2 45 cycles and Index 1 6 cycles. Adapters were trimmed using CutAdapt[75] and mapped to loci of interest using BWA. Filled reads were obtained, and filled reads filtered as indicated were visualized using a custom Python script.[6] For in vitro AbiFr ncRNA processing the reactions were treated and analyzed as mentioned above with the following modification: the reads were filtered based on reads length (40-60 nt).

For native expression, *P. donghaensis* was obtained from BCCM as a freeze-dried culture, resuspended in the appropriate media and cultured in Medium 14 at 25°C for 3 days. 5-mL cultures were spun down, resuspended in 750 μL TRI reagent (Zymo) and incubated for 5 minutes at room temperature. 0.5 mm zirconia/silica beads (BioSpec Products) were added, and the culture was vortexed for approximately 1 minute to mechanically lyse cells. 200 μL chloroform (Sigma Aldrich) was then added, and the above protocol was followed with the following modifications: prior to T4 PNK treatment, ribosomal RNA was removed using an NEBNext rRNA depletion kit (Bacteria) (NEB).

Electrophoretic mobility shift assay (EMSA)

Following small RNA sequencing, the AbiFr ncRNA sequence was determined. The AbiFr ncRNA sequence and random RNA of the same length were purchased from IDT. 50 pmol of the RNA was heated for 3 minutes at 65°C, then labeled with 100 pmol of pCp-Cy5 (Jena Biosciences) using 68 U T4 RNA ligase 1, ssRNA ligase (High Concentration) (NEB) in a 1X buffer of 50 mM Tris-HCl pH 7.5, 10 mM MgCl2, 2 mM DTT, 2 mM ATP, and 10% DMSO at 4°C overnight and purified using a Zymo RNA Clean & Concentrator-5 column as per the manufacturer's instructions. 5 nM of labeled RNA was incubated with increasing concentrations of PbAbiF protein (0, 0.5, 5, 25, 50, 150, 250, 500, 750, 1000, 1500, 2000, 2500, 3500 and 5000 nM) in NEBuffer r3.1 supplemented with 5% glycerol, 1μg/mL heparin and 0.01% triton. Samples were incubated at 37 °C for 1 hour. 5XHi-Density buffer was added to the samples and RNA was resolved by gel electrophoresis on Novex 10% TBE polyacrylamide gels (Thermo Fisher Scientific). The gels were pre-runed at a constant voltage (50) for 30 minutes in 0.5% TBE at 4°C. Following sample loading, the gels were run at a constant voltage (100) for 110 minutes in 0.5% TBE at 4°C. The gels were imaged using a BioRad ChemiDoc imaging system, and the area under the curve for the bound and unbound peaks was quantified using ImageJ. The fraction of bound RNA was calculated using the following formula: $fraction\ bound\ RNA = \frac{bound\ RNA}{bound\ RNA+unbound\ RNA}$

Hill equation fit for Kd estimation

EMSA data was quantified and converted to RNA bound fractions ranging from 0 to 1, as described above. A Hill equation was fitted using scipy's curve_fit function fitting both the Kd and n (Hill coefficient).[77] R-squared values were computed using the standard R-square formula with the prediction's residuals. Error estimates for the fitted parameters were calculated using the standard covariance formulation assuming normally distributed error estimate (1.96*sqrt(cov[param,param])).

Liquid growth assay

Vectors were transformed into BL21(DE3). A single colony was grown in TB at 37°C overnight. All samples were diluted to a final OD_{600}= 0.01 and plated in a clear bottom 96-well Clear Flat Bottom Plate (Corning). OD_{600} was measured using Biotek Synergy Neo 2 (Agilent). 1 mM IPTG was added to all samples at T=240 minutes.

Spotting assays

Vectors were transformed into BL21(DE3). A single colony was grown in TB at 37°C overnight. Six serial dilutions of the overnight culture were prepared in TB. 5 μL of each dilution were induced on LB plates with the appropriate antibiotics and inducer and then incubated for 16 hours at 37°C. Plates were imaged with a BioRad ChemiDoc imaging system.

For the pre-induction of RNA, the assay was carried out as similar to the above with the following modifications: the overnight culture was incubated with or without 2 mM arabinose.

Viable count assay

Vectors were transformed into BL21(DE3). A single colony was grown in TB at 37°C overnight. The next day, 1 mM IPTG was added to the media. At various time points, small aliquots were sampled from each condition. The samples were serial diluted and plated on LB agar plates with IPTG and incubated for 16 hours at 37°C. Following incubation, the number of colonies on each plate was counted.

Plasmid maintenance assay

Vectors were transformed into BL21(DE3). A single colony was grown in TB with 100 mg/mL Ampicillin and 0.1 mM IPTG at 37°C overnight. The overnight culture was then diluted into fresh TB without antibiotics and with 0.1 mM IPTG and allowed to grow overnight. The process of diluting the overnight culture in fresh media with IPTG was repeated for 10 days. Each day before diluting the culture, small aliquots were taken from each sample for serial dilutions. 5 μL of each dilution was plated on an LB agar plate supplemented with 100 mg/mL Ampicillin and incubated for 16 hours at 37°C.

In vitro AbiFr ncRNA processing

RNA was either purchased from IDT or produced by annealing oligo DNA to T7P primer and *in vitro* transcription (HiScribe T7 High Yield RNA Synthesis Kit, NEB). The RNA was then purified using Zymo RNA Clean & Concentrator-25 per the manufacturer's instructions. The RNA was then 3'-end labeled as described above. The ncRNA processing reaction contained 1 μM of RNA and 1 μM of AbiF in NEBuffer 3.1. The reaction was allowed to proceed at 37°C for 1 hour. Reactions were then treated with 1.6 U of Proteinase K (NEB) at room temperature for 15 minutes. RNA was resolved by gel electrophoresis on Novex 10% TBE-Urea polyacrylamide gels (Thermo Fisher Scientific) as specified and imaged using a BioRad ChemiDoc imaging system. For RNA sequencing, the RNA was extracted and sequenced as described above with the following modifications: 30 μL reaction was mixed with 90 μL of TRI reagent (Zymo) and 18 μL chloroform (Sigma Aldrich).

Cleavage of ssRNA, dsRNA, ssDNA, and dsDNA

ssRNA and its complement oligos as well as ssDNA and its complement oligos were all ordered from IDT. For the formation of dsRNA and dsDNA, the complementary oligos were mixed in a 1:1 ratio in the presence of 10 mM Tris pH 8, 50 mM NaCl and 10 mM EDTA and heated to 95°C for 2 minutes with a gradual decrease of temperature to room temperature. For the cleavage reaction, 1 μM of the oligos was incubated with 1 μM of AbiF or 20 μg of RNase A (NEB) for ssrRNA and dsRNA or 20 U of DNase I (RNase-free) (NEB) for ssDNA and dsDNA. All the reactions were carried out in NEBuffer r3.1 (NEB) and incubated at 37°C for 1 hour. The DNA and RNA were resolved by gel electrophoresis on 4% E-Gels (Invitrogen) and imaged using a BioRad ChemiDoc imaging system.

AbiF activity assays

All the activity assay reactions were carried out with approximately 1 μM of apo AbiF or AbiF RNPs. Oligonucleotides were added in a 1:1 or 1:10 ratio (AbiF:RNA) respectively. The assay reactions contained 1X NEBuffer r3.1, 20U of SUPERase In RNase inhibitor (Invitrogen) and 1 μL of RNaseAlert sensor resuspended in 10 μL ultra-pure water. The reactions were plated on a Corning Low Volume 384-well Black/Clear Flat Bottom Plate and fluorescence measurements over time were taken using a Biotek Synergy Neo 2 (Agilent).

For the buffer screening assay, the reactions were carried out as described above with the following buffers: NEBuffer 1, NEBuffer 2, NEBuffer r3.1, and NEBuffer 4 (NEB).

Activity assays with NTPs, rNTPs, NDPs, AMPs, and cAMPs were carried out as described above with a 10:1 ratio of nucleotides:PbAbiF.

Cryo-EM

For cryo-EM grid preparation, a freshly glow-discharged (60 s at 25 mA) Cu300 R1.2/1.3 holey carbon grid (Quantifoil) was mounted in the chamber of a Vitrobot Mark IV (Thermo Fisher Scientific) maintained at 12°C and 100% humidity. 4 μL of AbiF RNP complex was applied and immediately blotted using Ø55 grade 595 filter paper (Ted Pella) and plunged into liquid ethane. Cryo-EM data were collected using the Thermo Scientific Titan Krios G3i cryo-TEM at MIT.nano using a K3 direct detector (Gatan) operated in super-resolution mode with 2-fold binning, and an energy filter with slit width of 20 eV. Micrographs were collected automatically using EPU in AFIS mode, yielding 4,272 movies for PbAbiF fractionated into 40 frames and a flux of 18.5 e^-/pix/s giving a total fluence per micrograph of 42.2 $e^-/Å^2$. Cryo-EM data were processed using RELION.[89] Movies were corrected for motion using the RELION implementation of MotionCor2, with 4x4 patches and dose-weighting. CTF parameters were estimated using CTFFIND-4.1. For PbAbiF, particle picking was initially performed using the Topaz general model,[90] yielding 1,137,288 particles. A subset of 332,794 particles selected after 2D and 3D classifications was then used to train a Topaz picking model. From this round of picking, 1,142,600 particles were extracted, from which 727,709 particles were selected by 2D classification then 142,581 particles by 3D classification. These initially refined to 3.91 Å resolution. Per-particle defocus refinement and particle polishing improved the reconstruction to 3.62 Å resolution. 3D classification without alignment (T=8) yielded a 12,790 particle subset that was refined to produce a map at 3.56 Å resolution with more clearly defined features. C2 symmetry was applied throughout data processing. For EcAbiF, 1,651,017 particles were picked using the Topaz general model, which were filtered to 976,664 particles by 2D classification and 335,336 particles by 3D classification. Per-particle defocus refinement and particle polishing yielded a reconstruction at 3.27 A resolution. D2 symmetry was applied throughout data processing for the EcAbiF RNP. Resolution is reported using the gold-standard Fourier Shell

Correlation with 0.143 cutoff. For model building, AlphaFold2 models were docked into the cryo-EM density and then adjusted and rebuilt using Coot.[91] RNAs were built de novo using coot. The models were refined first using ISOLDE[92] then with PHENIX real_space_refine,[93] just performing one macro-cycle and with nonbonded_weight=2000. Figures were generated using UCSF ChimeraX.

Cas13e targeted activity

RNA was produced by PCR amplification and *in vitro* transcription (HiScribe T7 High Yield RNA Synthesis Kit, NEB). The target and non-target RNA were amplified from the 3N degenerate library plasmid, which was constructed as mentioned above. The collateral RNA was then 3'-end labeled as described above. All the activity assay reactions were carried out with approximately 1 μM of Cas13e RNP and target or non-target RNA were added in a 1:0.5 ratio (RNP:RNA). Collateral RNA was added in a 1:0.5 ratio (RNP:RNA). The assay reactions contained 1X NEBuffer r3.1, 20 U of SUPERase In RNase inhibitor (Invitrogen) The reaction was allowed to proceed at 37°C for 4 hr. RNA was resolved by gel electrophoresis on Novex 10% TBE-Urea polyacrylamide gels (Thermo Fisher Scientific) as specified and imaged using a BioRad ChemiDoc imaging system.

Cas13e collateral activity assay

RNA was produced by PCR amplification and *in vitro* transcription (HiScribe T7 High Yield RNA Synthesis Kit, NEB). The target and non-target RNA were amplified from the 3N degenerate library plasmid, which was constructed as mentioned above. All the activity assay reactions were carried out with approximately 1 μM of apo Cas13e and 45nM of apo LwaCas13a (Custom order from New England Biolabs). crRNA was added at a 1:1 ratio (protein:RNA) and target or non-target RNA were added in a 1:0.5 ratio (protein:RNA). The assay reactions contained 1X NEBuffer r3.1, 20 U of SUPERase In RNase inhibitor (Invitrogen) and 1 μL of RNaseAlert sensor resuspended in 10 μL ultra pure water. The reactions were plated on a Corning Low volume 384-well Black/Clear Flat Bottom plate, and fluorescence measurements over time were taken using Biotek Synergy Neo 2 (Agilent), at 37°C.

Endonuclease activity assay

RNA was purchased from IDT with 3 constitutive phosphorothioate bonds at the 5' or the 3' end. For the time course assay the RNA was modified with 3 constitutive phosphorothioate bonds at the 5' and 3' end The reactions were carried out with 1:1 ratio of PbAbiF:RNA and contained 1X NEBuffer r3.1, 20 U of SUPERase In RNase inhibitor (Invitrogen). The reaction was allowed to proceed at 37°C for 1 hour and for the time course assay, small aliqutes were taken at 0, 2, 5, 10, 20, 40 and 60 minutes. Reactions were then treated with 1.6 U of Proteinase K (NEB) at room temperature for 15 minutes. RNA was resolved by gel electrophoresis on Novex 10% TBE-Urea polyacrylamide gels (Thermo Fisher Scientific) as specified and imaged using a BioRad ChemiDoc imaging system.

Permissive sequence-based Cas13 homolog search

A collection of non-redundant Cas13 alignments generated from a previous Cas13 search,[2] as well as non-redundant c13c1 sequences (from an HMM search from a seed sequence[27]) aligned with MAFFT were curated. The alignments were then split into two halves, one half covering the HEPN 1 from each protein, the other covering the HEPN 2. These sets of alignments were then converted to HH-suite HMM profiles. Protein IDs for this analysis have the format GENOMEID_CONTIGID_START_END_STRAND where STRAND is -1 or 1, and START/END are the genomic coordinates where START is 0 indexed while END is the position after the last included base pair. Profiles were then searched using HHBlits against a database of clustered protein sequences described previously (clustered at 30% sequence identity with each cluster aligned using MUSCLE[83]) ignoring all clusters that either contains a single element or does not contain any protein $\geq$ 200 aa to reduce database contamination by partial/incomplete proteins.

HHblits scores and pairwise profile-profile alignments were then parsed. Positions of each HEPN catalytic site (R and H in R-X(4,6)-H) were then mapped onto the profile positions. For each pairwise HMM-HMM alignment, the positions of catalytic residues in the query alignment were then propagated to the positions in the target alignment. Any target alignment that did not have an R and H at the corresponding positions was discarded for lacking the HEPN active site. 2486 clusters passed this stage.

All representative sequences from the 30% id clusters passing the previous filter were then collected into a new database. Then, all representative sequences from the 30% id clusters were searched against this database using jackhammer with 8 iterations, inclusion expect threshold of 1e-10, expect threshold of 1e-10, and domain expect threshold of 1e-10, and subsequently aligned using MAFFT automatic mode without discarding/sampling any sequences. These alignments were then converted into HH-suite HMM profiles in an HHsuite database. Each profile was then searched against all other profiles in this database using HHblits with 1 iteration and a probability cutoff of 0.01 with 5000 max alignments. These pairwise alignments between 30% cluster ids were used to generate distance matrices for the profile-profile UPGMA HHsuite distance tree described below.

The major HH-suite HMM profiles of each Cas13 protein half were then used to search against HHM profiles of representative HEPN domains from PFAM using hhalign, with the resulting scores shown as a heatmap.

Profile-Profile UPGMA dendrogram

A matrix of similarity scores S was constructed using the profile-profile alignment scores from HHblits described above. A default score of 10 was assigned if no pairwise alignment was found, which is equivalent to an approximate noise floor. The score matrix S was symmetrized by averaging S with its transpose. A distance-like matrix, D, was then created from the score matrix as follows. For every valid i,j integer pair ($0 \leq i <$ S.shape[0], $0 \leq j <$ S.shape[0]), then D[i,j] = max(0, log ((S[i,j]+1e-8) / s_min[i,j]) / 2), where

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

s_min[i,j] = min(S[i,i],S[j,j]). This is essentially a normalized log transform of the similarity scores. While this matrix is not a true distance matrix due to the triangle inequality not necessarily being satisfied, methods like neighbor-joining trees would not work. In this case, the UPGMA method is used to infer a tree from the distance-like matrix D. The UPGMA method was performed using scipy's linkage module with "average" linkage. This tree's nodes were then clustered into 457 HEPN-containing superclusters (gene families) using scipy's fcluster with a distance threshold of 0.5. This dendrogram was only used to create superclusters.

Structure prediction and structure-based HEPN dendogram / Structure-based Cas13 ancestor search

For each passing supercluster, the longest 30% sequence identity cluster representative was then selected as the representative for the entire supercluster. Each representative sequence was used to generate a multiple sequence alignment using MMseqs2 from the colab server, and then a 3D protein structure for the monomer form was predicted using AlphaFold2.[23] Only representative sequences with an average pLDDT of $\geq$70 for the predicted structure were considered to avoid making comparisons with poor quality structures, which meant some superclusters would not be considered further. Structures were then aligned to each other using DALI,[19] storing both Z-scores and pairwise structural alignments. The Z scores were symmetrized by taking the average of the Z-score matrix with its transpose. For any supercluster, if the mean of the top 5 pairwise Z-scores (including the self Z-score) was below 5, we excluded the supercluster from analysis due to extreme divergence from all other HEPN-containing proteins. We repeated this until convergence. Additional superclusters were removed if manual inspection of their pairwise structural alignments revealed that it did not in fact contain a HEPN domain with an alignable active site. Dali scores were then converted into a distance-like matrix by taking D[i,j] = 1 / max(Z[i,j], Z[j,i]) where Z[i,j] is the DALI score between supercluster i and supercluster j. The distance-like matrix was then used to infer a UPGMA structure tree of HEPN-domain containing proteins with scipy's linkage module with "average" linkage.

An additional structure search was performed by searching solved Cas13 structures (in addition to the Cas13c and Cas13e models) against the set of the structures above along with the set of following structures: HEPN-containing structures from the PDB, and best of 10 (by pLDDT) AlphaFold2 models of representative HEPN-containing families without solved structures.[12] DALI was used to perform the search, and scores were visualized via heatmap.

Phage-linked RNA association analysis

1,124,303 prokaryotic genomes and draft genomes were obtained from NCBI and NCBI WGS. For genomes without gene predictions, genes were predicted with prodigal, resulting in a total of 4,594,609,314 proteins.[94] All predicted proteins were clustered with a cascaded clustering procedure using Linclust (100%, 98%, 95%, 90%, 70%, and finally 50% with 80% sequence coverage).[95] Due to the maximum dataset size limitation of Linclust/MMseqs2, the initial protein dataset was sharded into 2 randomly generated fragments and individually clustered (at 100% sequence identity). These shards were rejoined when clustering at 90%, when the combined database size was sufficiently small. Cluster ids were then mapped onto all proteins within each DNA contig. For each DNA contig, intergenic regions were determined as follows: A coverage vector P was created, which contains a 1 for any position where a CDS $\geq$80aa exists on the contig and 0 otherwise. All consecutive runs of 0 in P were then calculated as a list of tuples (start, end). Tuples within 1500 bp of the contig edge were removed. Tuples covering fewer than 80 bp were removed. Remaining tuples were then expanded 200 bp on either side, and the corresponding sequences in the original contig were denoted intergenic sequences. All intergenic sequences were searched versus Rfam using cmsearch with an evalue threshold of 1e-3 and an inclusion evalue threshold of 1e-4. All Rfam hits were then remerged with the database so that each contig had information about Rfam hits.

Due to the nature of horizontally transferred genes having separate evolutionary rates compared to core host genes, redundancy reduction was performed using a genome level redundancy reduction as follows. For each genome, an RpoB (a typically single copy, signature gene) amino acid sequence was extracted using the high scoring HMMER gene from all genes in that genome. RpoB sequences were then clustered at 95% sequence identity using MMSeqs2. Each gene in the genome was then assigned to the RpoB sequence from its genome and its corresponding cluster. Importantly, redundancy was not reduced by simply picking representative genomes, as many genomes have diverse horizontally transferred genes.

For the RNA association analysis, all protein/Rfam pairs within 10 kb in any contig were collected, resulting in 5,065,653,542 pairs. All protein/Rfam pairs were then grouped according to the following key: (rj_acc, pi_c50_id, pj_c90_id, rpob_c95_id), where rj_acc is the Rfam accession, pi_c50_id is the 50% sequence identity protein cluster id, pi_c90_id is the 90% protein cluster id, and rpob_c95_id is the corresponding RpoB cluster id. The protein with the longest distance between its corresponding CDS and the contig edge was selected as the representative for the group, with tiebreakers formed by the protein id name. The remaining pairs were then grouped by the key (pi_c50_id, rj_acc). Importantly, this analysis required multidimensional redundancy reduction, both via the RpoB gene clusters to reduce species redundancy, and via the 90% protein cluster to reduce sequence redundancy. To achieve this, within each group, each entry was considered a row, and the bipartite matching algorithm was employed to obtain the maximum number of pj_c90_id values that each match to a distinct rpob_c95_id value. Rows not corresponding to matches were discarded, resulting in 'non-redundant' positive pairs.

For each 50% identity cluster, the non-redundant protein count was calculated as the number of 90% sequence identity clusters within that cluster. 50% clusters with < 5 non-redundant proteins were discarded. A normalizer list was created for each cluster, which is a list of 90% non-redundant clusters with "default" selections for which 90% cluster representatives to include. This list

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

is generated analogously to the non-redundant positive pairs, except the grouping key is just (pi_c50_id) (without any reference to any rj_acc).

Then for each (pi_c50_id, rj_acc) key, the list of non-redundant positive pairs is combined with the normalizer list. For each pi_c90_id in the non-redundant positive pairs, the corresponding entry in the normalizer list is replaced with that of the non-redundant positive pair, resulting in a 'replaced normalizer list' that additionally now contains a token for each row describing if that protein is associated with rj_acc. This procedure allows different (pi_c50_id, rj_acc) keys to use different representative sequences for the pi_c90_id values without causing normalization errors. Next, for each non-redundant protein in the replaced normalizer list, a basal probability p is computed as max(min*(L/icity_width, 1.0)* a_min, a_min_min*]) where L is the distance to the closest contig edge, and where *a_min=0.25*, and *a_min_min=a_min*0.1*. These basal probabilities were then placed in an ordered list P. Then, using the list of probabilities P, the number of positive events (number of non-redundant entries in the replaced normalizer list that are associated with rj_acc), a PoissonBinomial test was performed, resulting in a P-value calculation which can be used to filter out spurious associations. Next, the Benjamini-Hochberg procedure was performed with an alpha_fdr of 0.05, resulting in a p_critical value of 0.00076. Using this critical threshold resulted in 175,510 significant (pi_c50_id, rj_acc) key pairs. All passing pairs had the corresponding protein component predicted for functional domains using hmmsearch[80] with the Pfam[88] database using an evalue threshold of 1e-3 and domain evalue threshold of 1e-3.

A database of bacteriophage sequences was downloaded from NCBI and converted into a BLAST+ database.

We next estimated phage interactions for each statistically significant (pi_c50_id, rj_acc) key pair using the following strategy. The non-redundant positive pairs belonging to the key pair were obtained (up to 10000 items), and the corresponding contigs from the original RNA genes were retrieved. Each of the Rfam hits were then expanded 50 bp upstream and downstream to form an expanded Rfam profile sequence. These sequences were then aligned with MAFFT automatic mode. Next, we identified conserved regions in the alignment as follows. The alignment was converted into an M length list (M is the number of columns in the alignment) called P, where P is the sequence identity (between 0 and 1) at each position P, not-including gaps in the calculation. A list G of length M was also created which contains the fraction of gap characters at each position. Positions in both G and P where $G < 0.25$ were discarded. Then a moving average was calculated over P with a window size of 15 (using 'same' padding at the edges) producing the list *Pma*. Positions in the original sequence corresponding to positions in the processed *Pma* where *Pma*>0.6 were considered conserved. All expanded Rfam sequences were then searched against phage using BLASTN[81] with an e-value threshold of 1e-3. For each hit, if the fraction of the query sequence overlapping with the conserved positions was $\geq$ 0.4, the hit was discarded as being too similar to the conserved regions of the RNA. Thus, all hits remaining are from variable regions of the Rfam expanded sequence that are not simply homologs of the core Rfam RNA in bacteriophage. Lastly, the strand of the match was calculated as follows. For the target hit, the mean value of T was calculated where $T[i] = 1$ if i corresponds to a (+) strand CDS, $T[i] = -1$ if i corresponds to a (–) strand CDS, $T[i] = 0$ if i does not correspond to a CDS in the phage. If mean(T) < 0, the strand preference was considered to be 1, and if mean(T) > 0, the strand preference was considered to be -1. P-values were computed as follows. For phage match, the number of phage matches *N_p* and the total number of non-redundant loci in the replaced normalizer list, N were calculated. Then the probability of this sample was calculated using a one-sided binomial test (with number of positive events being *N_p* and number of total draws being N) with a null hypothesis that the true match probability is 1e-3 and an alternative hypothesis that the true match probability is greater than 1e-3 (same as the e-value threshold of the BLAST search). A second test for match preference was also computed. In this case, the number of matches to a positive CDS strand was calculated *N_plus*, and similarly for the number of matches to a negative CDS strand, *N_minus*. Then, the match strand preference test was calculated using a two-sided binomial test (number of positive events being *N_plus* and number of total draws being *N_plus* + *N_minus*) with the null hypothesis being 0.5 chance of observing a positive strand and 0.5 chance of observing a negative strand in the match. All results were filtered using a p-value threshold of 1e-12 for the phage match test and a p-value threshold of 1e-3 for the match strand preference test.

Structure predictions for structure comparison

In the case of Swt1-HEPN, we used a 1:1 dimeric structure prediction with AlphaFold2 via ColabFold.[96] For Cas13e, we first created an MSA of all Cas13e direct repeats (DRs) that had an alignable DR, assuming the DR was expressed in the same orientation as the Cas13e. We then aligned all Cas13e's from these loci using MAFFT-einsi and created a paired multiple sequence alignment. We then created a 'pseudo' spacer region by taking an arbitrary sequence of length 14, padding it with two 'A' bases on each side and placing it immediately downstream from the Cas13 DR in the multiple sequence alignment. Spacers across all other orthologs are randomized. The Cas13e RNP structure was then predicted for each non-redundant (at 70% sequence identity) Cas13e ortholog using the aforementioned multiple sequence alignment generation procedure, and the one with the best i_PAE was used. Cas13c and F13a2 were used from the previous structure prediction that was performed to do structural comparisons. For visualization in Figure 6, some structures were generated using the AlphaFold3 online server, as RNP prediction became available with the AF3 model.

Sequence-based phylogenetic analysis of Cas13/AbiF/F13a1/Fa13a2

All 70% non-redundant sequences were included for analysis for Cas13a, Cas13b/b-t, Cas13c/c-t, Cas13d, Cas13e, AbiF, and F13a1/2, whereas for AbiF, only representative sequences for the 30% clusters were included for the analysis, due to the high degree of sequence diversity AbiF covers. To compare AbiF, which contains one HEPN domain, with the two HEPN domain-containing proteins, we duplicated the AbiF HEPN domain to model the gene duplication event that likely gave rise to the two HEPN-domain

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

systems. This assumption is supported by a single HEPN tree of AbiF and F13a1/2's HEPN-1 and HEPN-2, which show that the two HEPN domains of F13a1/2's likely originate from related AbiFs (Figure S3).

Due to the high sequence divergence between various Cas13s, it is not possible to use a standard alignment method to achieve a suitable alignment for phylogenetic analysis. We therefore performed an alignment consisting of only extracted HEPN domain helices as follows. Structures and structure predictions (from Figure 6) were used to annotate the HEPN domain using pymol's cealign structural alignment followed by comparison against reference HEPN domains.[97] The HEPN helix 1-4 were annotated for each structure and structure prediction using these alignments. For each class (Cas13a, Cas13b/b-t, Cas13c/c-t, Cas13d, Cas13e, AbiF, and F13a1/2), sequence alignments were performed within the class using MAFFT including the structure/predicted structure reference. The HEPN helices from the structure references were then mapped to the alignment columns corresponding to HEPN helix 1-4 (for both HEPN domains if applicable). For each class, the alignment regions surrounding HEPN helix 1-4 were then extracted, allowing for an extension of up to 30aa to allow for errors in annotation. This resulted in HEPN1 helix 1, 2, 3, 4 and HEPN 2 helix 1, 2, 3, 4 segments for all classes, except for AbiF, which only has HEPN helix 1, 2, 3, and 4. For AbiF, its HEPN helix 1, 2, 3, 4 segments was duplicated to form 8 segments: HEPN 1 helix 1, 2, 3, 4 (the original HEPN segments) and HEPN 2 helix 1, 2, 3, 4 (the duplicated segments). For each segment (e.g., HEPN 2 helix 3) all protein classes' segment corresponding to this segment (e.g., Cas13b HEPN 2 helix 3) were grouped and realigned using MAFFT. Non-conserved N- and C-terminal regions that were not present for all classes were trimmed from the segment's alignment. Finally, all segments were concatenated into a single alignment consisting of HEPN 1 and HEPN 2. Gap positions were then trimmed by removing alignment columns with 50% gap or more in python, and alignment blocks that did not have homology across all protein classes were removed.

A substitution model for the phylogenetic analysis was selected using ModelFinder from IQ-Tree2,[82] resulting in the Q.pfam+I+G4 substitution model being the best using the Bayesian Information Criterion. Phylogenetic tree reconstruction was then performed using IQ-Tree2[36] with the following options: "–safe" flag to avoid underflow, "–ninit 1000 –ntop 100 –nbest 50 -nstop 500" tree search parameters to search hard likelihood landscapes, and 5000 ultrafast bootstraps. Five independent tree inferences were performed, and the tree with the best likelihood was selected as the final tree for subsequent evolutionary analyses. The root of this tree was selected as the divergent group of AbiF proteins with smaller protein lengths.

Protein sequence lengths were determined based on the original input sequences without any filtering or processing. CRISPR association was determined using the combination of 6 CRISPR finder tools as described previously,[28] and CRISPRs within 3 kb of the coding sequence were considered associated. For AbiF, CRISPR associations were aggregated by cluster, calculating the fraction of CRISPR-associated 70% non-redundant loci in the cluster. For AbiFr ncRNA associations, 35 sets of profiles were considered: the Rfam AbiFr ncRNA profile (which we refer to as AbiFr to avoid ambiguity), the PbAbiF seeded custom profile (AbiF5_iter3.CM), the set of 29 ncRNA branched profiles from custom searching (diverse_rna_xinsi.CM), F13a1 ncRNA (F13a1.CM), and Cas13e DR (Cas13e_DR_rc.CM), each using an e-value threshold of 1e-4 when compared against loci using cmsearch. For cmsearch based searches, the region between S+100 and E-100 was deleted, where S,E correspond to the coding sequence start and end sites. A maximum distance threshold of 500 bp was used for all cmsearch-based searches except for the Cas13e DR, for which the search window was 3000bp.

Generation of PbAbiFr ncRNA profile

The conserved ncRNA region upstream from PbAbiF was extracted and used to create a covariance model and consensus secondary structure as follows: starting from the initial seed sequence, its secondary structure was predicted using UFold[85] (with lone pairs removed), which was then used as an initial secondary structure for creating a covariance model with cmbuild.[87] The covariance model was calibrated and then searched against all 70% non-redundant loci from the Cas13 ancestor search using an expect value of 1e-1 and an inclusion expect value of 1e-1. The resulting matching sequences were then retrieved from the original loci and extended by 30 bp upstream and downstream. The sequences were then aligned using the secondary structure-aware MAFFT-xinsi.[98] The resulting alignment was then assigned a consensus secondary structure using cacofold from R-scape with the –cacofold flag and an expect value of 1e-1. The resulting alignment and consensus secondary structure prediction was then used to create a covariance model with cmbuild (AbiF5_iter1.CM). The covariance model was calibrated and then searched against all 70% non-redundant loci from the Cas13 ancestor search using an expect value of 1e-1 and an inclusion expect value of 1e-1. The resulting matching sequences were then retrieved from the original loci and extended by 30 bp upstream and downstream. The sequences were then aligned using MAFFT-linsi due to the large number of sequences. The resulting alignment was then assigned a consensus secondary structure using cacofold from R-scape using the –cacofold flag, an expect value of 1e-3, and the –Rfam flag to remove extraneous pair predictions. The resulting alignment and consensus secondary structure prediction was then used to create a covariance model with cmbuild (AbiF5_iter2.CM). The process was repeated again to arrive at the third iteration resulting in a covariance model (AbiF5_iter3.CM). This alignment, covariance model, and predicted consensus secondary structure via cacofold was used as the PbAbiF covariance model and consensus secondary structure.

Generation of diverse AbiFr ncRNA profiles

From the HEPN1+2 sequence tree, AbiF branches with moderate branch support were selected, forming 30 mutually exclusive groups that covered most of the tree. For each clade, all CDS from the 70% non-redundant proteins from that clade were collected and their upstream regions were curated (the region between 250 bp upstream from the start site to 50 bp downstream from the start

Please cite this article in press as: Zilberzwige-Tal et al., Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.034

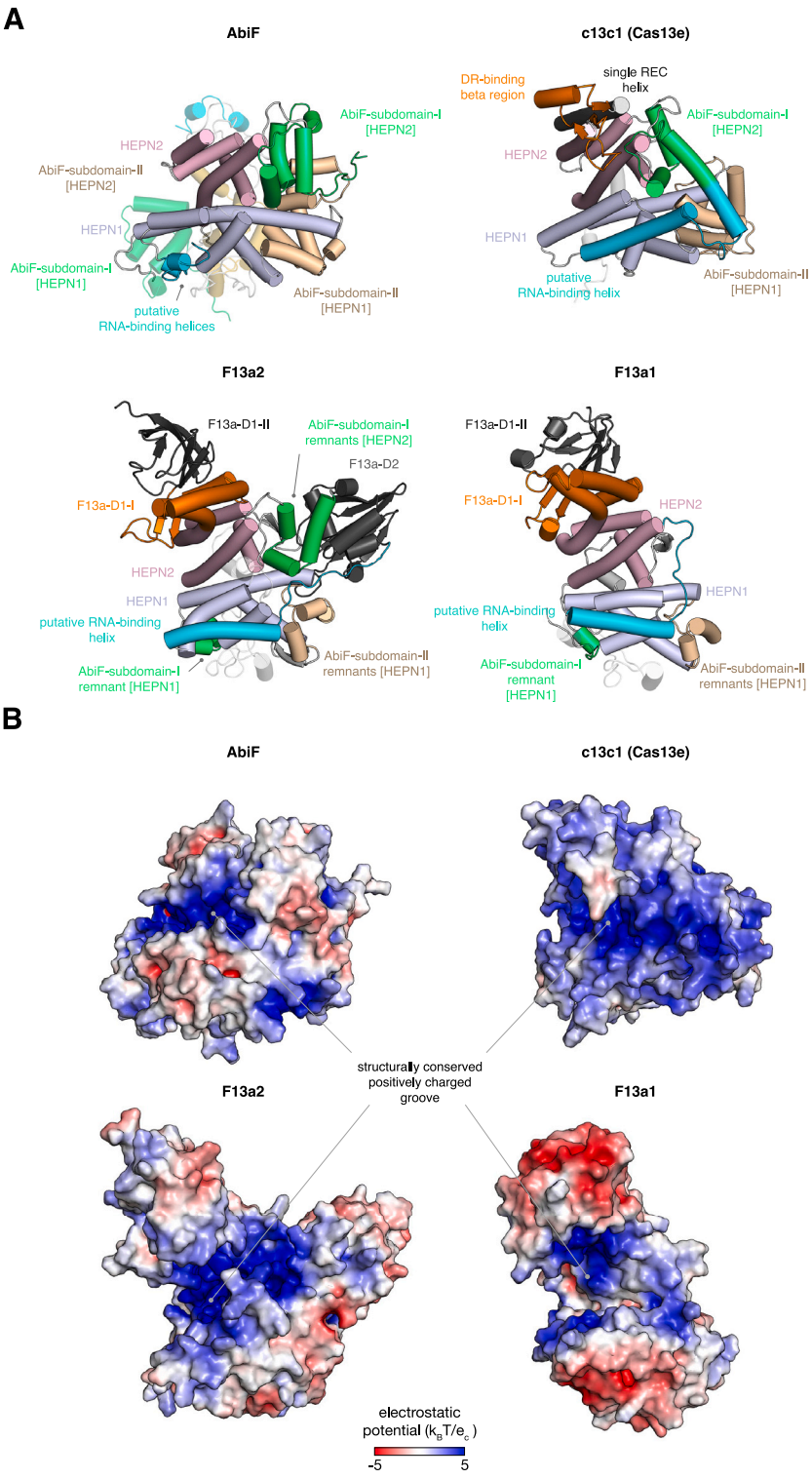
site), and similarly their downstream regions (the region between 50 bp upstream from the stop site to 250 bp downstream from the stop site). These regions were then aligned via MAFFT-einsi. All alignments were then manually inspected for signs of conservation outside the coding regions, and alignments without substantial conservation outside coding regions were discarded. The passing alignments were then trimmed to their conserved regions, excluding the coding sequence regions, and then realigned with MAFFT-ginsi. The resulting new alignments were then inspected for conserved blocks. For alignments without strong conservation following MAFFT-ginsi, a tree was built, and major branches were extracted to identify if a subpopulation of the non-coding region had strong conservation. If subpopulations with conservation were identified, they were split into a separate alignment group for downstream processing. All resulting alignments were then aligned using MAFFT-xinsi, producing secondary structure aware alignments. Sequences with poor alignment to the consensus sequence of the alignment were removed to reduce noise from unrelated sequences. Sequences were then further trimmed to remove long non-conserved 5' and 3' ends. Consensus secondary structures were then identified using R-scape with the –cacofold flag, an expect threshold of 1e-1, and the –Rfam flag. The consensus structure predicted by R-scape with cacofold was then used along with the sequence alignment to create a covariance model using cmbuild and cmcalibrate. All covariance models were then concatenated into a single file diverse_rna_xinsi.CM.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis was performed using GraphPad Prism 9.5 (GraphPad Software) by student t test. p values of less than 0.05 were considered significant.

CellPress
OPEN ACCESS

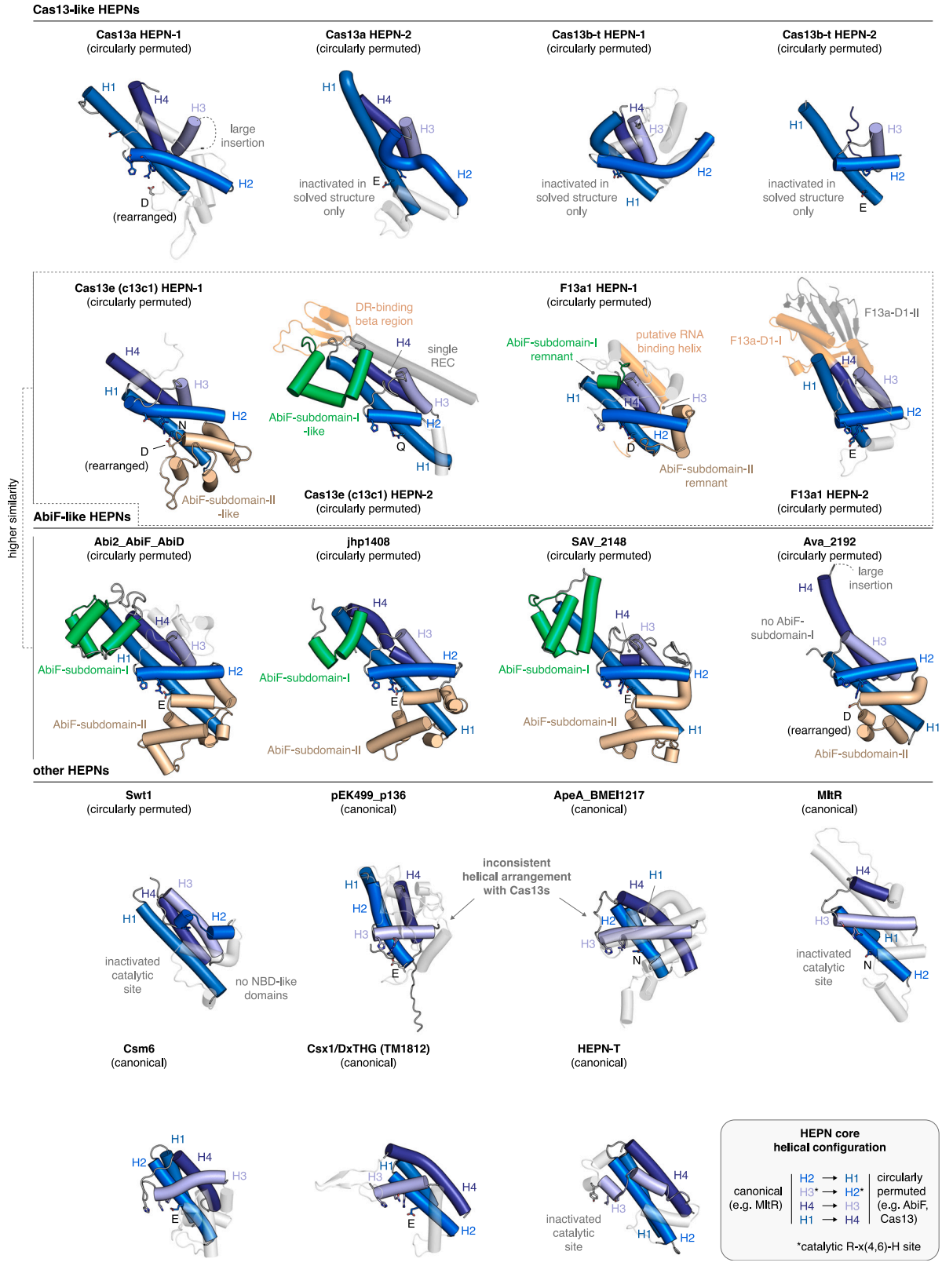
Supplemental figures



(legend on next page)

Figure S1. Comparisons of AlphaFold predictions of the Cas13/AbiF HEPN group, related to Figure 1

(A) Cartoon diagrams of the protein components of AbiF (dimer, as predicted by AF3), c13c1 (as predicted by AF3, not showing ncRNA), F13a2 (as predicted by AF2, due to not having known associated ncRNAs), and F13a1 (as predicted by AF3, not showing ncRNAs). Some regions are shown as translucent for clarity. (B) APBS electrostatic surface calculation of the protein surface of proteins from (A). Units shown are electrostatic potential ($k_B\,T\,e_c^{-1}$ where k_B is the Boltzmann constant, T is temperature in Kelvin, and e_c is the unit of electron charge).



(legend on next page)

Figure S2. Comparison of HEPN domain structures, related to Figures 1 and 6

Comparison of predicted and solved HEPN domain structures of select Cas13-like HEPN domains and other HEPN domains most related to Cas13, as determined by structural and sequence similarity. HEPN domain names are shown in bold above each structure. HEPN configuration type (either canonical or circularly permuted) is shown below each HEPN domain label. Helices are labeled H1–H4 (corresponding to helix 1 through helix 4) according to the order the helices appear in the corresponding protein from the N terminus to the C terminus and are color coded in shades of blue/purple. Active site residues are shown. Conserved glutamic acid (or corresponding rearrangements/replacements) are also shown with a 1 letter amino acid label. Bottom right: schematic of the mapping between the canonical helix ordering and the circularly permuted helix ordering found in AbiF-like HEPNs, Swt1, F13a1, and Cas13s.

Solved structures used in this figure: Cas13a (PDB: 5XWP), Cas13b-t (PDB: 7VTI), Swt1 (PDB: 4PQZ), MltR (PDB: 3BRJ), HEPN-T (PDB:3O10), Csm6 (8JBB), and Csx1 (6R9R). Other structures shown are AlphaFold2 predictions, except for Cas13e (c13c1) and F13a1, which are from AF3 RNP predictions with their corresponding RNAs.

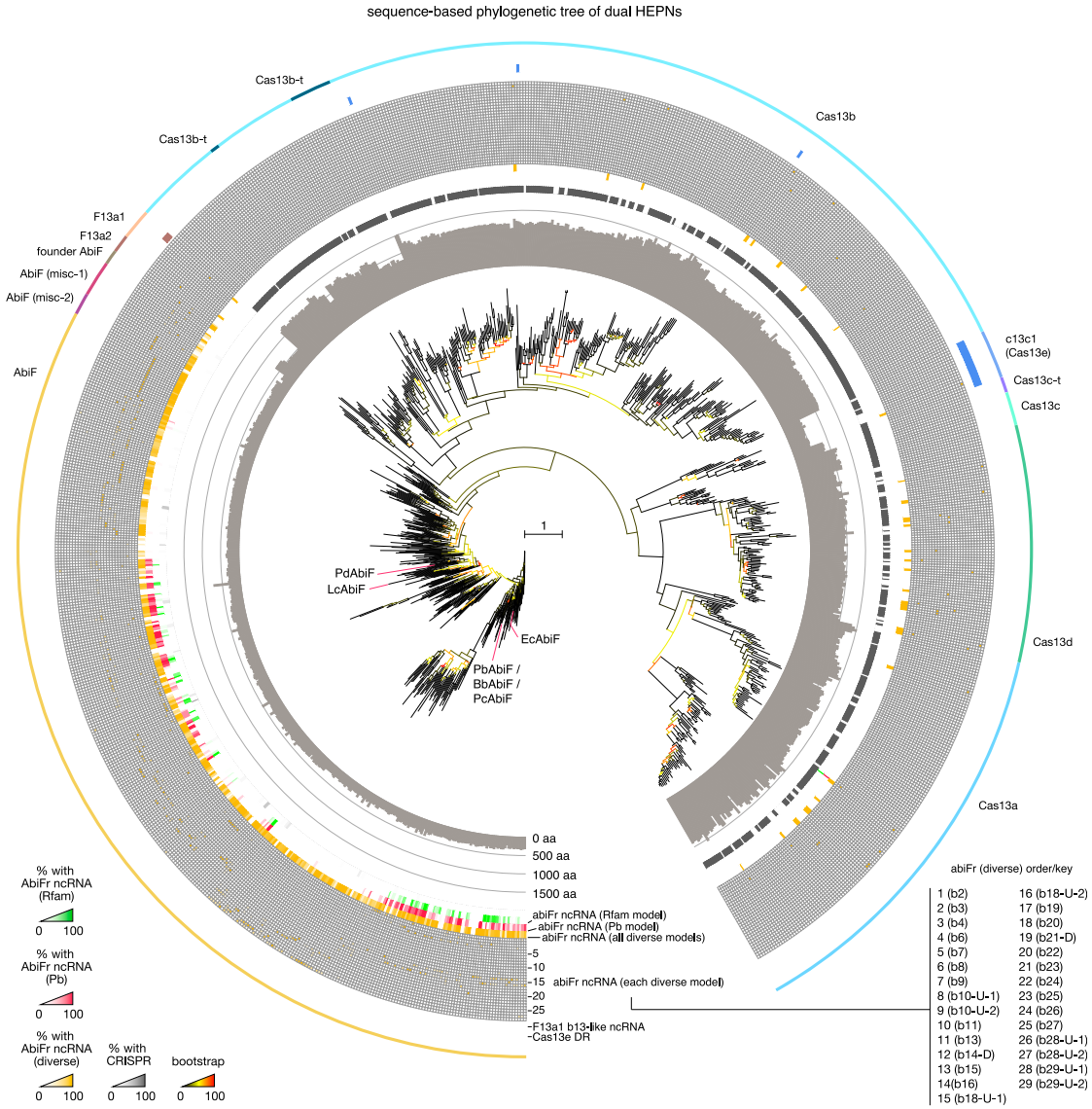


Figure S3. Extended phylogenetic analyses of Cas13-like HEPN domains, related to Figures 1, 2, 5, and 7

IQ-Tree2 (best of 5) phylogenetic tree of dual HEPNs with Cas13a-d, c13c1 (Cas13e), F13a1, F13a2, and AbiF with an expanded panel of abiFr ncRNA models. The tree was made using the Q.pfam+I+G4 substitution model as identified by ModelFinder. ncRNA presence was assessed using cmsearch with an E-value threshold of 1e−4 for the following model sets: (1) the Rfam abiF motif RNA model (abiFr Rfam), (2) the custom PbAbiF-like ncRNA model (abiFr Pb), (3) the union of the diverse ncRNA profiles generated in this study (abiFr all diverse models), (4) individual ncRNA models 1,...29 generated in this study with their corresponding names (abiFr each diverse model), (5) F13a1 b13-like ncRNA model, and finally (6) c13c1 (Cas13e) DR model. Additionally shown are CRISPR associations, and protein lengths. For AbiF, F13a1 ncRNA, c13c1 (Cas13e) DR, and CRISPR associations, associations are tabulated as a fraction of the cluster. Displayed protein lengths are average values across each cluster. For Cas13, and F13a1/2, individual loci (singleton clusters) are used, whereas for AbiF proteins, 30% sequence identity clusters are used. For all abiFr ncRNAs, vicinity is defined to be within 500 bp beyond the protein start and end site and not more than 100 bp inside the protein-coding sequence. For the c13c1 (Cas13e) DR model, vicinity is defined to be within 3,000 bp beyond the protein start and end site and not more than 100 bp inside the protein-coding sequence. For CRISPR, vicinity is defined to be within 3,000 bp of the protein-coding sequence.

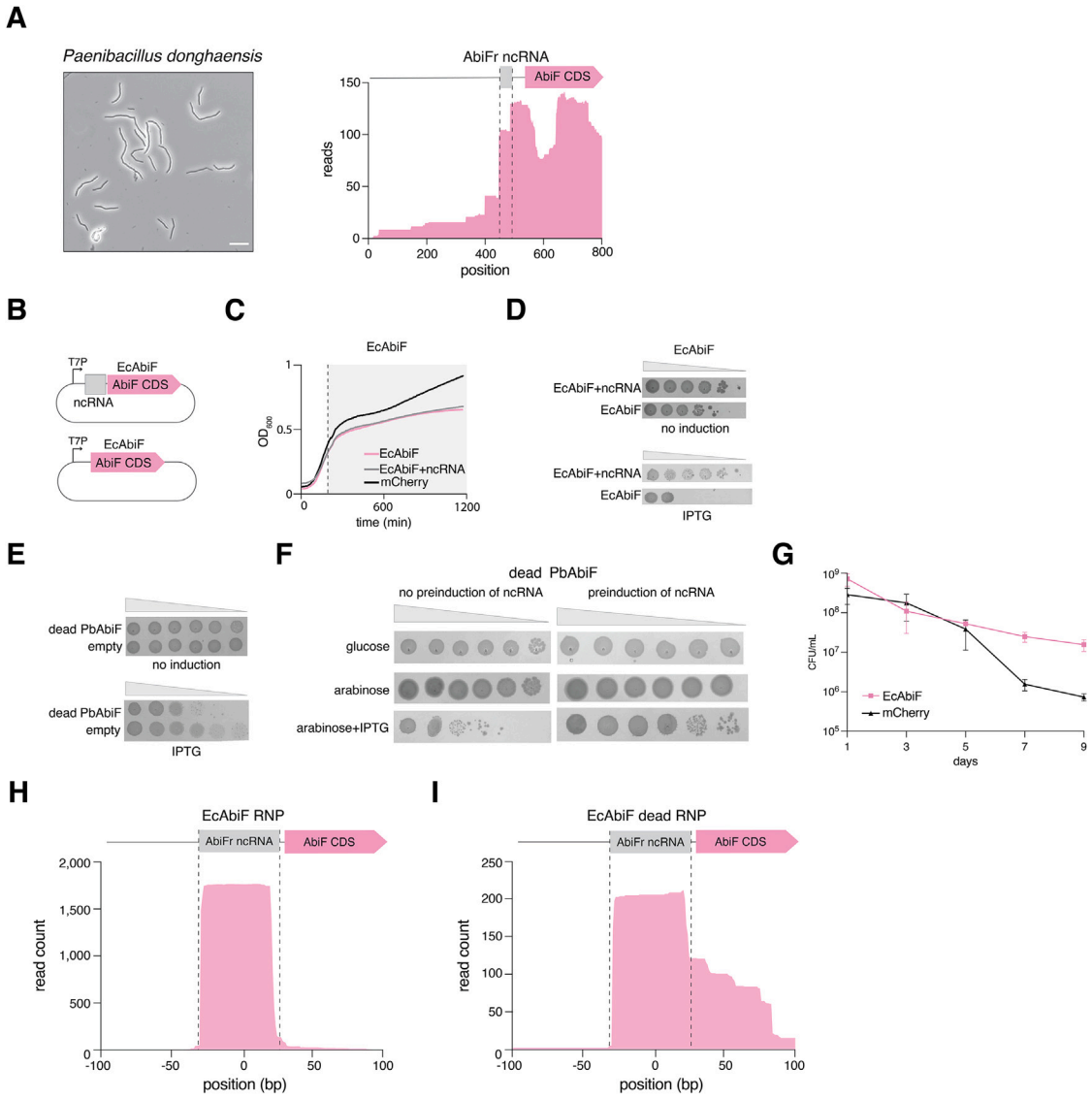


Figure S4. Extended sRNA-seq data and growth assays of PdAbiF, EcAbiF, and PbAbiF, related to Figures 4 and 5

(A) Left: brightfield image of *P. donghaensis*, scale bar is 10 μm. Right: RNA sequencing reads from *P. donghaensis* aligned to the *abiF* locus.

(B) Schematic representation of the *E. coli* expression vectors for EcAbiF and its ncRNA.

(C) Growth assays of EcAbiF induction, with and without EcAbiFr ncRNA. Dashed line marks the induction time for the growth assay.

(D) Spotting assay of the EcAbiF alone or EcAbiF and EcAbiFr ncRNA with and without induction.

(E) Spotting assay of the catalytically inactive PbAbiF form with or without induction.

(F) Spotting assay of the PbAbiF with or without the pre-induction of its ncRNA. Bacteria were grown in a liquid culture with and without arabinose and were then plated on a plate with the relevant induction.

(G) Plasmid addiction assay. Bacteria transformed with either EcAbiF and its ncRNA or mCherry plasmids harboring the ampicillin resistance gene were grown overnight in liquid culture containing both Amp and IPTG, and then one aliquot was plated on an Amp plate. The bacteria were then diluted into a liquid culture containing only IPTG grown overnight and plated on an Amp plate and a second aliquot was diluted into a liquid culture containing only IPTG was repeated for 10 days. Right: CFU counts from Amp plates over time.

(H) RNA sequencing reads from EcAbiF RNP pull-down aligned to the EcAbiF locus.

(I) RNA sequencing reads from dead EcAbiF (R213A, H218A) RNP pull-down aligned to the EcAbiF locus.

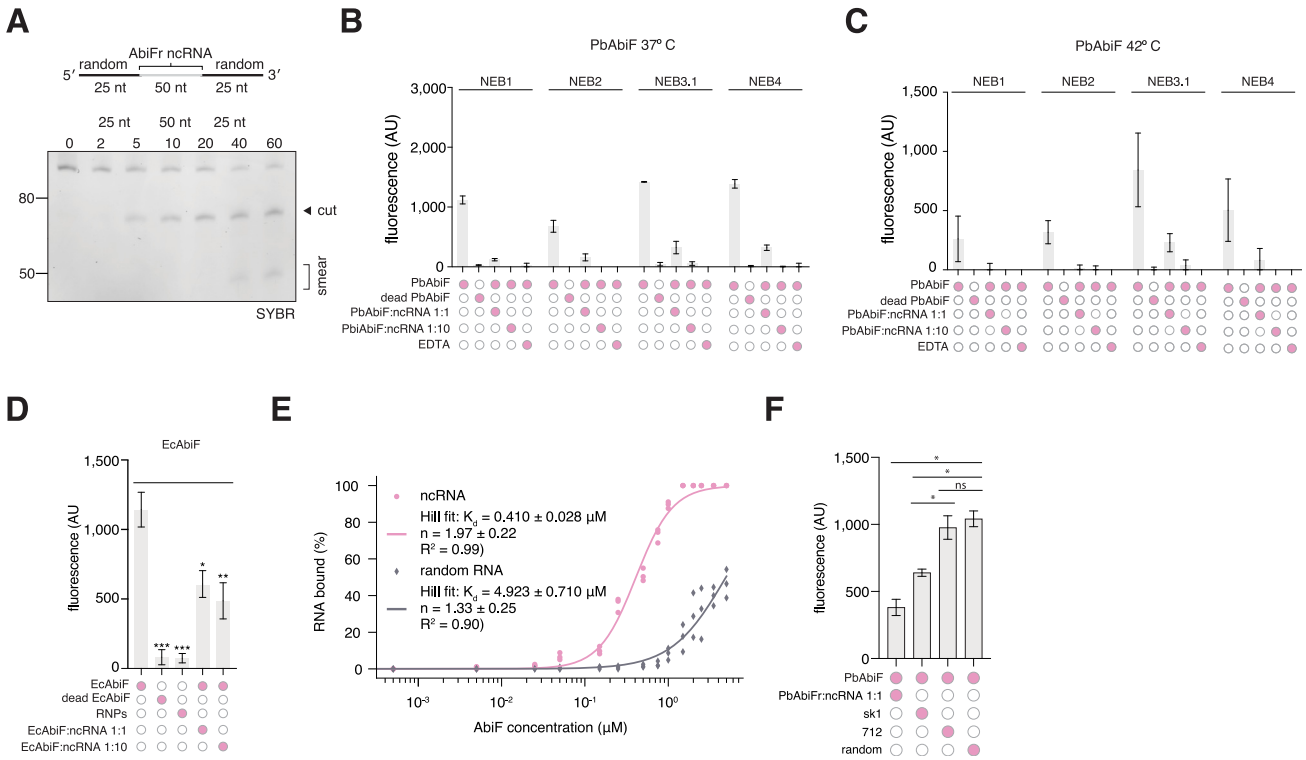


Figure S5. Extended PbAbiF and EcAbiF RNaseAlert reporter assay, PbAbiF endonuclease activity, and inhibition by phage's hairpin, related to Figure 5

(A) Time course assay of AbiFr ncRNA processing by the PbAbiF protein. PbAbiF was incubated with ssRNA composed of the AbiFr ncRNA and 25 random nucleotides upstream and downstream of the ncRNA sequence. Aliquots were taken at specified time points and the RNA was visualized using a SYBR stain.

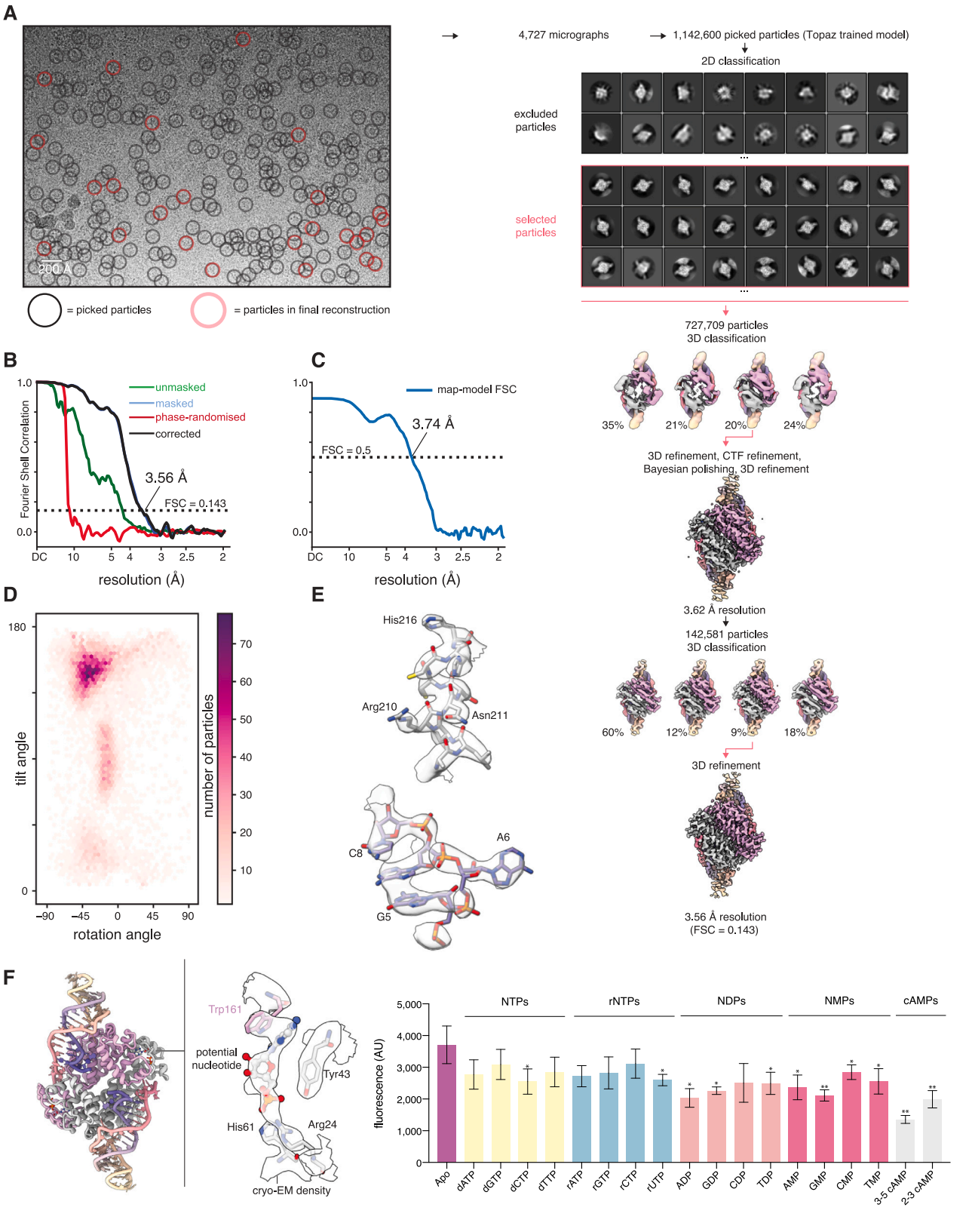
(B) PbAbiF activity assay using the RNaseAlert reporter system, wherein reporter RNA was incubated with PbAbiF in different buffers at 37°C.

(C) Same as (B) except at 42°C.

(D) EcAbiF RNA activity assay using the RNaseAlert reporter system. The reporter RNA was incubated with either EcAbiF and different ssRNA sequences or EcAbiF-EcAbiFr RNPs. *$p < 0.05$, **$p < 0.0,1$ ***$p < 0.001$ (two-tailed Student's t test, $n = 3$).

(E) EMSA assay binding curves of bound RNA fraction as a function of PbAbiF protein concentration for either PbAbiFr ncRNA or random RNA. Binding curves parameters were estimated using least squares for the Hill equation, resulting in K_D and Hill coefficient (n) estimates with estimate errors derived as 95% confidence intervals assuming normally distributed parameter errors.

(F) PbAbiF RNase activity using the RNaseAlert reporter system. PbAbiF was incubated with the reporter RNA and the SK1 or 712 hairpin. *$p < 0.05$, **$p < 0.01$ (two-tailed Student's t test, $n = 3$).



(legend on next page)

Figure S6. Extended cryo-EM data, related to Figure 6

(A) Left: example cryo-EM micrograph of the purified R2Bm TPRT complex. Right: 2D class averages of the PbAbiF RNP and a flowchart outlining how cryo-EM data were processed.

(B) Gold-standard Fourier Shell Correlation curve for the final reconstruction.

(C) Map-to-model Fourier Shell Correlation as calculated in PHENIX, softly masking the map around the fitted model.

(D) Orientation distribution plot for the final reconstruction.

(E) Example cryo-EM densities for different parts of the structure.

(F) Left: PbAbiF cryo-EM density that may be accounted for by a nucleotide (AMP is modeled as an example). Right: Apo PbAbiF RNaseAlert reporter assay in the presence of nucleotides, nucleosides, and cyclic nucleotides in a 10:1 ratio (protein:nucleotide). $*p < 0.05$, $**p < 0.01$ (two tailed Student's t test, $n = 3$).

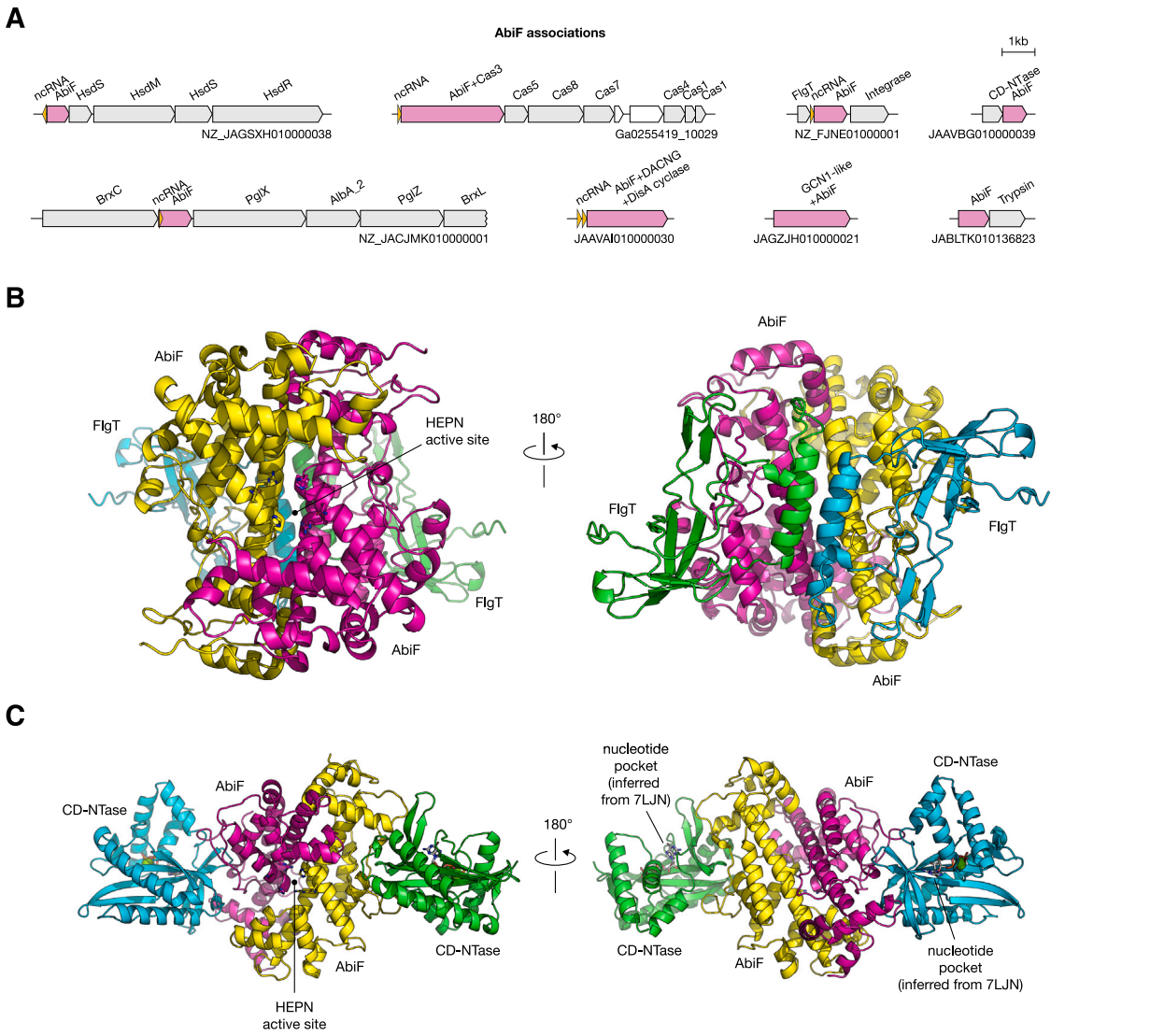


Figure S7. Diverse associations with AbiF-like systems

(A) Prokaryotic genomic loci depicting AbiF-like genes in diverse contexts.

(B) AlphaFold2 structure prediction of the AbiF-FlgT system in a 2:2 molar ratio.

(C) AlphaFold2 structure prediction of the AbiF-CBASS system in a 2:2 molar ratio.

Overall, the figure shows the additionally identified diverse AbiF systems. We found associations with various defense systems, such as restriction modification systems and bacteriophage exclusion (BREX). In these cases, AbiF could safeguard the expression and genomic maintenance of the associated defense systems, as shown for other TA modules[99] (A). We also found phage integrase AbiF systems, which often contain a flagellar *flgT*-like gene, essential for flagellar motility.[100] AlphaFold2 multimer prediction suggests FlgT-like protein interacts with AbiF, suggesting AbiF may have been co-opted for an alternative function (B). We also identified groups of large AbiF genes, one of which does not appear to have an associated ncRNA and is fused to GCN1, a ribosome-interacting protein.[101] Another system encodes an AbiF fusion to a DACNG domain and a DisA-like di-adenylate cyclase which is known to produce signaling c-di-AMP molecules involved in DNA damage response.[102] In another AbiF group, the ncRNA antitoxin appears to be replaced by a trypsin-like protease which may serve as a protein antitoxin by degrading the AbiF protein. We also found an AbiF associated with a CBASS-like secondary messenger-producing NTase. In this case, AlphaFold2 structure prediction indicates they form a complex where the NTase is predicted to bind to AbiF in a similar location to where the ncRNA antitoxin binds to PbAbiF (C). A structural alignment of the CBASS protein with PDB 7LJN shows that the AbiF subdomain of AbiF and the NTase are predicted to be in close proximity. Beyond AbiF, we found various STAND and NACHT ATPases fused to or associated with RxxxxH motif HEPN domains (Figure 1), which may be RNases allosterically triggered by invading pathogens.[103] There was also a tyrosine recombinase family that includes a HEPN-containing protein and a PD(D/E)xK nuclease, suggesting a potentially unique mechanism of transposition. Lastly, there were groups containing two HEPNs in the pEK499_p136 family, similar in domain architecture to F13a1/2 suggesting HEPN duplication may have occurred independently in different evolutionary lineages.