Made available through Montana State University's ScholarWorks

Protein-primed homopolymer synthesis by an antiviral reverse transcriptase

Stephen Tang, Rimantė Žedaveinytė, Nathaniel Burman, Shishir Pandey, Javier Mancilla-Ramírez, Louie M Kulber, Tanner Wiegand, Royce A. Wilkinson, Yanzhe Ma, Dennis J. Zhang, George D. Lampe, Mirela Berisa, Marko Jovanović, Blake Wiedenheft, Samuel H. Sternberg

Accessibility Disclaimer:

For a more accessible version of this document, please submit an accessibility request form through the Montana State University Library website.

Article

Protein-primed homopolymer synthesis by an antiviral reverse transcriptase

https://doi.org/10.1038/s41586-025-09179-5

Received: 7 March 2025

Accepted: 20 May 2025

Published online: 28 May 2025

Check for updates

Stephen Tang[1,7], Rimantė Žedaveinytė[1,7], Nathaniel Burman[2,7], Shishir Pandey[2,7], Josephine L. Ramirez[1], Louie M. Kulber[1], Tanner Wiegand[1,3], Royce A. Wilkinson[2], Yanzhe Ma[4], Dennis J. Zhang[1,6], George D. Lampe[1,3], Mirela Berisa[5], Marko Jovanovic[4], Blake Wiedenheft[2]✉ & Samuel H. Sternberg[1,3]✉

Bacteria defend themselves from viral predation using diverse immune systems, many of which target foreign DNA for degradation[1]. Defence-associated reverse transcriptase (DRT) systems provide an intriguing counterpoint to this strategy by using DNA synthesis instead[2,3]. We and others recently showed that DRT2 systems use an RNA template to assemble a de novo gene that encodes the antiviral effector protein Neo[4,5]. It remains unclear whether similar mechanisms of defence are used by other related DRT families. Here, we show that DRT9 systems defend against phage using DNA homopolymer synthesis. Viral infection triggers polydeoxyadenylate (poly-dA) accumulation in the cell, driving abortive infection and population-level immunity. Cryo-electron microscopy structures reveal how a non-coding RNA serves as both a structural scaffold and reverse transcription template to direct hexameric complex assembly and poly-dA synthesis. Notably, biochemical and functional experiments identify tyrosine residues within the reverse transcriptase itself that probably prime DNA synthesis, leading to the formation of protein–DNA covalent adducts. Synthesis of poly-dA by DRT9 in vivo is regulated by the competing activities of phage-encoded triggers and host-encoded silencers. Collectively, our study identifies a nucleic-acid-driven defence system that expands the paradigm of bacterial immunity and broadens the known functions of reverse transcriptases.

Bacteria encode diverse mechanisms for antiviral defence, the majority of which degrade invading nucleic acids using either programmable, sequence-specific or non-specific nucleases[1]. Also common are enzymes that act on host nucleic acids—synthesizing second messenger molecules[6], converting nucleotides into suicide/terminator analogues[7] or depleting nucleotides altogether[8,9]. Against this backdrop, reverse transcriptase (RT) enzymes, which polymerize rather than degrade DNA, would seem unlikely contributors to antiphage defence. Yet, multiple families of RTs have recently been implicated in defensive functions spanning the adaptive and innate arms of the bacterial immune system[2–5,10,11].

RTs associated with CRISPR–Cas comprise one family of defence-related systems, using RNA-templated DNA synthesis to facilitate spacer acquisition directly from viral RNA transcripts[10]. However, whereas CRISPR-associated RTs reverse-transcribe foreign RNA, other RT systems encode and reverse-transcribe their own non-coding RNA (ncRNA). Retron-encoded ncRNAs template the synthesis of hybrid RNA–DNA molecules—known as multicopy single-stranded DNA (msDNA)—that function as anti-toxins against host-encoded toxins[12]. After viral infection, a cascade of events leads to msDNA modification or degradation and toxin release, killing the infected cell and thereby pre-empting further viral propagation[12,13]. This strategy, known as abortive infection, protects the larger bacterial population at the cost of the infected cell, and is ubiquitous among innate immune mechanisms[1]. Abortive infection also characterizes a third group of antiviral RTs typified by AbiK, AbiA and AbiP2, which, in contrast to CRISPR-associated RTs and retrons, have been shown to synthesize random DNA polymers independently of any RNA template[14,15]. Notably, recent structural and biochemical studies revealed that DNA polymerization by Abi RTs is primed by a tyrosine residue within the RT itself and requires the formation of a homo-oligomeric RT complex[15,16]. These studies have demonstrated the presence of highly unusual biochemical behaviours across the Abi RT family, but how these activities relate to phage infection, in terms of their regulation and their effects on the phage and/or the host, remains unresolved.

The Abi RTs belong to a large and diverse family of systems collectively designated as the Unknown Group (UG)[17]. Pioneering work[2,3] revealed an expanded set of UG systems that function in antiviral immunity, termed DRTs. We and others recently resolved the molecular mechanism of the DRT2 defence system[4,5]. Like retron systems, DRT2 operons encode a conserved, structured ncRNA that provides the template for complementary DNA (cDNA) synthesis by the RT. Rolling-circle reverse transcription of the ncRNA leads to the synthesis of concatemeric cDNA (ccDNA) products, assembling intact promoters

[1]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA. [2]Department of Microbiology and Cell Biology, Montana State University, Bozeman, MT, USA. [3]Howard Hughes Medical Institute, Columbia University, New York, NY, USA. [4]Department of Biological Sciences, Columbia University, New York, NY, USA. [5]Metabolomics Core, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [6]Present address: Section of Microbiology, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [7]These authors contributed equally: Stephen Tang, Rimantė Žedaveinytė, Nathaniel Burman, Shishir Pandey. ✉e-mail: bwiedenheft@gmail.com; shsternberg@gmail.com

and a nearly endless open reading frame (*neo*) spanning ccDNA repeats. During phage infection, ccDNA is transcribed and translated to yield Neo proteins that drive cell growth arrest in a mechanism akin to abortive infection. Together, these studies have provided an early glimpse into the unique molecular functions of UG-family RTs, although it is unresolved whether similar mechanisms exist across the vast diversity of unexplored UG systems.

We set out to identify conserved mechanisms of RNA-templated de novo gene assembly in other bacterial immune systems. In this regard, we focused on DRT9 owing to its nearby placement on the phylogenetic tree of UG RTs and its similar genetic architecture to DRT2 (Fig. 1a). What we instead found was a fundamentally distinct mechanism of RT-mediated innate immunity. Here we show that DRT9 systems are triggered by phage infection to produce long, uninterrupted chains of poly-dA. Poly-dA synthesis is primed by tyrosine residues within the RT itself, and is templated by a structured ncRNA containing a universally conserved poly-uridine tract. Cryo-electron microscopy (cryo-EM) data reveal extensive interactions between protein and RNA components that mediate the assembly of a hexameric RT–ncRNA complex, in which the templating uridines reside proximal to the RT active site. The production of poly-dA is sensitively regulated by a 'tug of war' between host- and phage-encoded factors, and its accumulation in cells drives cell death and abortive infection. In addition to revealing an elegant and intricate mechanism by which DNA synthesis confers antiviral immunity, our work highlights an unusual synthetic pathway and biological function of a nucleic acid homopolymer.

DRT9 synthesizes DNA homopolymers

DRT9 systems encode an RT gene and an upstream ncRNA, occasionally in association with a SLATT-domain-containing protein[3] (Extended Data Fig. 1a and Supplementary Table 1). We selected ten diverse systems for gene synthesis, expressed them in *Escherichia coli*, and assessed their antiviral activity against a panel of well-characterized *E. coli* phages (Extended Data Fig. 1b and Supplementary Table 2). DRT9 systems from *Salmonella enterica* and *Photobacterium sanctipauli* (hereafter *Sen*DRT9 and *Psa*DRT9) defended against multiple phages, particularly those from the *Tequatrovirus* and *Tequintavirus* genera, and this protection was abolished by mutation of catalytic aspartate residues in the YADD active site (Extended Data Fig. 1c,d). Next, to determine whether DRT9 provides defence through direct inhibition of phage replication or by abortive infection, we performed phage infection assays in liquid culture. When cultures of *Sen*DRT9-expressing cells were infected with phage T5 at a low multiplicity of infection (MOI), they grew similarly to uninfected cultures, but a high MOI of T5 inhibited cell growth (Extended Data Fig. 1e). Quantification of viable phage particles and cells from high-MOI infections revealed that infected cells expressing DRT9 successfully inhibited phage replication, but were unable to recover from infection, indicating that DRT9 activation leads to programmed cell death and abortive infection (Extended Data Fig. 1f).

We next sought to investigate DRT9 RT activity in vivo using a Flag-tagged *Sen*DRT9 RT (*Sen*RT). Notably, epitope tagging at the *Sen*RT C terminus, but not the N terminus, severely impaired phage defence (Extended Data Fig. 1g). This observation was consistent with AlphaFold 3 modelling of *Sen*RT, which positioned the C terminus proximal to the predicted catalytic residues, suggesting a potential role for the C terminus in *Sen*RT enzymatic function (Extended Data Fig. 1h). Next, having confirmed that N-terminally Flag-tagged *Sen*RT retains wild-type (WT) defence function, we used the tagged RT for RNA immunoprecipitation (RIP–seq) and cDNA immunoprecipitation (cDIP–seq) sequencing[4] in uninfected and T5-phage-infected cells (Fig. 1b). As with DRT2, the ncRNA encoded immediately upstream of the RT was strongly enriched by RIP–seq but, in contrast to DRT2, cDIP–seq analysis did not detect a reverse-transcription product (Fig. 1c,d). Given the genetic evidence that reverse-transcription activity is essential for phage defence (Extended Data Fig. 1d), we reasoned that DRT9 indeed produces a cDNA, although its identification might require an alternative approach.

Inspired by reports that AbiK RTs synthesize untemplated, random DNA polymers[14,16], we hypothesized that analysing cDIP–seq reads that did not map to the reference genome might reveal unexpected cDNA products. Inspection of alignment statistics revealed a marked and specific increase in unmapped reads for phage-infected samples expressing WT *Sen*RT (Fig. 1e). This observation prompted us to manually inspect the unmapped reads, which revealed an unanticipated result—long, uninterrupted stretches of poly-deoxythymidylate (poly-dT) and poly-deoxyadenylate (poly-dA), a pattern further corroborated by motif detection analysis (Fig. 1f and Extended Data Fig. 2a). RIP–seq and cDIP–seq experiments with the related *Psa*DRT9 system similarly revealed strong enrichment of the upstream ncRNA, an absence of cDNA species mapping to the ncRNA and a significant overrepresentation of poly-dT motifs within unmapped reads (Extended Data Fig. 2b,c). These results provide evidence of an RT naturally synthesizing a DNA homopolymer and implicate poly-dT and/or poly-dA products in the mechanism of antiphage defence.

We next quantified reads containing homopolymers (at least 25 nucleotides in length) of each of the four deoxynucleotides across all cDIP–seq datasets. Poly-dT and poly-dA were detected exclusively in phage-infected cells expressing WT *Sen*DRT9, with poly-dT outnumbering poly-dA by a factor of about ten, while poly-dG and poly-dC products were not detectable above background levels (Fig. 1g). To account for potential sequencing artefacts affecting homopolymer detection and quantification, we turned to a classical molecular biology approach to independently assess poly-dT and poly-dA production. Southern blotting experiments using oligo-dT and oligo-dA probes corroborated our cDIP–seq results, revealing abundant poly-dA and poly-dT species, respectively, that were again restricted to phage-infected cells expressing WT *Sen*DRT9 (Fig. 1h and Extended Data Fig. 2d). Two additional observations from Southern blotting were notable. First, poly-dA and poly-dT species ranged from about 1,000 to more than 10,000 bases in length, relative to a double-stranded DNA (dsDNA) ladder (Fig. 1h). Second, poly-dA products were substantially more abundant than poly-dT products on the basis of Southern blotting, in contrast to the relative proportions observed by sequencing (Fig. 1g,h). To investigate this discrepancy, we performed control experiments in which synthetic dT_{100} and dA_{100} oligonucleotides were spiked into *E. coli* cell lysates at varying ratios before library preparation and sequencing (Extended Data Fig. 2e). These experiments revealed a strong bias against poly-dA capture and sequencing relative to poly-dT (Extended Data Fig. 2f), suggesting that Southern blotting data more accurately reflect the true intracellular distribution of homopolymer products. High-throughput sequencing nevertheless provided insights that would have been challenging to detect otherwise, including the presence of chimeric species containing dA_{10}–dT_{10} transition sequences (Fig. 1i). Notably, these species were more abundant than their inverse counterparts (dT_{10}–dA_{10}), suggesting that initial cDNA synthesis by DRT9 may exclusively produce poly-dA, which occasionally serves as a template and a primer to generate a complementary poly-dT strand.

Collectively, these findings indicate that phage infection triggers the synthesis of poly-dA by the DRT9 immune system, while poly-dT probably arises from low-level second-strand synthesis by the RT itself or another DNA polymerase. Notably, RNA-seq datasets did not reveal any evidence of poly-A or poly-U species (Extended Data Fig. 2g), therefore arguing against any coding role for DRT9 cDNA synthesis—a notable divergence from DRT2. This observation, along with the distinct nature of the cDNA sequence, underscores the functional and molecular diversity of DRT enzymes and their reverse transcription products.

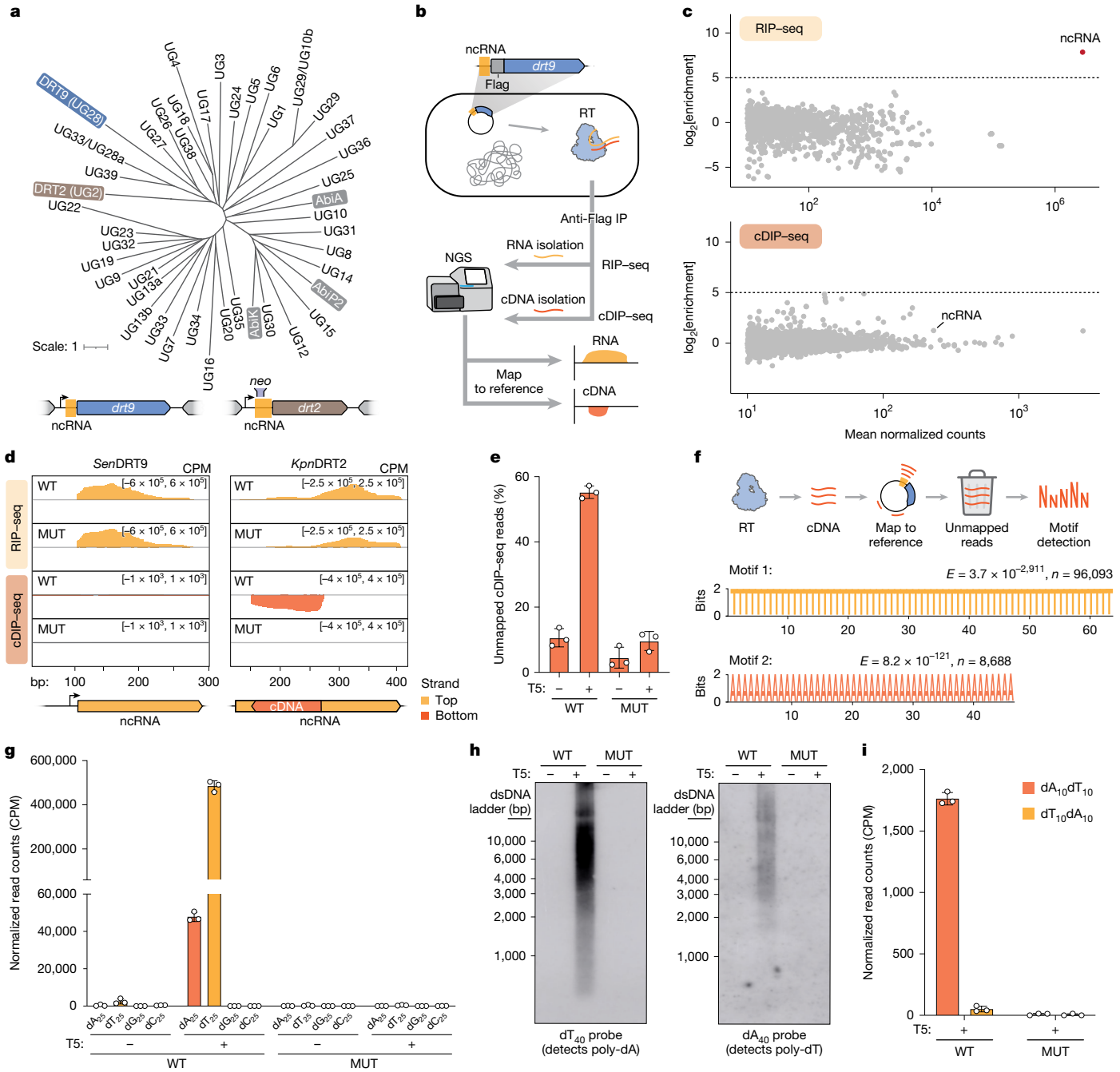


Fig. 1 | Systematic identification of DRT9 reverse transcription substrates and products in vivo. a, Phylogenetic tree of bacterial RT homologues within the UG family, based on ref. 3. DRT9 (UG28) and DRT2 (UG2) are highlighted, as are other experimentally studied systems; the genetic architectures of archetypal systems are shown below. **b**, Schematic of the RIP–seq and cDIP–seq approaches to identify the nucleic acid templates and products of *Sen*DRT9. NGS, next-generation sequencing. **c**, *MA* plots showing enriched RNA (top) and DNA (bottom) loci from RIP–seq and cDIP–seq experiments, respectively, relative to the input controls; the red dots indicate transcripts with $\log_2$[enrichment] > 5 and false-discovery rate (FDR) < 0.05. **d**, RIP–seq and cDIP–seq coverage tracks for WT or catalytically inactive (MUT) RTs in T5-phage-infected cells expressing *Sen*DRT9 (left) or *Kpn*DRT2 (right). Genomic locus schematics are shown below; data are normalized for sequencing depth and plotted as counts per million reads (CPM). **e**, Percentage of unmapped cDIP–seq reads from WT or MUT *Sen*DRT9 cDIP–seq datasets with or without T5 phage infection. **f**, Schematic of the unmapped read analytical pipeline (top), and the MEME results that revealed poly-dT and poly-dA motifs enriched in unmapped reads from the WT + T5 cDIP–seq dataset in **e** (bottom). *E* represents the *E*-value significance; *n* values show the number of contributing sites. **g**, Normalized homopolymer counts from WT and MUT cDIP–seq datasets with or without T5 phage infection. **h**, Southern blot analysis of total DNA isolated from cells expressing WT or MUT *Sen*DRT9 with or without T5 phage infection. Duplicate blots were probed with either oligo-dT_{40} (left) or oligo-dA_{40} (right); sizes from a double-stranded DNA (dsDNA) ladder are marked. **i**, Analysis of chimeric $dA_{10}dT_{10}$ and $dT_{10}dA_{10}$ counts from WT and MUT cDIP–seq datasets in the presence of T5 phage infection. For **e**,**g**,**i**, data are mean ± s.d. of $n = 3$ independent biological replicates.

The DRT9 ncRNA templates poly-dA synthesis

We set out to examine how an RT synthesizes kilobase-length polymers composed almost exclusively of adenosine nucleotides. We initially considered the hypothesis that the RT might directly select dATP substrates through protein-mediated recognition rather than using a nucleic acid template, akin to the reported mechanism of untemplated DNA synthesis by AbiK[16]. However, closer inspection of the DRT9

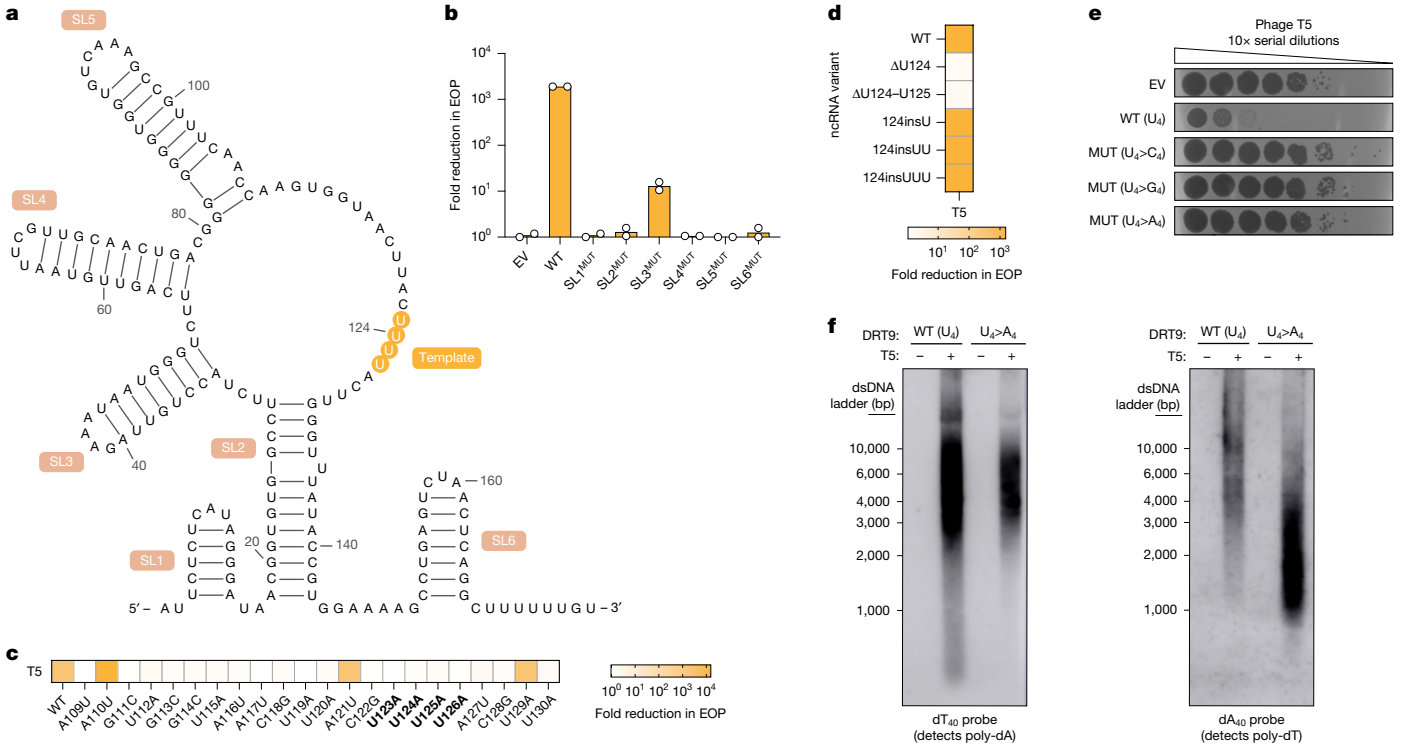


Fig. 2 | ncRNA sequence determinants of *Sen*DRT9-mediated phage defence and poly-dA synthesis. a, The predicted secondary structure of the *Sen*DRT9 ncRNA. SL regions and uridine nucleotides implicated in RNA-templated DNA synthesis are labelled; coordinates are numbered based on the mature ncRNA species identified by RIP–seq. **b**, Quantification of *Sen*DRT9 defence activity against T5 phage for scrambled ncRNA SL mutants, quantified as the fold reduction in efficiency of plating (EOP) relative to an EV control. Data are from $n = 2$ technical replicates. **c**, Quantification of *Sen*DRT9 defence activity for the indicated ncRNA point mutations, quantified as described in **b**. Uridine nucleotides implicated in RNA-templated DNA synthesis are highlighted in bold. Data are the mean of $n = 2$ technical replicates. **d**, Quantification of *Sen*DRT9 defence activity for the indicated ncRNA mutations, shown as described in **c**. **e**, Plaque assay showing the loss of *Sen*DRT9 defence activity against T5 phage for substitutions within the putative ncRNA template region; residues U123–U126 (U_4) were mutated to each of the indicated nucleotides. **f**, Southern blot analysis of total DNA isolated from cells expressing the WT or U_4>A_4 MUT ncRNA from **e**, with or without T5 phage infection. Duplicate blots were probed with either oligo-dT_{40} (left) or oligo-dA_{40} (right); sizes from a dsDNA ladder are marked. Representative data are shown for experiments repeated at least twice with similar results.

ncRNA sequence and predicted secondary structure suggested an alternative hypothesis. In addition to conserved stem-loop (SL) elements that loosely resemble the overall architecture of DRT2 ncRNAs, DRT9 ncRNAs contain a universally conserved uridine-rich stretch positioned similarly to the DRT2 reverse-transcription start site (Fig. 2a and Extended Data Fig. 3a,b). We envisioned that this stretch of uridines could serve as a template for poly-dA synthesis, potentially through iterative synthesis, melting and realignment of the growing poly-dA strand across from complementary uridine nucleotides.

To investigate the potential structural and templating functions of the ncRNA, we scrambled each of the conserved SL regions and assessed the effect of SL perturbation on antiphage defence activity. All SLs were essential for defence against T5 phage except for SL3, the disruption of which had a mild effect (Fig. 2b and Extended Data Fig. 3c). We then introduced single-nucleotide substitutions within a prominent loop region containing the putative U-rich template. Most mutations in this loop severely impaired defence, suggesting a critical role for this region in ncRNA function (Fig. 2c). Notably, deletion of a single uridine residue (U124) completely abolished defence, whereas similar perturbations in a secondary U-rich region at the 3′ end of the ncRNA had no effect (Fig. 2a,d and Extended Data Fig. 3d). Importantly, RIP–seq experiments on a subset of ncRNA mutants revealed that the observed loss of defence activity was in most cases not attributable to defects in RT–ncRNA binding (Extended Data Fig. 3e). The effects of these ncRNA perturbations tested against T2 phage closely mirrored those observed with T5 (Extended Data Fig. 3f), further supporting the functional importance of ncRNA structural and sequence motifs.

We further investigated the putative templating role of the ncRNA by testing additional mutations to the U-rich region. Mutating residues 123–126 (U_4) to alternative nucleotides (C_4, G_4 or A_4) abolished phage defence in all cases (Fig. 2e), initially suggesting loss of DNA synthesis with these mutants. However, Southern blot analysis of the U_4>A_4 mutant revealed an inversion of the WT homopolymer distribution: instead of primarily producing poly-dA with lower levels of poly-dT, the A_4 template primarily generated poly-dT with lower levels of poly-dA (Fig. 2f and Extended Data Fig. 3g). Technical challenges related to the chemical synthesis of guanosine- and cytidine-only oligonucleotides prevented us from testing whether the U_4>G_4 and U>C_4 variants could similarly template complementary homopolymers by Southern blotting.

These results implicate the U-rich loop as an RNA template for cDNA synthesis. Like DRT2 systems, which perform rolling-circle reverse transcription of an approximately 120-nucleotide template to generate repetitive cDNA products[4,5], *Sen*DRT9 also spools out repetitive cDNA products. However, in contrast to DRT2, *Sen*DRT9 templates a single nucleotide identity, yielding uninterrupted homopolymers.

DRT9 protein priming and oligomerization

We next sought to determine the biochemical basis for RNA-templated DNA homopolymer synthesis by DRT9. We initially purified *Sen*RT as a His_6–GST fusion protein, which co-purified with an RNA species corresponding to the first approximately 150 nucleotides of the full-length ncRNA transcript (Extended Data Fig. 4a–c). Given that His_6–GST-tagged RT constructs exhibited reduced phage defence

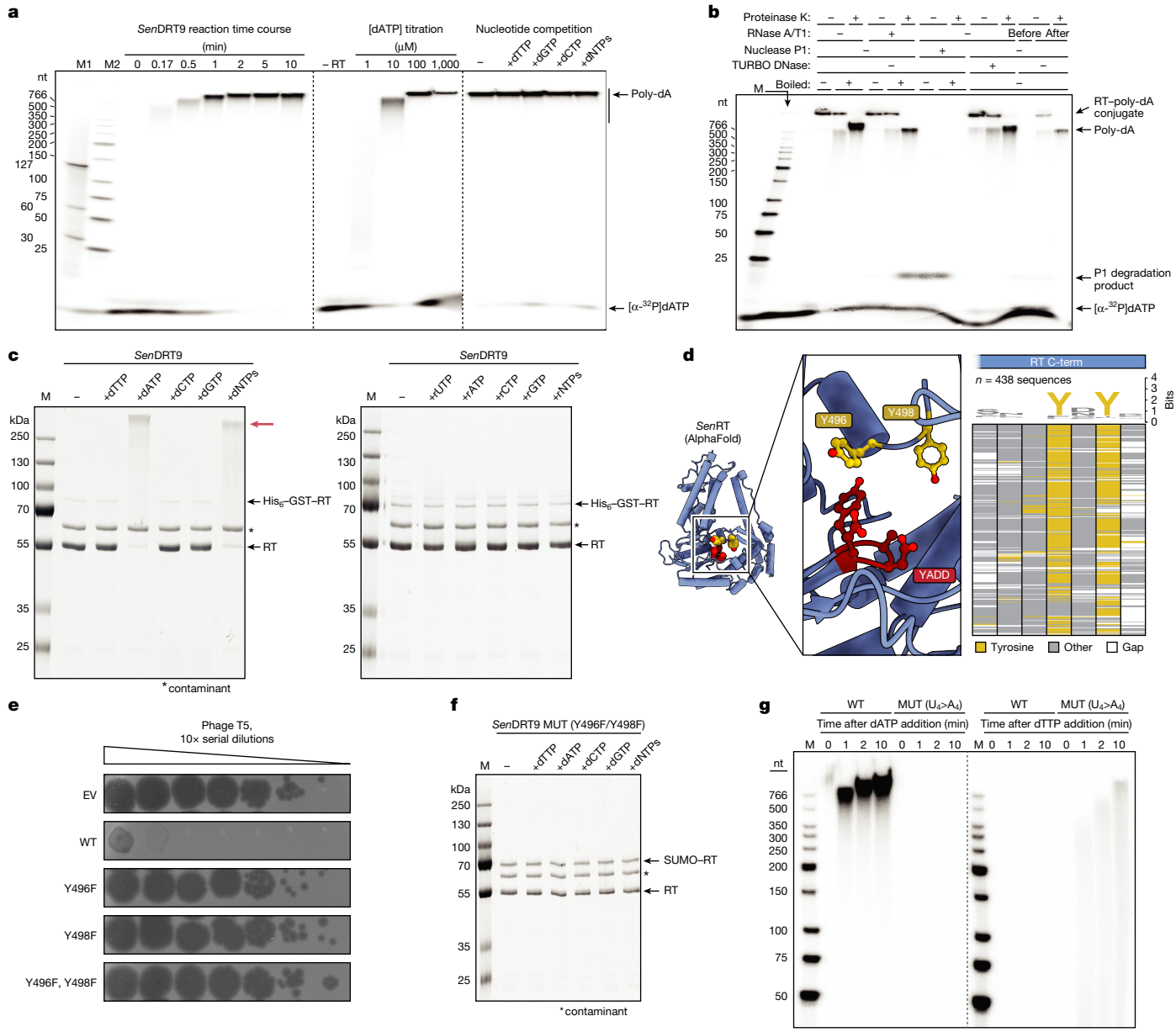


Fig. 3 | The *Sen*DRT9 RT–ncRNA complex performs protein-primed, RNA-templated DNA homopolymer synthesis. a, Denaturing 5% urea–PAGE analysis of DNA polymerization assays testing the effects of time (left), dATP concentration (middle) and nucleotide (nt) competition (right) on poly-dA synthesis. All of the reactions contained 20 nM RT–ncRNA complex. Left, the reactions contained 100 μM dATP. Middle, the reactions were incubated for 10 min. Right, the reactions were incubated for 10 min with 100 μM radiolabelled dATP and unlabelled nucleotides at 100 μM. M1 and M2 denote DNA ladder markers. **b**, Denaturing 5% urea–PAGE analysis of DNA polymerization assays, after the reactions were treated with the indicated proteinase or nuclease reagents with or without boiling. All of the reactions contained 20 nM RT–ncRNA and 100 μM [α-^{32}P]dATP, and were incubated at 37 °C for 10 min before proteinase/nuclease addition, with the exception of lane 15, which was pretreated with RNase A/T1 before dATP addition. **c**, SDS–PAGE analysis of DNA polymerization assays, in which RT–ncRNA complexes (0.15 μM) were incubated with the indicated dNTPs or rNTPs (0.9 mM) for 60 min at 37 °C. The red arrow denotes high-molecular-weight protein–DNA conjugates in reactions that contained dATP. The His_6–GST–RT band arises from incomplete tag removal. The asterisk indicates purification contaminant. M, protein marker. **d**, *Sen*RT AlphaFold 3 structure prediction (left), with a magnified view (middle) showing the close proximity of C-terminal residues Tyr496 and Tyr498 (yellow) to the YADD active site (red). The multiple-sequence alignment (right) highlights their strong conservation. **e**, Plaque assay showing a loss of *Sen*DRT9 defence activity against T5 phage for the indicated mutants. **f**, SDS–PAGE analysis of DNA polymerization assays as described in **c**, but with a Y496F/Y498F RT mutant. **g**, Denaturing 5% urea–PAGE analysis of DNA polymerization assays with either WT or U4>A4 mutant ncRNA in the presence of [α-^{32}P]dATP (left) or [α-^{32}P]dTTP (right). All of the reactions contained 20 nM RT–ncRNA complexes and 100 μM [α-^{32}P]dATP or [α-^{32}P]dTTP. For **a**–**c**,**f**,**g**, representative data are shown for experiments repeated at least twice with similar results.

activity (Extended Data Fig. 4d), we reasoned that the affinity tag and solubilization domain might impair RT activity. Notably, after protease cleavage of the N-terminal tag, we observed a shift in the gel-filtration retention volume to a monodisperse, higher-molecular-weight species, consistent with oligomerization (Extended Data Fig. 4e). Incubation of the oligomeric RT–ncRNA complex with [α-^{32}P]dATP, followed by denaturing urea–PAGE analysis, revealed robust and highly processive poly-dA synthesis activity that completely consumed dATP in the reaction, and that required no additional reaction components beyond the RT and ncRNA (Fig. 3a). Poly-dA products were hundreds of nucleotides

to over 1,000 nucleotides in length and insensitive to the addition of alternative nucleotide substrates (Fig. 3a and Extended Data Fig. 4f), corroborating the conclusions from cDIP–seq and Southern blot analyses that DRT9 produces long, homopolymeric DNA.

Additional observations from analyses of RT reaction products provided insights into the mechanism of homopolymer synthesis. Single-strand-specific nuclease P1 completely degraded poly-dA, whereas TURBO DNase—which preferentially degrades dsDNA—had little effect (Fig. 3b). Meanwhile, proteinase K treatment altered the mobility of poly-dA, whereas RNase treatment had no effect (Fig. 3b). A similar effect on mobility was observed with Southern blotting of proteinase-K-treated poly-dA products isolated from cells (Extended Data Fig. 5a). Notably, when RT–ncRNA complexes were incubated with unlabelled dNTP/NTP substrates and then analysed using SDS–PAGE, we observed a substantial mobility shift of the RT protein in reactions that contained dATP (Fig. 3c). Consistent with these results, gel-filtration analysis of the RT–ncRNA complex after dATP incubation revealed conversion into a distinct species that exhibited a heightened absorbance at 260 nm (A_{260})/A_{280} ratio and ran in the void volume, indicating increased nucleic acid content and molecular weight (Extended Data Fig. 5b). Taken together, these results suggest that poly-dA products are covalently linked to the RT protein itself, potentially through priming by a hydroxyl-containing amino acid side chain.

To identify candidate RT residues involved in protein priming of poly-dA synthesis, we immunoprecipitated the WT or catalytically inactive (MUT) RT from T5-infected cells and analysed RT peptides using mass spectrometry (MS). We hypothesized that a covalent DNA linkage would interfere with peptide detection, leading to apparent depletion of any poly-dA-linked peptide. Notably, only two peptides were significantly depleted in WT samples relative to the MUT condition (Extended Data Fig. 5c). Closer inspection of the AlphaFold-predicted RT monomer structure revealed that one of these peptides, derived from the C-terminal tail, contains two highly conserved tyrosine residues positioned in the immediate vicinity of the active site (Fig. 3d). Given the established role of tyrosine residues in protein-primed DNA synthesis by other RTs[16,18,19], we hypothesized that one or both of these tyrosine residues could similarly serve as primers for reverse transcription by the DRT9 RT.

To test this hypothesis, we individually or simultaneously mutated each of the putative priming tyrosines to phenylalanine and assessed the effect on DRT9 activity. Phage defence against T5 was completely abolished when either tyrosine residue was mutated (Fig. 3e). By contrast, loss of defence against T2 required substitution of both tyrosines (Extended Data Fig. 5d), suggesting that either residue can serve as a priming site. RIP–seq experiments in T5-infected cells showed that Tyr-to-Phe (Tyr>Phe) RT mutants exhibited mildly decreased ncRNA binding relative to the WT RT (Extended Data Fig. 5e), suggesting that the observed loss of defence may have also resulted in part from weakened RT–ncRNA interactions. Nonetheless, both in vitro polymerization reactions and in vivo experiments with phage-infected cells showed severe attenuation of poly-dA synthesis with the double Tyr>Phe mutant (Fig. 3f and Extended Data Fig. 5f,g). The residual presence of low levels of poly-dA product suggests that additional hydroxyl-containing residues in the RT active site could perhaps be used as alternative primers, albeit with considerably reduced efficiencies. Together, these results implicate C-terminal tyrosines within the *Sen*DRT9 RT as primers for poly-dA synthesis.

Finally, we sought to further assess our model for uridine templating of poly-dA synthesis through in vitro biochemical experiments. We purified RT–ncRNA complexes containing the same U_4>A_4 substitution tested in phage defence assays and assessed their polymerization activity. Whereas the WT RT–ncRNA complex efficiently polymerized dATP but not dTTP, the opposite substrate preference was observed with the U_4>A_4 ncRNA mutant (Fig. 3g and Extended Data Fig. 5h). Similar experiments using a U_4>G_4 mutant did not show evidence of dCTP polymerization (Extended Data Fig. 5i), and the U_4>C_4 ncRNA did not co-purify with the RT (Extended Data Fig. 5j), consistent with our earlier RIP–seq results (Extended Data Fig. 3e). These data indicate that templated homopolymer synthesis by DRT9 is reprogrammable, albeit limited to certain template sequences, potentially reflecting a combination of factors that could include ncRNA folding, base pairing energetics for repeated rounds of polymerization and melting, and structural flexibility of the homopolymer product.

Cryo-EM structure of the *Sen*DRT9 complex

To better understand the molecular and structural basis for poly-dA synthesis by *Sen*DRT9, we determined the structure of the His_6–GST-tagged RT and ncRNA using cryo-EM (Extended Data Fig. 6a–i). A 3.0 Å-resolution reconstruction enabled unambiguous assignment of 98% of the residues in the RT and 85% of nucleotides in the ncRNA (Extended Data Table 1). The RT contains the three characteristic finger, thumb and palm domains of canonical RTs, and an N-terminal extension (residues 1–58) reaches towards the C-terminal thumb domain, turning the conventional right-handed shape of the RT into a triangular architecture (Fig. 4a and Supplementary Fig. 2). The ncRNA forms a complex series of stem loops that interact with every domain of the protein (Fig. 4b,c). The 3′ end of the ncRNA curls around the thumb, threading through the centre of the triangle and forming extensive, non-base pairing interactions with the 5′ end of the ncRNA (Fig. 4b). The RT–ncRNA complex assembles into a C_3 symmetric trimer (Extended Data Fig. 6g) stabilized by kissing-loop interactions between SL5 (C93) of one molecule and SL5.1 (G113) of the adjacent subunit.

While the structure explains RT–ncRNA interactions and the role of the ncRNA in driving trimer assembly, the affinity-tagged protein is inactive for poly-dA synthesis in vitro and non-functional for defence against T5 phage in vivo (Extended Data Figs. 4d and 6j). We therefore reasoned that removal of the affinity tag would probably clarify the structural basis for RT–ncRNA function. After protease cleavage of the affinity tag, we used cryo-EM to determine a 2.6 Å-resolution reconstruction, which revealed a hexameric architecture—a back-to-back dimer of trimers (Fig. 4d, Extended Data Figs. 7 and 8a and Extended Data Table 1). The hexameric assembly is reminiscent of AbiK[16], but the mechanisms of assembly are distinct (Supplementary Fig. 2). Assembly of the AbiK hexamer relies on an α-helical repeat domain (that is, αRep), while assembly of the DRT9 hexamer relies on a combination of RNA–RNA (SL5–SL5.1 and SL4–SL4), RNA–protein and protein–protein interactions, resulting in a buried surface area of 3,000 $Å^2$ between trimers (Fig. 4d–f and Extended Data Fig. 8b,c). Although the structures of the trimer and hexamer are nearly equivalent, the hexamer contains additional density near the active site that is not present in the trimer reconstruction (Extended Data Fig. 8d). This unmodelled density may correspond to a small stretch of poly-dA product, but repeated efforts to sort these particles failed to separate distinct conformational states. Notably, the U_4 motif (123–126) is positioned around 8 Å from the YADD active site, consistent with a role in templating poly-dA synthesis (Fig. 4g and Extended Data Fig. 8e). Overall, these data reveal that the DRT9 RT–ncRNA complex forms a D_3 symmetric hexamer that performs RNA-templated poly-dA synthesis.

Host and viral factors modulating DRT9

Our biochemical investigation into DRT9 reverse transcription activity illuminated key molecular details of poly-dA synthesis, but also introduced an apparent discrepancy: in vitro poly-dA synthesis occurs constitutively, whereas in vivo poly-dA synthesis is strictly phage dependent. This suggested the presence of additional regulatory mechanisms that control DRT9 activation in the cellular context. To identify potential viral triggers of DRT9 activity, we screened

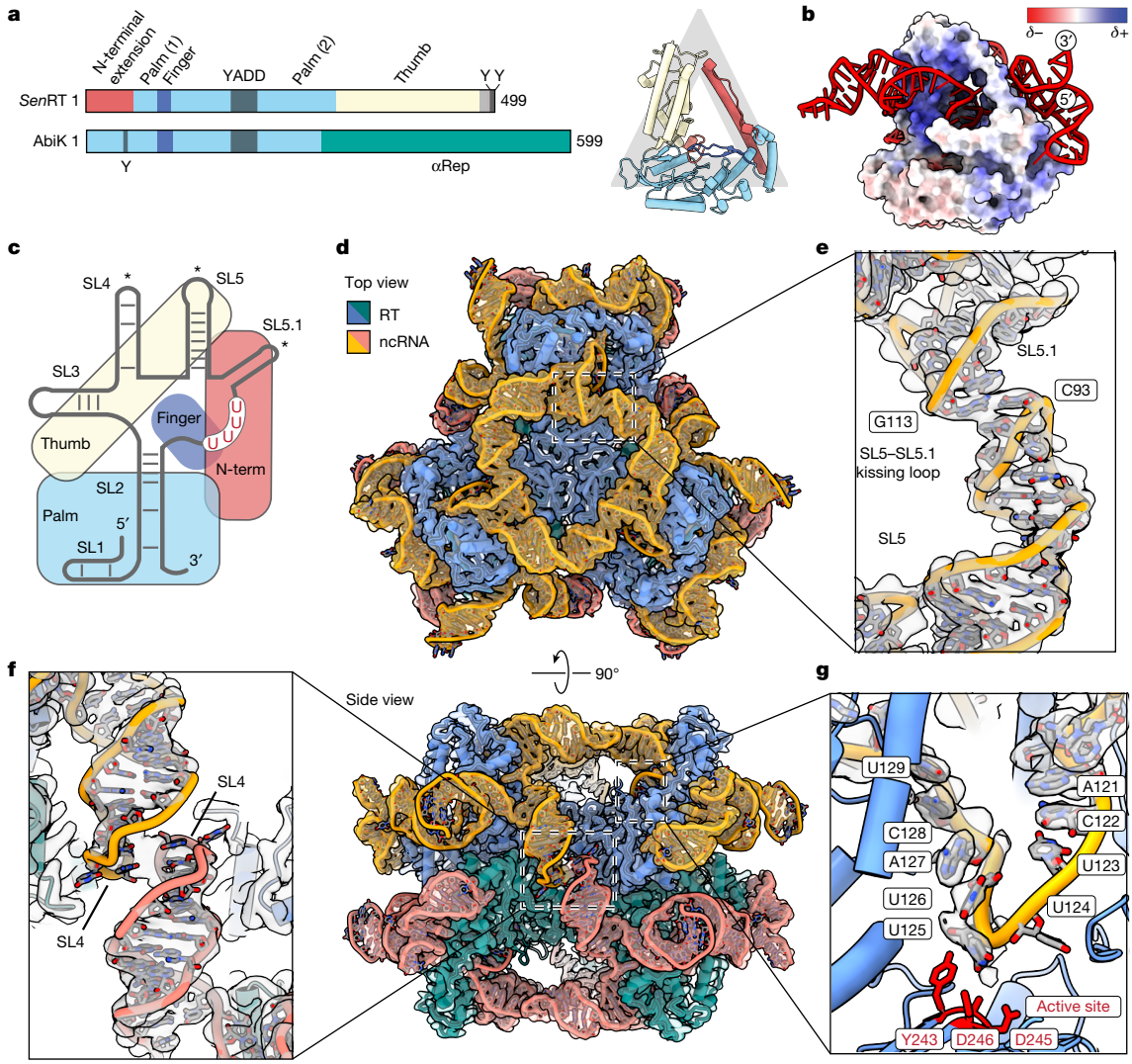


Fig. 4 | Cryo-EM structure of the hexameric *Sen*DRT9 RT–ncRNA complex. **a**, The domain architecture (left) of *Sen*DRT9 RT (PDB: 9NLX; this study) compared with AbiK (PDB: 7R06); the *Sen*RT has an N-terminal (N-term) extension (red) that reaches toward the thumb domain (right), creating a triangular architecture. **b**, Surface representation of a single RT monomer coloured by relative electrostatic potential; the 3′ end of the ncRNA (red ribbon) threads through the centre of the RT triangle. **c**, Cartoon representation of the ncRNA secondary structure observed in the cryo-EM density; features involved in intersubunit interactions (kissing loops) are highlighted with asterisks. The ncRNA interacts with each domain of the RT, represented as rectangles coloured as in **a**. **d**, A 2.6 Å semi-transparent map and the resulting model of the hexameric *Sen*DRT9 RT–ncRNA complex (PDB: 9NLV, this study). The top view looks down the three-fold symmetric axis of one trimer, and the side view reveals the back-to-back arrangement of two trimers. RTs and ncRNAs in the two trimers are coloured in blue/green and yellow/salmon, respectively. **e**, Semi-transparent density map and model of the kissing-loop interactions between flipped out bases in SL5 (C93) and SL5.1 (G113) of adjacent ncRNAs. **f**, Semi-transparent density map and model showing tetraloops in SL4. **g**, A magnified view of the templating nucleotides U123–U126 and the polymerase active site (red, YADD); density for the RT is hidden for simplicity.

for 'escaper' phages with mutations that confer resistance to DRT9 immunity. Using a well-established workflow to isolate escaper phages (Fig. 5a), we identified 41 independent T5 phage variants (11 unique genotypes) that evaded *Sen*DRT9, and 1 T5 phage variant and 13 independent Bas37 phage variants (4 unique genotypes) that evaded *Psa*DRT9. Phages that infected DRT9 strains similarly to an empty vector (EV) control were whole-genome sequenced to identify mutations linked to immune evasion (Extended Data Fig. 9a and Supplementary Table 3). Sequencing revealed diverse frameshift and missense mutations in *nrdA* (T5, Bas37) and *nrdB* (Bas37), which encode the large and small subunits of viral ribonucleotide reductase (RNR) complexes that catalyse ribonucleotide-to-deoxyribonucleotide conversion. This finding implicated nucleotide availability as an important facet of DRT9 immunity, as escaper mutations in *nrdA* and *nrdB* could limit deoxynucleotide levels in infected cells and decrease substrate availability for poly-dA synthesis. To examine how phage infection affects the nucleotide pool, either in the absence or presence of the DRT9 immune system, we performed metabolomics experiments with uninfected or T5-infected cells and found that phage T5 induced a sharp increase in adenosine nucleotide levels (Extended Data Fig. 9b). The presence of *Sen*DRT9 attenuated this increase (Extended Data Fig. 9b), consistent with the rapid consumption of dATP observed in our in vitro polymerization experiments (Fig. 3a). Notably, dATP levels in the context of DRT9 activation did not decrease below the levels observed in uninfected cells, arguing against the possibility that DRT9 activation might lead to abortive infection through nucleotide depletion. This was further corroborated by experiments showing that nucleoside supplementation had no discernible effect on DRT9 immunity or phage replication (Extended Data Fig. 9c), suggesting a more complex regulatory mechanism underlying DRT9 activation.

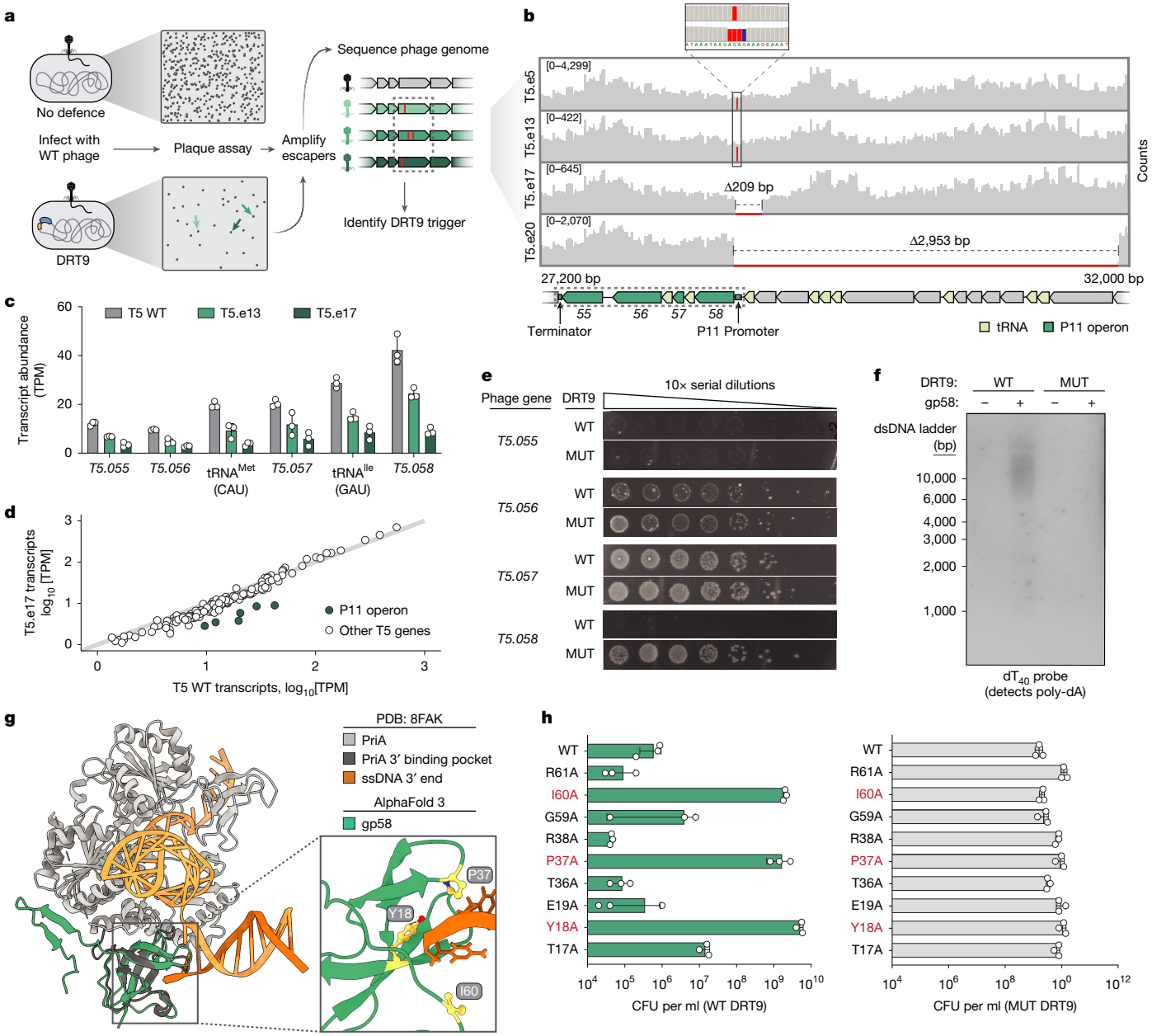


Fig. 5 | Identification of a viral trigger that activates *Sen*DRT9-mediated abortive infection. a, Schematic of the workflow to isolate and sequence phage variants that escape detection and/or elimination by a *Sen*DRT9 immune response. **b**, Representative coverage tracks from whole-genome sequencing of T5 escaper phage variants, shown above the corresponding genomic locus; all mutation classes shown perturb the putative P11 promoter controlling the expression of four protein-coding genes (green). **c**, Quantification of P11 operon transcript levels during infection of *Sen*DRT9-expressing cells with WT or escaper T5 phage variants, measured using RNA-seq and quantified as transcripts per million (TPM). **d**, Comparison of transcriptome-wide expression levels between WT T5 and T5.e17 escaper phage during infection of *Sen*DRT9-expressing cells, quantified as described in **c**. P11 operon transcripts are coloured in green. **e**, Colony-formation assays to evaluate cell viability after co-expression of WT or RT-inactive *Sen*DRT9 with candidate phage genes regulated by the P11 promoter. **f**, Southern blot analysis of total DNA isolated from cells expressing WT or MUT *Sen*DRT9 with or without gp58 induction. DNA was probed with oligo-dT_{40} to detect poly-dA species; sizes from a dsDNA ladder are marked. Representative data are shown from experiments repeated at least twice with similar results. **g**, The predicted AlphaFold 3 structure of T5 gp58 (green) superimposed onto the structure of *E. coli* PriA (grey) bound to a replication fork DNA substrate (orange; PDB: 8FAK). The magnified inset highlights gp58 residues predicted to bind to DNA 3′ ends[23]. Root mean squared deviation = 0.95 Å over 47 Cα atoms. **h**, Quantification of cell viability in colony-forming units (CFU) after co-expression of the indicated gp58 alanine substitution variants with WT (left) or MUT (right) *Sen*DRT9; mutations predicted to affect DNA 3′ end recognition are indicated with red text. Data are mean ± s.d. of $n = 3$ technical replicates.

A larger set of T5 escaper mutations clustered within a well-defined genomic region and consisted primarily of deletions spanning 200–15,000 bp (Supplementary Table 3). This tRNA-rich region, known as del-1, is a hotspot for large genomic deletions due to repeat-associated recombination[20,21]. However, the identification of additional escaper genotypes containing only intergenic point mutations within this region redirected our attention to the P11 promoter, which drives expression of a six-gene operon that includes two tRNAs (Fig. 5b). RNA-seq analysis revealed specific downregulation of this operon during infection with escaper phages bearing P11 promoter mutations, as compared to WT T5 (Fig. 5c,d). We therefore hypothesized that one or more genes within this operon might activate the DRT9 immune response, and we

cloned expression plasmids for each of the four protein-coding genes to test the effects of DRT9 and phage gene co-expression. Notably, a single gene, *T5.058*, prevented cell growth when co-expressed with WT *Sen*DRT9 but not the RT-inactive mutant (Fig. 5e). This was supported by Southern blotting analysis, which showed accumulation of poly-dA products in cells co-expressing *T5.058* with WT *Sen*DRT9 (Fig. 5f and Extended Data Fig. 9d), albeit at lower levels than those induced by phage infection. These results strongly implicate gp58, the *T5.058* gene product, as a necessary and sufficient T5 phage factor for triggering DRT9-mediated abortive infection.

We predicted the structure of gp58, an uncharacterized protein, and identified a high degree of structural similarity to the 3′-DNA-binding domain of PriA (Fig. 5g), a primosomal protein involved in DNA replication restart at stalled or damaged replication forks[22]. Co-immunoprecipitation (co-IP) experiments revealed an interaction between the RT and gp58 in cells expressing the WT *Sen*DRT9 system, but not the RT-inactive mutant (Extended Data Fig. 9e), implicating gp58 in RT- and/or poly-dA-dependent binding. To further examine the role of gp58, we used its structure prediction, alongside previous studies of DNA binding by PriA[23], to design gp58 variants predicted to disrupt DNA interactions. Mutating gp58 in this manner restored normal growth in cells co-expressing gp58 and *Sen*DRT9, indicating failed triggering of DRT9-mediated cell death (Fig. 5h). Thus, we conclude that gp58 probably recognizes the DNA homopolymer synthesized by *Sen*DRT9 through direct binding to its 3′ end.

Our finding that DRT9 activity is triggered by a phage protein predicted to bind DNA 3′ ends dovetailed with a separate insight that arose from literature review. A comprehensive study of *E. coli* gene essentiality previously found that *sbcB*, which encodes the 3′-to-5′ exodeoxyribonuclease I (ExoI), is an essential gene in strains expressing UG17 (an RT system related to DRT9), suggesting that ExoI mitigates the otherwise toxic effects of UG17[24]. Notably, we were unable to obtain transformants of a *ΔsbcB* strain with WT *Sen*DRT9, whereas the RT-inactive mutant was well tolerated (Fig. 6a), revealing a similar conditional essentiality relationship. This finding suggested a model in which constitutive poly-dA degradation by ExoI in uninfected cells is out-competed by poly-dA 3′-end binding by gp58 in T5-infected cells, leading to homopolymer accumulation and abortive infection. Importantly, this mechanism would resolve the apparent discrepancy between our in vivo observation that poly-dA synthesis occurs only after phage infection, and our in vitro observation that poly-dA synthesis activity occurs constitutively. In excellent agreement with this model, DRT9-derived poly-dA and poly-dT products were efficiently degraded by recombinant ExoI in vitro, but were protected from degradation in the presence of gp58 (Fig. 6b and Extended Data Fig. 9f). Notably, gp58 did not protect against endonuclease activity (Fig. 6b), further supporting its role in cDNA 3′ end protection.

Finally, we performed co-IP proteomics experiments using the Flag-tagged RT in the absence or presence of T5 infection to identify additional viral or host factors that are involved in DRT9 immunity (Fig. 6c). As expected, gp58 was significantly enriched in cells expressing WT *Sen*DRT9 but not the RT-inactive mutant, corroborating our earlier findings based on western blotting. Additional T5 proteins implicated in DNA binding and genome replication were also uniquely enriched by the WT RT, including T5 single-stranded DNA binding protein (SSB) (Fig. 6d,e). Notably, immunoprecipitation of the WT RT also strongly enriched two host-encoded nucleoid-associated proteins, H-NS and StpA (Fig. 6d,e). Both proteins preferentially bind to AT-rich DNA regions and are key regulators of transcription and nucleoid architecture[25], providing an elegant link between DRT9-catalysed cDNA synthesis and associated host factors.

Discussion

Our study reveals a mechanism of antiviral immunity mediated by DRT9 defence systems, in which a ncRNA templates the synthesis of long, kilobase-length homopolymers of deoxyadenylate to drive abortive infection (Fig. 6f). Biochemical and structural data demonstrate that a hexameric RT–ncRNA complex polymerizes dATP using a protein primer and poly-uridine template. Genetic screening revealed a phage-encoded DNA-binding protein, gp58, that drives poly-dA accumulation in vivo by blocking its degradation by a host exonuclease, ExoI. Importantly, our finding that gp58 alone is sufficient to induce cellular toxicity in a DRT9-expressing strain argues against poly-dA synthesis acting to specifically antagonize phage infection. Instead, we propose that poly-dA accumulation drives a phage-extrinsic mechanism of abortive infection.

The mechanism by which poly-dA accumulation leads to cell death remains unclear. In vitro reverse-transcription reactions with *Sen*DRT9 rapidly consumed dATP, leading us to initially suspect that nucleotide depletion might inhibit both host and phage genome replication by starving the cell of adenosine nucleotides. If true, this mechanism would provide an elegant counterpart to other phage defence systems that similarly deplete deoxyribonucleotides[8,9]. However, while nucleotide metabolomics measurements in infected cells revealed that dATP levels were lower in the presence of DRT9, supplementation of deoxyadenosine nucleosides did not suppress the immune response (Extended Data Fig. 9b,c). An alternative hypothesis is that poly-dA accumulation in cells leads to the sequestration of specific DNA-binding proteins, as suggested by our co-IP experiments (Fig. 6d,e). However, previous studies of strains mutated in *hns* and/or *stpA* do not indicate that their dysfunction would be lethal to the cell[26,27], suggesting that additional factors may be involved. A third hypothesis arises when considering that DRT9 systems sometimes associate with SLATT-domain-containing proteins (Extended Data Fig. 1a), which contain transmembrane helices predicted to form toxic, membrane-embedded pores[28]. Considering recent reports of other phage defence systems that use oligonucleotides to modulate ion-channel activity[29,30], nucleic-acid-mediated regulation of membrane proteins may be a broader theme in antiviral immunity. However, whether *Sen*DRT9 poly-dA products similarly engage membrane proteins in *E. coli* requires future investigation. Resolving the abortive-infection mechanism of poly-dA synthesis will be a focus of future effort and will require pursuing these, and potentially additional, molecular hypotheses.

The unusual mechanism of RNA-templated homopolymer synthesis reported here sets DRT9 RT enzymes apart from all other known polymerases. Although nucleotidyltransferase enzymes that polymerize RNA substrates—such as poly(A) and poly(U) polymerase—also produce nucleic acid homopolymers, they do so without a template, relying instead on protein-mediated nucleotide selection within the active site[31]. Similarly, terminal deoxynucleotidyl transferase can add homopolymeric stretches of DNA when a given nucleotide is in excess, but lacks the specificity conferred by a nucleic acid template[32]. DRT9 enzymes are, to our knowledge, the only polymerases that harness an RNA template to direct the synthesis of DNA homopolymers, yielding kilobase-length products in a highly processive manner. Notably, these homopolymers remain covalently tethered to the enzyme itself through a protein-priming mechanism dependent on conserved tyrosine residues.

The structural architecture of DRT9 RT–ncRNA complexes is equally distinctive, especially in comparison to DRT2-family enzymes. The DRT2 RT–ncRNA complex behaves as a functional monomer[5], whereas the DRT9 complex adopts a hexameric architecture comprising a dimer of trimers. This arrangement echoes the hexameric and trimeric assemblies of two other UG-family RTs, AbiK and Abi-P2 (ref. 16), but is uniquely stabilized by RNA-mediated interactions that form a tightly interlinked cage to encapsulate the enzyme subunits. The proximity of the U-rich template region to the catalytic pocket provides a structural framework for understanding the mechanism of poly-dA synthesis. Indeed, the proposed mechanism by which DRT9 RTs iteratively engage the same

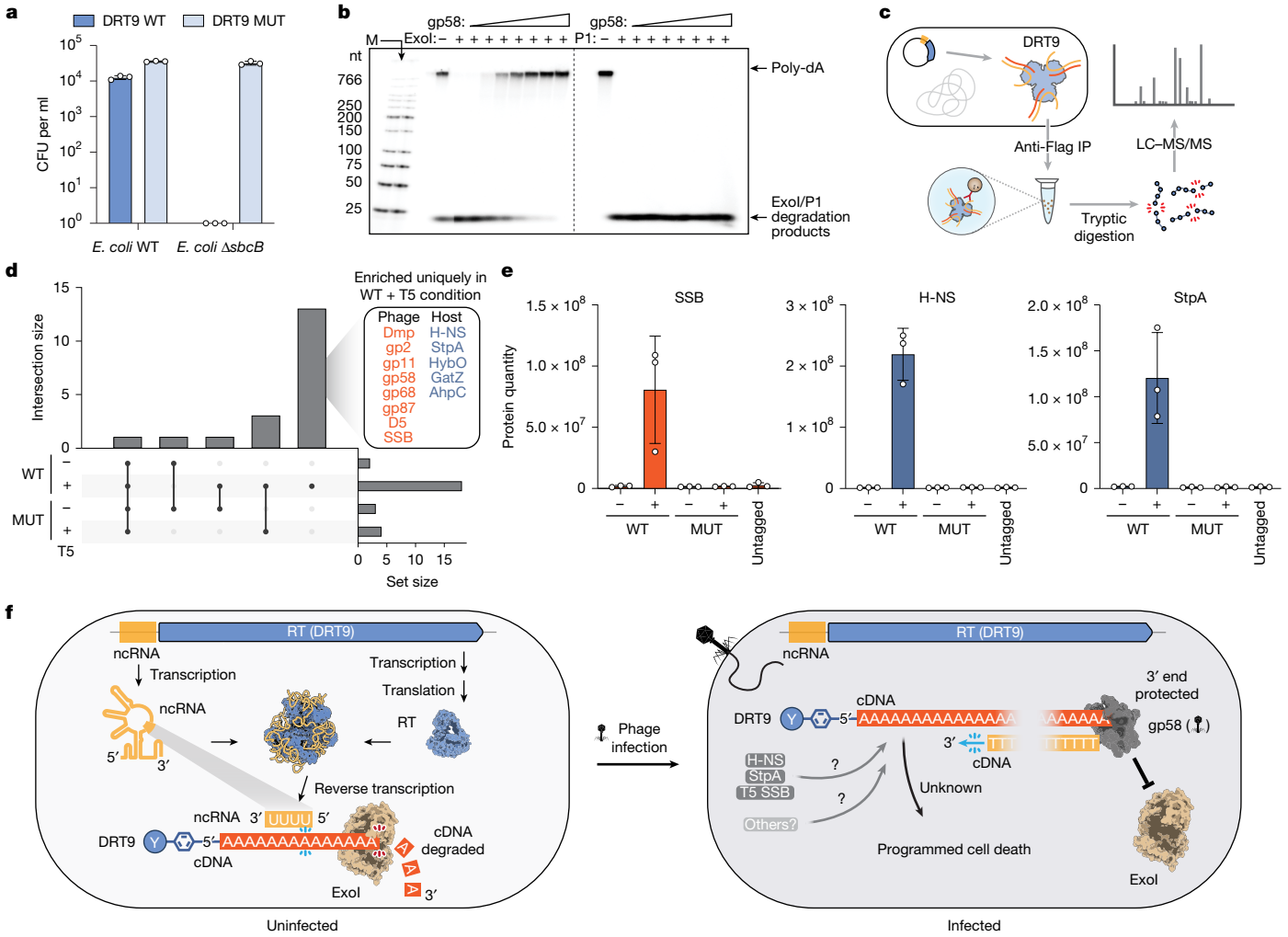


Fig. 6 | Host regulation of and response to poly-dA synthesis by DRT9 immune systems. a, Quantification of cell viability in CFU after transformation of WT or *ΔsbcB E. coli* with WT or RT-inactive (MUT) *Sen*DRT9, demonstrating that the *sbcB* gene product (ExoI) neutralizes the otherwise toxic properties of DRT9 synthesis products. Data are mean ± s.d. of $n = 3$ technical replicates. **b**, Denaturing 5% urea–PAGE analysis of DNA polymerization assays after the reactions were treated with ExoI (left) or nuclease P1 (right) in the presence of increasing concentrations of T5 gp58. The reactions were incubated at 37 °C for 10 min before addition of gp58 and ExoI or nuclease P1. Representative data are shown for experiments repeated at least twice with similar results. **c**, Schematic of co-IP–MS experiments to identify protein interactors of *Sen*DRT9 and its DNA synthesis products. **d**, UpSet plot showing the number of overlapping and unique protein interactors identified from co-IP–MS experiments with WT or MUT *Sen*DRT9 and with or without T5 phage infection; significantly enriched proteins were identified as those exhibiting greater than 20-fold enrichment with FDR < 0.05. Phage proteins (orange) and host proteins (blue) enriched uniquely in the WT + T5 condition are labelled. **e**, Normalized protein intensity values for T5 SSB and *E. coli* H-NS and StpA across the indicated co-IP–MS experiments from **d**, including an untagged control. Data are mean ± s.d. of $n = 3$ independent biological replicates. **f**, Model for the antiphage defence mechanism of DRT9 systems. The RT–ncRNA hexamer constitutively synthesizes poly-dA using a U-rich template within the ncRNA. Host ExoI degrades poly-dA in uninfected cells, but phage infection leads to expression of a trigger protein (gp58 in T5) that competes for 3′-end binding, leading to poly-dA accumulation along with partial second-strand synthesis of poly-dT. Additional viral and host factors (SSB, H-NS and StpA) bind to poly-dA and/or poly-dT products, leading to programmed cell death.

template region for repetitive DNA synthesis is highly reminiscent of repeat addition processivity, a hallmark feature of telomerase reverse transcription activity[33]. Future studies will be needed to determine the precise number of uridines required for templated deoxyadenosine incorporation, how the nascent cDNA is realigned within the active site and, importantly, how kilobase-sized homopolymers extrude from the body of the hexameric complex while remaining covalently tethered at the 5′ end and engaged within the active site at the 3′ end.

As with DRT2, insights into DRT9 arose from questioning conventional assumptions about RT enzymes and their cDNA synthesis products. In both cases, a departure from standard approaches in high-throughput sequencing data analysis enabled us to identify unexpected cDNA products. Notably, our initial interest in DRT9 stemmed from the hypothesis that its properties would closely resemble DRT2, given their close phylogenetic relationship, but our findings here illuminate a markedly different mechanism of antiviral immunity. As such, we anticipate that further exploration of RT diversity will continue to uncover unexpected DNA synthesis mechanisms, and that the full scope of nucleic acid functions in antiviral defence is only beginning to come into view.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-09179-5.

1. Georjon, H. & Bernheim, A. The highly diverse antiphage defence systems of bacteria. *Nat. Rev. Microbiol.* **21**, 686–700 (2023).
2. Gao, L. et al. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077–1084 (2020).
3. Mestre, M. R. et al. UG/Abi: a highly diverse family of prokaryotic reverse transcriptases associated with defense functions. *Nucleic Acids Res.* **50**, 6084–6101 (2022).
4. Tang, S. et al. De novo gene synthesis by an antiviral reverse transcriptase. *Science* **386**, eadq0876 (2024).
5. Wilkinson, M. E., Li, D., Gao, A., Macrae, R. K. & Zhang, F. Phage-triggered reverse transcription assembles a toxic repetitive gene from a noncoding RNA. *Science* **386**, eadq3977 (2024).
6. Duncan-Lowey, B. & Kranzusch, P. J. CBASS phage defense and evolution of antiviral nucleotide signaling. *Curr. Opin. Immunol.* **74**, 156–163 (2022).
7. Bernheim, A. et al. Prokaryotic viperins produce diverse antiviral molecules. *Nature* **589**, 120–124 (2021).
8. Tal, N. et al. Bacteria deplete deoxynucleotides to defend against bacteriophage infection. *Nat. Microbiol.* **7**, 1200–1209 (2022).
9. Hsueh, B. Y. et al. Phage defence by deaminase-mediated depletion of deoxynucleotides in bacteria. *Nat. Microbiol.* **7**, 1210–1220 (2022).
10. Silas, S. et al. Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase–Cas1 fusion protein. *Science* **351**, aad4234 (2016).
11. Millman, A. et al. Bacterial retrons function in anti-phage defense. *Cell* **183**, 1551–1561 (2020).
12. Bobonis, J. et al. Bacterial retrons encode phage-defending tripartite toxin–antitoxin systems. *Nature* **609**, 144–150 (2022).
13. Carabias, A. et al. Retron-Eco1 assembles NAD^+-hydrolyzing filaments that provide immunity against bacteriophages. *Mol. Cell* **84**, 2185–2202 (2024).
14. Wang, C. et al. A reverse transcriptase-related protein mediates phage resistance and polymerizes untemplated DNA in vitro. *Nucleic Acids Res.* **39**, 7620–7629 (2011).
15. Gapińska, M. et al. Structure-functional characterization of *Lactococcus* AbiA phage defense system. *Nucleic Acids Res.* **52**, 4723–4738 (2024).
16. Figiel, M. et al. Mechanism of protein-primed template-independent DNA synthesis by Abi polymerases. *Nucleic Acids Res.* **50**, 10026–10040 (2022).
17. González-Delgado, A., Mestre, M. R., Martínez-Abarca, F. & Toro, N. Prokaryotic reverse transcriptases: from retroelements to specialized defense systems. *FEMS Microbiol. Rev.* **45**, fuab025 (2021).
18. Zoulim, F. & Seeger, C. Reverse transcription in hepatitis B viruses is primed by a tyrosine residue of the polymerase. *J. Virol.* **68**, 6–13 (1994).
19. Yushenova, I. A. & Arkhipova, I. R. Biochemical properties of bacterial reverse transcriptase-related (rvt) gene products: multimerization, protein priming, and nucleotide preference. *Curr. Genet.* **64**, 1287–1301 (2018).
20. Wang, J. et al. Complete genome sequence of bacteriophage T5. *Virology* **332**, 45–65 (2005).
21. Glukhov, A. S., Krutilina, A. I., Kaliman, A. V., Shlyapnikov, M. G. & Ksenzenko, V. N. Bacteriophage T5 mutants carrying deletions in tRNA gene region. *Mol. Biol.* **52**, 3–9 (2018).
22. Marians, K. J. PriA: at the crossroads of DNA replication and recombination. *Prog. Nucleic Acid Res. Mol. Biol.* **63**, 39–67 (1999).
23. Sasaki, K. et al. Crystallization and preliminary crystallographic analysis of the N-terminal domain of PriA from *Escherichia coli*. *Biochim. Biophys. Acta* **1764**, 157–160 (2006).
24. Rousset, F. et al. The impact of genetic diversity on gene essentiality within the *Escherichia coli* species. *Nat. Microbiol.* **6**, 301–312 (2021).
25. Dillon, S. C. & Dorman, C. J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* **8**, 185–195 (2010).
26. Johansson, J. & Uhlin, B. E. Differential protease-mediated turnover of H-NS and StpA revealed by a mutation altering protein stability and stationary-phase survival of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **96**, 10776–10781 (1999).
27. Srinivasan, R., Scolari, V. F., Lagomarsino, M. C. & Seshasayee, A. S. N. The genome-scale interplay amongst xenogene silencing, stress response and chromosome architecture in Escherichia coli. *Nucleic Acids Res.* **43**, 295–308 (2015).
28. Burroughs, A. M., Zhang, D., Schäffer, D. E., Iyer, L. M. & Aravind, L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res.* **43**, 10633–10654 (2015).
29. Tan, J. M. J. et al. A DNA-gated molecular guard controls bacterial Hailong anti-phage defence. *Nature* https://doi.org/10.1038/s41586-025-09058-z (2025).
30. Tak, U., Walth, P. & Whiteley, A. T. Bacterial cGAS-like enzymes produce 2′,3′-cGAMP to activate an ion channel that restricts phage replication. Preprint at *bioRxiv* https://doi.org/10.1101/2023.07.24.550367 (2023).
31. Martin, G., Doublié, S. & Keller, W. Determinants of substrate specificity in RNA-dependent nucleotidyl transferases. *Biochim. Biophys. Acta* **1779**, 206–216 (2008).
32. Motea, E. A. & Berdis, A. J. Terminal deoxynucleotidyl transferase: the story of a misguided DNA polymerase. *Biochim. Biophys. Acta* **1804**, 1151–1166 (2010).
33. Lai, C. K., Miller, M. C. & Collins, K. Roles for RNA in telomerase nucleotide and repeat addition processivity. *Mol. Cell* **11**, 1673–1683 (2003).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025

Methods

Plasmid and *E. coli* strain construction

All strains and plasmids used in this study are described in Supplementary Tables 4 and 5, respectively. DRT9 operons, including their native upstream and downstream flanking sequences, were chemically synthesized (Twist Bioscience or GenScript) and cloned into a pACYC184 plasmid backbone by Gibson Assembly. Derivative plasmids were cloned using a variety of methods, including around-the-horn PCR, restriction digestion–ligation and Golden Gate assembly. All plasmids were cloned and propagated in *E. coli* strain NEB Turbo (sSL0410). Clones were verified by Sanger sequencing or whole-plasmid sequencing. Substitution and insertion mutations to the *Sen*DRT9 ncRNA are numbered relative to the first nucleotide of the ncRNA, with insertion numbering referencing the nucleotide immediately upstream of the inserted bases (for example, A109U indicates mutation of A at position 109 to U, and 124insU indicates insertion of U downstream of position 124). Experiments were performed in *E. coli* K-12 strain MG1655 unless otherwise indicated. The Δ*sbcB* strain from the Keio collection[34], which contains a single-gene deletion in *E. coli* K-12 strain BW25113, was a gift from S. Tavazoie.

Phylogenetic analyses

The phylogenetic tree of UG RT homologues in Fig. 1a was visualized using data from a previous study[3]. For visualization, each UG/DRT clade was collapsed into a single representative, using the longest branch within each clade. To build the phylogenetic tree of DRT9 RT homologues in Extended Data Fig. 1a, a BLASTp search of a local copy of the NCBI NR database (downloaded 4 April 2023) was queried with 124 DRT9 (that is, UG28) sequences identified previously[3] (-evalue 0.01 -max_target_seqs 1,000,000), resulting in the identification of 438 unique DRT9 homologues. These proteins were aligned with MAFFT[35] (LINSI option) and a phylogenetic tree was constructed from the resulting alignment with FastTree (-wag -gamma options)[36]. Genomes encoding these DRT9 homologues were then downloaded using NCBI's Batch-Entrez function, and loci comprising *drt9* plus 10 kb of upstream and downstream nucleotide sequence were extracted. Next, to identify DRT9 systems that encode a SLATT-domain-containing gene, we predicted open reading frames (ORFs) in each DRT9 locus using Eggnogg Mapper[37] and used HMMER[38] to search the resulting ORFs with a previously built hidden Markov model (PFAM: PF18160). Finally, ncRNA associations were annotated using the CMsearch function of Infernal[39] to search DRT9 loci with a CM built from DRT9-associated ncRNAs described below.

Sequence identity matrices

Sequence identity between selected DRT9 homologues was analysed by performing MAFFT alignment of RT amino acid sequences using the default settings. Accession numbers for RT proteins are listed in Supplementary Table 2.

Phage propagation and plaque assays

Phages T2, T5 and λ-vir were gifts from M. Laub. BASEL collection phages were gifts from A. Harms. All phages were propagated in liquid culture by picking single plaques into LB medium, supplemented with 5 mM $CaCl_2$ and 5 mM $MgSO_4$, containing *E. coli* MG1655 cells diluted 1:100 from overnight cultures. After 4–5 h of incubation with shaking at 37 °C, chloroform was added to a final concentration of 5% to lyse the residual bacteria, and lysates were centrifuged at 4,000*g* for 10 min to pellet the cell debris. The supernatants were passed through a sterile 0.22-μm filter and stored at 4 °C.

Small-drop plaque assays were performed as previously described[4]. In brief, *E. coli* K-12 strain MG1655 (sSL0810) (strain descriptions and genotypes are provided in Supplementary Table 4) was transformed with the indicated plasmids (plasmid descriptions and sequences are provided in Supplementary Table 5). Transformants were inoculated in liquid LB medium containing the appropriate antibiotic and grown overnight at 37 °C with shaking. The next day, overnight cultures were mixed with molten soft agar (0.5% agar in LB medium supplemented with 5 mM $CaCl_2$ and 5 mM $MgSO_4$) at 45 °C and poured over solid-bottom agar (1.5% agar in LB medium containing the appropriate antibiotic) in a Petri dish. Then, 10× serial dilutions of phage in LB medium were spotted onto the surface of the soft agar lawn. Plates were incubated at 37 °C for 8–16 h to allow plaque formation. Plaque-forming units (PFU) per ml were calculated using the following formula: $\frac{\text{number of plaques}}{0.003\ \text{ml} \times \text{dilution factor}}$. When individual plaques were indistinguishable, a count of 50 plaques was assigned at the lowest dilution showing visible lawn clearance. Phage defence activity was assessed by calculating the fold reduction in EOP, which was determined by dividing the PFU per ml obtained on a lawn of EV-transformed control cells by the PFU per ml obtained on a lawn of defence-system-expressing cells.

Plaque- and colony-formation time course analysis

Phage replication and cell viability over the course of T5 infection were measured as previously described[4]. *Escherichia coli* MG1655 cells were transformed with plasmids encoding WT or MUT *Sen*DRT9 and grown to an optical density at 600 nm (OD_{600}) of 0.4 in LB medium supplemented with chloramphenicol (25 μg ml^{-1}). For the $t = 0$ timepoint, an aliquot of T5 phage lysate was taken for plaque counting, and an aliquot of uninfected culture was taken for colony counting. Infections were initiated by adding phage at an MOI of 5. The cultures were incubated with shaking at 37 °C for 90 min, and aliquots of each infection were taken every 30 min for plaque and colony counting.

For plaque-forming unit (PFU) measurements, 100 μl of each culture was taken at the indicated timepoint and mixed with chloroform (4% final concentration) to lyse cells and terminate infections. Lysates were centrifuged at 4,000*g* for 3 min to pellet cell debris. Supernatants containing phages were serially diluted in LB, and the PFU per ml was measured using the plaque assay protocol described above.

For CFU measurements, 200 μl of each culture was collected at the indicated timepoint and centrifuged at 4,000*g* for 3 min. The pellets were washed in 1 ml LB and centrifuged again to remove residual phage particles, and then resuspended in fresh LB. Serial dilutions of cells were prepared and 5 μl of each dilution was spotted onto an LB-agar plate supplemented with chloramphenicol (25 μg ml^{-1}). The plates were incubated overnight at 37 °C and colonies were counted the next day. The CFU per ml was calculated using the following formula: $\frac{\text{number of colonies}}{0.005\ \text{ml} \times \text{dilution factor}}$.

AlphaFold structural modelling

Protein sequences were submitted to the AlphaFold 3 webserver[40] and the resulting structural predictions were rendered using ChimeraX[41].

RIP–seq and cDIP–seq

RIP–seq and cDIP–seq were performed as previously described[4]. *Escherichia coli* K-12 strain MG1655 (sSL0810) cells were transformed with plasmids encoding N-terminally 3×Flag-tagged DRT9 systems with a WT or mutant (YAAA) RT active site (plasmid sequences are provided in Supplementary Table 5). Individual colonies for each replicate experiment were inoculated in 40 ml liquid LB and grown at 37 °C to an OD_{600} of 0.3–0.4. For experiments performed in the absence or presence of phage infection, cultures were split in half and phage T5 was added to one half at a MOI of 5. Cultures were further incubated at 37 °C with shaking for 45 min before collection. For experiments performed only in uninfected cells, 20 ml cultures were grown to OD_{600} of 0.5 and collected. Cells were collected by centrifugation at 4,000*g* for 10 min at 4 °C and then washed with 5 ml of cold TBS (20 mM Tris-HCl pH 7.5 at 25 °C, 150 mM NaCl). Cells were centrifuged again at 4,000*g* for 5 min at 4 °C. The pellets were washed with 1 ml of cold TBS and centrifuged at 10,000*g* for 5 min at 4 °C. After the supernatants were removed, pellets were flash-frozen in liquid nitrogen and stored at −80 °C.

To prepare antibody–bead complexes for immunoprecipitation, protein G Dynabeads (Thermo Fisher Scientific) were washed three times in 1 ml IP lysis buffer (20 mM Tris-HCl pH 7.5 at 25 °C, 150 mM KCl, 1 mM $MgCl_2$, 0.2% Triton X-100), resuspended in 1 ml IP lysis buffer, combined with anti-Flag antibody (Sigma-Aldrich, F3165) and rotated for more than 3 h at 4 °C. Then, 60 µl of beads and 20 µl of antibody mix were prepared per sample. Antibody–bead complexes were washed twice to remove unbound antibodies and resuspended in IP lysis buffer to a final volume of 60 µl per sample.

Flash-frozen pellets were thawed on ice and resuspended in 1.2 ml IP lysis buffer supplemented with 1× Complete Protease Inhibitor Cocktail (Roche) and 0.1 U μl^{-1} SUPERase•In RNase Inhibitor (Thermo Fisher Scientific). Cells were lysed by sonication and lysates were cleared by centrifugation at 21,000*g* for 15 min at 4 °C. The supernatants were transferred to new tubes and two 10 µl aliquots of each sample (one for RIP–seq and one for cDIP–seq) were set aside and stored at −80 °C as input controls. The remainder of each sample was combined with 60 µl of antibody–bead complex and rotated overnight at 4 °C. The next day, each sample was washed three times with 1 ml ice-cold IP wash buffer (20 mM Tris-HCl pH 7.5 at 25 °C, 150 mM KCl, 1 mM $MgCl_2$). During the final wash, each sample was split into two 500 µl volumes for downstream RIP or cDIP processing.

RIP elution was performed by removing the supernatants and resuspending the beads in 750 µl TRIzol (Thermo Fisher Scientific). After 5 min incubation at room temperature, the supernatants containing eluted RNA were transferred to new tubes. The samples were then mixed with 150 µl chloroform, incubated at room temperature for 2 min and centrifuged at 12,000*g* for 15 min at 4 °C. RNA was isolated from the upper aqueous phase using the RNA Clean & Concentrator-5 kit (Zymo Research). RNA from input samples was isolated in the same manner using TRIzol and column purification. cDIP elution was performed by removing supernatants, resuspending beads in 90 µl IP wash buffer, and treating the samples with 5 µg RNase A (Thermo Fisher Scientific) and 25 µg proteinase K (Thermo Fisher Scientific). The supernatants containing eluted DNA were transferred to new tubes. DNA was isolated using the Monarch Spin PCR and DNA Cleanup kit (NEB), according to the Oligonucleotide Cleanup protocol. Input samples were also treated with RNase A and proteinase K before DNA isolation by column purification.

For RIP–seq library preparation, RNA was diluted in NEBuffer 2 and heated at 92 °C for 2 min for fragmentation by alkaline hydrolysis. The samples were then treated with TURBO DNase (Thermo Fisher Scientific), RppH (NEB) and T4 PNK (NEB) to remove DNA and prepare RNA ends for adapter ligation. RNA was purified using the Zymo RNA Clean & Concentrator-5 kit. Adapter ligation, reverse transcription and indexing PCR were performed using the NEBNext Small RNA Library Prep kit. Libraries were sequenced on an Element AVITI system in paired-end mode with 75 or 150 cycles per end.

For cDIP–seq library preparation, the samples were heated at 95 °C for 2 min and then immediately placed on ice to denature DNA. Adapter ligation and conversion of ssDNA to dsDNA were performed using the xGen ssDNA & Low-Input DNA Library Prep Kit (IDT). Indexing PCR was performed using NEBNext Ultra II Q5 Master Mix. Libraries were sequenced on the Element AVITI system in paired-end mode with 75 or 150 cycles per end.

RIP–seq analysis

RIP–seq datasets were processed as previously described[4]. Cutadapt[42] (v.4.2) was used to remove adapter sequences, trim low-quality ends from reads and filter out reads shorter than 15 bp. Reads were mapped to reference files containing the MG1655 genome (NCBI: NC_000913.3), relevant plasmid sequence and T5 (NCBI: NC_005859.1) genome using bwa-mem2[43] (v2.2.1) with the default parameters. SAMtools[44] (v.1.17) was used to sort and index alignments. Coverage tracks for top- and bottom-strand alignments were generated using bamCoverage[45] (v.3.5.1) with normalization for sequencing depth (based on the total number of reads passing initial trimming and length filtering). Coverage tracks were visualized in IGV[46].

For transcriptome-wide analysis of RNAs enriched by RIP–seq, aligned reads were assigned to annotated transcriptome features using featureCounts[47] (v.2.0.2) with -s 1 for strandedness. Counts matrices were analysed using DESeq2[48] to calculate the enrichment (fold-change) and false-discovery rate for each transcript compared between input and IP samples. Transcriptome-wide comparisons were visualized as MA plots, using ggplot2 to plot baseMean (mean-normalized counts across all conditions) against $\log_2$[enrichment]. All comparisons included three independent biological replicates.

Total RNA-seq

Escherichia coli MG1655 cells expressing WT *Sen*DRT9 were grown in liquid culture until reaching an OD_{600} of 0.4. Cells were infected with WT phage T5 or the indicated escaper phage (T5.e13 or T5.e17; Supplementary Table 3) at MOI = 5. The samples were collected after 10 min of infection by centrifuging 1 ml of culture at 4,000*g* for 3 min, removing the supernatant and freezing the pellet in liquid nitrogen. To extract RNA, the pellet was resuspended in 750 µl of TRIzol, incubated at room temperature for 5 min and mixed with 150 µl chloroform. After incubation at room temperature for 2 min, the samples were centrifuged at 12,000*g* for 15 min at 4 °C. RNA was isolated from the upper aqueous phase using the RNA Clean & Concentrator-5 kit (Zymo Research). Library preparation, sequencing and read processing were performed as described above for RIP–seq. After removing low-abundance transcripts (mean counts < 10), gene expression was quantified by dividing counts by gene length and scaling to 1 million total counts per sample, resulting in TPM expression values.

cDIP–seq analysis

Adapter trimming, quality trimming and length filtering of cDIP–seq and corresponding input (total DNA) datasets were performed as described above for RIP–seq experiments. Trimmed and filtered reads were mapped to combined reference files, sorted, indexed and plotted onto coverage tracks as described above. For transcriptome-wide analysis of cDIP–seq data, alignments over annotated transcriptome features were counted using featureCounts with -s 2 for strandedness. Count matrices were processed by DESeq2 and plotted as described above. The *lacZ* and *racR* loci were masked from this analysis, as they were previously found to show artifactual enrichment by cDIP–seq[4]. All transcriptome-wide comparisons were performed using three independent biological replicates.

Unmapped read analysis

For analysis of unmapped reads in cDIP–seq and total RNA-seq datasets (using the input controls from RIP–seq experiments), reads were first stringently trimmed and filtered to remove all residual adapter sequences. Cutadapt was used to process reads, adapting a previously described method for tiling across the 3′ adapter[49], with error, overlap, quality and length thresholds set at -e 0.2 -O 10 -q 30 -m 45. The fraction of unmapped reads for each sample was calculated using SAMtools flagstat. Unmapped reads were then extracted for downstream analysis using SAMtools fasta. Motif analysis was performed using MEME[50] (v.5.5.7) in differential enrichment mode with minimum width 45 (-minw 45) and allowing any number of motif repetitions (-mod anr), using unmapped reads from cDIP–seq of the catalytically inactive RT (YAAA) as the control sequence set. Quantification of homopolymer-containing reads was performed using countPattern from the Biostrings package (v.2.70.3) in R. Reads were analysed for the presence of at least 25 consecutive A, T, G or C bases, or for the presence of $A_{10}T_{10}$ or $T_{10}A_{10}$ sequences, allowing 1 mismatch. Reads containing homopolymers were counted and normalized according to the total number of filtered reads for each sample.

Oligonucleotide spike-in sequencing

To assess potential biases in library preparation and sequencing of poly-dA and poly-dT, 100-nucleotide single-stranded DNA homopolymers (oligo-dA and oligo-dT) were chemically synthesized (IDT) and analysed by next-generation sequencing. Then, 250 pmol oligo-dA and oligo-dT were each 5′-phosphorylated by treatment with T4 PNK (NEB) for 30 min at 37 °C. Reactions were cleaned up using the Monarch Spin PCR and DNA Cleanup Kit (NEB) and eluted in 20 µl water. Phosphorylated oligos were added to about 5 pmol of total DNA, extracted from *E. coli* K-12 strain MG1655 cells expressing catalytically inactive *Sen*DRT9, in various amounts (50 fmol oligo-dA only, 50 fmol oligo-dT only, 50 fmol each of oligo-dA and oligo-dT, or 500 fmol of oligo-dA and 50 fmol of oligo-dT). Spiked-in DNA samples were subjected to the same library preparation and sequencing workflow as described above for cDIP–seq.

Southern blotting

DNA samples were collected from either uninfected or phage T5 infected cells. Cultures from a single colony were grown in 20 ml liquid culture until reaching an OD_{600} of 0.3. For experiments performed in presence of phage infection, phage T5 was added at a MOI of 5. For uninfected cells, an equal volume of LB was added. Cultures were further incubated at 37 °C with shaking for 60 min before collection. For experiments containing the gp58 trigger, cells were grown in 20 ml LB supplemented with chloramphenicol (25 µg ml^{-1}), spectinomycin (100 µg ml^{-1}) and 2% glucose until they reached an OD_{600} of 0.3. Cells were pelleted by centrifugation at 4,000*g* for 5 min, washed with 10 ml LB and resuspended in 20 ml LB with chloramphenicol (25 µg ml^{-1}), spectinomycin (100 µg ml^{-1}) and 0.2% L-arabinose to induce gp58 expression. Cultures were incubated at 37 °C with shaking for 45 min before collection. Cells were collected by centrifugation at 4,000*g* for 10 min at 4 °C and then washed with 5 ml of cold TBS. Cells were centrifuged again at 4,000*g* for 5 min at 4 °C. The pellets were washed with 1 ml of cold TBS and centrifuged at 10,000*g* for 5 min at 4 °C. Next, the supernatants were removed and the pellets were stored at −20 °C.

Genomic DNA samples were extracted using the Wizard Genomic DNA Purification Kit (Promega). Then, 3 µg of gDNA was loaded onto an agarose gel (0.8% agarose, 1× TAE buffer) and run for 2 h at 90 V. The gel was soaked for 10 min in denaturation buffer (0.5 M NaOH, 1.5 M NaCl), and DNA was transferred to a Hybond-N+ membrane (GE Healthcare) by upward capillary transfer in alkaline transfer buffer (0.25 M NaOH, 1.5 M NaCl). The next day, DNA was denatured for probe hybridization by incubating the membrane in 0.4 M NaOH for 10 min, then rinsed in Na_2HPO_4 (0.5 M, pH 7.3) and deionized water. DNA transfer was confirmed by staining the membrane with methylene blue and imaging on the Amersham Imager 600 (GE Healthcare). The membrane was pre-hybridized in ULTRAhyb-Oligo buffer (Thermo Fisher Scientific) for 1 h at 42 °C. A biotinylated oligonucleotide probe of poly-dT_{40} or poly-dA_{40} was added to the hybridization buffer at a final concentration of 5 nM and hybridization was performed overnight at 42 °C. The next day, the membrane was washed twice with Southern blot wash buffer I (2× SSC with 0.1% SDS) and twice with Southern blot wash buffer II (0.1× SSC with 0.1% SDS). The membrane was developed using the Chemiluminescent Nucleic Acid Detection Module Kit (Thermo Fisher Scientific) and imaged with an Amersham Imager 600 (GE Healthcare). Probe sequences are provided in Supplementary Table 6. Uncropped gel and blot images are provided in Supplementary Fig. 1.

For the samples subjected to proteinase K treatment, 3 µg of gDNA were incubated with 1 µl of proteinase K (NEB) in 1× rCutsmart buffer (NEB) at 55 °C for 30 min. Enzyme inactivation was performed by heating the reaction to 95 °C for 10 min. The reactions were mixed with 6× purple gel loading dye (NEB) and loaded onto an agarose gel. Southern blotting was performed as described above.

ncRNA covariance modelling

Related homologues of DRT9 were identified using the amino acid sequence of a *S. enterica* homologue (*Sen*DRT9, EAO1508900.1) as the seed query in a BLASTp search on the NR protein database (max target sequences = 100). Nucleotide sequences 1 kb upstream and downstream of RT genes were retrieved, clustered at 99.9% sequence identity to remove replicates using CD-HIT[51] (v.4.8.1), and aligned using MAFFT[52] (v.7.505). The resulting alignment was trimmed at the 5′ and 3′ ends to the precise boundaries of the ncRNA as determined by RIP–seq on *Sen*DRT9. These putative ncRNA sequences were clustered at 95% sequence identity using CD-HIT and realigned using mLocARNA[53] (v.1.9.1) with the default parameters. The resulting structure-based multiple alignment was used to build and calibrate a covariance model (CM) using the Infernal suite[39] (v.1.1.4). The CMsearch function of Infernal was then used to scan through nucleotide sequences of RT and 1-kb flanking windows, generated by an expanded BLASTp search (max target sequences = 5,000) seeded on *Sen*DRT9 and clustered at 85% sequence identity using CD-HIT. The final CM was evaluated for statistically significant co-varying base pairs with R-scape[54] at an *E*-value threshold of 0.05, and incorporated a total of 201 homologous DRT9 systems, including *Sen*DRT9.

ncRNA polyuridine conservation analysis

A window encompassing the polyuridine tracks was extracted from the alignment of DRT9-associated ncRNAs described above, using the *Sen*DRT9 ncRNA as a reference. Next, columns composed of more than 50% gaps were removed with Trimal[55]. A weblogo was then generated from this alignment, before uridine occupancies were calculated for each position with the Biostrings package (v.2.74.0) in R and plotted with ggplot2 (v.3.5.1).

*Sen*DRT9 RT–ncRNA complex purification

To purify the His_6–GST-tagged RT in complex with its ncRNA, *E. coli* BL21-AI cells (NEB) were transformed with the expression vector and cells were grown in LB broth at 37 °C to an OD_{600} of 0.6. Cultures were induced with 0.2 mM IPTG and 0.1% arabinose for overnight expression at 16 °C. Cells were pelleted by centrifugation at 3,000*g* for 25 min at 4 °C, before lysis by sonication in lysis buffer I (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 5% glycerol, 10 mM imidazole, 0.1% Triton-X-100, 1 mM TCEP). The lysate was clarified by centrifugation at 10,000*g* for 30 min at 4 °C. Clarified lysates were passed through a pre-equilibrated Ni-NTA column (Cytiva) using the AKTA purification system (GE Healthcare), then rinsed with 10 column volumes of wash buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 5% glycerol, 20 mM imidazole, 1 mM TCEP) before elution with lysis buffer I supplemented with 300 mM imidazole. The collected fractions were combined and concentrated (100 kDa MWCO PES Spin-X UF concentrator, Corning), before size-exclusion chromatography using the Sup200 16/600 gel-filtration column (Cytiva) pre-equilibrated in protein storage buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 5% glycerol, 1 mM TCEP). The fractions containing the complex were pooled and treated with TEV protease overnight before another separation in the same gel filtration column.

Purification of His_6–2×Strep–SUMO-tagged RT in complex with its ncRNA was performed using the same culture and induction steps. The supernatants collected after cell lysis were passed through a Strep-Tactin column (IBA) pre-equilibrated in lysis buffer II (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 5% glycerol, 1 mM TCEP) and eluted with 2.5 mM desthiobiotin (IBA) in lysis buffer II. Pooled fractions were concentrated and further purified using a Superose 6 10/300 column (Cytiva). The fractions containing the complex were pooled and treated with Ulp1 protease overnight, before another separation by the same column. The final fractions were pooled, concentrated, aliquoted, flash frozen in liquid nitrogen and stored at −80 °C.

To purify the dATP homopolymer associated *Sen*DRT9-ncRNA complex, we first incubated 150 nM *Sen*DRT9–ncRNA with 0.9 mM dATP for 30 min at 37 °C. The mixture was analysed using the Superose 6 10/300 (Cytiva) column, equilibrated in protein storage buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 5% glycerol, 1 mM TCEP).

Analysis and sequencing of ncRNA co-purifying with RT

Nucleic acids co-purifying with *Sen*DRT9-encoded RT were extracted by treating the RT–ncRNA complex with buffer-saturated phenol (pH 8.5). After vortexing, the sample was centrifuged at 12,000*g* for 4 min, after which the aqueous phase was transferred to a fresh tube and mixed with chloroform. After an additional round of centrifugation, the recovered aqueous phase was clarified using the Monarch RNA Cleanup kit (NEB). For analytical RNase/DNase treatments, the purified ncRNA (200 ng) was treated with either 100 U ml^{-1} TurboDNase (Invitrogen) or 250 U ml^{-1} RNase A/T1 (Thermo Fisher Scientific) in a 10 µl reaction and incubated for 10 min at 37 °C. Reactions were quenched using 2× RNA Loading Dye (NEB) and heated at 70 °C for 10 min, before electrophoretic separation on a denaturing 10% urea–PAGE gel. Gels were stained with SYBR Gold. RNA-seq library preparation was performed as described above for RIP–seq samples. In brief, RNA was diluted in NEBuffer 2 and heated to 92 °C for 2 min for fragmentation by alkaline hydrolysis, before treatment with TURBO DNase (Thermo Fisher Scientific), RppH (NEB) and T4 PNK (NEB). Adapter ligation, reverse transcription and indexing PCR were performed using the NEBNext Small RNA Library Prep kit. Libraries were sequenced on an Element AVITI in paired-end mode with 150 cycles per end.

T5 phage gp58 purification

T5 phage-encoded gp58 was purified similarly to the *Sen*DRT9 RT–ncRNA complex, with the following exceptions. Gp58 was expressed as an N-terminal His_6–2×Strep–SUMO fusion, and was overexpressed in *E. coli* BL21-AI cells (NEB). Cells were induced at an OD_{600} of 0.5 with 0.3 mM IPTG and 0.1% arabinose. After overnight induction at 16 °C, cells were centrifuged at 3,000*g* for 25 min at 4 °C, sonicated in lysis buffer III (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5% glycerol) and subjected to affinity chromatography using Strep-Tactin column (IBA). Elution was done in a lysis buffer III supplemented with 250 mM desthiobiotin (IBA). The fractions were pooled together, treated with Ulp1 overnight and passed through a Ni-NTA column equilibrated in the lysis buffer III. The flowthrough was collected, concentrated and was further purified by size-exclusion chromatography using a Superdex 75 column (Cytiva) pre-equilibrated in gp58 storage buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5% glycerol). The fractions were pooled, concentrated, aliquoted, flash-frozen in liquid nitrogen and stored at −80 °C.

Biochemical poly-dA synthesis assays

Reverse transcription reactions generally contained *Sen*DRT9-encoded RT–ncRNA complexes and nucleotide substrates in polymerization buffer I (50 mM Tris-HCl, pH 8.5, 100 mM NaCl, 2 mM $MgCl_2$, 5 mM TCEP), at concentrations indicated in associated figure legends. Reactions were initiated by the addition of dATP, typically present at a concentration of 100 µM, and were incubated at 37 °C for 10 min, unless otherwise stated. Radioactive experiments contained trace [α-^{32}P]dATP together with unlabelled (cold) dATP. The reactions were quenched and treated with various nuclease or proteinase reagents before electrophoretic separation, as indicated in figure legends. Experiments involving nuclease P1 digestion were instead incubated in polymerization buffer II (50 mM Tris-HCl, pH 7.0, 100 mM NaCl, 2 mM $MgCl_2$, 5 mM TCEP) to achieve more optimal conditions for nuclease P1 cleavage, which loses activity at alkaline pH.

Reactions investigating the fate of unlabelled or radiolabelled [α-^{32}P]dATP substrates were analysed by denaturing urea–PAGE at 5–10% acrylamide concentrations. The samples were mixed in equal volumes with 2× RNA loading dye (NEB) before denaturation by heating, followed by electrophoretic separation. Radioactive gels were exposed to a phosphor screen for 5 h at −20 °C and imaged using the Typhoon imaging system (GE). The reactions investigating the fate of the *Sen*DRT9-encoded RT enzyme were analysed using 10% SDS–PAGE gels. The samples were mixed with 6× SDS loading dye before denaturation by heating and electrophoretic separation. Gels were stained with Coomassie Blue. Uncropped gel and blot images are provided in Supplementary Fig. 1.

Biochemical competition assays with *Sen*DRT9, ExoI, nuclease P1 and gp58

Competition assays involving gp58 and ExoI were performed in polymerization buffer I. Reactions initially contained 20 nM RT–ncRNA complex and 100 µM dATP, and were first incubated at 37 °C for 10 min, before treatment with ExoI in the absence or presence of gp58. ExoI was present at 0.015 U $µl^{-1}$, and gp58 was titrated into reactions at a range of 0.1–10 µM. Competition assays involving gp58 and nuclease P1 were performed in polymerization buffer II. Reactions initially contained 20 nM RT–ncRNA complex and 100 µM dATP, and were first incubated at 37 °C for 10 min, before treatment with nuclease P1 in the absence or presence of gp58. Nuclease P1 was present at 2 U $µl^{-1}$, and gp58 was titrated into reactions at a range of 0.1–10 µM. Competition assays using the U_4>A_4 mutant ncRNA were performed using 20 nM RT–ncRNA complex and 100 µM [α-^{32}P]dTTP supplemented with cold dTTP. The same procedure was followed as described above, with the exception that the polymerization reaction was carried out for 60 min at 37 °C. The reactions were quenched by proteinase K treatment, before addition of 2× RNA loading dye (NEB), denaturation by heating and electrophoretic separation by denaturing 5% urea–PAGE. Uncropped gel and blot images are provided in Supplementary Fig. 1.

Cryo-EM sample preparation

For all samples, Quantifoil copper grids with a mesh size of 300 and R1.2/1.3 hole spacing were glow discharged using a Pelco EsiGlow at 15 mA for 45 s. Then, 3 µl of 6 µM purified GST–DRT9 (trimer) or 2.5 µM DRT9 (hexamer) were applied to prepared grids in a Vitrobot Mk IV (Thermo Fisher Scientific) set to 100% humidity and 4 °C. The grids were subjected to double-sided blotting with a force of 5 for 6 s before plunge-freezing into liquid ethane. Clipped grids were loaded into autogrid boxes and stored in liquid nitrogen before imaging.

Cryo-EM data collection

Both datasets were collected on a Talos Arctica (Thermo Fisher Scientific) equipped with a Gatan K3 direct electron detector at Montana State University's Cryo-EM core facility using SerialEM[56] (v.4.2.0) under the control of SmartScope[57] (v.0.9.4) for automated data collection. In total, 10,450 and 15,594 micrographs were collected for the hexamer and trimer datasets, respectively. All particles were imaged with a pixel size of 0.9061 Å and apparent magnification of ×45,000. The total dose per exposure was calculated as 59.5 (trimer) and 59.8 (hexamer) e per $Å^2$ per exposure.

Cryo-EM image analysis, *Sen*DRT9 RT–ncRNA trimer

All image analysis was completed in cryoSPARC[58] (v.4.6.2). After filtering by CTF-fit < 8 Å and max in-frame motion of 5, 9,814 exposures remained. cryoSPARC Live's blob picker (50–150 Å particle diameter) was used to identify particles, which were then extracted with a box size of 496 px and Fourier cropped to 248 px (binx2). In total, 1,878,603 particles were selected from streaming 2D classification and processed to yield a stack of 610,556 particles that were used to generate a 3.7-Å-resolution reconstruction without imposing symmetry (Extended Data Fig. 6b). This volume was used to generate de novo templates for particle picking which yielded 5,235,086 initial picks. Then, 2,530,970 particles were selected from 2D classification

(Extended Data Fig. 6c) and fed into a four-class ab initio reconstruction (Extended Data Fig. 6d), followed by heterogenous refinement. Non-uniform refinement of the best class produced a 3.7 Å volume containing 2,226,754 particles. After re-extracting unbinned particles and an additional round of 2D classification and 3D sorting by ab initio and heterogenous refinement, an initial consensus volume was obtained from 2,107,864 particles. A five-class 3D classification of these particles without a mask and with force hard classification enabled the removal of additional junk, leaving 1,368,451 particles for further processing. A final round of multiclass ab initio reconstruction, followed by heterogeneous refinement (Extended Data Fig. 6e) yielded a final stack of 367,640 particles that were refined to 3.02 Å resolution by C_3 non-uniform refinement (EMD-59295) and used for model building (Extended Data Fig. 6g–i).

Cryo-EM image analysis, *Sen*DRT9 RT–ncRNA hexamer

After filtering by CTF-fit < 8 Å resolution, 9,529 exposures remained. In total, 151,584 particles were picked on-the-fly using cryoSPARC Live (100–200 Å particle diameter) and extracted with a box size of 496 px and binned ×4 to expedite processing. A preliminary reconstruction was generated from 48,338 (re-extracted, unbinned) particles that refined to 3.32 Å when D_3 symmetry was imposed (3.73 Å, C_1). De novo templates were produced using this volume, and template picking on the full dataset yielded 6,011,113 picks. After two rounds of 2D classification, 1,441,369 particles (Extended Data Fig. 7c) were re-extracted without binning and a box size of 496 px before removing 206,158 duplicate particles. The remaining 1,227,994 particles were subjected to seven-class ab initio reconstruction (Extended Data Fig. 7d) followed by heterogenous refinement. The best class from heterogenous refinement contained 584,077 particles. These particles were re-extracted to obtain better estimates of particle centres before a final round of ab initio and heterogenous refinement yielded a stack of 309,844 particles that refined to 2.9 Å resolution by C_1 refinement. After volume alignment, imposing D_3 symmetry during refinement increased the resolution to 2.59 Å (Extended Data Fig. 7f–h). This map (EMD-59293) was used for model building the hexameric DRT9 ribonucleoprotein complex.

Cryo-EM model building

An initial model for the DRT9 trimer was generated using several copies of an AlphaFold 3 (ref. 40) structure prediction of the RT monomer and fitting them into the final reconstructions using ChimeraX's[41] fit in map command. To model the ncRNA, density maps were submitted to ModelAngelo[59] (v.1.0.13), along with the ncRNA sequence identified from RIP–seq experiments described above. Although ModelAngelo was unable to build a complete chain for the ncRNA, it correctly modelled a poly-G stretch (80–86) (Extended Data Fig. 8f), which was used as a seed sequence to manually build the rest of the chain in COOT[60]. After building the ncRNA, the complete model for the trimer was refined into the raw map using Phenix (v.1.21.2-5419) RealSpaceRefinement[61] with reference model, secondary structure and NCS restraints enabled. Iterative editing in COOT (v.0.9.8.1) and refinement in Phenix was used to produce the final model (PDB: 9NLX). Sharpened maps from Phenix's anisotropic half-map sharpening tool were used to help with modelling difficult regions. All outlier residues in the validation report were inspected manually before submission.

The Trimer PDB was used as a starting point for model building of the hexameric form of the complex. After fitting two copies of the trimer structure into the map of the hexamer using ChimeraX's fit in map command, the model was subjected to RealSpaceRefinement as described above. Iterative editing in COOT and refinement in Phenix were used to produce the final model for the DRT9 hexamer (PDB: 9NLV). Sharpened maps from Phenix's anisotropic half-map sharpening tool were used to aid in modelling difficult regions. All outlier residues were manually inspected before submission. To better visualize unmodelled density near the active site, a 3.0 Å calculated map was generated using ChimeraX's molmap command and subtracted from the experimental density (Extended Data Fig. 8d).

RT complex interactome maps

RT interactome maps were generated by assessing contacts between protomeric and ncRNA subunits with the contacts command in ChimeraX[41], using the default settings. Interactions were then plotted with BioCircos[62].

Escaper phage isolation

To isolate escaper phages, serial dilutions of phages T5 or Bas37 were mixed in molten soft agar with *E. coli* K-12 strain MG1655 expressing *Sen*DRT9, *Psa*DRT9 or an EV control, and plated onto LB-agar plates supplemented with chloramphenicol (25 μg ml^{-1}). The plates were incubated overnight at 37 °C to allow plaques to form. Individual plaques were picked into 1 ml of MG1655 culture at an OD_{600} of around 0.1. Phages forming plaques on *Sen*DRT9 or *Psa*DRT9 plates were inoculated into respective MG1655 cultures expressing *Sen*DRT9 or *Psa*DRT9, while plaques from empty-vector plates were inoculated into MG1655 carrying the EV. Cultures were incubated at 37 °C with shaking for 3 h, and then 800 μl of each culture was transferred to new tubes and mixed with chloroform (5% final volume) to lyse residual bacteria. Cell debris and chloroform were removed by centrifugation for 5 min at 4,000*g*, and 5 μl of each phage lysate was propagated further in fresh cultures at OD_{600} of around 0.1. Phages were amplified in this manner for a total of three rounds. Escaper phage phenotypes were validated by small-drop plaque assays. T5 escaper phages were isolated from *Sen*DRT9 plates and *Psa*DRT9 plates, and Bas37 escaper phages from *Psa*DRT9 plates; attempts to isolate Bas37 escapers from *Sen*DRT9 were unsuccessful.

Escaper phage whole-genome sequencing

Phage DNA was isolated by treating 88 μl of phage supernatant with 1 μl of DNase I (200 U ml^{-1}, NEB) and 1 μl of RNase A (10 mg ml^{-1}, Thermo Fisher Scientific) in 10 μl 1× DNase buffer (NEB). The reactions were incubated at 37 °C for 1 h, and enzymes were inactivated by heating at 75 °C for 10 min. Phage capsids were digested by adding 1 μl of proteinase K (20 mg ml^{-1}, Thermo Fisher Scientific) and 99 μl of phage lysis buffer (10 mM Tris-HCl, 10 mM EDTA, 0.5% SDS), followed by incubation at 37 °C for 30 min and 55 °C for 30 min. DNA clean-up was performed using the Mag-Bind Total Pure NGS beads (Omega), with a bead ratio of 0.9× sample volume.

Phage genomic DNA was tagmented using TnY (a homologue of Tn5) purified in-house according to previously described methods[63]. A total of 10 ng of purified genomic DNA (gDNA) was tagmented with TnY preloaded with Nextera read 1 and read 2 oligos (Supplementary Table 6), followed by proteinase K treatment (NEB, final concentration 16 U ml^{-1}) and column purification. PCR amplification and Illumina barcoding was done for 13 cycles with KAPA HiFi Hotstart ReadyMix, with an annealing temperature of 63 °C and an extension time of 1 min. The PCR reactions were then pooled and resolved on a gel. A smear from 400–800 bp was extracted with a Qiaquick Gel Extraction kit (Qiagen) for sequencing on an Element AVITI in paired-end mode with 150 cycles per end.

The resulting datasets were processed using cutadapt[42] (v.4.2) to remove Illumina adapter sequences, trim low-quality ends from reads and filter out reads shorter than 15 bp. Reads were mapped to the T5 (NC_005859.1) genome or Bas37 genome (MZ501089.1), using Bowtie2 (ref. 64) (v.2.2.1) with the default parameters. SAMtools[44] (v.1.17) was used to sort and index alignments, and coverage tracks were visualized in IGV[46] (v.2.17.4). Mutations were identified with breseq[65] (v.0.39.0) using alignments from Bowtie2 and the same T5 or Bas37 reference genomes as above. T5 and Bas37 escaper mutations are listed in Supplementary Table 3.

Infection response growth curves

Escherichia coli K-12 strain MG1655 cells transformed with plasmids encoding WT or MUT *Sen*DRT9 were grown to OD_{600} of 0.2 in LB medium supplemented with chloramphenicol (25 μg ml^{-1}). Then, 180 μl of each culture was transferred into wells of a 96-well optical plate containing 20 μl of T5 lysate diluted to result in a final MOI of 5 or 0.05, or 20 μl of LB for the uninfected condition. The plate was incubated for 4 h at 37 °C with shaking. OD_{600} values were recorded every 10 min using the Synergy Neo2 microplate reader (Biotek).

Cell viability assays with trigger and host factor candidates

To assess the effects of *Sen*DRT9 co-expression with phage genes of interest on cell growth, candidate trigger genes identified by escaper screening were cloned onto an expression vector under the control of an arabinose-inducible araBAD promoter. *E. coli* K-12 strain MG1655 cells expressing either WT or MUT *Sen*DRT9 were transformed with trigger plasmids and grown under repressive conditions in LB medium supplemented with 2% glucose. Overnight cultures were centrifuged at 4,000*g* for 5 min and cell pellets were washed with fresh LB to remove residual glucose. Cells were pelleted again as before and resuspended in fresh LB. 10× serial dilutions of each culture were prepared and 5 μl were spotted on LB agar plates supplemented with chloramphenicol (25 μg ml^{-1}), spectinomycin (100 μg ml^{-1}) and arabinose (0.2%). Plates were incubated overnight at 37 °C and colonies were counted the next day. The CFU per ml was calculated as described above for the colony-formation time course analysis.

To assess the cell viability of Δ*sbcB* MG1655 cells expressing *Sen*DRT9, a Δ*sbcB* knock-out strain was transformed with either WT *Sen*DRT9 or MUT *Sen*DRT9. Cells were recovered in 10 ml of LB medium at 37 °C for 4 h. After recovery, cultures were diluted 1:10 in fresh LB, and 100 μl were plated on LB agar plates supplemented with chloramphenicol and kanamycin. The CFU per ml was calculated using the following formula: $\frac{\text{number of colonies}}{0.1\text{ ml} \times \text{dilution factor}}$.

Nucleotide quantification by LC–MS/MS

Escherichia coli K-12 strain MG1655 cells were transformed with EV or *Sen*DRT9 plasmids. Individual colonies were inoculated into LB medium supplemented with chloramphenicol (25 μg ml^{-1}) and grown overnight at 37 °C with shaking. The next day, cultures were diluted 1:100 in 20 ml fresh LB medium and grown to OD_{600} of 0.3. Cultures were split in half and T5 phage was added to one half at an MOI of 2. Uninfected and infected cultures were grown for an additional 45 min at 37 °C with shaking. Cells were centrifuged at 4,000*g* for 5 min at 4 °C and pellets were washed with 1 ml of cold 1× TBS. Cells were centrifuged again at 15,000*g* for 1 min at 4 °C. After supernatants were removed, pellets were flash-frozen in liquid nitrogen and stored at −80 °C.

Cell lysates were prepared as previously described[8]. Flash-frozen pellets were thawed on ice and resuspended in 600 μl of 100 mM sodium phosphate buffer (10:1 mixture of Na_2HPO_4 and NaH_2PO_4) supplemented with 1 mg ml^{-1} lysozyme. Cells were mechanically lysed using a bead beater (2.5 min × 2 cycles at 4 °C). Lysates were centrifuged at 4,000*g* for 15 min at 4 °C and supernatants were transferred to new tubes. The samples were transferred to Amicon Ultra-0.5 3 kDa filter units (Merck Millipore) and centrifuged at 12,000*g* for 45 min at 4 °C. Filtrates containing metabolites were collected and stored at −80 °C. After retrieval from −80 °C, and before use for injection in LC–MS/MS, filtrates were additionally purified by centrifugation at 20,000*g* for 20 min at 4 °C. To confirm that each condition contained equal amounts of cellular material, the protein concentration in filtration supernatants was measured using the BCA assay (Thermo Fisher Scientific). Protein concentrations varied by less than 10% across all samples.

Nucleotide and deoxynucleotide profiling analysis was carried out with ion-paired reversed-phase liquid chromatography using an HPLC (1290 Infinity II, Agilent Technologies) system coupled to a triple quadrupole mass spectrometer (6495D, Agilent Technologies) with electrospray ionization operated in negative mode. The column was a ZORBAX RRHD Extend-C18 (2.1 × 150 mm, 1.8 μm pore size; 759700-902, Agilent Technologies). Then, 5 μl of the experimental samples was injected. Mobile phase A was 3% methanol (in H_2O), and mobile phase B was 100% methanol. Both mobile phases contained 10 mM of the ion-pairing agent tributylamine (90780, Sigma-Aldrich), 15 mM acetic acid, and 5 μM medronic acid (5191-4506, Agilent Technologies). The LC gradient was: 0–2.5 min 100% A, 2.5–7.5 min ramp to 80% A, 7.5–13 min ramp to 55% A, 13–20 min ramp to 99% B, 20–24 min hold at 99% B. The flow rate was 0.25 ml min^{-1}, and the column compartment was heated to 35 °C. The column was then backflushed with 100% acetonitrile (0.25 ml min^{-1} flow rate 24.05–27 min, followed by 0.8 ml min^{-1} flow rate 27.5–31.35 min and 0.6 ml min^{-1} flow rate, 31.35–31.50 min) and re-equilibrated with 100% A (0.4 ml min^{-1} flow rate 32.25–40 min). The conditions of the MRM transitions for dATP and dTTP were as follows (collision energy (V)): dATP, 490 > 391.9 (24), 490 > 158.9 (32); dTTP, 481 > 383.1 (20), 481 > 158.8 (36). Fragmentor and CAV were kept constant at 166 V and 4 V, respectively. For other nucleotides and deoxynucleotides measured, conditions of the MRM transitions are shown in Supplementary Table 7. Moreover, the identity of each compound was confirmed with subsequent injection and acquisition of a pure chemical standard. Data analysis, including peak area integration and signal extraction, was performed with Skyline-daily (v.24.1.1.398).

Nucleoside-supplementation assays in plaque assays

Nucleoside supplementation was performed using *E. coli* MG1655 cells transformed with either EV or DRT9 expression plasmids. Cultures were grown in LB supplemented with chloramphenicol (25 μg ml^{-1}) until reaching an OD_{600} of 0.3. Deoxythymidine, deoxyadenosine, deoxycytidine or deoxyguanosine were added to 4 ml cultures at a final concentration of 20 μM. Cultures were pre-incubated at 37 °C with shaking for 5 min, before adding phage T5 at MOI of 2.5, and grown for another hour. Phages were isolated by adding chloroform to a final concentration of 4%, and the samples were centrifuged at 3,000*g* for 3 min. The supernatant was used to make 10× serial dilutions used for small-drop plaque assays, performed as described previously. EOP was calculated by spotting the phage on soft agar containing *E. coli* MG1655 cells transformed with an EV; the graph shows the results from two technical replicates.

Co-IP western blot

Escherichia coli K-12 strain MG1655 cells were transformed with plasmids encoding either untagged *Sen*DRT9, *Sen*DRT9 with mutated RT catalytic residues (YAAA), N-terminally 3×Flag-tagged *Sen*DRT9 or N-terminally 3×Flag-tagged *Sen*DRT9 with mutated RT catalytic residues (YAAA), as well as a plasmid encoding C-terminally V5-tagged gp58. Individual colonies were inoculated into LB medium supplemented with chloramphenicol (25 μg ml^{-1}), spectinomycin (100 μg ml^{-1}) and 2% glucose and grown overnight at 37 °C with shaking. The next day, cultures were diluted 1:100 in 25 ml fresh LB medium with continued 2% glucose repression and grown to an OD_{600} of 0.4. The cultures were centrifuged at 4,000*g* for 5 min and the cell pellets were washed with fresh LB to remove residual glucose. Cells were pelleted again as before, and resuspended in fresh LB with chloramphenicol (25 μg ml^{-1}), spectinomycin (100 μg ml^{-1}) and 0.2% arabinose, and grown at 37 °C with shaking for 30 min. Cells were centrifuged at 4,000*g* for 10 min at 4 °C and pellets were washed with 1 ml of cold 1× TBS. After another round of centrifugation at 15,000*g* for 1 min at 4 °C, the supernatants were removed and the pellets were flash-frozen in liquid nitrogen and stored at −80 °C.

To prepare antibody–bead complexes for immunoprecipitation, protein G Dynabeads (Thermo Fisher Scientific) were washed three times in 1 ml IP–MS lysis buffer (50 mM Tris-HCl pH 7.5 at 25 °C, 150 mM NaCl, 5% glycerol, 0.2% Triton X-100). The washed beads were resuspended again in 1 ml IP–MS lysis buffer, combined with anti-Flag antibody

(Sigma-Aldrich, F3165), and rotated for >3 h at 4 °C. Then, 75 µl of beads and 10 µl of antibody were prepared per sample. Antibody–bead complexes were washed twice to remove unbound antibodies and resuspended in IP–MS lysis buffer to a final volume of 75 µl per sample.

Flash-frozen pellets were thawed on ice and resuspended in 1.2 ml IP–MS lysis buffer supplemented with 1× Complete Protease Inhibitor Cocktail (Roche) and 0.1 U μl^{-1} SUPERase•In RNase Inhibitor (Thermo Fisher Scientific). Cells were lysed by sonication (2 s on, 5 s off, amplitude 20%, 1.5 min total time) and the lysates were cleared by centrifugation at 21,000*g* for 15 min at 4 °C. The supernatants were transferred to new tubes and the protein concentrations were measured using the BCA assay (Thermo Fisher Scientific). For each sample, 1 mg of protein was combined with 75 µl of antibody–bead complex and rotated overnight at 4 °C. Then, 50 µl of each cleared lysate was set aside as input control and stored at −80 °C. The next day, each sample was washed twice with 1 ml cold IP–MS wash buffer 1 (50 mM Tris-HCl pH 7.5 at 25 °C, 150 mM NaCl, 5% glycerol, 0.02% Triton X-100) followed by another two washes with cold IP–MS wash buffer 2 (50 mM Tris-HCl pH 7.5 at 25 °C, 150 mM NaCl, 5% glycerol).

For elution, beads were resuspended in 75 µl 2× SDS loading buffer and incubated at 50 °C for 10 min. Input samples were thawed, and both input and IP samples were prepared for SDS–PAGE by combining 15 µl protein with 18.75 µl 2× SDS loading buffer and 3.75 µl 0.1 M DTT. The samples were denatured at 95 °C for 5 min and immediately transferred to ice. The samples were loaded onto 12% Criterion XT Bis-Tris Protein Gels and subjected to electrophoresis in XT MES running buffer at 80 V for 5 min followed by 200 V for 30–40 min until the dye front reached the bottom of the gel. Proteins were transferred to nitrocellulose membranes using iBlot. The membranes were blocked in TBS-T (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Tween-20) containing 5% BSA for 1 h at room temperature. The membranes were incubated overnight at 4 °C with V5 primary antibody (V5-Tag (D3H8Q), Cell Signaling Technologies) diluted 1:2,000 in the blocking buffer. After three 5 min washes with TBS-T, the membranes were incubated with the secondary antibody (mouse anti-rabbit IgG (L27A9)) diluted 1:5,000 in a 1:1 mixture of blocking buffer and TBS-T for 1 h at room temperature. After three 5 min washes with TBS-T, signals were detected using the West Dura Signal kit (Thermo Fisher Scientific) and imaged using the chemiluminescence setting on an Amersham Imager 600 (GE Healthcare). Uncropped gel and blot images are provided in Supplementary Fig. 1.

Co-IP–MS

Escherichia coli K-12 strain MG1655 cells were transformed with plasmids encoding either untagged *Sen*DRT9, N-terminally 3×Flag-tagged *Sen*DRT9, or N-terminally 3×Flag-tagged *Sen*DRT9 with mutated RT catalytic residues (YAAA). Individual colonies were inoculated into LB medium supplemented with chloramphenicol (25 µg ml^{-1}) and grown overnight at 37 °C with shaking. The next day, the cultures were diluted 1:100 in 40 ml fresh LB medium and grown to an OD_{600} of 0.3. The cultures were split in half and T5 phage was added to one half at an MOI of 5. Uninfected and infected cultures were grown for an additional 45 min at 37 °C. Cells were centrifuged at 4,000*g* for 10 min at 4 °C and the pellets were washed with 1 ml of cold TBS. After centrifuging again at 15,000*g* for 1 min at 4 °C and removing the supernatants, the pellets were flash-frozen in liquid nitrogen and stored at −80 °C.

To prepare antibody–bead complexes for immunoprecipitation, Protein G Dynabeads (Thermo Fisher Scientific) were washed three times in 1 ml IP–MS lysis buffer (50 mM Tris-HCl pH 7.5 at 25 °C, 150 mM NaCl, 5% glycerol, 0.2% Triton X-100). Washed beads were resuspended again in 1 ml IP–MS lysis buffer, combined with anti-Flag antibody (Sigma-Aldrich, F3165), and rotated for >3 h at 4 °C. Then, 100 µl of beads and 10 µl of antibody were prepared per sample. Antibody–bead complexes were washed twice to remove unbound antibodies and resuspended in IP–MS lysis buffer to a final volume of 100 µl per sample.

Flash-frozen pellets were thawed on ice and resuspended in 1.2 ml IP–MS lysis buffer supplemented with 1× Complete Protease Inhibitor Cocktail (Roche) and 0.1 U μl^{-1} SUPERase•In RNase Inhibitor (Thermo Fisher Scientific). Cells were lysed by sonication and the lysates were cleared by centrifugation at 21,000*g* for 15 min at 4 °C. The supernatants were transferred to new tubes and the protein concentrations were measured using the BCA assay (Thermo Fisher Scientific). For each sample, 1 mg of protein was combined with 100 µl of antibody–bead complex and rotated overnight at 4 °C. The next day, each sample was washed twice with 1 ml cold IP–MS wash buffer 1 (50 mM Tris-HCl pH 7.5 at 25 °C, 150 mM NaCl, 5% glycerol, 0.02% Triton X-100) followed by another two washes with cold IP–MS wash buffer 2 (50 mM Tris-HCl pH 7.5 at 25 °C, 150 mM NaCl, 5% glycerol). For elution and on-bead tryptic digestion, beads were resuspended in 80 µl Tris-urea buffer (50 mM Tris-HCl pH 7.5 at 25 °C, 2 M urea, 1 mM DTT) with 0.5 µg ml^{-1} sequencing grade modified trypsin (Promega) and incubated 1 h at 25 °C with regular agitation. The supernatants were transferred to new tubes and beads were resuspended again in 60 µl Tris-urea buffer without trypsin. The supernatants were combined with those from the first elution step. This step was repeated once more, resulting in a total eluate volume of 200 µl per sample. The eluates were centrifuged at 5,000*g* for 1 min and the supernatants were transferred to new tubes to remove any residual beads. The eluates were flash-frozen in liquid nitrogen and stored at −80 °C before further processing.

IP–MS eluates were reduced with 5 mM DTT at 25 °C with 600 rpm agitation for 45 min, after which they were alkylated in the dark with 10 mM iodoacetamide at 25 °C and 600 rpm for 45 min. The samples were then cleaned up using the previously described SP3 protocol[66]. Specifically, 500 µg of an equal mixture of carboxylate-modified hydrophilic beads (Cytiva, 45152105050250) and hydrophobic beads (Cytiva, 65152105050250) was added to each sample. To induce the binding of proteins to the SP3 beads, 100% ethanol was added to the sample at a 1:1 ratio, and the samples were incubated at 25 °C and 1,000 rpm for 15 min. After incubation, the beads were washed three times with 80% ethanol and reconstituted in freshly prepared 100 mM ammonium bicarbonate. The samples were digested off the beads using 1 µg of sequencing-grade modified trypsin (Promega V5111). After 16 h of incubation at 25 °C and 600 rpm, the samples were taken off the magnetic beads, dried using a Thermo Savant SpeedVac and reconstituted in 3% acetonitrile and 0.2% formic acid.

Digested peptides from the IP–MS eluates were analysed on the Waters M-Class UPLC system using a 15 cm × 75 µm IonOpticks C18 1.7 µm column coupled to a benchtop Thermo Fisher Scientific Orbitrap Q Exactive HF mass spectrometer. Peptides were separated at a 400 nl min^{-1} flow rate with a 90-min gradient, including sample loading and column equilibration times. Data were acquired in data-dependent mode using Xcalibur (v.4.5.474.0.); each cycle's 12 most intense peaks were selected for MS2 analysis. MS1 spectra were measured with a resolution of 120,000, an AGC target of 3×10^6 and a scan range from 300 to 1,800 *m*/*z*. MS2 spectra were measured with a resolution of 15,000, an AGC target of 1×10^5, a scan range from 200–2,000 *m*/*z* and an isolation window width of 1.6 *m*/*z*.

Raw data were searched against a combined reference proteome that included the *E. coli* K-12 strain MG1655 proteome (NCBI RefSeq assembly GCF_000005845.2), T5 phage proteome (NCBI RefSeq assembly GCF_000858785.1) and the *Sen*DRT9 RT sequence, using MaxQuant[67] (v.2.0.3.0). One or more unique/razor peptides were required for protein identification. All of the subsequent analyses were performed in R (v.4.4.0).

For peptide-level analysis to compare peptide detection between the immunoprecipitated WT and MUT RT samples, the data were first filtered to remove peptides with LFQ intensity of 0. Values were then normalized to the sum of LFQ intensities for each condition. After $\log_2$ transformation, significantly depleted peptides were determined using

unpaired two-tailed *t*-tests, followed by multiple-comparison correction using the Benjamini–Hochberg method.

For protein-level analysis to identify DRT9 interactors, protein groups with more than five MS/MS spectral counts in at least one condition were retained in the normalized LFQ intensities. A random value between 1×10^6 and 1% of the lowest data point was added as a pseudocount to all datapoints, followed by $\log_2$-transformation. The data were then used to determine significantly enriched proteins with an unpaired two-tailed *t*-test, with correction for multiple-comparisons using the Benjamini–Hochberg method. Significantly enriched interactors were identified using cut-offs of fold-enrichment of >20 and adjusted $P < 0.05$, relative to a control IP from cells expressing untagged WT *Sen*DRT9. IP–MS hits are described in Supplementary Table 8.

Phylogenetic analysis of DRT9 RT and other RTs

Sequences of DRT9, a group II intron RT, KpnDRT2, AbiA, AbiK, AbiP2 and the HIV-1 RT were used to predict monomeric and apo structures of each protein with AlphaFold 3 (ref. 40). These predicted structures were then aligned with FoldMason[68], and the regions corresponding to the Palm and Finger domains were extracted from the alignment. FastTree[36] was then used to build the phylogenetic tree shown in Supplementary Fig. 2.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Next-generation sequencing data are available at the NCBI Gene Expression Omnibus (GEO) under accession numbers GSE295895, GSE295896 and GSE295897. The published genomes used for bioinformatics analyses were obtained from the NCBI (Supplementary Table 1). Cryo-EM maps and experimentally determined models were deposited at the appropriate public database (that is, EMDB, PDB, respectively) under accession numbers EMD-49525 and 9NLX (GST–DRT9 trimer), and EMD-49523 and 9NLV (DRT9 hexamer). Previously determined protein structures used for structural comparisons were obtained from the PDB (7R06 and 8FAK). The phylogenetic tree of UG RT homologues was obtained from ref. 3. MS data are available in the MassIVE database under accession number MSV000097825. All other data supporting the findings of this study are available in the Article and its Supplementary Information. Source data are provided with this paper.

Code availability

No custom scripts were developed for this Article.

34. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
35. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
36. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
37. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
38. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
39. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
40. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
41. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci. Publ. Protein Soc.* **27**, 14–25 (2018).
42. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* **17**, 10–12 (2011).
43. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In *Proc 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* 314–324 (IEEE, 2019).
44. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
45. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
46. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
47. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
48. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
49. Blue, S. M. et al. Transcriptome-wide identification of RNA-binding protein binding sites using seCLIP-seq. *Nat. Protoc.* **17**, 1223–1265 (2022).
50. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
51. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
52. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
53. Will, S., Joshi, T., Hofacker, I. L., Stadler, P. F. & Backofen, R. LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA* **18**, 900–914 (2012).
54. Rivas, E., Clements, J. & Eddy, S. R. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* **14**, 45–48 (2017).
55. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
56. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
57. Bouvette, J. et al. Automated systematic evaluation of cryo-EM specimens with SmartScope. *eLife* **11**, e80047 (2022).
58. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
59. Jamali, K. et al. Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024).
60. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
61. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
62. Cui, Y. et al. BioCircos.js: an interactive Circos JavaScript library for biological data visualization on web applications. *Bioinformatics* **32**, 1740–1742 (2016).
63. Liscovitch-Brauer, N. et al. Profiling the genetic determinants of chromatin accessibility with scalable single-cell CRISPR screens. *Nat. Biotechnol.* **39**, 1270–1277 (2021).
64. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
65. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol. Biol.* **1151**, 165–188 (2014).
66. Hughes, C. S. et al. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **14**, 68–85 (2019).
67. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
68. Gilchrist, C. L. M., Mirdita, M. & Steinegger, M. Multiple protein structure alignment at scale with FoldMason. Preprint at *bioRxiv* https://doi.org/10.1101/2024.08.01.606130 (2024).

Acknowledgements We thank A. I. Palmieri, S. Kang and A. J. Robinson for laboratory support; P. Sims for discussions regarding sequencing data analysis; S. Tavazoie, M. Liu and P. Oikonomou for sharing Keio knockout strains; D. Puleston and R. Sorek for metabolomics support; L. E. Berchowitz for Southern blotting support; the members of the JP Sulzberger Columbia Genome Center for next-generation sequencing support; and L. F. Landweber for qPCR and gel imager instrument access. Research in the Wiedenheft laboratory is supported by the National Institutes of Health (NIH R35GM134867), the M.J. Murdock Charitable Trust and the Montana Agricultural Experimental Station. Funding for the Montana State University Cryo-EM Core Facility (RRID: SCR_026324) was contributed by the National Science Foundation (DBI-1828765), the M.J. Murdock Charitable Trust, NIGMS (P30GM140963) and the MSU Office of Research, Economic Development and Graduate Education. Image processing was performed using the Tempest High Performance Computing System, operated and supported by University Information Technology Research Cyberinfrastructure (RRID: SCR_026229) at Montana State University. S.T. was supported by a Ruth L. Kirchstein Individual Predoctoral Fellowship (F30AI183830) from the NIH; N.B. by a Ruth L. Kirschstein Individual Predoctoral Fellowship (GM153146) from the NIH and received support from Montana INBRE (P20GM103474); L.M.K. by an NIH Medical Scientist Training Program grant (T32GM145440); M.J. by NIH/NIGMS grant R35GM152258; S.H.S. by NSF Faculty Early Career Development Program (CAREER) Award 2239685, a Pew Biomedical Scholarship, an Irma T. Hirschl Career Scientist Award, a Mallinckrodt Scholarship, the Howard Hughes Medical Institute Investigator Program and a generous startup package from the Columbia University Irving Medical Center Dean's Office and the Vagelos Precision Medicine Fund.

Author contributions S.T., R.Ž. and S.H.S. conceived the project. S.T. and R.Ž. performed RIP–seq and cDIP–seq experiments, Southern blotting experiments and phage genetics experiments. S.T. and M.B. performed metabolomics experiments. S.T. and Y.M. performed proteomics experiments, with supervision from M.J. R.Ž. performed nucleoside-supplementation and oligo-spike-in experiments. N.B. prepared cryo-EM samples, collected and processed cryo-EM data, and built atomic models. S.P. purified SenDRT9-encoded RT–ncRNA complexes and performed all biochemical experiments, with assistance from R.A.W. J.L.R. performed plaque assays with ncRNA mutants, performed phage trigger and host factor experiments,

and cloned and tested gp58 variants. L.M.K. cloned and screened DRT9 homologues for activity, isolated escaper phages, and performed and analysed phage whole-genome sequencing. T.W. performed phylogenetic and bioinformatics analyses and assisted in data interpretation and visualization. D.J.Z. constructed the ncRNA covariance model and performed growth curve experiments. G.D.L. assisted in the design and execution of phage whole-genome sequencing. S.T., R.Ż., N.B., S.P., J.L.R., L.M.K., T.W., B.W. and S.H.S. discussed the data and wrote the manuscript, with input from all of the authors.

Competing interests S.H.S. is a co-founder and scientific advisor to Dahlia Biosciences, a scientific advisor to CrisprBits and Prime Medicine, and an equity holder in Dahlia Biosciences and CrisprBits. B.W. is the founder of SurGene and is listed as an inventor on patent applications related to CRISPR–Cas systems and applications thereof. The other authors declare no competing interests.

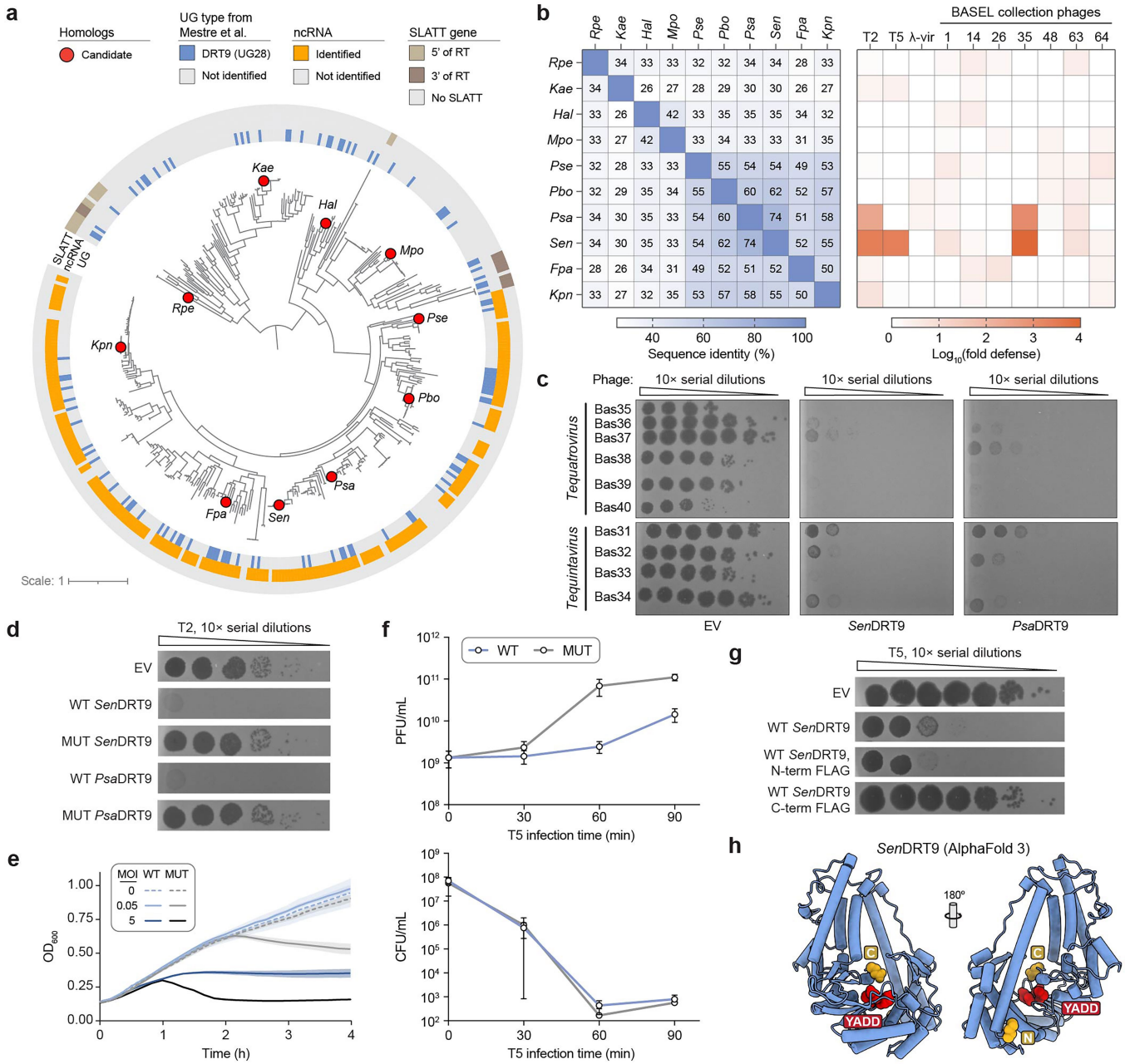
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-09179-5.

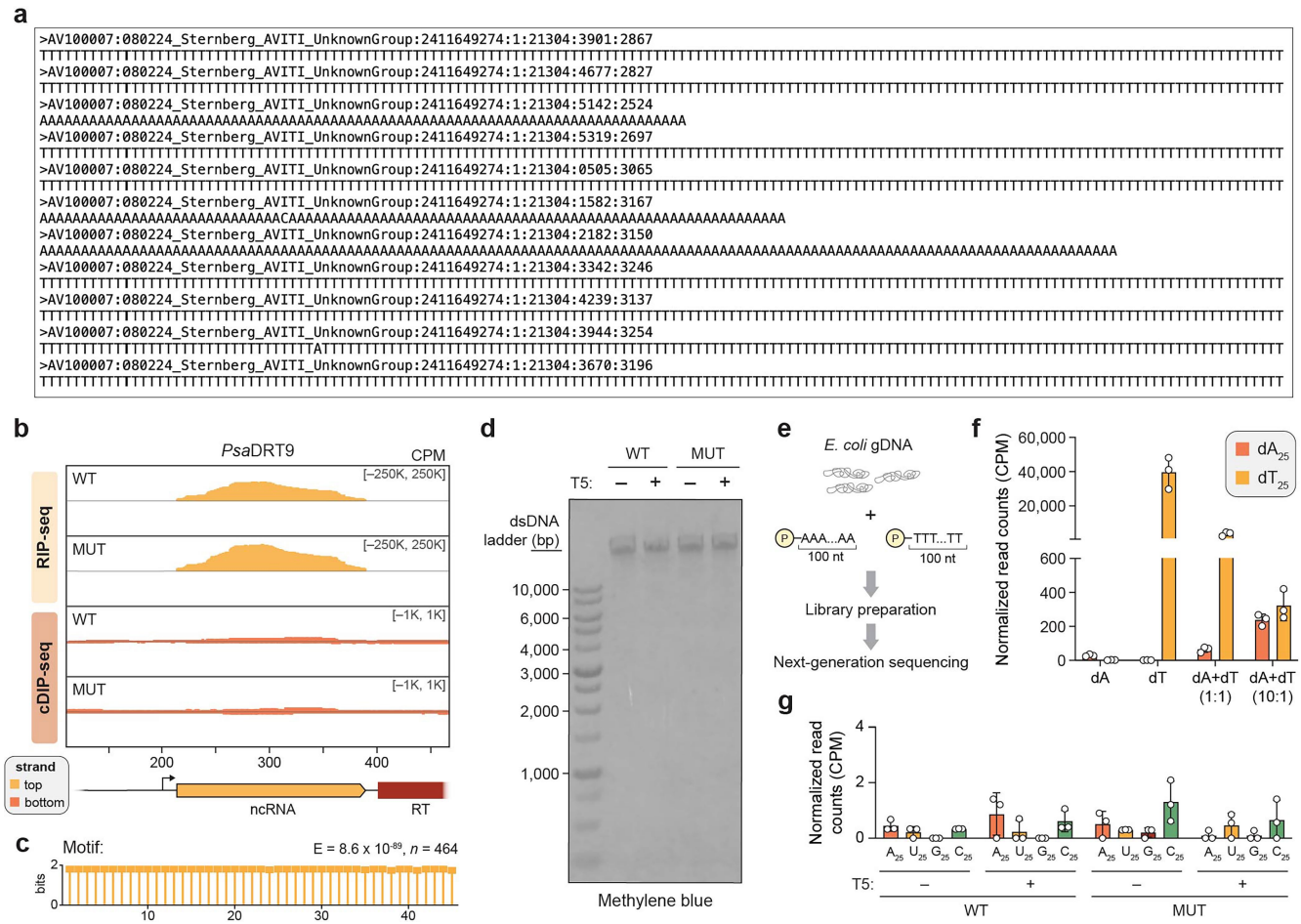
Correspondence and requests for materials should be addressed to Blake Wiedenheft or Samuel H. Sternberg.

Peer review information *Nature* thanks Jeff Miller and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

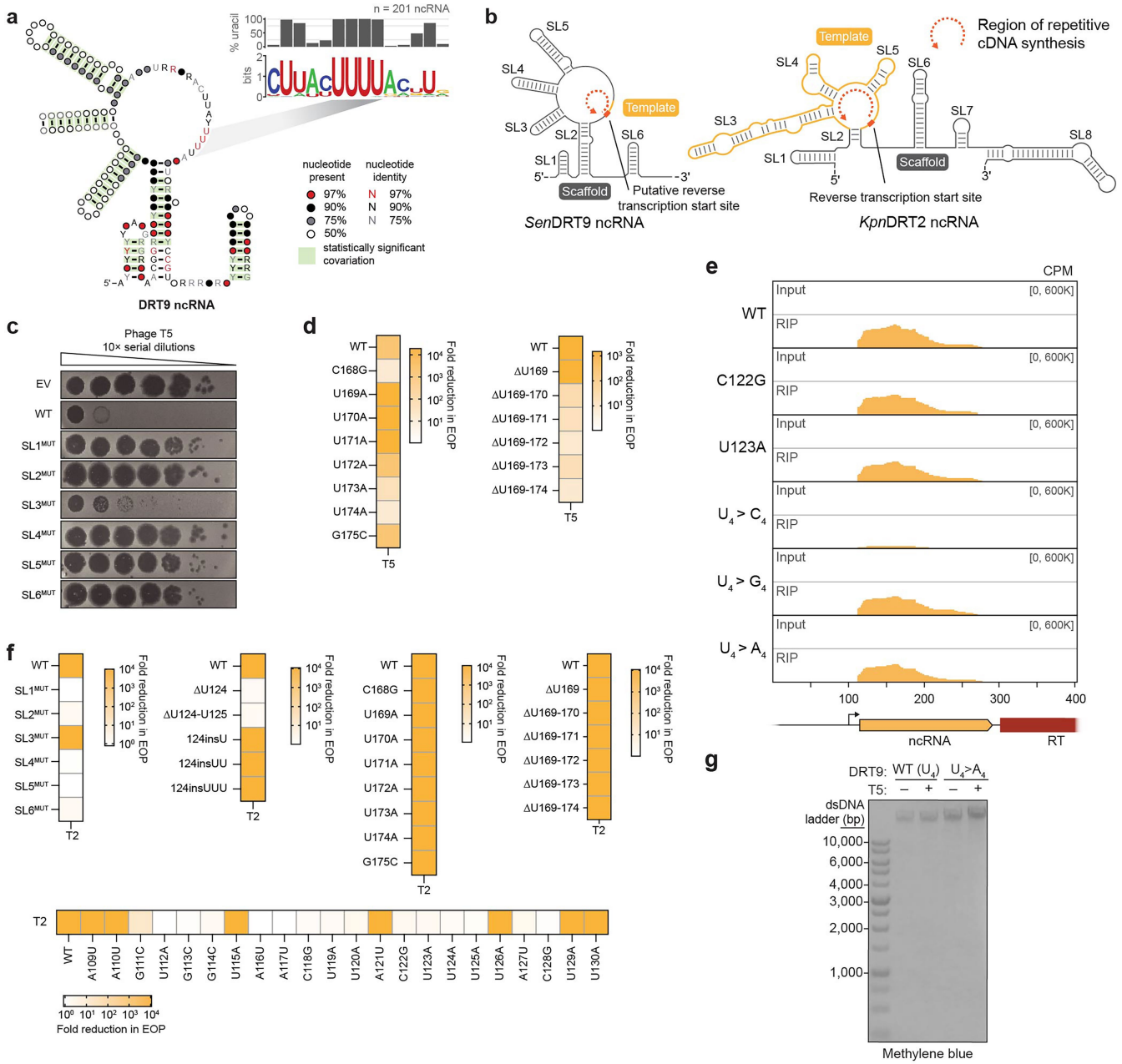
Reprints and permissions information is available at http://www.nature.com/reprints.



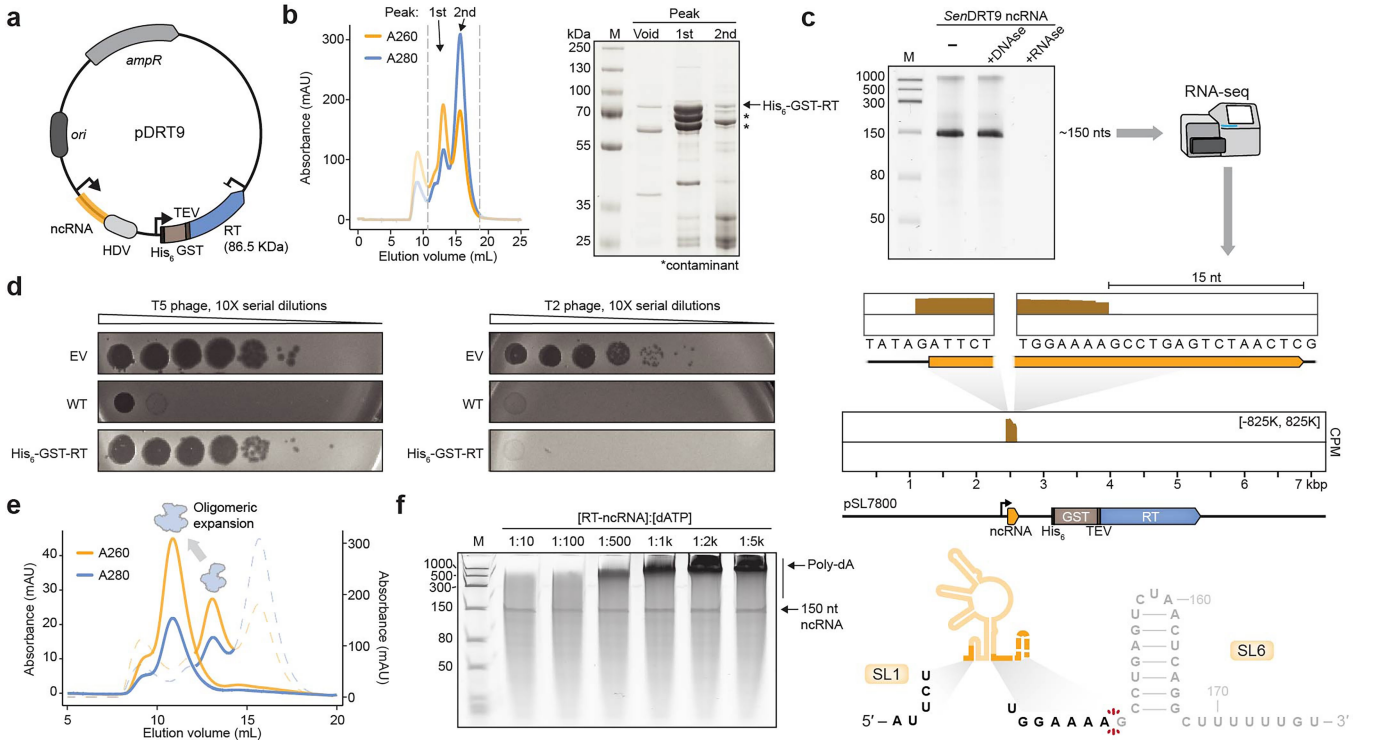
Extended Data Fig. 1 | Screening and identification of active DRT9 immune systems. a, Phylogenetic tree of DRT9-encoded RT homologues. Outer rings show SLATT protein and ncRNA association within nearby genomic neighbourhoods, and the inner ring shows previously identified DRT9 (UG28) homologues[3]; systems selected for experimental testing are indicated with red circles. **b**, Heatmap of pairwise amino acid sequence identity percentages among DRT9-encoded RT homologues tested in this study (left), and heatmap of phage defence activity for the same DRT9 systems tested against 10 diverse *E. coli* phages (right); RT proteins encoded N-terminal FLAG tags. **c**, Representative plaque assays demonstrating that *Sen*DRT9 and *Psa*DRT9 exhibit broad defence against phages from the *Tequatrovirus* and *Tequintavirus* genera, as compared to an empty vector (EV) control. **d**, Plaque assays demonstrating that defence activity against T2 phage is completely abolished for both *Sen*DRT9 and *Psa*DRT9 systems encoding catalytically inactive RT mutants (MUT). **e**, Growth curves of cells expressing WT or MUT *Sen*DRT9 +/− T5 phage at the indicated multiplicity of infection (MOI). Data are shown as mean ± s.d. for n = 3 independent biological replicates. **f**, Plaque forming unit (PFU; top) and colony forming unit (CFU; bottom) measurements from an infection time course experiment in which cells expressing WT or MUT *Sen*DRT9 were infected with T5 phage at a high MOI. Cells expressing the WT defence system attenuated phage replication but were unable to recover from infection, indicating that defence activation leads to abortive infection and cell death. Data are shown as mean ± s.d. for n = 3 independent biological replicates. **g**, Plaque assays demonstrating that *Sen*DRT9 encoding an RT with an N-terminal FLAG, but not C-terminal FLAG, retains WT defence against T5 phage. **h**, AlphaFold 3 structure prediction of the RT monomer from *Sen*DRT9, highlighting the predicted positions of the N- and C-termini (orange spheres) and YADD active site (red spheres).



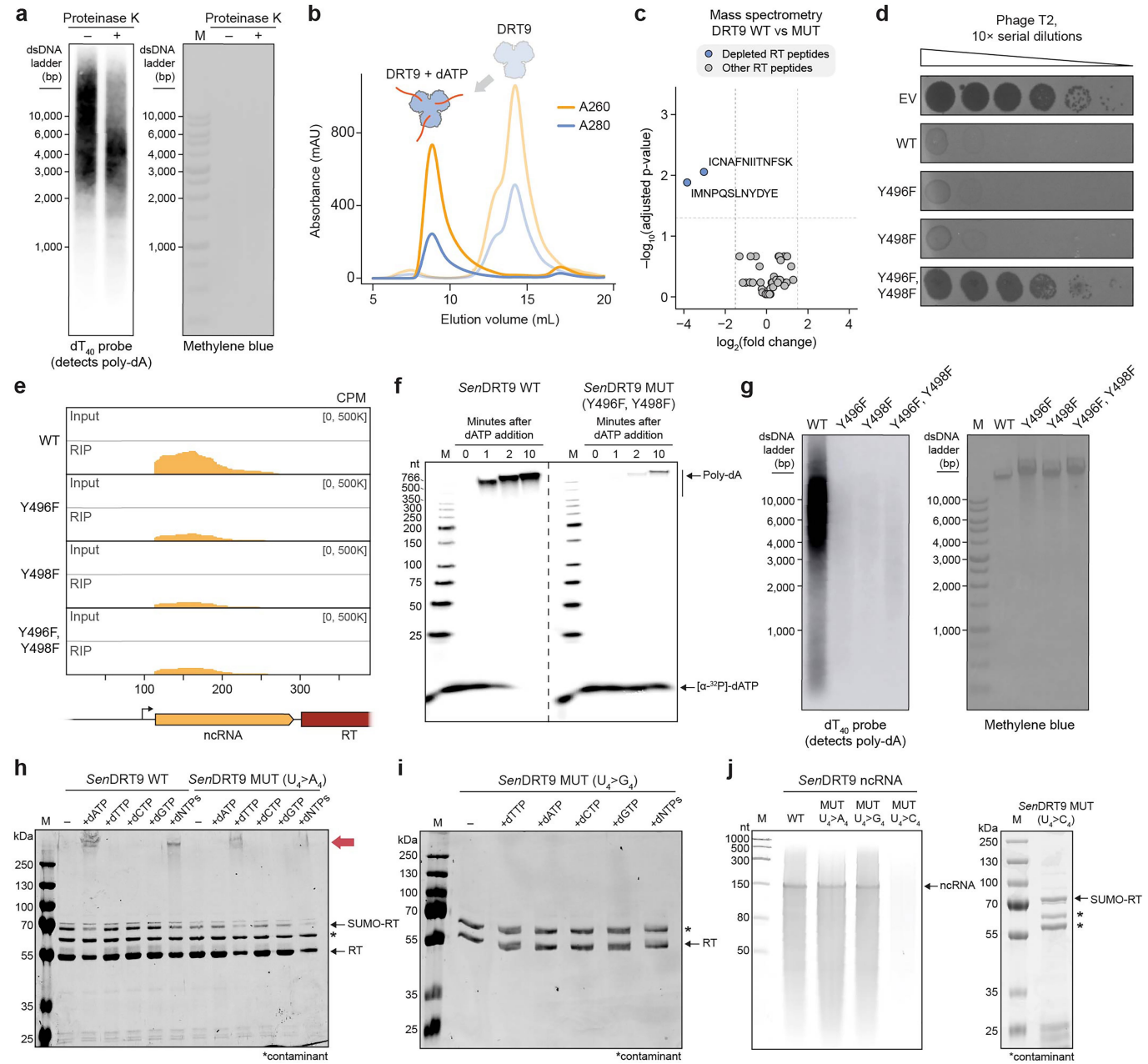
Extended Data Fig. 2 | Discovery and characterization of homopolymeric DNA products elicited by DRT9 immune systems. a, Unmapped reads from a cDIP-seq dataset of *Sen*DRT9-expressing cells infected with T5 phage, showing uninterrupted strings of poly-dT and poly-dA. **b**, RIP-seq and cDIP-seq coverage tracks (top to bottom) for either WT or RT-inactive (MUT) *Psa*DRT9, in the presence of T5 phage infection. A schematic of the genomic locus is shown below; data are normalized for sequencing depth and plotted as counts per million reads (CPM). **c**, MEME analysis results revealing a poly-dT motif enriched in unmapped reads from the WT + T5 cDIP-seq dataset in **b**. *E*, E-value significance; *n*, number of contributing sites. **d**, Methylene blue-stained membrane used for the Southern blot shown in Fig. 1h. Representative data are shown for experiments repeated at least two times with similar results. **e**, Schematic of oligo spike-in experiment to address potential bias in next-generation sequencing-based detection of poly-dA and poly-dT on the AVITI platform; P, phosphorylated 5′ end. **f**, Bar graph of dA_{25} and dT_{25} counts from total DNA sequencing of oligo spike-in experiments schematized in **e**; the apparent bias against poly-dA capture and sequencing leads to artificially elevated levels of dT_{25}-containing reads relative to dA_{25}-containing reads. **g**, Bar graph of normalized homopolymer counts from total RNA-seq datasets of WT and MUT *Sen*DRT9-expressing cells +/− T5 phage infection. These data demonstrate that poly-dA and poly-dT cDNAs are not transcribed, in contrast to the cDNA products of DRT2. Data in **f**,**g** are shown as mean ± s.d. for n = 3 independent biological replicates.



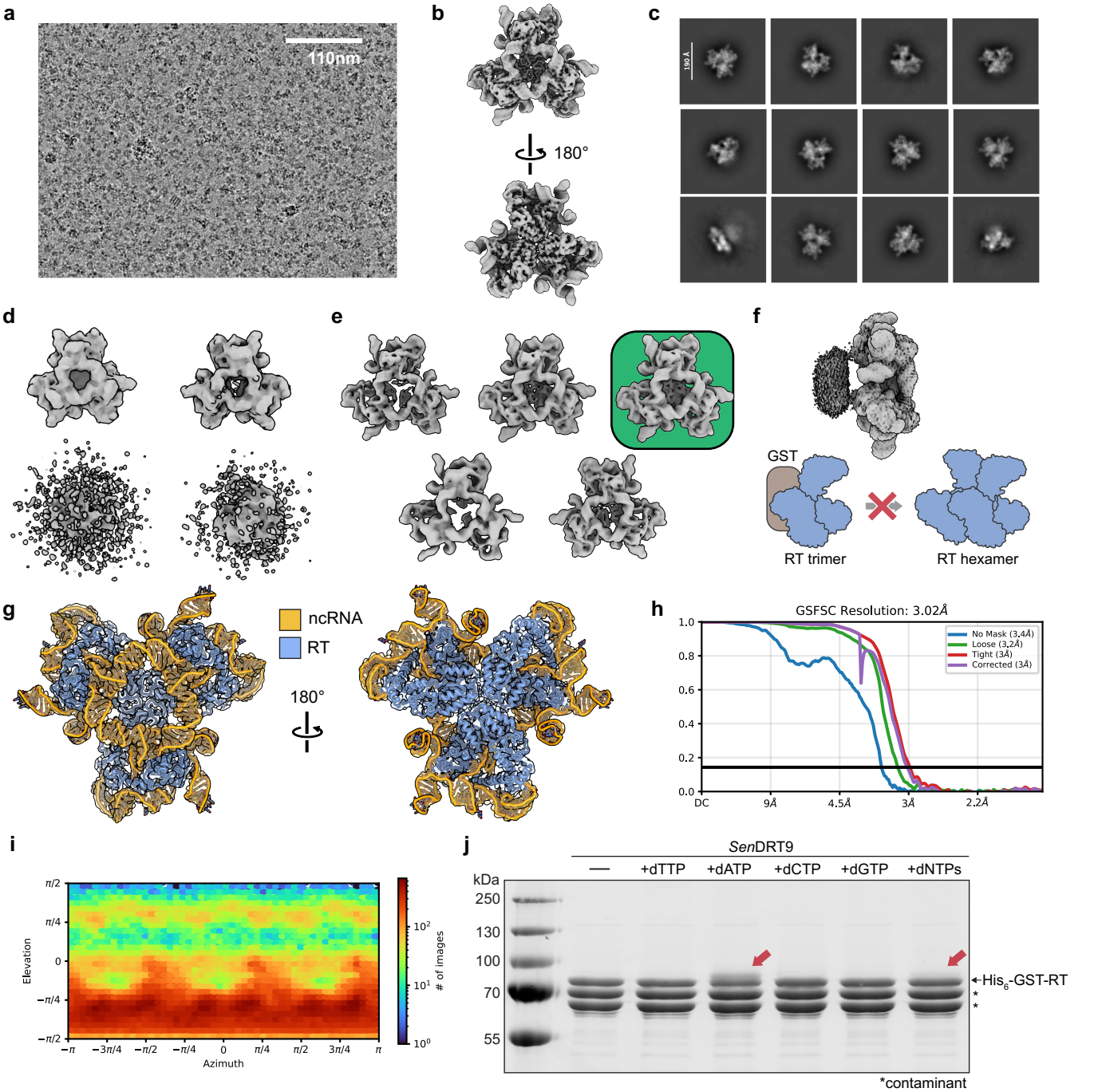
Extended Data Fig. 3 | Additional characterization of ncRNA sequence perturbations on *Sen*DRT9 defence activity. a, Covariance model of the DRT9 ncRNA from an analysis of 201 homologous systems (left), and WebLogo from a multiple sequence alignment centered around the putative U-rich template region (top right). **b**, Comparison of *Sen*DRT9 (left) and *Kpn*DRT2 (right) ncRNAs, highlighting the scaffold (grey) and template (orange) regions. Both ncRNAs template cDNA synthesis from a similar location relative to SL2 (reverse transcription start site, in red), and program repetitive cDNA synthesis across the template region. **c**, Representative plaque assays for the data shown in Fig. 2b. **d**, Heat map quantifying *Sen*DRT9 defence activity against T5 phage for the indicated ncRNA mutations and deletions within the 3′-proximal SL6 and U-rich region, quantified as the fold reduction in EOP relative to an empty vector (EV) control; data are shown as the mean of n = 2 technical replicates. **e**, RIP-seq coverage tracks for *Sen*DRT9 with WT ncRNA or the indicated ncRNA mutations in uninfected cells; the bottom three variants are mutated in the putative template region (residues 123–126). A schematic of the genomic locus is shown below; data are normalized for sequencing depth and plotted as counts per million reads (CPM). **f**, Heat map quantifying *Sen*DRT9 defence activity against T2 phage for the same ncRNA mutations tested against T5 phage in Fig. 2b–d and panel **d**; data are shown as in **d**. **g**, Methylene blue-stained membrane used for the Southern blot shown in Fig. 2f. Representative data are shown for experiments repeated at least two times with similar results.



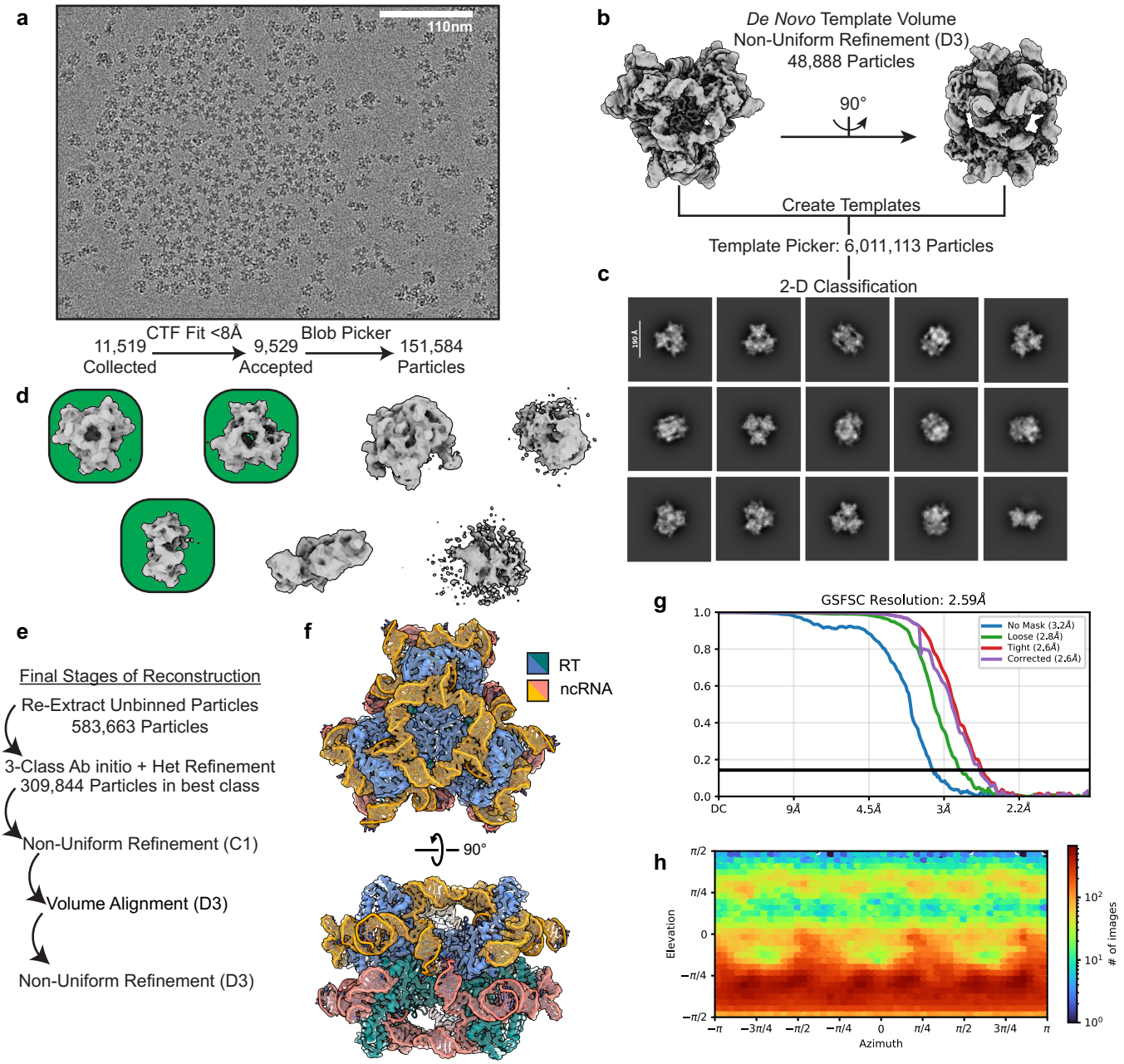
Extended Data Fig. 4 | Purification and characterization of the *Sen*DRT9-encoded RT-ncRNA complex. a, *E. coli* expression vector design for the *Sen*DRT9-encoded ncRNA and His_6-GST-tagged RT. **b**, Size-exclusion chromatogram of the His_6-GST-tagged RT-ncRNA complex separated on a Superdex 200 10/300 column (left), and SDS-PAGE analysis of the void volume and labelled peaks (right); the high A_{260}:A_{280} ratio is consistent with a protein-nucleic acid complex. **c**, Denaturing, SYBR Gold-stained 10% urea-PAGE analysis of the ~150-nt ncRNA species co-purifying with *Sen*RT after RNase or DNase treatment (top), and RNA-seq analysis of the ncRNA (middle). The mature ncRNA carries an extraneous 5′-G resulting from the T7 promoter and lacks SL6 at the 3′ end (bottom), suggesting that SL6 may be involved in transcriptional termination in vivo while being dispensable for RT-mediated poly-dA synthesis. **d**, Plaque assay showing loss of *Sen*DRT9 defence activity for a His_6-GST-tagged RT variant against T5 phage (left), but not T2 phage (right); EV, empty vector. **e**, Overlaid chromatograms from gel filtration experiments with RT-ncRNA complexes before and after TEV protease treatment of the His_6-GST-RT fusion protein, revealing a shift to earlier retention volume and thus increased oligomeric state; the persistent high A_{260}:A_{280} ratio suggests that the RT-ncRNA interaction remains intact. **f**, Denaturing 8% urea-PAGE analysis of DNA polymerization assays that contained 150 nM RT-ncRNA complex and increasing amounts of dATP per RT monomer, as indicated. Reactions were incubated at 37 °C for 60 min, followed by proteinase K treatment and phenol-chloroform extraction prior to electrophoretic separation; the gel was stained with SYBR Gold. For **b,c,f**, representative data are shown for experiments repeated at least two times with similar results.



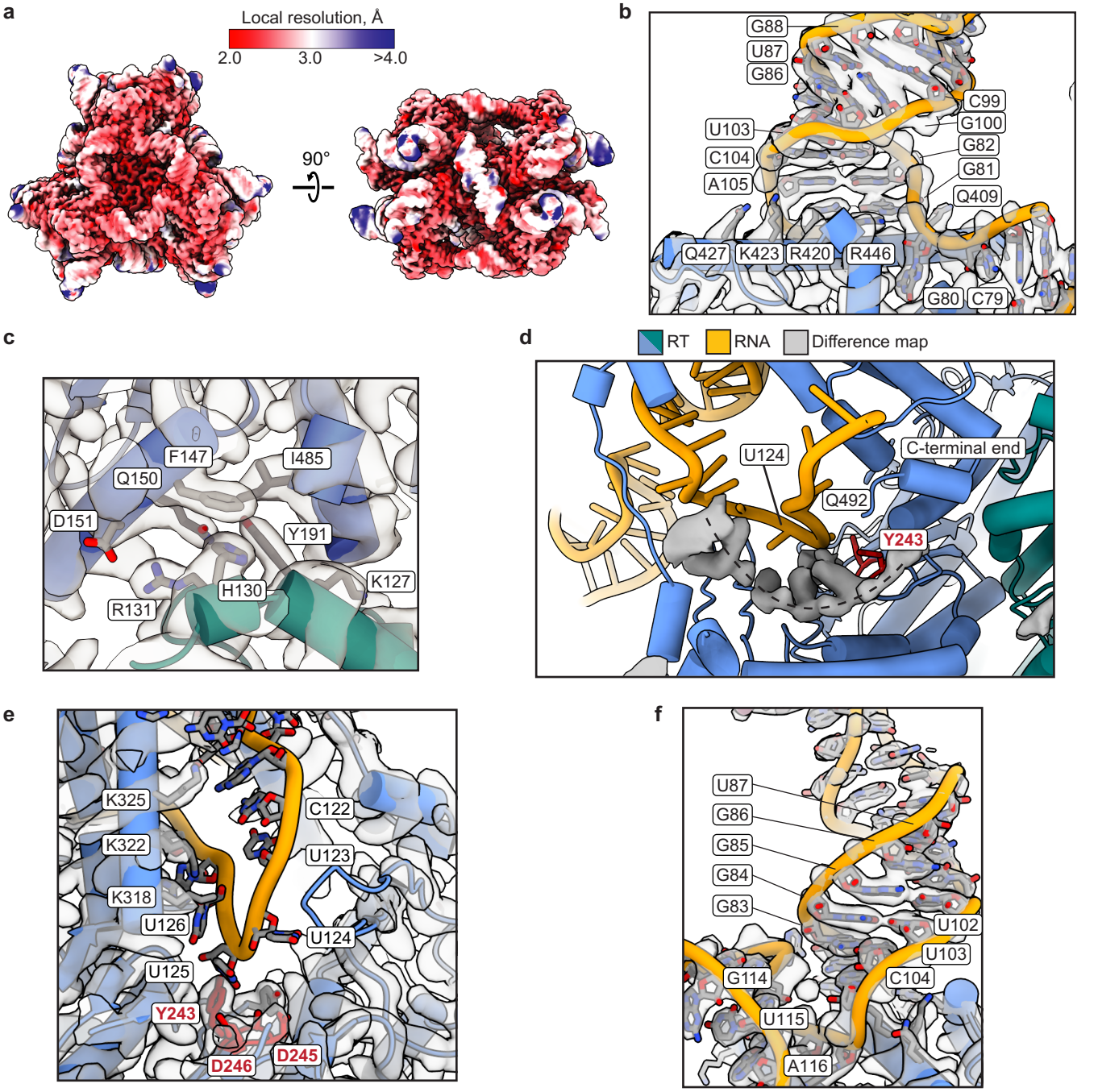
Extended Data Fig. 5 | Roles of C-terminal tyrosine residues in protein-primed reverse transcription and DRT9 phage defence. a, Southern blot analysis of total DNA isolated from T5-infected cells expressing WT *Sen*DRT9 +/− proteinase K treatment. The blot was probed with oligo-dT_{40} (left) to detect poly-dA species; the methylene blue-stained membrane after transfer is shown at right; M denotes a ladder marker. **b**, Overlaid chromatograms from gel filtration analysis of RT-ncRNA complexes +/− dATP incubation, revealing an increased A_{260}:A_{280} ratio and a dramatic shift in retention volume in the presence of dATP. **c**, Volcano plot of differential peptide abundance from mass spectrometry analysis of WT versus catalytically inactive (MUT) RT proteins immunoprecipitated from T5-infected cells expressing *Sen*DRT9. Each dot indicates a *Sen*RT peptide, and blue dots indicate significantly depleted peptides with $\log_2$(fold change) ≤ 1.5 and adjusted p-value < 0.05, as determined by unpaired two-tailed t-test with correction for multiple comparisons using the Benjamini-Hochberg method; the most depleted peptide, IMNPQSLNYDYE, contains two tyrosine residues likely involved in priming of poly-dA synthesis. **d**, Plaque assay showing loss of *Sen*DRT9 defence activity against T2 phage for a double Y496F,Y498F mutant, but not single Y496F or Y498F mutants; EV, empty vector. **e**, RIP-seq coverage tracks for *Sen*DRT9 in T5-infected cells with WT and single or double Y496F/Y498F RT mutants, as indicated. A schematic of the genomic locus is shown below; data are normalized for sequencing depth and plotted as counts per million reads (CPM). **f**, Denaturing 5% urea-PAGE analysis of DNA polymerization assays with WT (left) and Y496F,Y498F mutant (right) RT. All reactions contained 20 nM RT-ncRNA and 100 µM [α-^{32}P]-dATP; M denotes a DNA ladder marker. **g**, Southern blot analysis of total DNA isolated from T5-infected cells expressing *Sen*DRT9 with WT and single or double Y496F/Y498F RT mutants, as indicated. The blot was probed with oligo-dT_{40} (left) to detect poly-dA species; the methylene blue-stained membrane after transfer is shown at right; M denotes a DNA ladder marker. **h**, SDS-PAGE analysis of DNA polymerization assays with either WT *Sen*DRT9 or a $U_4 > A_4$ mutant ncRNA. RT-ncRNA complexes (0.15 µM) were incubated with the indicated dNTPs (0.9 mM). The red arrow denotes high-molecular weight (MW) protein-DNA conjugates; the SUMO-RT band arises from incomplete tag removal; *, purification contaminant; M, protein ladder marker. **i**, SDS-PAGE analysis of DNA polymerization assays as in **h**, but with a $U_4 > G_4$ mutant ncRNA; no high-MW species are observed. **j**, Denaturing, SYBR Gold-stained 10% urea-PAGE analysis (left) of WT, $U_4 > A_4$, and $U_4 > G_4$ ncRNAs co-purifying with *Sen*RT. The absence of a band for the $U_4 > C_4$ ncRNA, despite a clear band corresponding to the RT by SDS-PAGE (right), indicates that the $U_4 > C_4$ mutation disrupts RT-ncRNA complex formation. For **a**,**f-j**, representative data are shown for experiments repeated at least two times with similar results.



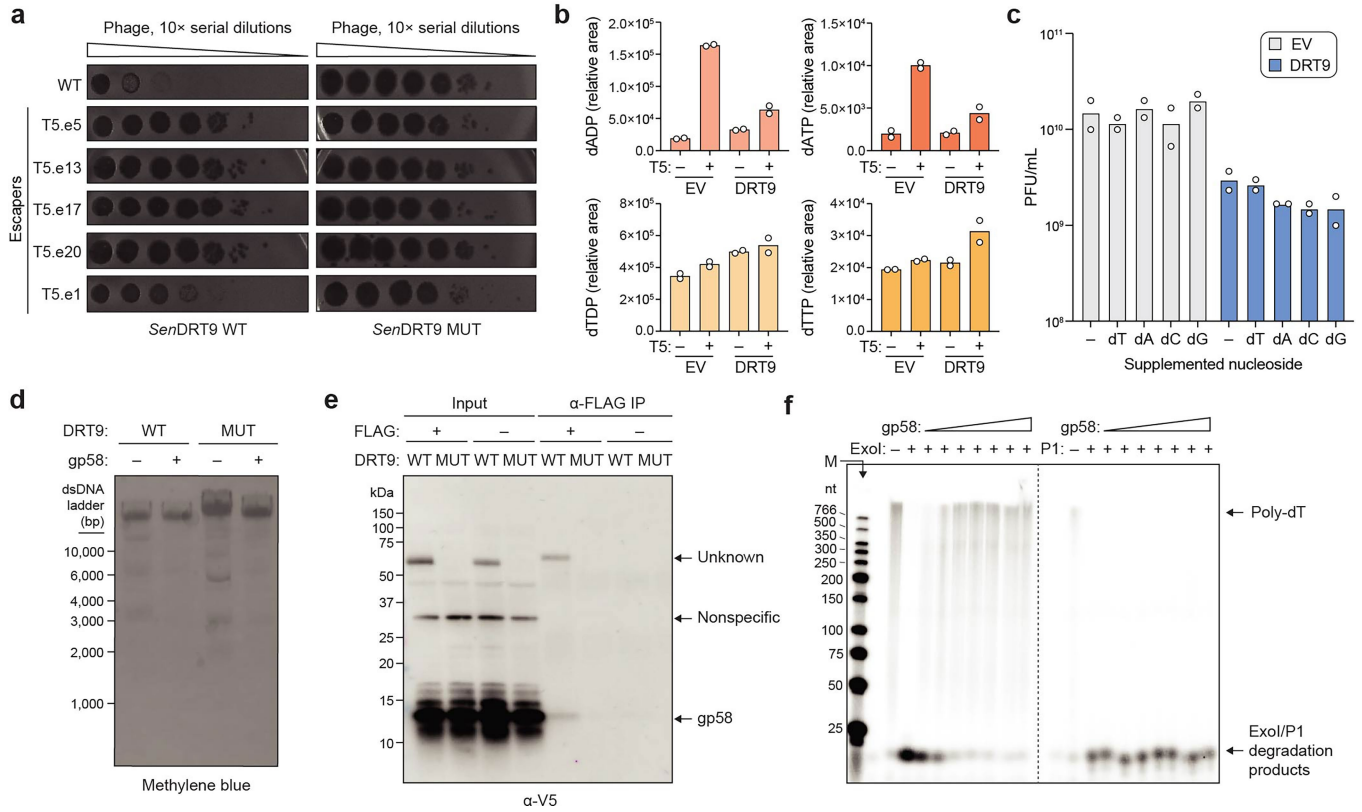
Extended Data Fig. 6 | Cryo-EM analysis and image processing workflow for trimeric *Sen*DRT9 RT-ncRNA complex. a, Representative micrograph from *Sen*DRT9 RT-ncRNA complex data collection with the His_6-GST fusion protein, revealing a densely-packed field of particles in vitreous ice. **b**, Initial consensus volume after processing blob-picked particles, which refined to 3.7 Å resolution (C3-reconstruction) for de novo template generation and subsequent particle picking. **c**, Representative selected 2D classes from 5,235,086 initial particles picked using the template in panel **b**. **d**, Four-class *ab initio* reconstruction from 2,530,970 particles in selected classes to remove junk particles. **e**, Results of a five-class 3D classification without a mask and filtered at 8 Å resolution, with a class similarity of 0.1 to remove remaining poorly aligned particles. The best volume, highlighted by a green background, contained 367,640 particles that were selected for further processing. **f**, View of the final reconstruction in **g** shown at high contour (top), revealing noisy density at the N-terminus of each RT monomer corresponding to the His_6-GST fusion used for expression and purification. We hypothesized that the GST fusion would prevent the formation of higher-order oligomers (bottom). **g**, Final (C3) reconstruction that refined to 3 Å resolution and was used for building the trimeric RT-ncRNA complex model. The map is shown with partial transparency and coloured according to the model, with RT subunits in blue and ncRNA subunits in orange. **h**, Plot of Cryosparc's gold standard Fourier shell correlation for the map in **g**. **i**, Plot of particle view distribution reveals mild anisotropy but no missing views. **j**, SDS-PAGE analysis of DNA polymerization assays with His_6-GST-RT sample, in which RT-ncRNA complexes were incubated with the indicated dNTP(s) before reactions were quenched and resolved electrophoretically. Red arrows indicate preliminary evidence of protein-DNA conjugates in reactions that contained dATP. Representative data are shown for experiments repeated at least two times with similar results.



Extended Data Fig. 7 | Cryo-EM analysis and image processing workflow for hexameric *Sen*DRT9 RT-ncRNA complex. a, Representative micrograph from the *Sen*DRT9 RT-ncRNA hexameric complex, revealing a monolayer of evenly distributed particles. **b**, Initial reconstruction from on-the-fly image analysis that was used to generate de novo templates used for particle picking. **c**, Representative 2D Classes from 6,011,113 template-picked particles. **d**, Initial volumes from a 7-class *ab initio* reconstruction of 1,411,369 particles. Selected volumes for downstream processing are shown with a green background. **e**, Final steps used to sort selected particles in **d** to yield the final reconstruction. **f**, Final (D3) reconstruction of the *Sen*DRT9 RT-ncRNA hexamer that refined to 2.6 Å and was used for model building. The map is shown as a partially transparent surface, with RT subunits shown in blue/cyan and ncRNA subunits shown in orange/salmon. **g**, Plot of Cryosparc's gold standard Fourier shell correlation for the map in **f**. **h**, Plot of particle view distribution reveals mild anisotropy but no missing views.



Extended Data Fig. 8 | Additional features of the *Sen*DRT9 hexamer visualized by cryo-EM. a, Local resolution map of the *Sen*DRT9 RT-ncRNA hexamer (FSC cutoff 0.143), showing that resolution falls off near the periphery of the complex in flexible regions such as SL3 and SL4. **b**, Semi-transparent density highlighting a network of protein-RNA interactions that stabilize SL5. **c**, Semi-transparent density highlighting protein-protein interactions that stabilize the back-to-back assembly of trimers. **d**, Difference map generated by subtracting a calculated map of the model from the experimental density. Unmodelled density (grey) is visible in the map directly adjacent to the YADD active site (red), which may correspond to a small stretch of poly-dA product; repeated efforts to sort these particles failed to separate distinct conformational states. **e**, A close-up view of the *Sen*RT active site (YADD residues in red). Density for the RT is shown as a semi-transparent surface, and the density map for the ncRNA is omitted here for simplicity. **f**, Semi-transparent density map highlighting the ncRNA poly-G track (residues 80-86) that was used to establish register for ModelAngelo.



Extended Data Fig. 9 | Activation of DRT9 immunity by phage-encoded factors. a, Representative plaque assays with WT or escaper T5 phage variants tested in strains expressing either the WT or RT-inactive (MUT) *Sen*DRT9 system. **b**, Bar graphs of relative nucleotide levels for the indicated species quantified by LC-MS/MS, in lysates from cells expressing *Sen*DRT9 or an empty vector (EV) control +/− T5 phage infection. Data are shown as the mean for n = 2 independent biological replicates. **c**, Bar graphs of plaque forming units per mL (PFU/mL) of T5 phage lysates after liquid culture infection of cells expressing *Sen*DRT9 or an EV control, in the presence of the indicated supplemented nucleosides; data are shown as the mean for n = 2 independent biological replicates. **d**, Methylene blue-stained membrane of total DNA from cells expressing WT or MUT *Sen*DRT9 +/− gp58 induction, for the Southern blot shown in Fig. 5f. **e**, Western blot analysis of *Sen*RT co-immunoprecipitation with gp58. Immunoprecipitation was performed with a FLAG antibody on lysates from cells co-expressing *Sen*RT (WT or MUT, FLAG-tagged or untagged) with V5-tagged gp58, followed by Western blot analysis using a V5 antibody. The unique presence of a band corresponding to gp58 in the IP eluate from cells expressing WT *Sen*DRT9 indicates a DNA-dependent interaction between the RT and gp58. **f**, Denaturing 5% urea-PAGE analysis of DNA polymerization assays, after reactions were treated with ExoI (left) or Nuclease P1 (right) in the presence of increasing concentrations of T5 gp58. All reactions contained 20 nM RT-ncRNA ($U_4 > A_4$) and 100 μM [α-^{32}P]-dTTP and were incubated at 37 °C for 60 min prior to addition of gp58 and ExoI or Nuclease P1. For **d**-**f**, representative data are shown for experiments repeated at least two times with similar results.

Extended Data Table 1 | Cryo-EM data collection, refinement, and validation statistics

	#1 GST-DRT9 Trimer (EMDB-49525) (PDB 9NLX)	#2 DRT9 Hexamer, Pre-Polymerization State (EMDB-49523) (PDB 9NLV)
Data collection and processing		
Magnification	45,000	45,000
Voltage (kV)	200	200
Electron exposure (e–/Å^2)	59.5	59.8
Defocus range (μm)	-0.8, -2.5	-0.8, -2.5
Pixel size (Å)	.9061	.9061
Symmetry imposed	C3	D3
Initial particle images (no.)	5,235,086	6,029,711
Final particle images (no.)	367,640	309,844
Map resolution (Å)	3.02	2.59
FSC threshold	0.143	0.143
Map resolution range (Å)	2.6-6.0Å	2.0-4.5
Refinement		
Initial model used (PDB code)	AlphaFold3, ModelAngelo	AlphaFold3, ModelAngelo
Map sharpening methods	Phenix Half-Map	Phenix Half-Map
Refinement Package	Phenix	Phenix
Refinement Method	RealSpaceRefinement	RealSpaceRefinement
Model composition		
Non-hydrogen atoms	20,076	41,424
Protein residues	1,443	2,952
Nucleotide Residues	441	882
Ligands	0	0
***B* factors (Å^2)**		
Protein	54.73	86.41
Nucleotide	75.05	145.79
Ligand	N/A	N/A
R.m.s. deviations		
Bond lengths (Å)	0.003 (0)	0.004 (0)
Bond angles (°)	0.772 (0)	0.694 (0)
Validation		
MolProbity score	1.67	1.33
Clashscore	8	9
Poor rotamers (%)	0.5(5)	0(1)
Ramachandran plot		
Favored (%)	99	98
Allowed (%)	1	2
Disallowed (%)	0	0

This table lists statistics related to cryo-EM structures of the *Sen*DRT9 RT-ncRNA complex.

nature portfolio

Corresponding author(s): Samuel H. Sternberg, Blake Wiedenheft

Last updated by author(s): May 9, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Next-generation sequencing data was collected using an Element Biosciences AVITI platform. Cryo-EM data was collected using SerialEM (v4.2.0) under the control of Smartscope (v.0.9.4). Mass spectrometry data was collected using on Waters M-Class UPLC using Xcalibur software (v4.5.474.0). Metabolite detection was performed using using an HPLC (1290 Infinity II, Agilent Technologies) coupled to a triple quadrupole mass spectrometer (6495D, Agilent Technologies).
Data analysis	Structural figures were generated using ChimeraX (v1.9). Protein structures were predicted using AlphaFold (v3). Cryo-EM image analysis was carried out in cryoSPARC (v4.6.2). Modeling of the RT-ncRNA complex was performed using ModelAngelo (v1.0.13) and COOT (v0.9.8.1), and model refinement was carried out in Phenix (v1.21.2-5419). For DRT9 ncRNA searches, covariance models were built with LocaRNA (v1.9.1) and searches were executed with Infernal (v1.1.4). For analysis of DRT9 loci, MAFFT (v7.505) was used for alignments, FastTree (v2.1.11) was used to build phylogenies, and iTOL (online) was used to visualize trees. Clustering prior to alignment for both DRT9 ncRNA and DRT9 loci was performed using CD-HIT (v4.8.1). DNA and RNA sequencing data was processed with cutadapt (v4.2), aligned with bwa-mem2 (v2.2.1) , filtered and converted to BAM with SAMtools (v1.17), and converted to bigwig files with bamCoverage (v3.5.1). Coverage tracks were viewed in IGV (v2.17.4). Transcript counts were calculated with featureCounts (v2.0.2) and log2 (fold changes) were calculated with DESeq2 (v1.40.2). Unmapped reads were extracted for downstream analysis using SAMtools fasta. Motif analysis was performed using MEME (v5.5.7). Phage escaper genome alignment was performed using Bowtie2 (v2.2.1), and mutations were identified with breseq (v.0.39.0). Mass spectrometry data were searched against a proteome reference using MaxQuant (v2.0.3.0). Nucleotide quantification by LC-MS/MS data analysis was performed with Skyline-daily (v24.1.1.398). No custom code was developed for this manuscript.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Next-generation sequencing data are available in the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) under the accessions GSE295895, GSE295896, and GSE295897. The published genomes used for bioinformatics analyses were obtained from NCBI (Supplementary Table 1). Cryo-EM maps and experimentally determined models were deposited to the appropriate public database (i.e. EMDB, PDB) under the accessions EMD-49525, PDB:9NLX (GST-DRT9 Trimer) and EMD-49523, PDB:9NLV (DRT9 Hexamer). Previously determined protein structures used for structural comparisons were obtained from the PDB (7R06; 8FAK). The phylogenetic tree of UG RT homologs was obtained from https://doi.org/10.1093/nar/gkac467. Mass spectrometry data are available in the MassIVE database under the accession MSV000097825. Source data is available for Figures 1e, 1g, 1i, 2b, 2c, 2d, 5c, 5d, 5h, 6a, 6e, and Extended Data Figures 1b, 1e, 1f, 2f, 2g, 3d, 3f, 4b, 4e, 5b, 5c, 9b, 9c. All other data are available in the manuscript or in Supplementary Information.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were chosen based on previous studies in the phage defense field and are reported in the figure legends. Standard sample sizes were followed for next generation sequencing (Wiegand et al., 2024), biochemistry (Hobbs et al., 2022), mass spectrometry (Tal, Morehouse et al., 2021, Parnas, Jovanovic, Eisenhaure et al., 2015), and bacterial and phage growth assays (Stokar-Avihail et al., 2023, Gao et al., 2020).
Data exclusions	No data were exluded.
Replication	All data could be reproduced. Experiments presented in the manuscript were repeated at least two times with similar results.
Randomization	Randomization was not applicable for this study design, as it does not involve selecting samples from a larger population.
Blinding	Investigators were not blinded as it was not applicable for this study design. There were no subjective analyses performed that could be affected by the interpretation of the data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Monoclonal ANTI-FLAG M2 antibody (mouse), Sigma Aldrich, catalog number: F1804, clone M2;
Monoclonal V5-tag (D3H8Q) antibody (rabbit), Cell Signaling Technology, catalog number 13202;
Monoclonal mouse anti-rabbit IgG (L27A9) , Cell Signaling Technology, catalog number 5127.

Validation

According to the manufacturer, the "monoclonal ANTI-FLAG® M2 may be used in IP [immunoprecipitation] procedures when used in conjunction with an insoluble carrier matrix, such as a Protein G resin" (https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/175/747/f1804bul-ms.pdf). As suggested, the ANTI-FLAG M2 antibody was used together with Dynabeads Protein G resin (Thermo Fisher) in this study.
According to the manufacturer, V5-tag and IgG antibodies are suitable for Western Blotting. It is recommended to incubate membrane with diluted primary antibody [V5-tag] in 5% w/v BSA, 1X TBS, 0.1% Tween® 20 at 4°C with gentle shaking overnight (https://www.cellsignal.com/products/primary-antibodies/v5-tag-d3h8q-rabbit-mab/13202?srsltid=AfmBOoo_m-toXk8LNLQ2eO42KbyEgA1VnzmpocWpYSh9lm1AZKlqVK8v), and the procedure was followed in this study.

Plants

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A