Structural Basis for Retron Co-option of Anti-phage ATPase-nuclease

Bing Wang[1], Renee Hoffman[1], Ya-Ming Hou[2], and Hong Li[1]*

[1]Department of Structural Biology, Van Andel Institute, Grand Rapids, MI 49503

[2]Department of Biochemistry, Thomas Jefferson University, Philadelphia, PA, USA.

Highlights

- Ec78 reverse transcriptase and msDNA sequester PtuAB toxicity
- Ec78 PtuAB heterocomplex targets host $tRNA^{Tyr}$
- ATPase and HNH activities are required for Ec78 PtuAB function
- Ec78 PtuA evolved novel domains for binding retron and for self-assembly
- Ec78 requires distinct ncRNA and msDNA processing for binding PtuAB
- Ec78 retron facilities efficient human genome editing

*Corresponding author: hong.li@vai.org, Department of Structural Biology, Van Andel Institute, Grand Rapids, MI 49503

Summary

Retrons are recently identified bacterial defense systems that employ a tripartite of reverse transcriptase, non-coding RNA (ncRNA) and its derived multi-copy single stranded DNA (msDNA) to sequester effector activity. Phage invasion activates retrons, triggering effector activity and inducing abortive infection and cell growth arrest. Ec78 distinguishes itself from other retrons by leveraging the Septu defense system, a stand-alone ATPase-nuclease pair (PtuAB), by reshaping its phage sensing and molecular assembly processes. To elucidate the evolutionary adaptation of PtuAB by Ec78, we determined electron cryomicroscopy structures of Ec78 as well as the retron-displaced PtuAB at near atomic resolutions. We show that the Ec78-associated ATPase, PtuA, acquired unique elements that enable its interactions with the reverse transcriptase and the msDNA, as well as self-assembly when displaced by the retron. Supported by biochemical and mutational analysis, we also show that the retron-displaced PtuAB forms a tetramer, unlike its stand-alone counterpart, that indiscriminately restricts the host but, in the presence of the retron, confers a well-controlled immune response, eliciting ATP hydrolysis- and msDNA-regulated targeting to host $tRNA^{Tyr}$. Our studies reveal an evolutionary principle for retrons to co-opt conserved enzyme modules for defense in response to different cellular needs.

43 **Introduction**

44

45 Once considered selfish genetic elements,[1] the widely distributed bacterial genetic elements known
46 as retrons have recently been shown to play a role in anti-phage defense.[2,3]. Retrons protect the
47 bacterial hosts from invading phages through a broad spectrum of effector proteins that include
48 membrane proteins, DNA-binding proteins, metabolic enzymes, ATPases and nucleases.[3]
49 Investigating how retrons and their effectors contribute to this collaborative process under different
50 selection pressures would reveal their evolutionary capacity and better serve the retron-based
51 biotechnology applications.

52

53 Retrons encode a reverse transcriptase (RT) and a non-coding RNA (ncRNA, also referred to as
54 msrRNA) that serves both as the template and the primer for the synthesis of the a multi-copy
55 single-stranded DNA (msDNA, also referred to as RT-DNA) by the RT, leading to expression of
56 the RT, the ncRNA and the msDNA that form the tripartite system[4]. The msDNA is covalently
57 linked to the initiation guanine nucleotide downstream of the conserved UUA motif of the ncRNA
58 via a 2′-5′ linkage (branched). In some cases, the msDNA is separated from the ncRNA by the
59 host XseA-XseB exonuclease (ExoVII) prior to the retron-mediated defense[4,5]. Furthermore, the
60 template region of the ncRNA is processed by the host RNase H, a process required for the retron-
61 mediated defense[6], leaving behind a short template:primer duplex with the msDNA. The mature
62 RT-ncRNA-msDNA tripartite sequesters and neutralizes the cytotoxic activities of the effectors
63 that are often encoded as polycistrons along with retrons themselves [3,4,7]. Upon being activated by
64 phage infections, the effector activities lead to abortive infection and cell death. Like CRISPR-
65 Cas, the retron systems provide bacteria with another non-coding RNA-mediated defense against

66 invading phages[8]. The unique production of msDNA by retrons has been harnessed for genome
67 engineering as well as prime-editing, in conjunction with CRISPR-Cas9, across diverse
68 systems.[9,10]
69
70 Retron Ec78 (retron-Eco7) is a Type I-A retron. Of the 11 known types of retrons, Type I-A retrons
71 are unique in partnering with the stand-alone Septu defense system that is an aggregate of PtuA
72 and PtuB proteins (collectively PtuAB) [2,3,11]. The recruitment of PtuAB by Ec78 or other Type I-
73 A retrons suggests that the Septu- and the Ec78-mediated defense are regulated differently. Indeed,
74 unlike Septu that is sensitized by the phage tail fiber[12], Ec78 produces msDNA, which is consistent
75 with its response to DNA targeting proteins[13]. Despite this difference, the Septu- and the Ec78-
76 associated PtuAB share a high degree of similarity. Both PtuA proteins belong to the family of
77 ABC ATPases[14] whereas both PtuB proteins share a conserved HNH domain characteristic of the
78 ββα-Me (metal) superfamily of nucleases[15]. The ABC ATPase domain links PtuA to the
79 overcoming lysogenization defect (OLD) nucleases that also function in anti-phage defense and
80 the Mr11-Rad50 proteins for DNA repair[14]. Like a recently characterized OLD nuclease complex,
81 GajA-GajB, whose nuclease function is inhibited by ATP[16,17], the stand-alone PtuAB also shows
82 reduced DNA nicking activity as ATP level increases[18]. However, the dependence of Ec78-
83 associated PtuA on ATP remains unknown and its PtuAB targets host $tRNA^{Tyr}$ [19]. The distinct
84 hosts substrates may explain in part why the stand-alone PtuAB targets a broad spectrum of
85 phages[11,12] whereas the Ec78-associated PtuAB is specific for the T5 phage[12]. The molecular basis
86 underlying the defense function of the Ec78 retron remains unclear. Importantly, the evolutionary
87 process by which PtuAB was co-opted into the more complex retron system under the control of
88 the RT, the ncRNA and the msDNA tripartite assembly is elusive.

89

90 To elucidate the evolutionary basis for retrons to acquire the stand-alone defense systems, we
91 performed electron cryomicroscopy (cryoEM) and biochemical analysis on the *E.coli* ECONIH5
92 retron Ec78 (**Fig. 1a & 1b**). Ec78 encodes an unprocessed ncRNA of 144 nucleotides (nts) in
93 length, including 8-nt flanking inverted repeats (IRa1 and IRa2) and four RNA stem loops (stem
94 loops I-IV) of varying sizes. The largest stem loop, stem loop IV, with a 22 base-paired stem,
95 serves as the primary template for msDNA synthesis (**Fig. 1a**). Following msDNA synthesis, stem
96 loop IV is removed by RNase H, leaving behind the ncRNA without stem loop IV, a short
97 template:primer duplex, and the msDNA (**Fig. 1a**). We showed that the Ec78 ncRNA is further
98 processed on retron, leaving behind 40 nts without stem loop I and the two inverted repeats. In
99 addition, the msDNA is debranched, which enables PtuAB to bind. We also showed that, unlike
100 the stand-alone PtuAB, the PtuAB heterocomplex is the cytotoxic unit of the Ec78 retron requiring
101 both the ATPase and HNH activities. We identified the structural elements acquired by PtuAB that
102 enables its switch from the stand-alone to the retron-based defense, thereby illuminating the
103 adaptative process for the Ec78 retron.

104

105 **Results**

106

107 *PtuAB complex is the retron-sequestered toxin unit*

108

109 It was previously shown that the PtuAB components of Ec78 induce cell growth arrest[19]. To
110 understand how each of the PtuAB enzymatic functions contributes to this activity, we conducted
111 growth assays on cells expressing different combination of Ec78 components and their ATPase

112 and HNH deficient variants. We constructed expression plasmids encoding the intact retron, RT-
113 ncRNA, PtuAB, PtuA, PtuB, and their ATPase or HNH deficient variants respectively. Individual
114 plasmids were transformed into the DH10B cells and selected under appropriate antibiotics (**Fig.**
115 **1c**). Consistently, expression of PtuAB, but not the intact retron or PtuA and PtuB individually, in
116 DH10B cells severely impaired growth, indicating that PtuAB is the toxin unit (**Fig. 1c**).
117 Furthermore, mutation of the Walker B motif residue Glu396 (E396Q) of PtuA or the catalytic
118 residue His57 of the HNH domain (H57A) of PtuB in PtuAB co-expression cells rescued the
119 growth defects (**Fig. 1c**), suggesting that both the ATPase activity and the nuclease activity are
120 required for cell growth arrest. These results confirm PtuAB as the functional defense unit and
121 both the ATPase and HNH nuclease activities are necessary for the function, a conclusion that is
122 in stark contrast to that of the stand-alone PtuAB system whose activity is strictly controlled prior
123 to phage invation[18,20]. Therefore, in the Ec78 retron, the RT-ncRNA-msDNA entity exerts control
124 over the adapted PtuAB for its defense function. Upon phage infection, the PtuAB displaced by
125 Ec78 indiscriminately impairs host targets, leading to abortive immunity.
126
127 *The PtuAB complex targets host tRNA^{Tyr}*
128
129 Previous efforts to identify phage determinants that evade Ec78 defense identified tRNA^{Tyr} as a
130 potential target [19]. To confirm that PtuAB targets host tRNA^{Tyr} specifically, we co-transformed
131 plasmids that overexpress tRNA^{Tyr}-GTA-1, tRNA^{Tyr}-GTA-2, tRNA^{Ala}, tRNA^{His}, tRNA^{ser}, tRNA^{Arg},
132 tRNA^{Met}, or no tRNA (null) along with that expressing PtuAB into BL21(DE3) cells to increase
133 the expression levels (**Fig. 2a**). We showed that overexpression of both tRNA^{Tyr} (GTA-1 and
134 GTA-2) molecules, but not the null or other tRNAs, notably reduced PtuAB toxicity, indicating

135 PtuAB specifically targets $tRNA^{Tyr}$ (**Fig. 2a & Suppl. Fig. 1a**), consistent with the previous
136 finding that mutant phages overexpressing $tRNA^{Tyr}$ evade retron Ec78 [19]. The toxicity of PtuAB
137 in the non-expression cells DH10B can also be rescued by the expression of $tRNA^{Tyr}$ and $tRNA^{Ala}$
138 to a lesser degree (**Suppl. Fig. 1b**).
139
140 We further reconstituted the wild-type PtuAB complex by co-purifying them from the cells
141 overexpressing $tRNA^{Tyr}$ and the intact Ec78 complex including their catalytic variants and verified
142 that the wild-type PtuA or its complexes are active in ATP hydrolysis (**Suppl. Fig. 2**). We
143 subsequently performed *in vitro* tRNA cleavage assays. To our surprise, neither PtuAB nor PtuB
144 or the DNase I-treated Ec78 was able to degrade in vitro transcribed $tRNA^{Tyr}$-GTA-2 (**Suppl. Fig.**
145 **1c**). Furthermore, the DNase I-treated Ec78 also did not degrade the native $tRNA^{Tyr}$ isolated from
146 E coli[21] in the presence or absence of divalent ions or ATP (**Suppl. Fig. 1d**). It is possible that an
147 unidentified host factor is further required for the observed targeting activity in cells.
148
149 The tRNA targeting activity by Ec78 differs from the stand alone Septu system that has been
150 reported to target DNA[18,20]. The Ec78 PtuAB tRNA targeting activity is more consistent with that
151 of the ATPase-TOPRIM nuclease pair, AriA-AriB, of the phage antirestriction induced
152 system (PARIS) system, which is also a retron-independent abortive immunity system[22,23].
153 However, AriB contains the TOPRIM-like domain and functions when released from AriA[22,23]
154 whereas PtuB contains a HNH domain and requires PtuA for function.
155
156 *Molecular basis for Ec78 retron to sequester PtuAB*
157

158 To unveil the molecular basis for Ec78 to sequester PtuAB, we carried out cryoEM structural
159 analysis of the purified retron-effector complex. The. sample preparation process indicated a
160 sensitivity of assembly to ATP hydrolysis. Whereas the wild-type retron displayed significant
161 heterogeneity in assembly (**Suppl. Fig. 2a**), especially due to PtuB displacement, that containing
162 the ATPase deficient mutant, E396Q in PtuA showed reduced heterogeneity and allowed its
163 structural determination (**Suppl. Fig. 3 & Suppl. Table 1**).
164
165 The cryo-EM map revealed well-resolved density for both the ncRNA and the msDNA, enabling
166 unambiguous assignment of their nucleotides (**Fig. 3 & Suppl. Fig. 3**). Surprisingly, only stem
167 loop II (rU27 to rU48), a linker (rC49-rU58) and the template region (rG59-rU67) base paired with
168 msDNA (dA69-dC77) of the ncRNA are observed. Furthermore, the msDNA is debranched at its
169 2′-5′ linkage to the initiation guanine nucleotide and lacks the first four nucleotides (dT1-dA4)
170 (**Fig. 3a**), unlike that previously observed in the Ec86 retron[24,25]. We confirmed the processing of
171 ncRNA and msDNA by analyzing the nucleic acids co-purified with either the intact retron or RT
172 alone by examining the nucleic acids from the cell extracts or co-purified with proteins on PAGE
173 gels (**Fig. 3b**). We further showed that the processing of both nucleic acids does not require the
174 presence of PtuAB as the processed products are also present in the RT alone sample (**Fig. 3b**). To
175 assess if RT influences the processing activities, we mutated Lys143 and Arg146 of the RT
176 (RT(RK)) that are found near the msDNA cleavage site. We showed that the RT(RK) variant
177 reduced processing (**Fig. 3b**), and its retron assembly no longer sequesters the PtuAB toxicity
178 (**Suppl. Fig. 4**), suggesting that, though host enzymes are likely responsible for the observed
179 processing activities, RT exerts a notable influence on this process.
180

181 The retron-effector particles are classified into two major classes. The first is a 4:2:1 ratio of PtuA,
182 PtuB, and RT-ncRNA-msDNA, with four PtuA subunits, two PtuB subunits, and one RT-ncRNA-
183 msDNA complex (**Fig. 3c**). The second class exhibits a 4:2:2 stoichiometry, featuring an additional
184 RT-ncRNA complex at the distal end of the msDNA hairpin, where only the terminal 12 msDNA
185 nucleotides (dC65-dC77) are visible (**Fig. 3c**). This suggests that the complete the Ec78 retron
186 assembly accommodates a single msDNA molecule. There are other classes with either reduced
187 PtuB occupancy or different combinations of PtuB orientations (cis vs. trans to RT) without other
188 structural differences. These classes led to reconstructions at lower resolutions and are thus not
189 modeled (**Suppl. Fig. 3 & Suppl. Table 1**). Interestingly, though we used the E396Q PtuA variant
190 in structural determination, which showed significantly reduced ATPase activity (**Suppl. Fig. 2c**),
191 ADP instead of ATP was trapped in three of the four subunits (**Fig. 3c**). This result contrasts with
192 the stand-alone PtuAB structures in which ATP molecules are trapped, consistent with its lack of
193 ATP hydrolysis [18,20]. Since there are no other differences in molecular interactions among these
194 classes of structures, we focus the subsequent structural analysis primarily on the 4:2:1 assembly.
195
196 A single RT subunit forms the head of the Ec78 retron assembly from which a long msDNA hairpin
197 emerges to form the central spine of the assembly (**Fig. 3c**), unlike the previously observed Ec86
198 retron structure that is capped by RT dimers and bordered by two antiparallel msDNA[24,25]. In the
199 4:2:2 class, a second RT is positioned at the distal end of the msDNA hairpin but with its own
200 msDNA largely disordered, leading to an essentially same architecture (**Fig. 3c**). The lack of the
201 dimerization α-helix at the N terminus of Ec78 RT is responsible for the different oligomerization
202 state than that of Ec86.
203

204 The 3'-overhang of the msDNA is anchored to the RT-bound ncRNA via a short RNA:DNA
205 duplex (template:primer) that feeds into the catalytic Palm domain of RT while its hairpin extends
206 approximately 110 Å from the RT-ncRNA head (**Fig. 3a**). Four PtuA subunits, forming a dimer
207 of dimers, construct a diamond-shaped platform with a central groove that adheres along the length
208 of the msDNA hairpin, while two PtuB proteins cap the outer edges (**Fig. 3c**). Should the IRa1 and
209 stem loop I of the ncRNA and the 2′-5′ linkage of the msDNA is retained, they would obstruct the
210 assembly of PtuAB, suggesting that removal of these ncRNA elements, including msDNA
211 debranching, is a required mechanism for Ec78 to achieve sequestration of the toxicity.
212
213 The PtuA platform forms an extensive interface with the msDNA, totaling ~2000 $Å^2$ buried solvent
214 accessible area (**Fig. 4a**). However, the contacts are non-specific and mostly occur between the
215 loops in the helix-grip fold of PtuA and the minor groove of the msDNA (**Fig. 4b & 4c**). Most of
216 these amino acids are positively charged and are not conserved in Septu PtuA (**Suppl. Fig. 5a**). In
217 addition, a unique helix-loop-helix insertion to the RecA-like nucleotide binding domain of PtuA
218 forms a helical flap with the corresponding element of an opposing PtuA subunit that encircles the
219 msDNA (**Fig. 3c** & **Fig. 4b**). Thus, the assembly of PtuA with msDNA is based on shape
220 complementarity, similar to other ABC ATPases, such as the Rad50 ATPase[26], that act on DNA
221 nonspecifically. Importantly, loop forming a number of contacts with msDNA (m loop) is
222 significantly longer in the stand-alone PtuA, which would obstruct its assembly with DNA (**Fig.
223 4b & Suppl. Fig. 5a**). In addition, the helical flap encircling the msDNA is deleted in the stand-
224 alone PtuA (**Fig. 4b & Suppl. Fig. 5a**). The distinct elements in Ec78 PtuA evolved for binding
225 msDNA enables it to partner with Ec78, reflecting a unique evolutionary path for retron acquisition.
226

227 The association of PtuA with Ec78 retron is further enhanced by its contacts with RT, which buries
228 ~690 Å^2 solvent accessible surface (**Fig. 4c**). This interaction is largely mediated by the N-terminal
229 extension (NTE) of PtuA, an element also distinctly absent from the stand-alone PtuA (**Suppl. Fig.**
230 **5a**). The first α-helix of NTE is wedged between the characteristic Finger and the Thumb domains
231 of RT, forming several specific protein-protein contacts (**Fig. 4c**). In addition, the first α-helix of
232 the NTE also contacts the ncRNA (**Fig. 4c**), though, non-specifically, which further stabilizes the
233 PtuA-retron association. The observed NTE-RT-ncRNA interaction is another significant feature
234 acquired by the Ec78 PtuA.
235
236 *Mechanism of PtuAB activation*
237
238 To determine if there are distinct sequence features in the PtuAB proteins adapted to function with
239 Type I-A retrons, we performed phylogenetic analysis of both proteins. The HNH nuclease PtuB
240 revealed a somewhat scattered distribution between retron-associated and retron-independent
241 sequences (**Fig. 5a**), suggesting minimal evolutionary pressure for co-opting stand-alone PtuB to
242 function within the Type I-A retrons. In the retron structures, PtuB employs its helical C-terminal
243 domain (CTD) to interact with the C-terminal extension of PtuA (**Fig. 5b**). A pair of CTEs from
244 each PtuA dimer sandwiches one PtuB subunit, forming a shared nine-helix bundle where three
245 helices are from PtuB (**Fig. 5b**) Both CTD of PtuB and CTE of PtuA seem to be well conserved
246 between the Ec78 and Septu systems (**Suppl Fig. 5**). Furthermore, a comparison of the overall
247 structure (**Fig. 5c**) and the active site (**Fig. 5d**) of Ec78 retron PtuB with stand-alone PtuB and the
248 HNH domain of AceCas9 [27] indicates a high degree of structural similarity, which, however, seems
249 to contradict to their different nucleic acid targets.

250

251 Unlike PtuB, the ATPase PtuA exhibits distinct clustering between the retron-associated and
252 retron-independent sequences (**Fig. 6b**), suggesting that the evolution of the PtuA subunit from
253 the stand-alone Septu system facilitated its integration into retron function. This is further
254 supported by the observed structural features unique to the retron-associated PtuA for binding the
255 RT-ncRNA-msDNA tripartite complex that are distinctively absent from the stand-alone PtuA
256 (**Fig. 4b & Suppl. Fig. 5a**).

257

258 To further analyze the molecular basis for PtuAB activation, we obtained its structure in the
259 absence of the retron. We introduced the Gln361 to leucine mutation within the signature motif
260 (357-LSDGQR-362) of PtuA[14] (**Suppl. Fig. 5b**) to abrogate the wild-type toxicity, which
261 permitted protein purification and structure determination (**Suppl. Fig. 6 & Suppl. Table 1**).
262 Single particle reconstruction led to a major class of PtuAB particles that was refined to 3.2 Å (**Fig.**
263 **6b**).

264

265 The PtuAB complex form a similar diamond-shaped platform comprising four PtuA and two PtuB
266 subunits (**Fig. 6b**), significantly different from the hexameric and inflammasome-like assembly of
267 the stand-alone PtuAB[18]. The retron-displaced Ec78 PtuAB assembles largely through
268 dimerization of two helical flaps that otherwise encircle msDNA when assembled with retron (**Fig.**
269 **6b**). In comparison with the retron-sequestered PtuAB, the retron-displaced PtuAB displays a
270 substantially narrower central groove and a more compressed helical flap (**Fig. 6c**), both of which
271 would exclude msDNA from binding. Interestingly, the retron-displaced PtuAB bearing the
272 Q361L mutation in the ATPase domain also traps ADP as opposed to ATP on all four sites (**Fig.**

273 **6b**), consistent with its detectable ATPase hydrolysis activity (**Suppl. Fig. 2c**). Notably, the stand-
274 alone PtuA contains leucine instead of Gln361 at this position and lacks ATPase activity,
275 indicating that the retron-associated PtuA is adapted to have a strong ATPase activity.
276
277 The unique involvement of the helical flap in Ec78 PtuA oligomerization prevents it from forming
278 the same inflammasome-like hexameric assembly observed for the stand-alone PtuAB complex[18]
279 (**Fig. 6d**). This difference may sufficiently explain the differences in both ATP hydrolysis and
280 defense targets between the two PtuAB systems. It is remarkable that through addition of the
281 elements to the well-conserved ATPase core, the PtuAB proteins are co-opted for sensing different
282 phage determinants and targeting different host substrates.
283
284 The conformational closure of the retron-displaced PtuAB suggests a possible mechanism of its
285 activation (**Fig. 7**). In the presence of the msDNA and the associated RT-ncRNA, PtuA is locked
286 in a conformation that prevents rapid displacement of PtuB. In the absence of the msDNA, for
287 instance as a result of the phage attack, and in conjunction with ATP hydrolysis, PtuA activates
288 PtuB for its nuclease activity, leading to growth arrest. This msDNA-mediated sequestration
289 mechanism is consistent with the detected phage escape mutants in the D15 DNase[13].
290
291 *Ec78 retron facilities efficient genome editing in human cells*
292
293 To assess the ability of Ec78 retron for producing DNA to support genome editing in human cells,
294 we transformed Ec78 retron that encodes a msDNA with flanking homology arms to a EMX1 site
295 (**Fig. 8a**) into HEK293T cells that expresses an endogenous SpyCas9 targeting the same EMX1

site. Previous genome editing studies using over 100 retrons found that efficient production of msDNA led to replacement of the target site by the retron-synthesized msDNA. We thus included a highly efficient retron previously identified, Mestre-1531, a Type II-A retron, targeting the same site as the control. Next generation sequencing showed that Ec78 retron produced at least comparable genome editing efficiency as Mestre-153 (**Fig. 8b**). These results make Ec78 retron the only one that is highly effective in genome editing and with 3D structure information.

Discussions

We report structural and functional characterization of the first retron example that co-opts a stand-alone anti-phage system for retron-mediated defense. In contrast to the protein-controlled Septu PtuAB, the Ec78 PtuAB possesses strong ATP hydrolysis activity and is neutralized by the RT-ncRNA-msDNA tripartite. Upon being displaced, the PtuAB indiscriminately restricts the host, leading to growth arrest. The Ec78 PtuAB acquired three protein elements: the N-terminal extension to interact with the RT and ncRNA, a helical flap and a number of positively charged loops to entrap the msDNA. Acquisition of these tripartite interacting elements eliminates its ability to self-assemble into the inflammasome-like complex as its stand-alone counterpart in Septu that allows a protein-mediated control, such as sensing the phage tail fiber. The difference in PtuAB assembly is likely also responsible for the different host targets observed for the Septu PtuAB, which targets genomic DNA, and Ec78 that targets host $tRNA^{tyr}$. Interestingly, another type I-A retron, Ec83, that also partners with PtuAB, does not seem to target host tRNA[19], raising the possibility that the PtuAB system offers a versatile platform for sensing additional phage triggers and targeting different nucleic acids in response to different environmental pressures.

319

320 The function of Ec78 retron RT resembles that of Ec86 retron[24,25] and that of the diversity-
321 generating retroelements (DGR)[28], for which three dimensional structures are now available, in
322 using ncRNA to template DNA synthesis. Different from these latter RTs, however, Ec78 RT has
323 evolved features specific for its partner effector. The Ec78 RT protein adopts the characteristic
324 "right-hand" architecture typical of group II intron family of RTs, comprising three distinct
325 subdomains - Finger, Thumb and Palm – that are all preserved in the three RT enzymes. Structural
326 comparison reveals a high similarity between Ec78 and Ec86, with a root mean square deviation
327 (RMSD) of 1.155 over 121 Cα atoms. However, Ec78 RT exhibits two distinguishing features: the
328 absence of an N-terminal helix preceding the Finger domain and a particularly long, positively
329 charged loop within the Palm domain (Lys141-Lys146). In Ec86 RT, the N-terminal helix
330 facilitates RT dimerization, whereas its absence in Ec78 RT accounts for the monomeric state of
331 RT in the Ec78 retron system. The positively charged loop plays a critical role in stabilizing the 5'
332 end of the msDNA, with the side chain of Arg143 forming a hydrogen bond with the phosphate
333 group of terminal dG5, while Arg142, Arg143 and Lys146 engage in electrostatic interaction with
334 dG5 and dG6. We found that the R143A/K146A mutation impaired msDNA biogenesis required
335 for retron assembly, leading to cytotoxicity when the mutant retron is expressed in cells.

336

337 Consistent with the structural data, phylogenetic analyses also revealed co-evolved features that
338 support the specific assembly. We observed distinct clusters of RTs based on types, with only a
339 minor proportion scattered throughout the tree. For instance, the positively charged loop that
340 impacts msDNA debranching is highly conserved among Type I-A retrons, whereas the N-terminal
341 helix required for dimerization is a common feature in Type II retrons (such as Ec86). Notably,

342 this clustering pattern is more pronounced in larger subtypes with a greater number of members,
343 whereas smaller subtypes, with fewer members, exhibit greater dispersion. These findings suggest
344 that co-evolutionary selection plays a dominant role in retron evolution, while modular exchange
345 occurs to a lesser extent, particularly in smaller subtypes. These findings suggest a potential co-
346 evolutionary relationship between RT and its associated effector in retron system, highlighting
347 structural adaptations that contribute to functional specificity within each retron type.
348
349 The unique production of msDNA by retrons has been harnessed for genome applications, in
350 conjunction with CRISPR-Cas9, in bacteria, yeast and human cells[10,29]. Transfecting the plasmid
351 carrying the RT-msrRNA cassette in targeted cells produces msDNA by the RT, providing an
352 amplifiable single-stranded DNA that act as the homology dependent repair (HDR) template to
353 repair the double-stranded break generated by CRISPR-Cas9. Owing to the continuous template
354 production, the retron-mediated HDR is shown to be more efficient than that mediated by synthetic
355 DNA[10]. In addition, retron RTs have been repurposed for prime-editing, in which the fused nickase
356 CRISPR-Cas9 nicks the DNA while the RT produces the desired edited DNA from the prime
357 editing guide RNA (pegRNA).[29] Retron reverse transcriptases present a promising, compact
358 alternative for prime editing, achieving efficiencies comparable to the larger M-MLV RT through
359 protein engineering[29]. Our results indeed showed highly efficient genome editing with Ec78 retron.
360 However, the implications of the distinctive biogenesis of Ec78 ncRNA and msDNA in genome
361 editing remain unclear, highlighting a need for further research.
362
363 **Materials and Methods**
364

365 *Cloning, Protein Expression and Purification*

366

367 For protein expression, the operon sequences of *E.coli* ECONIH5, including its native promoter,
368 coding sequences of ncRNA, RT, PtuA and PtuB, were PCR amplified from the pLG008 plasmid[2].
369 The PCR product was cloned into a modified polycistronic expression plasmid, pST44[30], where a
370 C-terminal 3C protease-cleavable His-tag was introduced at C-terminus of PtuB by Gibson
371 assembly. For the co-expression of PtuA and PtuB, the coding sequences of PtuA and PtuB were
372 individually cloned into the first and second multiple cloning sites of a modified pRSF-Duet
373 expression vector, where an N-terminus of His-Sumo tag was introduced on PtuA. For the toxicity
374 assay, the sequences encoding PtuA and PtuB were cloned either separately or together into
375 expression vectors. When cloned separately, each gene was inserted into either the pBAD vector
376 or the pRSF-Duet vector. For co-expression, the ptuA and ptuB genes were cloned together either
377 in the pRSF-Duet vector or as part of a co-operonic construct in the pBAD vector. All the mutations
378 were performed by QuikChange mutagenesis method with Q5 DNA polymerases.

379

380 The expression plasmids were transformed into either the Nico21 (DE3) (NEB) or the BL21-
381 CondonPlus (DE3)-RIPL (Agilent) cells. Transformed cells were plated onto LB agar plates
382 containing appropriate antibiotics for selection. A single colony from the plate was used to
383 inoculate 50 ml of LB medium supplemented with the same antibiotics and the primary culture
384 was grown overnight at 37 °C. Eight milliliters of the overnight culture were transferred into 1 L
385 of LB medium with a total volume of 6 L prepared for subsequent purification. The cells were
386 grown at 37 °C until the OD_{600nm} reached 1.0. The temperature was lower to 19 °C, and after 30
387 minutes, the culture was induced with 0.2 mM isopropyl β-d-1-thiogalactopyranoside (IPTG).

388 Induction was allowed to proceed overnight. After overnight induction, cells were pelleted by
389 centrifugation and snap-frozen in liquid nitrogen and stored at -80 °C if not used for purification
390 immediately.
391
392 For protein purification of the retron complex, the cell pellets were resuspended and lysed by
393 sonication in 80 ml of lysis buffer (50 mM Tris-Cl, 500 mM NaCl, 5% glycerol, 10 mM Imidazole,
394 5 mM TCEP). The lysate was clarified by centrifugation, and the supernatant was loaded into a 5
395 ml HisTrap HP column (Cytiva) using a peristaltic pump. Subsequent washing and elution steps
396 were also performed with the peristaltic pump. The column was sequentially washed with 200 ml
397 of low-imidazole wash buffer (50 mM Tris-Cl, 500 mM NaCl, 5% glycerol, 25 mM Imidazole)
398 followed by50 ml of high-imidazole wash buffer (50 mM Tris-Cl, 500 mM NaCl, 5% glycerol, 40
399 mM Imidazole). Protein was then eluted with 20 ml of elution buffer (50 mM Tris-Cl, 500 mM
400 NaCl, 5% glycerol, 400 mM Imidazole, 5 mM TCEP). The eluted protein was then diluted with
401 low-salt buffer (50 mM Tris-Cl, 100 mM NaCl, 5% glycerol, 5 mM TCEP) to achieve a final salt
402 concentration of 300 mM NaCl. The diluted sample was then loaded onto a 5 ml HiTrap Q HP
403 anion exchange column (Cytiva) using a peristaltic pump. The column, connected to an ÄKTA
404 pure™ 25 chromatography system, was washed and eluted with a gradient of NaCl from 300 mM
405 to 2M in 25 mM Tris-Cl buffer (pH8.0) and 2 mM 1,4-dithiothreitol (DTT). The fractions of the
406 elution peak were combined, concentrated and further purified by size exclusion chromatography
407 (SEC) using a Superose 6 10/300 column in a buffer containing 25 mM Tris-Cl (pH8.0), 250 mM
408 NaCl, and 5 mM DTT. The intact complex was identified based on ultraviolet absorbance values
409 at 260 nm and 280 nm, as well as SDS-PAGE analysis. The fraction with highest protein

410 concentration was used immediately for cryo-EM grid preparation. All remaining fractions were
411 snap-frozen in liquid nitrogen and stored at -80 °C.
412
413 The protein purification of the PtuA and PtuB complex was carried out similarly as described
414 above. To avoid the toxicity of co-expressing PtuA and PtuB, we expressed the two proteins
415 separately and the cells were combined for co-purification. The wild-type PtuAB proteins could
416 also be expressed when $tRNA^{Tyr}$ is co-expressed. For the non-toxic PtuAB mutants, cells co-
417 expressing both proteins were harvested as described. After HisTrap affinity purification, the
418 eluted protein was loaded to a 5 ml HiTrap Heparin column instead of the HiTrap Q column. All
419 other steps remained unchanged.
420
421 *Cryo-EM sample preparation, data collection, and 3D reconstruction*
422
423 Cryo-EM grids were prepared using a Mark IV Vitrobot. 3 µl of the sample (3 mg/ml) were applied
424 to R 1.2/1.3 Carbon Quantfoil grids, which had been freshly glow-discharged in a Gatan Solarus
425 950. The grids were double-sided blotted for 4s with a constant force of 0, in a chamber maintained
426 at 100% relative humidity at 8 °C. They were then plunged into liquid ethane and stored in liquid
427 nitrogen before screening.
428
429 Grids were initially screened on an Arctica Talos microscope equipped with a K3 camera operated
430 at 200 kV. Good grids were subsequently transferred to a Titan Krios microscope, also equipped
431 with a K3 camera and operated at 300 kV. Micrographs were automatically collected using
432 SerialEM in low-dose mode at a nominal magnification of 130,000× in a super-resolution mode,

433 with an energy filter of 15 eV, leading to a corrected physical pixel size of 0.828 Å/pixel. A total
434 dose of 50 e^-/ $Å^2$ was applied over 50 frames, with a random defocus range of -1.0 to -1.8 µm.
435
436 The raw micrographs, initially in super-resolution, were frame-aligned using Patch Motion
437 Correction with cropping to half the original resolution (0.828 Å/pixel), followed by Patch CTF
438 Estimation. Particle picking was performed automatically using Blob Picker with a particle range
439 size set to 100 -180 Å. Subsequent particle inspection and extraction yielded a total of 639,730
440 particles. The extracted particles underwent two-rounds of 2D classification, resulting in 50,913
441 high-quality particles. These refined 2D templates were then used for particle picking via Template
442 Picker, followed by another round of particle inspection, extraction, and two-rounds of 2D
443 classification. Ultimately, 199,577 good particles were selected for 3D reconstruction using Ab-
444 initio reconstruction, Homogeneous Refinement, and Non-uniform Refinement. The resulting
445 volume and selected particles exported to Relion for further processing of the complete datasets,
446 where they served as initial model and templates for topaz training.
447
448 The complete set of 20,372 raw movies of the Ec78 complex was imported to Relion (v5.0)[31] for
449 processing. The raw micrographs, initially in super-resolution, were frame-aligned using
450 MotionCor2 (v1.6.4)[32] with a binning factor of 2. CTF estimation was then performed using Gctf
451 (v1.18)[33] within Relion (v4.0). Based on a CTF resolution cutoff of 6.0 Å, a total of 19,745
452 micrographs were selected for further analysis. A subset of 999 randomly selected micrographs
453 was used for topaz training. For topaz training, the cryoSPARC[34] particle file was first converted
454 to Relion-compatible star file using the PyEM package (csparc2star)[35] before being imported to
455 Relion. Subsequent 2D classification provided high-quality templates for automated particle

456 picking from the selected 999 micrographs. The picked particles were inspected and refined
457 through another round of 2D classification, and the resulting high-quality particles were used as
458 input for Topaz training. The trained topaz model was then applied for automated particle picking
459 across the entire dataset, yielding a total of 3,630,467 particles. The picked particles were extracted
460 and downscaled by a factor of 2, resulting in a final pixel size of 1.656 Å/pixel. The extracted
461 particles underwent three-rounds of 2D classification, yielding a total of 1,235,478 high-quality
462 particles for further processing. 3D classification was preformed using the initial volume from
463 cryoSPARC, applying alignment and Blush regularization. Similar maps were combined, and the
464 junk particles were discarded, resulting in two good classes: one consisting of 875,922 particles
465 and the other 294,359 particles. For high-resolution reconstruction, those particles were separately
466 re-extracted at the original physical pixel size of 0.828 Å/pixel. Subsequent 3D refinement and
467 post-processing of these two classes generated two maps, leading to final resolution of 2.75 Å and
468 2.69 Å, respectively, following Non-uniform Refinement.
469
470 A data set of 17040 raw micrographs was collected for the PtuAB complex. The same data
471 processing and refinement were carried out, leading to a major class of particles with a reported
472 resolution of 3.25 Å.
473
474 Both complexes were modeled based on the Septu PtuAB complex coordinates (8EE7) and the
475 Ec86 coordinates (7V9X), manually adjusted in COOT[36] and refined in PHENIX[37].
476
477 *Phylogenetic analysis*

478 The phylogenetic analysis was performed with an in-house snakemake pipeline modified
479 from https://github.com/vihoikka/Cas10_prober. Briefly, 91,902 annotated and marked as
480 'complete' genomes were downloaded on August 22, 2024. The genomes were analyzed through
481 Defensefinder (ref) to identify RT and PtuAB sequences. For retron RT analysis, all organisms
482 that annotated as retron systems were singled out, resulting in 8,143 sequences. These RT
483 sequences were clustered with CD-hit (similarity cutoff 0.99) resulting in 1,935 representative
484 sequences. To create a RT phylogenetic tree, proteins were aligned using Muscle 5.1 with the
485 Super5 algorithm intended for large data sets. The RT tree was constructed from the alignment
486 using FastTree 2.1.11 with the WAG+ CA model and Gamma20-based likelihood and visualized
487 in R4.1.1 using ggtree and ggplot.
488
489 PtuA and PtuB phylogenetic analyses were carried out similarly. A total of 6295 PtuA and PtuB
490 sequences from either Septu (5878) or Type I-A retrons (417) were extracted and clustered before
491 the phylogenetic tree construction as description above.
492
493 *ATPase assay*
494
495 ATPase activity was measured using the ADP-Glo Kinase assay (Promega). The reactions were
496 carried out with or without 500 μM ultrapure ATP in a reaction buffer containing 20 mM Tris
497 pH8.0, 150 mM NaCl, 2 mM DTT, 10 mM $MgCl_2$. Reactions containing 0.2 mg/ml of a protein
498 complex were incubated at 37 °C for 45 minutes before adding ADP-Glo reagent to remove free
499 ATP, followed by the detection reagent. Luminescence was measured on Bio-tek multiplate reader.
500 The reported luminescence reading for each sample was normalized against the readings from

501 reactions without ATP. Mean values obtained from three technical replicates with bars indicating
502 standard deviations.
503
504 *Cell growth spot assay*
505
506 An individual colony from previously transformed bacteria grown on agar plates with appropriate
507 antibiotic(s) was inoculated in LB medium containing the appropriate antibiotic(s) and allowed to
508 grow overnight at 37 °C with shaking. The next day, the overnight culture was used to inoculate a
509 fresh culture that grew until the OD600 reached 0.25. Tenfold serial dilutions of these cultures
510 were prepared in LB medium, and 7.5 µl of these dilutions was spotted onto LB agar plates
511 containing the appropriate antibiotic(s) and inducer (IPTG). The plates were incubated overnight
512 at 37 °C and were imaged the next day using a ChemiDoc Imaging System (Bio-Rad).
513
514 For tRNA rescue experiments, plasmids encoding E coli tRNAs were constructed as described in
515 Masuda et al.[21] and co-transformed with that encoding PtuAB.
516
517 *Non-coding RNA and msDNA analysis*
518
519 The nucleic acids present in either RT or the intact retron-expressing cells were extracted using a
520 Qiagen Plasmid Plus Mid kit and eluted. The nucleic acids co-purified with either RT or the intact
521 retron were extracted by denaturing the purified proteins. The resulting nucleic acids were resolved
522 on a 15% urea PAGE gel and visualized by SYBR Gold staining (ThermoFisher).
523

524 *tRNA cleavage*

525

526 Purified PtuAB, PtuB, and Ec78 at 1 mM concentration were incubated with either the native E
527 coli $tRNA^{Tyr21}$ to test their activity on tRNA cleavage. In vitro transcribed or the native $tRNA^{Tyr}$
528 was heated to 95°C for 5 minutes, and then immediately transferred to 65°C. After adding $MgCl_2$
529 to a final concentration of 10 mM, the tRNA was allowed to cool to room temperature. The
530 cleavage assay was performed in a buffer containing 10 mM HEPES (pH 7.5), 50 mM KCl, 10
531 mM $MgCl_2$, 1mM DTT and 10 μM $MnCl_2$ and at 37 °C for 1 hour. To test the effect of ATP on
532 tRNA cleavage, 1 mM ATP was added. The reaction was stopped by the addition of RNA loading
533 dye (Thermofisher) and heat-denatured at 95 °C for 3 minutes. The reaction products were resolved
534 on 15% urea-polyacrylamide gel electrophoresis gels and visualized by SYBR Gold staining and
535 imaged using the Bio-Rad ChemiDoc system.

536

537 *Plasmid construction for human cell transfection*

538

539 Plasmids used for human cell transfection were adapted from the design described by Khan et al.[10],
540 in which the reverse transcriptase (RT) is expressed under a CAG promoter, while the non-coding
541 RNA (ncRNA) and Cas9 guide RNA (gRNA) are driven by U6 and H1 promoters, respectively.
542 The RT gene strand was codon-optimized and synthesized with the associated ncRNA through
543 Twist Bioscience. The a1/a2 region of the ncRNA was extended to 18 bp, and the msd donor stem
544 was shortened to 9 bp, following modifications reported in the previous study[10]. The best-
545 performing retron from Khan et al. [10] (1531) was used as a performance benchmark. Both tested

546 retrons carried ncRNA sequences designed to introduce a precise three-base edit in the EMX1
547 locus, converting GAAGGG to AAAGTT.
548
549 *Human cell transfections*
550
551 HEK293T cells constitutively expressing SpCas9 from a CBh promoter at the AAVS1 Safe Harbor
552 locus (GeneCopoeia, SL502) were maintained in high-glucose DMEM supplemented with
553 GlutaMax (Thermo Fisher Scientific, 10569010). For retron transfections, 200,000 cells per
554 sample were nucleofected with 500 ng of plasmid using a Lonza 4D-Nucleofector™ X Unit (AAF-
555 1003X) with the SF Cell Line Kit and pulse code CM130.
556
557 For DNA repair inhibition experiments, AZD7648 (1 μM; Selleck Chemicals, S8843) and PolQi2
558 (3 μM; MedChem Express, HY-150279) were added to the culture media before seeding
559 nucleofected cells into 24-well plates. After 48 hours, genomic DNA (gDNA) was extracted using
560 the Quick-DNA Microprep Kit (Zymo Research) and eluted in 30 μl of ultra-pure, nuclease-free
561 water.
562
563 A two-step PCR was performed to amplify the target region and incorporate adapters for Illumina
564 sequencing. Amplicons were sequenced on an Illumina NovaSeq platform using paired-end
565 sequencing (150 bp read length) at a depth of 1.1–3.5 million reads per sample. Sequencing data
566 were analyzed by an in-house Python script to quantify the percentage of precise editing.
567
568 **Author contributions**

B.W. R.H. and H.L. designed the experiments, B.W. expressed and purified the proteins and performed all biochemical and cryoEM studies; R.H. designed and performed human genome editing experiments; Y.H. provided the native tRNA and tRNA-expression plasmids. B.W. R.H. and H.L. wrote the manuscript. All authors edited the manuscript and provided insightful comments.

Acknowledgments

Cryo-EM data were collected at the David Van Andel Advanced Cryo-Electron Microscopy Suite (RRID:SCR_023210) at the Van Andel Institute (Grand Rapids, MI). We thank the Van Andel Institute Cryo-EM Core staff for their assistance with data collection. Authors thank all lab members for their support and discussions.

Data availability

The atomic coordinates of the cryo-EM structures of Ec78 and PtuAB have been deposited in the Protein Data Bank under the identifiers 9NNB, 9NNH, 9NNK and in the Electron Microscopy Data Bank under the entries EMD-49570, EMD-49575, EMD-49576 respectively. The coordinates and map files are available at the following link prior to the release by PDB:

https://fsu-my.sharepoint.com/:f:/g/personal/hli4_fsu_edu/EmsiX1oYM9RHhr4b408dbaIBVYuZ6X6VZHF9RTcuJPPdDg?e=SOns4c

Funding

This work was supported by NIH grant R35 GM152081 to H.L.

References

1 Rice, S. A. & Lampson, B. C. Bacterial reverse transcriptase and msDNA. *Virus Genes* **11**, 95-104 (1995). https://doi.org:10.1007/BF01728651

2 Gao, L. *et al.* Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077-1084 (2020). https://doi.org:10.1126/science.aba0372

3 Millman, A. *et al.* Bacterial Retrons Function In Anti-Phage Defense. *Cell* **183**, 1551-1561 e1512 (2020). https://doi.org:10.1016/j.cell.2020.09.065

4 Bobonis, J. *et al.* Bacterial retrons encode phage-defending tripartite toxin-antitoxin systems. *Nature* **609**, 144-150 (2022). https://doi.org:10.1038/s41586-022-05091-4

5 Lima, T. M. & Lim, D. A novel retron that produces RNA-less msDNA in Escherichia coli using reverse transcriptase. *Plasmid* **38**, 25-33 (1997). https://doi.org:10.1006/plas.1997.1298

6 Shimamoto, T., Shimada, M., Inouye, M. & Inouye, S. The role of ribonuclease H in multicopy single-stranded DNA synthesis in retron-Ec73 and retron-Ec107 of Escherichia coli. *J Bacteriol* **177**, 264-267 (1995). https://doi.org:10.1128/jb.177.1.264-267.1995

7 Mestre, M. R., Gonzalez-Delgado, A., Gutierrez-Rus, L. I., Martinez-Abarca, F. & Toro, N. Systematic prediction of genes functionally associated with bacterial retrons and classification of the encoded tripartite systems. *Nucleic Acids Res* **48**, 12632-12647 (2020). https://doi.org:10.1093/nar/gkaa1149

8 Mayo-Munoz, D., Li, H., Mestre, M. R. & Pinilla-Redondo, R. The role of noncoding RNAs in bacterial immunity. *Trends in microbiology* (2024). https://doi.org:10.1016/j.tim.2024.09.006

9 Doman, J. L. *et al.* Phage-assisted evolution and protein engineering yield compact, efficient prime editors. *Cell* **186**, 3983-4002.e3926 (2023). https://doi.org:10.1016/j.cell.2023.07.039

10 Khan, A. G. *et al.* An experimental census of retrons for DNA production and genome editing. *Nature biotechnology* (2024). https://doi.org:10.1038/s41587-024-02384-z

11 Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359** (2018). https://doi.org:10.1126/science.aar4120

12 Stokar-Avihail, A. *et al.* Discovery of phage determinants that confer sensitivity to bacterial immune systems. *Cell* **186**, 1863-1876 e1816 (2023). https://doi.org:10.1016/j.cell.2023.02.029

13 Azam, A. H. *et al.* Viruses encode tRNA and anti-retron to evade bacterial immunity. *bioRxiv* (2023).

14 Dot, E. W., Thomason, L. C. & Chappie, J. S. Everything OLD is new again: How structural, functional, and bioinformatic advances have redefined a neglected nuclease family. *Mol Microbiol* **120**, 122-140 (2023). https://doi.org:10.1111/mmi.15074

15 Yang, W. Nucleases: diversity of structure, function and mechanism. *Quarterly reviews of biophysics* **44**, 1-93 (2011). https://doi.org:10.1017/S0033583510000181

16 Yang, X. Y. *et al.* Molecular basis of Gabija anti-phage supramolecular assemblies. *Nat Struct Mol Biol* **31**, 1243-1250 (2024). https://doi.org:10.1038/s41594-024-01283-w

17 Li, J. *et al.* Structures and activation mechanism of the Gabija anti-phage system. *Nature* **629**, 467-473 (2024). https://doi.org:10.1038/s41586-024-07270-x

18 Li, Y. *et al.* PtuA and PtuB assemble into an inflammasome-like oligomer for anti-phage defense. *Nat Struct Mol Biol* **31**, 413-423 (2024). https://doi.org:10.1038/s41594-023-01172-8

19 Azam, A. H. *et al.* Evasion of antiviral bacterial immunity by phage tRNAs. *Nature communications* **15**, 9586 (2024). https://doi.org:10.1038/s41467-024-53789-y

20 Wang, S. *et al.* Landscape of New Nuclease-Containing Antiphage Systems in Escherichia coli and the Counterdefense Roles of Bacteriophage T4 Genome Modifications. *Journal of virology* **97**, e0059923 (2023). https://doi.org:10.1128/jvi.00599-23

21 Masuda, I. *et al.* A genetically encoded fluorescent tRNA is active in live-cell protein synthesis. *Nucleic Acids Res* **45**, 4081-4093 (2016).

22 Burman, N. *et al.* A virally encoded tRNA neutralizes the PARIS antiviral defence system. *Nature* **634**, 424-431 (2024). https://doi.org:10.1038/s41586-024-07874-3

23 Deep, A., Liang, Q., Enustun, E., Pogliano, J. & Corbett, K. D. Architecture and activation mechanism of the bacterial PARIS defence system. *Nature* **634**, 432-439 (2024). https://doi.org:10.1038/s41586-024-07772-8

24 Wang, Y. *et al.* Cryo-EM structures of Escherichia coli Ec86 retron complexes reveal architecture and defence mechanism. *Nat Microbiol* **7**, 1480-1489 (2022). https://doi.org:10.1038/s41564-022-01197-7

25 Carabias, A. *et al.* Retron-Eco1 assembles NAD(+)-hydrolyzing filaments that provide immunity against bacteriophages. *Mol Cell* **84**, 2185-2202 e2112 (2024). https://doi.org:10.1016/j.molcel.2024.05.001

26 Gut, F. *et al.* Structural mechanism of endonucleolytic processing of blocked DNA ends and hairpins by Mre11-Rad50. *Mol Cell* **82**, 3513-3522 e3516 (2022). https://doi.org:10.1016/j.molcel.2022.07.019

27 Das, A. *et al.* Coupled catalytic states and the role of metal coordination in Cas9. *Nat Catal* (2023).

28 Handa, S. *et al.* RNA control of reverse transcription in a diversity-generating retroelement. *Nature* (2025). https://doi.org:10.1038/s41586-024-08405-w

29 Doman, J. L. *et al.* Phage-assisted evolution and protein engineering yield compact, efficient prime editors. *Cell* **186**, 3983-4002 e3926 (2023). https://doi.org:10.1016/j.cell.2023.07.039

30 Tan, S., Kern, R. C. & Selleck, W. The pST44 polycistronic expression system for producing protein complexes in Escherichia coli. *Protein Expr Purif* **40**, 385-395 (2005). https://doi.org:10.1016/j.pep.2004.12.002

31 Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530 (2012). https://doi.org:10.1016/j.jsb.2012.09.006

32 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature methods* **14**, 331-332 (2017). https://doi.org:10.1038/nmeth.4193

33 Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12 (2016). https://doi.org:10.1016/j.jsb.2015.11.003

34 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature methods* **14**, 290-296 (2017). https://doi.org:10.1038/nmeth.4169

35 Asarnow, D., Palovcak, E. & Cheng, Y. asarnow/pyem: UCSF pyem v0.5 (v0.5). *Zenodo* (2019). https://doi.org:https://doi.org/10.5281/zenodo.3576630

36 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004). https://doi.org:S0907444904019158 [pii]10.1107/S0907444904019158

37 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010). https://doi.org:10.1107/S0907444909052925

697 **Figure captions**

698

699 **Fig. 1.** Schematic of Ec78 retron coding sequence features and defense activities. **a**. key features
700 of the Ec78 retron components. Upper, amino acid residues of the motifs within the reverse
701 transcriptase (RT), PtuA, and PtuB relevant to the study are listed. Mutated residues are colored
702 in red. Lower, secondary structures of the non-coding RNA (ncRNA) and msDNA before and after
703 processing. Key regions are colored and labeled. **b**. Co-phylogenetic analysis of retron RTs and
704 their effectors. Retron types are color-coded and labeled. The effectors containing the HNH,
705 TOPRIM or ATPase domains are color-coded and labeled. Some known retrons are marked. **c**.
706 Results of cell growth analysis in DH10B cells transformed with plasmids encoding various Ec78
707 components and variants of PtuAB. "induced" indicates the plates containing IPTG; "Ec78"
708 denotes the transformation of the single plasmid encoding all retron-effector components;
709 "PtuAB+RT-ncRNA" denotes co-transformation of the plasmids encoding RT and ncRNA,
710 respectively. The top five and bottom three rows are two independent experiments and thus contain
711 respective PtuAB transformation controls.

712

713 **Fig. 2.** Cell growth assays reveal Ec78 targets host $tRNA^{Tyr}$. Results of cell growth analysis in
714 BL21(DE3)pLysS cells co-transformed with the plasmid encoding PtuAB and that encoding
715 various E. coli tRNAs. "induced" indicates the plates containing IPTG.

716

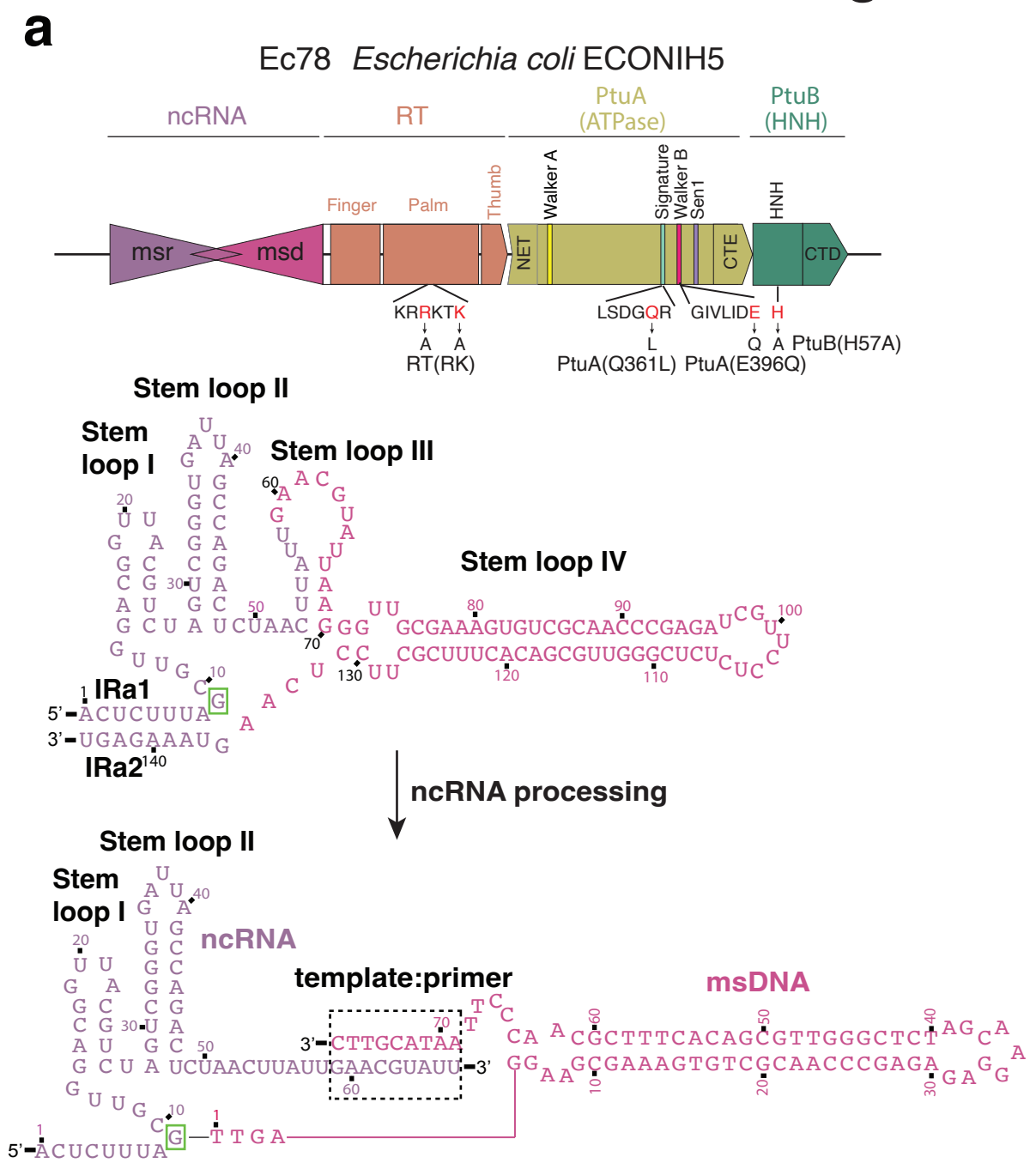
717 **Fig. 3.** Structure overview of the Ec78 retron and analysis on ncRNA and msDNA processing. **a**.
718 The secondary and tertiary structures of the ncRNA and msDNA observed in the Ec78 structures.
719 Only the reverse transcriptase (RT) is retained for simplicity with its characteristic domains labeled.

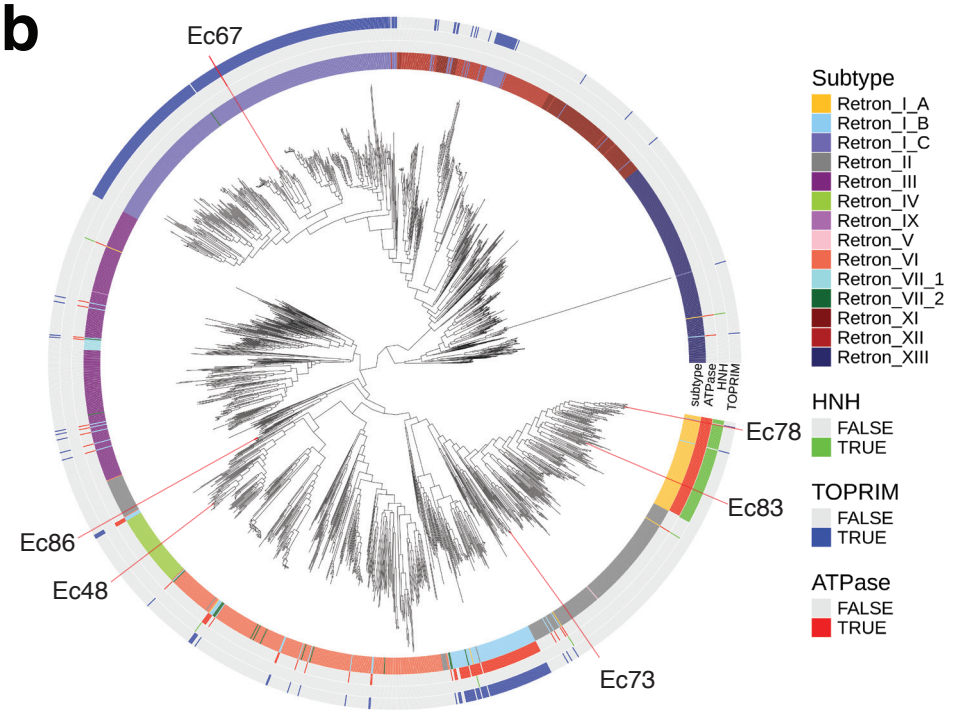
720 **b.** Gel analysis results of the nucleic acids extracted from the purified Ec78, RT-ncRNA, and
721 RT(R143A/K146A)-ncRNA (RT(RK)) (right) or from the RNase A-treated cell extract (left). **c**.
722 Density maps (upper) and cartoon representations (lower) of the Ec78 retron complex. The 4:2:1
723 complex is shown to the left with one PtuB in trans with respect to the top RT and the 4:2:2
724 complex is shown to the right with the corresponding PtuB in cis with respect to the top RT. Orange
725 dots indicate the bound ADP molecules. Insets show the close-up views of the density around the
726 two selected ADP molecules in the 4:2:1 complex.
727
728 **Fig. 4.** Interactions between PtuA and Ec78 components in Ec78 structures. **a.** Binding of msDNA
729 to the central groove formed by PtuA. RT, PtuA, and PtuB molecules are shown in surface
730 representation. The other dimer of PtuAB is omitted for clarity. **b**. Superimposition of the Ec78
731 PtuA to the Septu PtuA (gray ribbons) structures reveal unique protein elements in Ec78 PtuA
732 evolved for contacting msDNA (m loop and helical flap) and the RT-ncRNA complex (N-terminal
733 extension, NTE). The loop equivalent to the m loop in Septu PtuA would obstruct its ability to
734 contact msDNA in the same manner. **c**. PtuA residues, mostly positively charged, making close
735 contacts (within 3.5 Å) with the msDNA. These residues are indicated in Supplementary Figure 5.
736 **d.** Close-ups in two orthogonal views illustrating structural features of PtuA interactions with the
737 RT-ncRNA-msDNA tripartite. The characteristic domains of the RT are indicated and color coded.
738 Nucleotides of the ncRNA and msDNA are labeled. NTE denotes the N-terminal extension of
739 PtuA. Inset highlights the specific contacts of first α-helix of NTE with both the Finger and Thumb
740 domains.
741

742 **Fig. 5.** Phylogenetic and structural analysis of PtuB. **a.** Phylogenetic analysis of both retron Type
743 I-A (black) and Septu (orange) PtuB sequences. **b**. Close-up view of the C-terminal domain of
744 PtuB (CTD) interacting with a pair C-terminal extension (CTE) of PtuA. **c.** Superimposed PtuB
745 structures between the Ec78 (beige) and the Septu (gray, PDB code: 8EE7) systems. The catalytic
746 site is marked by a red dot. The region displaying the largest difference is marked by Ec78 PtuB
747 residues. **d.** Comparison of the HNH domain active site among Ec78 PtuB, Septu PtuB (PDB code:
748 8EE7) and the Cas9 from *Acidothermus cellulolyticus* (PDB code: 8D2K).
749
750 **Fig. 6.** Phylogenetic analysis of PtuA and structural analysis of the retron-free PtuAB. **a.**
751 Phylogenetic analysis of both retron Type I-A (black) and Septu (orange) PtuA sequences. **b.**
752 Density maps and the corresponding cartoon representations of the Ec78 retron-displaced PtuAB
753 complex in two orthogonal views. Orange dots indicate the bound ADP molecules. Inset shows
754 the density map around one of the bound ADP molecules. **c.** Structural comparison between the
755 retron-displaced (orange) and the retron-bound PtuAB (gray), both in tetrameric forms, when a
756 single PtuA subunit is aligned as indicated. The msDNA model is from the 4:2:1 Ec78 structure
757 and shows clashes with the retron-displaced PtuA. **d**. Schematic oligomers of the retron-displaced
758 (tetramer) and the Septu (hexamer) PtuAB. The hexameric structure of Septu PtuAB aligned with
759 a PtuA subunit is shown to the right. Note that PtuB molecules are included in the assembly but
760 not highlighted.
761
762 **Fig. 7.** Proposed models of activation for the Ec78 and Septu systems. Left, Ec78 activation model.
763 The RT-ncRNA-msDNA sequestered PtuAB undergoes some release of the PtuB molecules with
764 the help of ATP hydrolysis without causing cell growth arrest. Upon phage infection, the

765 tetrameric PtuAB is released from the tripartite and, in the presence of ATP hydrolysis and a yet
766 to be identified factor, PtuB is released to target $tRNA^{Tyr}$ molecules. Right, Septu activation model.
767 The hexameric PtuAB complex is activated by phage proteins, and upon the reduction of ATP
768 levels, releases PtuB molecules to target genomic DNA.
769
770 **Fig. 8.** Ec78 retron precise editing. **a.** Schematic of Ec78 retron precise editing targeting an EMX1
771 site. **b.** Quantification of Ec78 retron precise editing in comparison with no reverse transcriptase
772 control (No RT), in the absence (Ec78) or the presence (Ec78+inhib) of inhibitors, AZD7648 and
773 PolQi2, and the same experiments with the Mestre-1531 retron (RT 1531) with the same inhibitors.
774 Bars represent means ± standard deviations of two biological replicates.
775
776
777
778
779
780

Figure 1


a
Ec78 Escherichia coli ECONIH5
ncRNA
RT
PtuA (ATPase)
PtuB (HNH)
Finger
Palm
Thumb
Walker A
Signature
Walker B
Sen1
HNH
msr
msd
NET
CTE
CTD
KRRKTK
A A
RT(RK)
LSDGQR
L
PtuA(Q361L)
GIVLIDE
Q A
PtuA(E396Q)
H
PtuB(H57A)
Stem loop II
Stem loop I
Stem loop III
Stem loop IV
IRa1
IRa2
5'-ACUCUUUA
3'-UGAGAAAU
ncRNA processing
Stem loop II
Stem loop I
ncRNA
template:primer
msDNA
3'-CTTGCATAA
3'-GAACGUAUU
5'-ACUCUUUA
TTGA


b
Ec67
Ec78
Ec83
Ec86
Ec48
Ec73
Subtype
Retron_I_A
Retron_I_B
Retron_I_C
Retron_II
Retron_III
Retron_IV
Retron_IX
Retron_V
Retron_VI
Retron_VII_1
Retron_VII_2
Retron_XI
Retron_XII
Retron_XIII
HNH
FALSE
TRUE
TOPRIM
FALSE
TRUE
ATPase
FALSE
TRUE

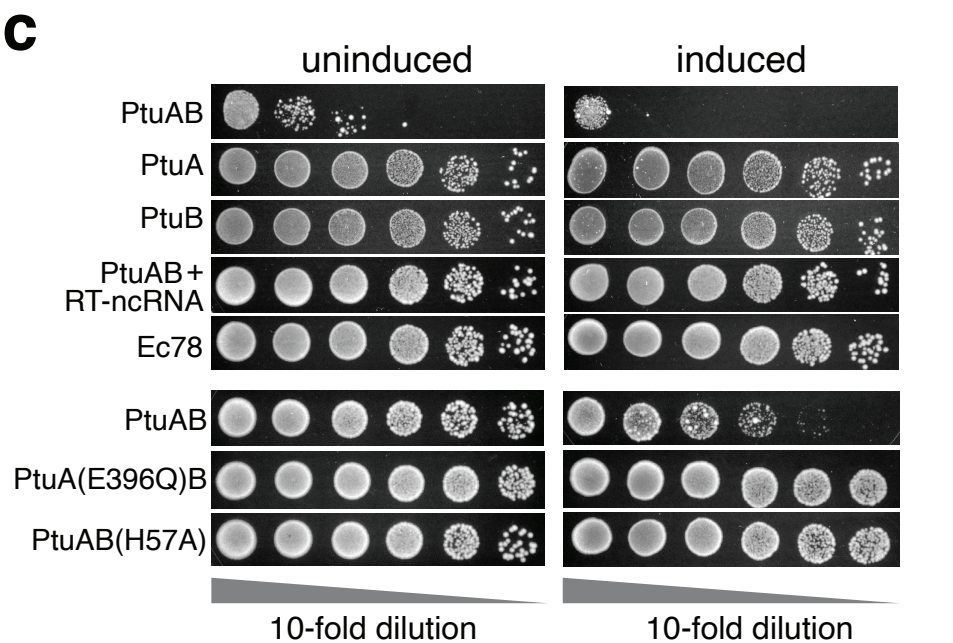

c
uninduced
induced
PtuAB
PtuA
PtuB
PtuAB+ RT-ncRNA
Ec78
PtuAB
PtuA(E396Q)B
PtuAB(H57A)
10-fold dilution
10-fold dilution

Figure 2

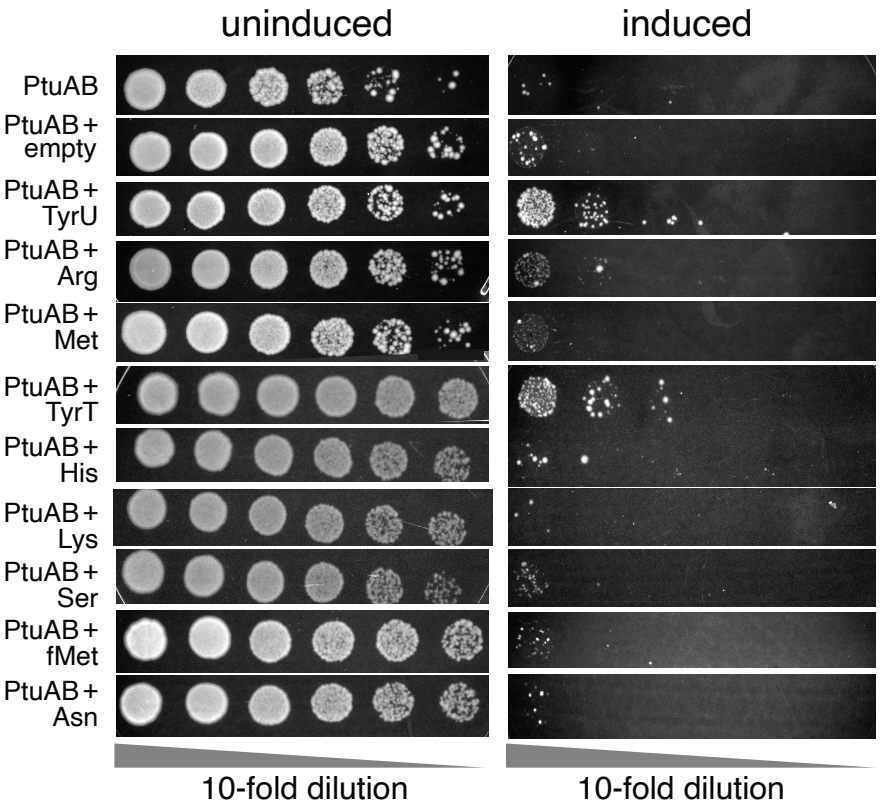

uninduced
induced
PtuAB
PtuAB + empty
PtuAB + TyrU
PtuAB + Arg
PtuAB + Met
PtuAB + TyrT
PtuAB + His
PtuAB + Lys
PtuAB + Ser
PtuAB + fMet
PtuAB + Asn
10-fold dilution
10-fold dilution

Figure 3

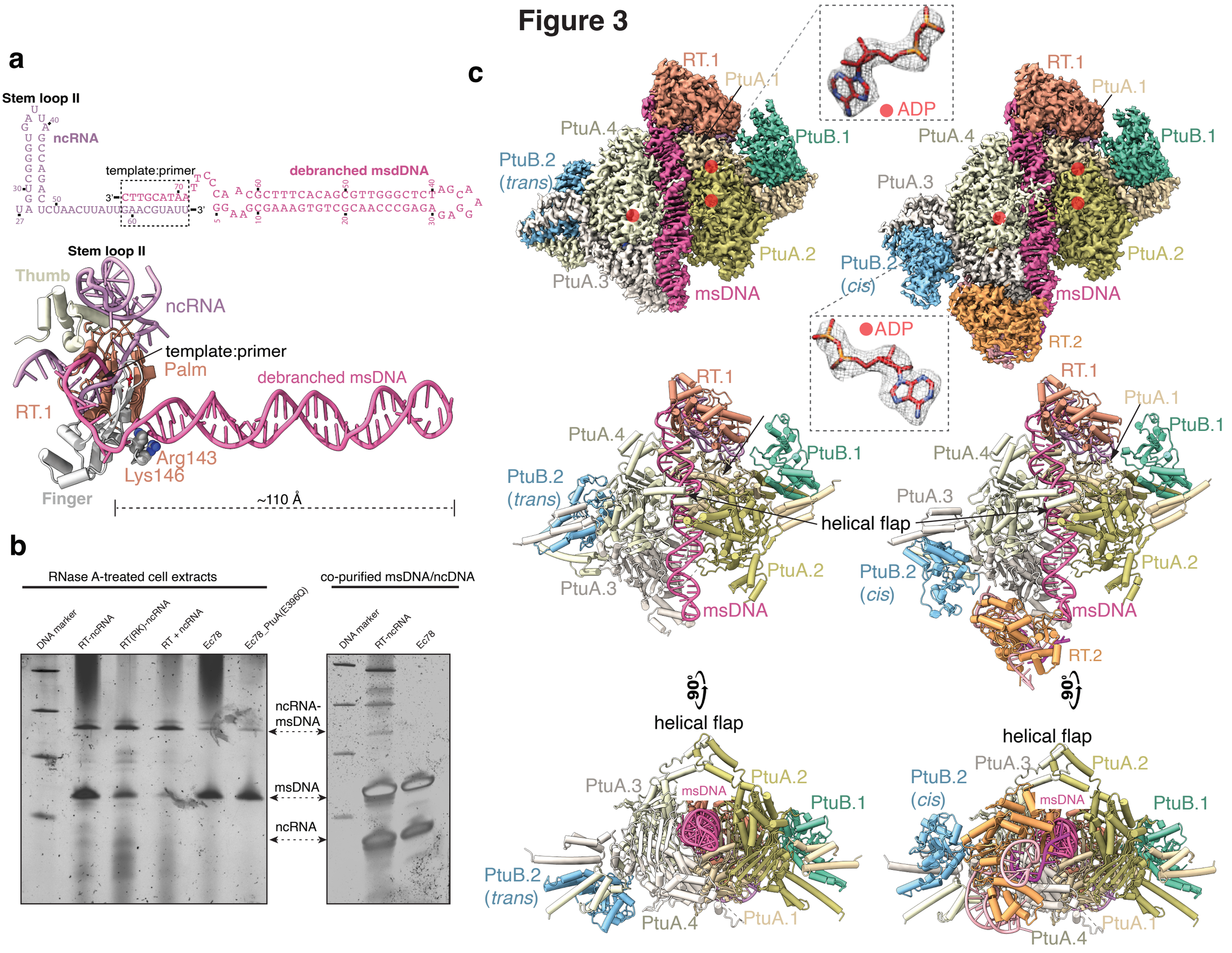

a
Stem loop II
ncRNA
template:primer
debranched msdDNA
3'-CTTGCATAA
UA UCUAACUUAUUGAACGUAUU-3'
Stem loop II
Thumb
ncRNA
template:primer
Palm
debranched msDNA
RT.1
Arg143
Lys146
Finger
~110 Å
b
RNase A-treated cell extracts
co-purified msDNA/ncDNA
DNA marker
RT-ncRNA
RT(RK)-ncRNA
RT + ncRNA
Ec78
Ec78_PtuA(E396Q)
DNA marker
RT-ncRNA
Ec78
ncRNA-msDNA
msDNA
ncRNA
c
RT.1
PtuA.1
ADP
PtuA.4
PtuB.1
PtuB.2 (trans)
PtuA.3
PtuA.2
msDNA
PtuB.2 (cis)
RT.2
helical flap
90°
PtuB.2 (trans)
PtuB.2 (cis)

Figure 4

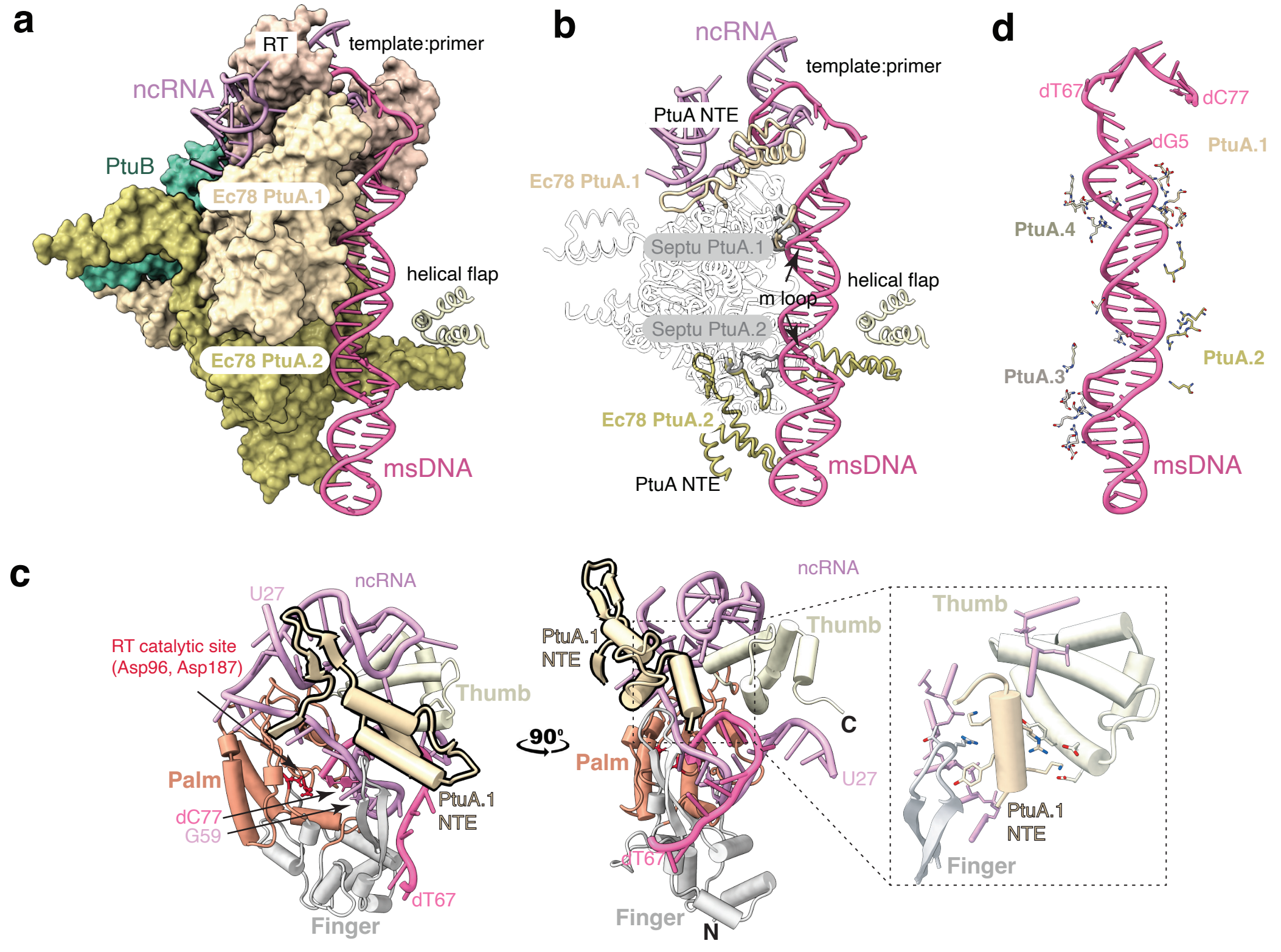

a
RT
template:primer
ncRNA
PtuB
Ec78 PtuA.1
helical flap
Ec78 PtuA.2
msDNA
b
ncRNA
template:primer
PtuA NTE
Ec78 PtuA.1
Septu PtuA.1
helical flap
m loop
Septu PtuA.2
Ec78 PtuA.2
msDNA
PtuA NTE
d
dT67
dC77
dG5
PtuA.1
PtuA.4
PtuA.2
PtuA.3
msDNA
c
U27
ncRNA
RT catalytic site
(Asp96, Asp187)
Thumb
Palm
dC77
G59
PtuA.1
NTE
dT67
Finger
90°
ncRNA
PtuA.1
NTE
Thumb
C
Palm
U27
dT67
Finger
N
Thumb
PtuA.1
NTE
Finger

Figure 5

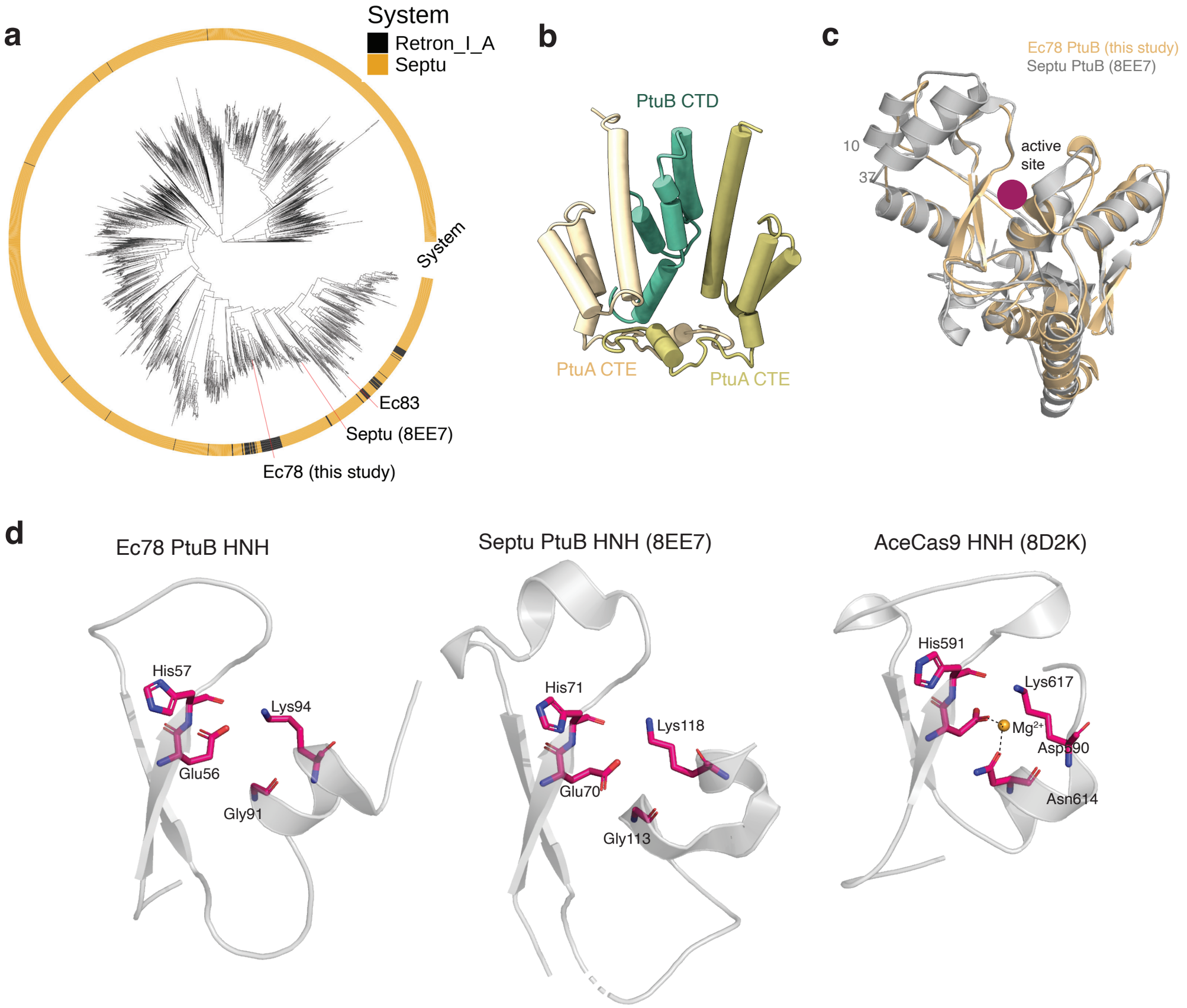

a
System
Retron_I_A
Septu
System
Ec83
Septu (8EE7)
Ec78 (this study)
b
PtuB CTD
PtuA CTE
PtuA CTE
c
Ec78 PtuB (this study)
Septu PtuB (8EE7)
10
37
active site
d
Ec78 PtuB HNH
His57
Lys94
Glu56
Gly91
Septu PtuB HNH (8EE7)
His71
Lys118
Glu70
Gly113
AceCas9 HNH (8D2K)
His591
Lys617
Mg2+
Asp590
Asn614

Figure 6

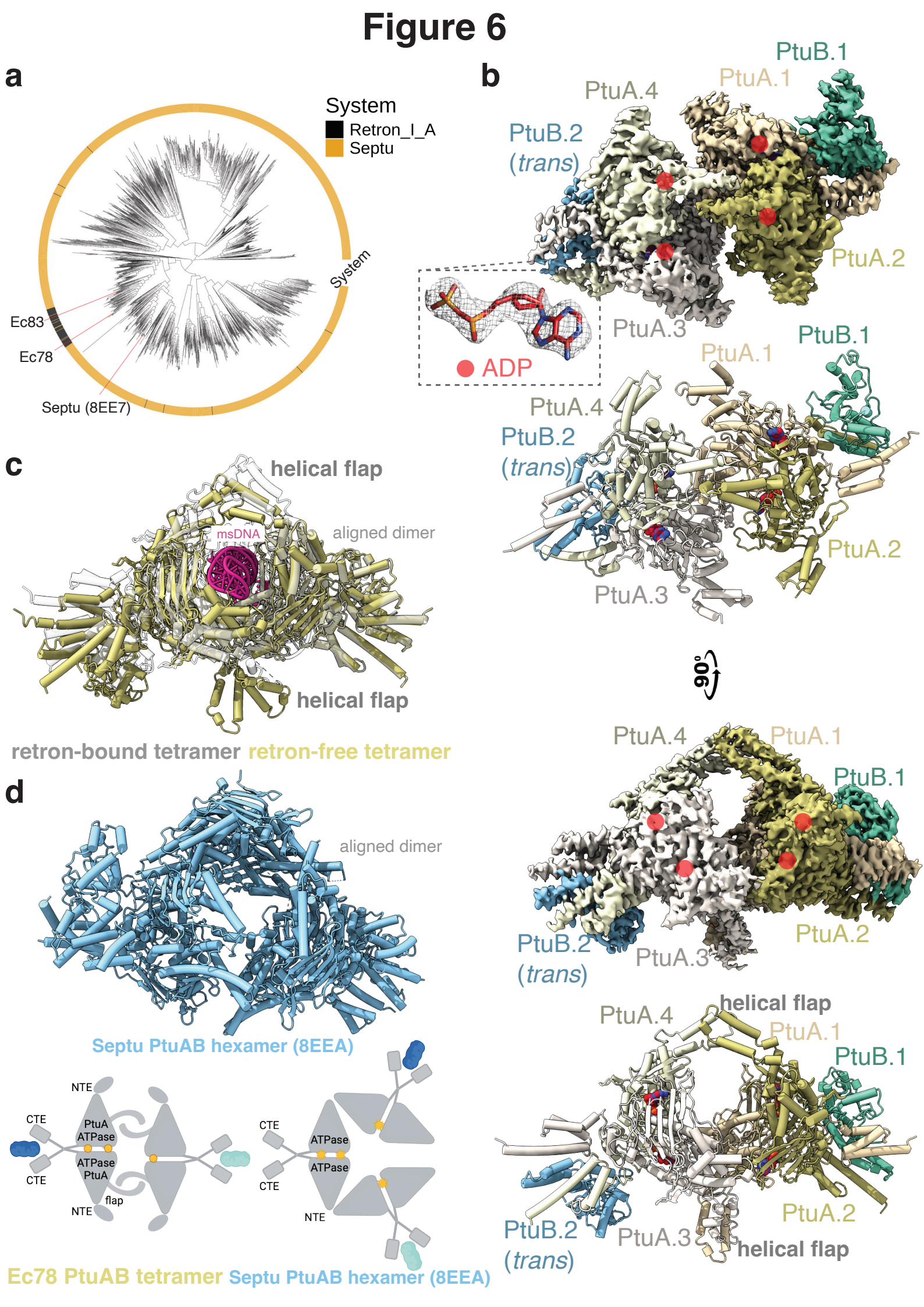

a
System
Retron_I_A
Septu
System
Ec83
Ec78
Septu (8EE7)
b
PtuB.1
PtuA.1
PtuA.4
PtuB.2
(trans)
PtuA.2
PtuA.3
ADP
PtuB.1
PtuA.1
PtuA.4
PtuB.2
(trans)
PtuA.2
PtuA.3
90°
PtuA.4
PtuA.1
PtuB.1
PtuB.2
(trans)
PtuA.3
PtuA.2
PtuA.4
helical flap
PtuA.1
PtuB.1
PtuB.2
(trans)
PtuA.3
PtuA.2
helical flap
c
helical flap
msDNA
aligned dimer
helical flap
retron-bound tetramer
retron-free tetramer
d
aligned dimer
Septu PtuAB hexamer (8EEA)
NTE
CTE
PtuA
ATPase
ATPase
PtuA
CTE
flap
NTE
CTE
ATPase
ATPase
CTE
NTE
Ec78 PtuAB tetramer
Septu PtuAB hexamer (8EEA)

Figure 7

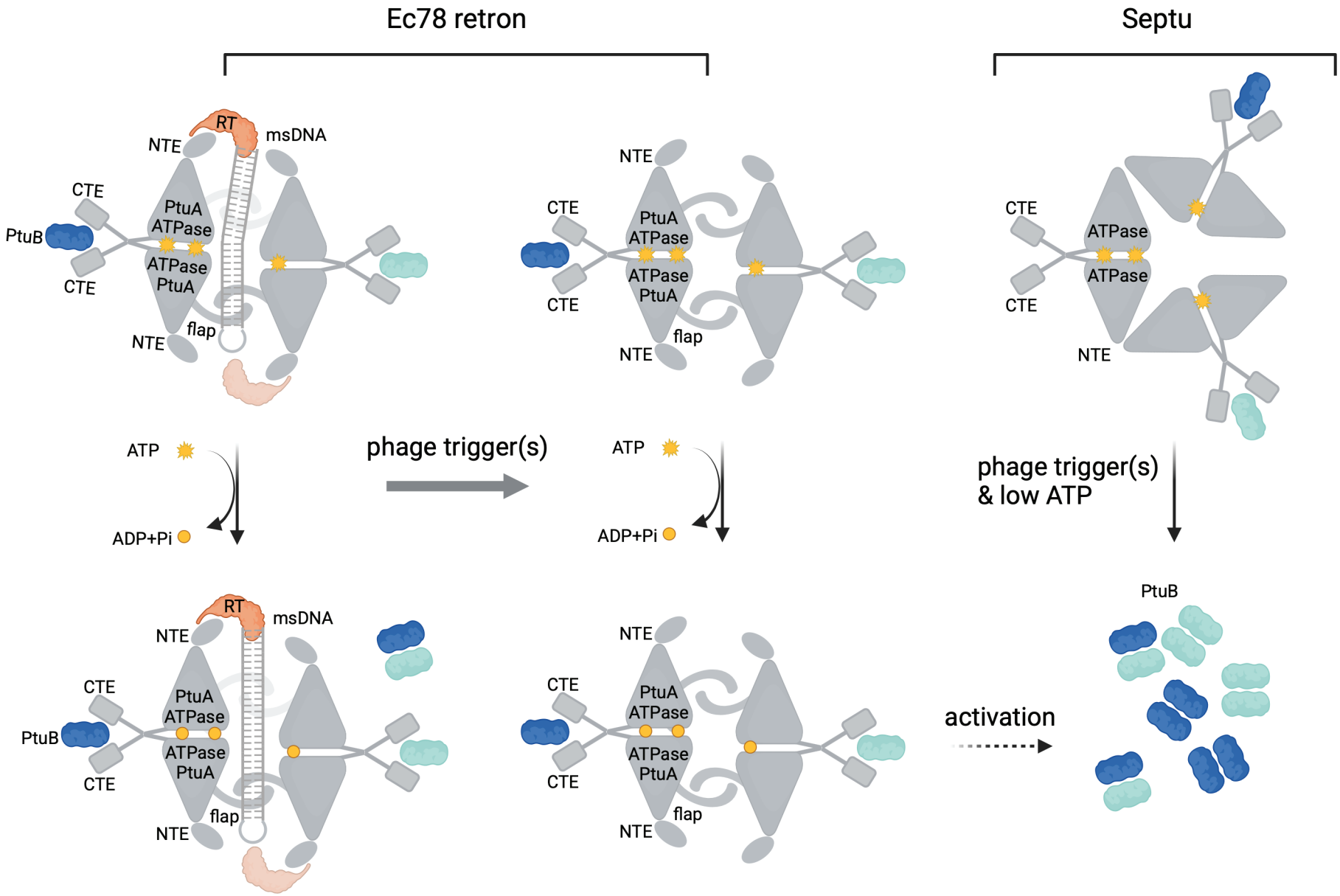
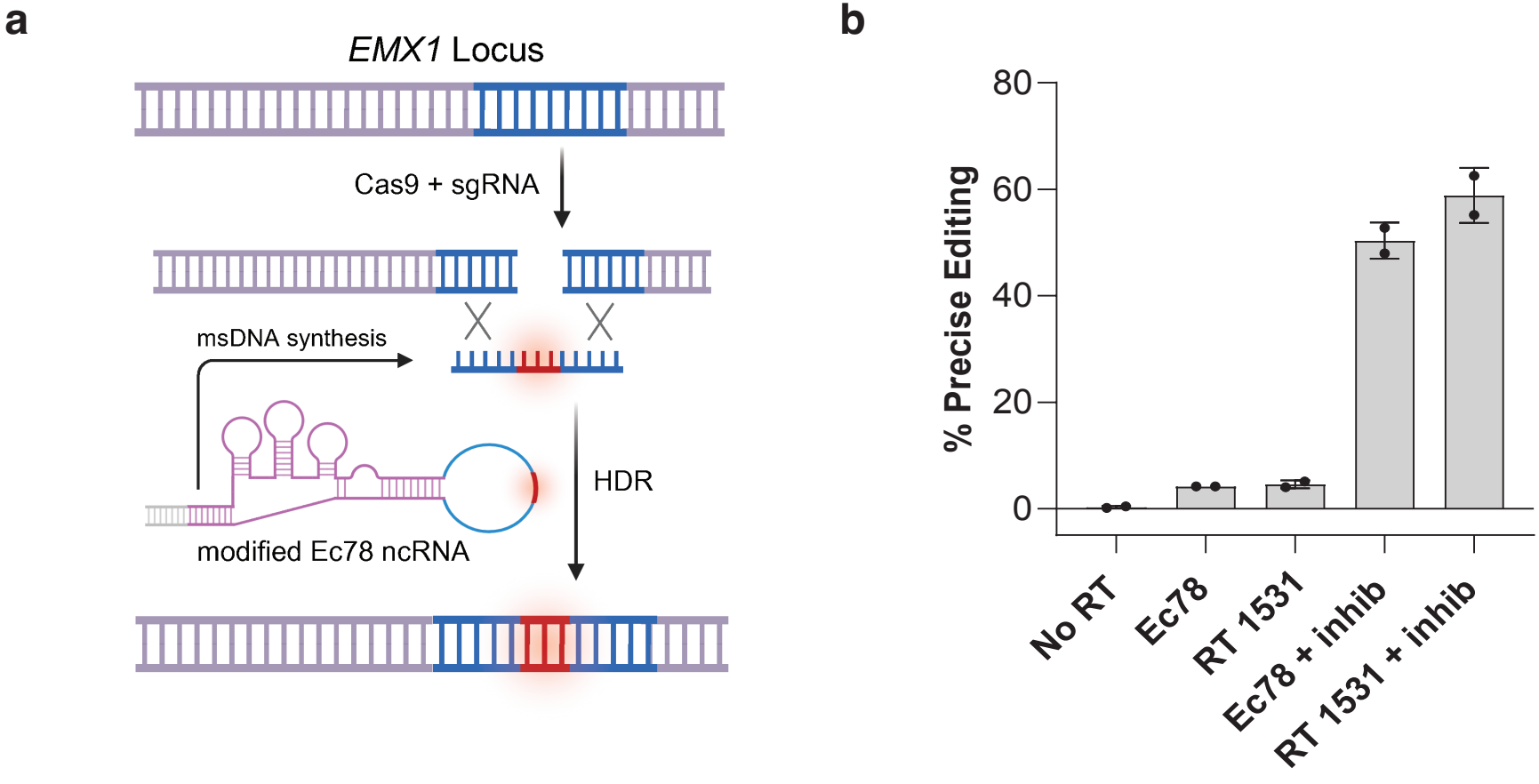

Ec78 retron
Septu
RT
msDNA
NTE
CTE
PtuB
PtuA
ATPase
ATPase
PtuA
flap
ATP
ADP+Pi
phage trigger(s)
phage trigger(s)
& low ATP
activation

Figure 8


a
EMX1 Locus
Cas9 + sgRNA
msDNA synthesis
HDR
modified Ec78 ncRNA
b
% Precise Editing
80
60
40
20
0
No RT
Ec78
RT 1531
Ec78 + inhib
RT 1531 + inhib