END nucleases: Antiphage defense systems targeting multiple hypermodified phage genomes

Wearn-Xin Yee[1], Yan-Jiun Lee[2], Timothy A. Klein[1], Alex Wirganowicz[1], Andres E. Gabagat[1], Bálint Csörgő[3], Kira S. Makarova[4], Eugene V. Koonin[4], Peter R. Weigele[2], Joseph Bondy-Denomy[1]

[1]Department of Microbiology and Immunology, University of California, San Francisco, CA 94158, USA

[2]Research Department, New England Biolabs, Ipswich, MA 01938, USA

[3]Synthetic and Systems Biology Unit, Institute of Biochemistry, HUN-REN Biological Research Centre, Szeged, Hungary

[4]Computational Biology Branch, Division of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Abstract

Prokaryotes carry clusters of phage defense systems in “defense islands” that have been extensively exploited bioinformatically and experimentally for discovery of immune functions. However, little effort has been dedicated to determining which specific system(s) within defense islands limit lytic phage reproduction in clinical bacterial strains. Here, we employed the CRISPR-based Cascade-Cas3 system to delete defense islands in a *Pseudomonas aeruginosa* clinical isolate to identify mechanisms of lytic phage antagonism. Deletion of one island in a cystic fibrosis-derived clinical isolate sensitized the strain to phages from the *Pbunavirus* family, which are commonly used as therapeutics. The causal defense system is a Type IIS restriction endonuclease-like protein (END^{PaCF1}), common in Pseudomonads, however it lacks an associated methyltransferase typical Type IIS R-M systems. END^{PaCF1} protects bacteria against phages with hypermodified DNA and is surprisingly agnostic to the specific structure of the modification, which is unlike typical type IV restriction endonucleases. In END^{PaCF1}, the endonuclease domain is fused to a catalytically inactive Endonuclease III (iEndoIII), a domain that recognizes non-canonical bases to repair DNA in prokaryotes and eukaryotes. We therefore propose that nucleases containing an i**En**doIII **d**omain (**END** nucleases) can sense diverse DNA hypermodifications. Our findings reveal modularity of the sensing and cleavage domains, as expected of a modification-dependent endonucleases. We further show that some hypermodified phages, including *Pbunavirus* family members and *Wrowclawvirus* family (Pa5oct-like) of jumbo phages, encode END nuclease inhibitors that directly bind to the nuclease, likely via the iEndoIII domain. These inhibitors are necessary for *Pbunavirus* to plaque on clinical isolates and sufficient to enable other hypermodified phages to plaque in the presence of this defense system.

Introduction

The recent discovery of numerous antiphage defense systems has opened up new research directions including the identification of a variety of novel nucleases[1], signaling and pattern-recognition based defenses[2–4], multiple programmed cell death mechanisms and more[5,6]. Many of these defense systems show direct evolutionary connections and/or functional parallels to animal and human immunity and thus the study of bacterial antiphage defense can yield insights into human innate immunity[7–9]. Furthermore, exploitation of bacterial antiphage systems has inspired essential tools for molecular biology and biotechnology, from restriction enzymes that revolutionized molecular cloning[10], to CRISPR which transformed gene editing[11]. Another impact of the study of antiphage defense system lies in phage therapy, given that defense systems likely limit the utility of phages for treatment.

Due to the biotechnological implications of the antiphage defense systems that have been previously characterized, much recent research has attempted to discover novel systems through bioinformatic analyses followed by experimental validation, often in non-native hosts. Because antiphage defense systems often co-localize into genomic defense islands, bioinformatic discovery approaches that use such co-localization to identify defense systems have been developed[6,12–15]. An alternative strategy that has not been nearly as widely employed involves manipulating immune islands directly in native hosts to identify active defense systems and determine their roles in phage tropism and regulation under native conditions[16,17].

The restriction modification (RM) systems are a major group of phage defense systems that can restrict phage and plasmid replication by recognizing and cleaving either unmodified or, conversely, modified DNA. RM systems are grouped into four types based on their mechanism by which the substrate is targeted[18,19]. Type II RM systems are the most common[20] and typically include a modification methyltransferase that methylates ‘self’ DNA, while the restriction endonuclease cuts unmethylated, ‘foreign’ DNA either at the recognition site, or a few bases away[19,21,22]. In contrast, type IV RM systems recognize and restrict specific hypermodified ‘foreign’ DNA, whereas the ‘self’ DNA is not hypermodified and therefore is not restricted[23,24]. Type IV restriction systems, such as GmrSD and SauUSI, are also known as modification dependent restriction endonucleases (MDREs) and lack a methylase gene[1,25].

MDREs are modular, containing a modification sensor domain and an effector nuclease domain. There is a wide range of sensor domains, of which the most common ones belong to the PseudoUridine synthase and Archaeosine transglycosylase (PUA) superfamily, including YTH, SRA, EVE and ASCH domains, all of which are thought to recognize specific types of modification[24,26–28]. Recently, a winged helix sensor domain was described[29], and there is likely a larger variety of modification sensor domains that are yet undiscovered. The nuclease domains of MDREs typically belong to the PD-DExK, GIY-YIG, HNH and phospholipase D-like (PLD) superfamilies, each of which has specific cleavage requirements, such as presence of divalent cations[21,29,30].

The incessant evolutionary arms race between phages and bacteria has produced a complex interaction network between multiple bacterial defense systems and phage-encoded inhibitors which impart immunity to these host systems[31,32]. For example, phage T4, which contains hypermodified cytosine in its genome, encodes multiple small-protein inhibitors of GmrSD and closely related endonucleases delivered during the DNA entry stage of infection[33,34].

Here, we identify a single-gene defense system in a clinical *P. aeruginosa* strain CF040 which we call END^{PaCF1} (see below). Unlike previously characterized Type IV modification- dependent restriction endonucleases, END^{PaCF1} appears to be agnostic to the exact modification of the foreign DNA and can target DNA molecules containing a wide variety of hypermodifications occurring on both pyrimidines and purines. We attributed this ability to broad recognize a wide range of phages with hypermodified DNA to the C-terminal domain (CTD) of END^{PaCF1} which is homologous to the eukaryotic Ogg1 DNA glycosylase and prokaryotic Endonuclease III, enzymes that detect and excise non-canonical bases. Comparative genomic analysis shows that END^{PaCF1} belongs to a larger group of nucleases with inactivated **En**doIII **d**omains (END) that are predicted to recognize a broad range of hypermodified bases. Using END^{PaCF1} we show that Pbunaviruses, a well-studied family of phages commonly used for *P. aeruginosa* phage therapy[35], have hypermodified genomes. Furthermore, some Pbunaviruses encode an inhibitor that is necessary and sufficient for END^{PaCF1} inhibition.

Results

Cascade Cas3 can be used to interrogate large defense islands to identify endogenously functional defense systems

Cascade-Cas3 has previously been developed to delete large parts of the *P. aeruginosa* chromosome[36]. We adapted the all-in-one vector carrying Cascade-Cas3 and the guide to delete large defense islands in *P. aeruginosa* clinical isolates. The clinical isolate CF040 was selected as it harbors two predicted defense islands which include well-characterized Type I RM systems (*hsdMSR*)[37]. Therefore, guides against the *hsdM/R* genes of each island were designed, introduced into the all-in-one vector and used to transform CF040. Transformants were then grown in the presence of gentamicin and the rhamnose inducer. After targeting, two surviving colonies with putative deletions in the same defense island were selected (Δ*di1*-1 and Δ*di1*-2), and whole genome sequencing revealed that Δ*di1*-1 (29kb) and Δ*di1*-2 (36kb) had different parts of defense island 2 deleted (Figure 1a; see Extended Data Table 1 for gene annotation). Plaque assays using these deletion strains showed that while the Pbunavirus phages F8 and PB-1 did not plaque well on CF040 and CF040Δ*di1*-1, there was a three-log increase in plaquing between clones CF040Δ*di1*-1 and CF040Δ*di1*-2 (Figure 1b). This finding suggests that clone CF040Δ*di1*-1 encodes a defense system(s) against the Pbunavirus phages that was deleted in clone CF040Δ*di1*-2. BLASTP searches of non-redundant (nr) protein sequences at the NCBI with each of the proteins encoded by the four genes deleted in CF040Δ*di1*-2 as queries showed that one of these proteins was annotated to contain a “PLDc-BfiI DEXD like” region (CDD: 197216) that is characteristic of type IIS restriction nucleases. Neither DefenseFinder nor PADLOC identified this gene as a defense system. Independent deletion of the gene encoding this putative type IIS-related enzyme showed that this single gene was responsible for inhibiting the *Pbunavirus* phages in CF040 (Figure 1b).

Thus, the all-in-one Cascade-Cas3 vector can be employed to efficiently interrogate defense islands in clinical isolates, and uncovered a novel, single-gene defense system which we name END^{PaCF1}.

END^{PaCF1} is an endonuclease represented in diverse bacteria

Restriction modification (RM) systems often consist of a methylase gene adjacent to the endonuclease, the function of which is to help distinguish self-DNA from non-self[19]. Therefore, to determine if END^{PaCF1} is associated with a methylase, we analyzed gene neighborhoods using WebFlaGs[38] and found that the gene encoding END^{PaCF1} is in defense islands in various bacteria, but no other gene appeared to be co-associated with it across genomes (Figure 1c). In particular, no adjacent gene encoding a methyltransferase was found, suggesting that this is not a canonical RM system.

A PSI-BLAST search of the non-redundant protein database using END^{PaCF1} as the query identified homologs of this protein in numerous bacteria, both clinical pathogens and environmental microbes (Figure 1d). All END^{PaCF1} homologs contained the active site HxK motif essential for catalysis of phosphodiester backbone cleavage, similar to type IIS enzymes BfiI and NgoAVII (Figure 1e)[39,40]. All of these endonucleases belong to the phospholipase D (PLD) superfamily of hydrolases that cleave diverse substrates including the phosphodiester bonds of dsDNA[30,41]. However, BfiI and NgoAVII are co-encoded with methyltransferases in canonical Type IIS R-M systems[22].

END^{PaCF1} acts early in infection, is non-abortive, and targets diverse phage families

To determine which families of phages are restricted by END^{PaCF1}, END^{PaCF1} was first introduced into the phage sensitive lab strain PAO1, which lacks an END^{PaCF1} homolog, on an episomal vector (pHERD30T, shortened to p30T) for overexpression or in the chromosome at the *att*Tn7 site[42], under an inducible promoter. We challenged PAO1 p30T$^{END\text{-}PaCF1}$ with a panel of diverse *P. aeruginosa* phages. Plaque assays with this overexpression strain showed that END^{PaCF1} inhibited growth of Yuavirus (YuA-like), Abidjanvirus (Ab18-like), Pbunavirus (F8 family of phages) and Wroclawvirus (Pa5Oct related jumbo phages) (Extended Data Figure 1a).

We next confirmed that PAO1 *att*Tn7::END^{PaCF1} also targeted all four phage families when induced (Extended Data Figure 1b). However, in the absence of induction, PAO1 *att*Tn7::END^{PaCF1} targeted F8 but not 14-1 (Figure 2a and Extended Data Figure 1b), which was similarly observed in CF040 (Extended Data Figure 1c). This finding suggests that expression of END^{PaCF1} in uninduced PAO1 *att*Tn7::END^{PaCF1}

166 approximates the expression level in the native CF040 strain. END$^{PaCF1\text{-}dead}$, in which
167 the essential catalytic residues of the HPK active sites were replaced with alanine
168 (APA), failed to confer protection against phages, confirming that phage targeting is
169 due to the endonuclease activity of END^{PaCF1} (Extended Data Figure 1b). Of note, no
170 escapers for phages F8, PB-1, YuA, PaMX11 and Ab18 were isolated on PAO1
171 *att*Tn7::END^{PaCF1}, suggesting that phages cannot easily escape the system ($\geq 10^5$
172 phages in the first spot) (Extended Data Figure 1b).

173 To determine whether END^{PaCF1} inhibits phage reproduction early or late in infection,
174 PAO1 and PAO1 *att*Tn7::END^{PaCF1} were infected with F8 and the bacteria were
175 collected every 15 min. qPCR results demonstrated that in the absence of END^{PaCF1},
176 the F8 genome replicated approx. 100-fold over 60 minutes; however, no replication
177 of the F8 genome was observed in PAO1 *att*Tn7::END^{PaCF1} (Figure 2b). Thus,
178 END^{PaCF1} appears to act early in infection, antagonizing DNA replication.

179 Given that a broad variety of bacterial defense systems induce dormancy or
180 programmed cell death (a phenomena often denoted as abortive infection)[43–47], we
181 sought to determine whether END^{PaCF1} was abortive. To this end, PAO1 was
182 transformed with p30T with or without END^{PaCF1} to intentionally overexpress the
183 system. Both strains were then infected with F8 at high multiplicity of infection (MOIs)
184 and tracked over 20 hours. We observed no decrease in the bacterial growth rate at
185 high MOIs and overexpression of END^{PaCF1} in the absence of phage did not decrease
186 growth rate of PAO1 (Figure 2c). These results indicate that END^{PaCF1} is not an
187 abortive infection system.

188 To test whether YuA (Yuavirus), Ab18 (Abidjanvirus), 14-1 and F8 (Pbunavirus) and
189 Pa5Oct (Wroclawvirus) were also targeted by END^{PaCF1} homologs in other strains,
190 BLASTP was used to identify strains carrying END^{PaCF1} homologs. The *P. aeruginosa*
191 lab strain PAK was found to encode a homolog of END^{PaCF1}, END^{PAK}, with 98% amino
192 acid identity. When END^{PAK} was deleted from their respective native strains, F8 and
193 PB-1 phages formed plaques. There were no visible changes upon infection with
194 Yuavirus, Abidjanvirus and Pa5Oct, suggesting that these phages were blocked
195 through another mechanism (Extended Data Figure 1c).

Additional homologs with different levels of amino acid identity as determined by BLASTP were introduced into PAO1 (see Extended Data Table 2 for full list of homologs tested). Plaque assays showed that END^{PaCF1} homologs END^{Pa3} (91% amino acid identity to END^{PaCF1} from *P. aeruginosa*) and END^{Pa4} (40% amino acid identity to END^{PaCF1} from *P. aeruginosa*) caused a 4-5 log reduction for phages 14-1, F8, YuA, Ab18 and Pa5Oct. In contrast, END^{Se} from *Salmonella enterica* (50% amino acid identity) resulted in a 3-log reduction for phages Ab18 and Pa5Oct (Extended Data Figure 1d). These results confirm that the family of END^{PaCF1} nucleases can target phages of these distinct families.

END homologs target diverse hypermodified phages

Of the targeted phages, Yuavirus M6 carries 5-(2-aminoethyl) uridine (5-NedU) instead of thymine in its genome[48]. The closely related Abidjanvirus Ab18 is predicted to also contain a hypermodified thymine[40,41]. However, the genomes of the other targeted phages targeted by END^{PaCF1} were not known to be hypermodified. To detect potential modifications, F8 and Pa5Oct genomes were subjected to *in vitro* digestion by various commercially available restriction enzymes with well-defined site specificities (Extended Data Table 3). Pa5Oct, a jumbo phage unrelated to phiKZ-nucleus forming phages[49], was resistant to EcoRI, BmrI, ApaLI, XbaI and BsaI, suggesting that it contains a modified G, in a GA and/or GG motif (Extended Data Table 2, Figure 3a). Analysis of the Pa5Oct genome with Domainator[50] followed by confirmation with Foldseek[51] also revealed genes associated with biosynthesis of 7-deaza-2'-deoxy-G such as gp162 (QueE; E-value 3.3 E^{-12}) and gp230 (6-pyruvoyl tetrahydropterin synthase; E-value 5.5 E^{-14}), further suggesting that Pa5Oct has a modified G.

We presumed that F8 phage contained no modifications because we have previously demonstrated digestion of its DNA *in vivo* with EcoRI[52]. However, we found that, of the ten restriction enzymes tested, only EcoRI, HindIII, XbaI, DraI, BsaI digested F8 DNA (Figure 3a). Virion DNA from phage F8 is not cleaved by BsgI, FspI, BmrI, ApaLI and NdeI. The enzymes which could not digest F8 all contain a TG dinucleotide within their recognition sequences, which is not present in the recognition sites of EcoRI, HindIII, XbaI, DraI, BsaI (Extended Data Table 3).

To identify the modified base, virion DNA purified phages F8 and 14-1 were enzymatically digested to free nucleosides and subjected to high-performance liquid chromatography/mass spectrometry (LC-MS). The chromatograms of F8 and 14-1 nucleoside digests revealed a novel nucleoside peak for each, with an apparent mass of 362 U and 390 U, respectively. Quantitation of peaks from canonical nucleotides shows a T:A ratio of less than one, indicating that these novel nucleosides to be thymidine derivatives replacing approximately one quarter of the canonical thymidines. The observed masses and the follow-up fragmentation analysis are consistent with an erythrose/threose sugar alcohol derivative in phage F8, and a circularized aldopentose derivative in phage 14-1 (Figure 3b, Extended Data Figure 2). Taken together, these restriction digests and analytical methods suggest that the Pbunavirus DNA previously thought to be unmodified, contains hypermodified thymine, in a TG motif. This confirms that the END nucleases appear to only target phages with hypermodified genomes, and that they can be used to reveal that a phage has a hypermodified genome when it is not previously known.

To determine if END nuclease homologs also recognized and/or targeted hypermodified DNA, previously identified homologs (END^{Pa3}, END^{Pa4}, END^{Se}) and additional homologs identified in other Gram-negative bacteria were expressed in *Escherichia coli* BW25113 for plaque assays. These assays confirmed that the *E. coli* homologs of END^{PaCF1}, specifically END^{Vc}, END^{Ec1} and END^{Se1}, targeted all hypermodified phages tested in *E. coli*, but not T5-like phages (Extended Data Figure 3 and Extended Data Table 2 for list of homologs). These hypermodified coliphages included *Nonagvirus*[53], *Seuratvirus*[53], T-even phages[54] and Mosigvirus[55,56], with a total of at least four different hypermodifications over two bases (G and C). T5 and related phages were used as a control as they lack hypermodifications in their genomes.

To further confirm that the phages were targeted due to the presence of hypermodification in the genome, we tested targeting of wildtype T4 containing glucose-hydroxymethyl-cytosine, T4 containing hydroxymethyl-cytosine (T4-HMC) and T4 with an unmodified genome (T4-C) in an *E. coli* DH10B strain to maintain the phage DNA modifications. The same phages were then used to infect DH10B carrying different END homologs. In this experiment, END^{Pa4} restricted both the wildtype T4 and T4-HMC but not T4-C. In contrast, homologs from *E. coli*, END^{Ec1} and the closely

related END^{Se}, caused a five-log decrease of plaquing efficiency for wildtype T4 but had modest protection against T4-HMC (Figure 3c), which is similar to the protection profile of type IV restriction endonuclease GmrSD[25]. END^{Vc} caused a four-log decrease in plaquing efficiency for T4 and a one-log decrease in plaquing efficiency of T4-HMC phage (Figure 3c). No END homologs restricted T4-C, which was in turn restricted by EcoRI as EcoRI targets non-modified DNA (Figure 3d)[57]. This confirms that END nuclease homologs recognize hypermodified DNA.

Overall, END^{Pa4} was found to target eight distinct hypermodified phage families with at least ten different known hypermodifications. Similarly, END^{Se} targeted all known modified phages in *E. coli*, as well as two groups of hypermodified phages in *P. aeruginosa*, Abidjanvirus and Wrowclawvirus. Other *P. aeruginosa* homologs, (END^{PAK}, END^{Pa3}, END^{PaCF1}) targeted all known hypermodified phages in *P. aeruginosa*, whereas END^{Ec1} and END^{Vc} targeted hypermodified phages in *E. coli*. Thus, all END homologs each targeted multiple hypermodifications (Figure 3e). None of the END homologs tested restricted growth of phages lacking hypermodified bases in their genomes in either *P. aeruginosa* or *E. coli* (Figure 3e). These results confirm that END homologs target only phages with hypermodified genomes, with some homologs targeting regardless of the specific nature of the hypermodification.

C-terminal domain of END^{PaCF1} determines substrate specificity and defines a larger group of END nucleases

We showed that END nucleases define a distinct family of promiscuous MDREs that appeared to be agnostic to which hypermodification they target (Figure 3e). We therefore sought to determine which domain was responsible for detecting the hypermodified genomes. HHpred analysis[58] revealed significant similarity of END^{PaCF1} CTD with endonuclease III family members, in particular, that from *Deinococcus radiodurans* (DrEndoIII) and eukaryotic Ogg1 DNA glycosylase (including human hOGG1) (Figure 4a, Extended Data Figure 4). Comparison of the AlphaFold3[59] structural model of END^{PaCF1} with the structures in the Protein DataBank confirmed similarity between the NTD and PLD superfamily enzymes and between CTD and ENDOIII / hOGG1 (Figure 4a), but not to the predicted structure of the Brig1 DNA glycosylase defense system recently characterized[60] (Figure 4b). DrEndoIII detects and targets oxidized T and C bases for base excision repair in prokaryotes, whereas

hOgg1 excises 8-oxo-7,8-dihydro-2'-deoxyguanosine in mammalian cells[61–63]. END^{PaCF1} lacks the catalytic residues of the endonuclease III family and therefore is likely to be enzymatically inactive (iEndoIII) (Extended Data Figure 4b). We therefore hypothesized that the CTD hOgg1/iEndoIII domain of END^{PaCF1} is responsible for detecting hypermodified genomes by recognizing the respective non-canonical bases. We first confirmed the conservation of the predicted structure and orientation of the CTD in the more distant homologs, END^{Pa4} and END^{Ec1} (approx. 40-50% amino acid identity to END^{PaCF1} and to each other) (Figure 4c). Next, to determine if CTD determines specificity, END^{Pa4} endonuclease was fused END^{Ec1} CTD, resulting in fusion protein END$^{Pa4\text{-}Ec1\text{-}fusion}$ (Figure 4d). All three nucleases END^{Pa4}, END^{Ec1} and END$^{Pa4\text{-}Ec1\text{-}fusion}$ fully targeted phages Bas21 and Bas47. However, similar to END^{Ec1} but not END^{Pa4}, END$^{Pa4\text{-}Ec1\text{-}fusion}$ preferentially targets T4 glu-HMC over T4-HMC (Figure 4e, Extended Data Figure 5). This result suggests that the CTD iEndoIII domain could be responsible for the specificity/substrate selection by END^{PaCF1}.

To further determine if specificity is determined by the CTD, we sought to identify other endonucleases containing a homologous PLD NTDs, but with different CTDs. Indeed, a PSI-Blast search identified PLD nucleases fused to several unrelated CTDs including EVE (of the PUA superfamily), helicases, helicase-associated domains previously shown to bind DNA[64], and pseudobarrel domains (Figure 4f). The NTD PLD nuclease fused to CTD pseudo-barrel or helicase domains have been identified as functional restriction enzymes BfiI and SauUSI, respectively[1,40]. Consistent with our hypothesis that substrate specificity is due to the CTD, SauUSI targets 5-methylcytosine DNA[1] whereas BfiI is encoded next to a methyltransferase and targets unmodified DNA[22,40]. SauUSI, in particular, is capable of restricting T4 phage infection[1]. Thus, END^{PaCF1} belongs to a larger group of defense endonucleases featuring an N-terminal PLD nuclease and a variable C-terminal modification-recognition domain.

In a reciprocal search, we identified the iEndoIII sensor fused to two distinct catalytic domains, a "NucS"-like PD-DExK superfamily nuclease, and the McrB-like GTPase. The PD-DExK-iEndoIII homolog from a metagenome (Accession MDV2503288 "END^{Mag}"), showed high similarity to archaeal Ogg1 enzymes that recognizes 8-oxo-guanine, similar to hOgg1[65], as demonstrated by HHpred search (Extended Data Figure 6). We cloned END^{Mag} into PAO1 and *E. coli* BW25113 and found that, in plaque

assays END^{Mag}, like END^{PaCF1}, targeted phages with modified G (Pa5Oct, Bas19, Bas20, Bas21, Bas25) but not phages with hypermodification on T or C or phages with unmodified genomes (Figure 4g). Thus, whereas Ogg1 and EndoIII target oxidized G and T/C bases respectively, END^{PaCF1} and END^{Mag} co-opted the inactivated hOgg1/iEndoIII domain as a sensor of modified bases to target hypermodified phage genomes (Figure 4h). Given that i**En**doIII **d**omain of these endonucleases appears to be responsible for their ability to recognize diverse hypermodified phage genomes, we propose to denote this group of defense enzymes collectively as **END** nucleases.

A single anti-END phage protein inhibits diverse END nuclease homologs

We previously observed that F8 and PB-1 were targeted by END^{PaCF1} in CF040, 14-1 was not (Figure 1c), but 14-1 is highly similar to F8 and PB-1 according to a gene cluster comparison (Figure 5a). To better understand what 14-1 gene facilitates escape from END^{PaCF1}, we made hybrid phages of 14-1 and F8 by co-infecting PAO1 with both phages followed by selection against parental 14-1 and F8 using Type I-C Cascade-Cas3 and END^{PaCF1} respectively. Analysis of the assembled hybrid phage genomes suggested three candidate regions that could harbor potential inhibitors. Sequences within each region were then individually queried for 14-1 and F8 SNPs across all hybrid phage genomes. It was found that all hybrid phages carried gp88 from 14-1, suggesting that this protein is an inhibitor of END^{PaCF1} (Figure 5b).

To test the prediction that 14-1 gp88 inhibits END^{PaCF1}, gp88 and its homolog identified using BLASTp, gp52 from jumbo phage Pa5Oct, were introduced into the native strain CF040. Expression of 14-1 gp88 or Pa5Oct gp52 increased F8 and PB-1 plaquing efficiency by five logs compared to wildtype. No difference in plaquing efficiency was observed for 14-1 (Figure 5c). To confirm that this effect was due to the inhibition of END^{PaCF1} and not other immune systems present in CF040, both inhibitors were expressed in the presence of END^{PaCF1} in PAO1 and facilitated plaquing of all families of phages. These results suggest that the inhibitor is not phage-family nor hypermodification specific (Figure 5d). Furthermore, Cascade-Cas3 targeting of gp88 enabled the isolation of a 14-1 phage mutant with an early stop codon in gp88 due to a frameshift mutation (14-1^{gp88*}). Compared to 14-1, 14-1^{gp88*} was susceptible to END^{PaCF1} in the native host but was rescued by either of the END^{PaCF1} inhibitors expressed *in trans* (Figure 5e). Additionally, these phage proteins also inhibited the *E.*

354 *coli* homolog END^{Ec2} (Figure 5f). Therefore, both 14-1 gp88 (anti-END1) and Pa5Oct
355 gp52 (anti-END1^{Pa5Oct}) blocks multiple END homologs regardless of the exact DNA
356 hypermodification targeted by END nucleases.

357 To test whether the identified inhibitors specifically bound END^{PaCF1}, anti-END1 and
358 FLAG-tagged anti-END1^{Pa5Oct} were cloned together with His-tagged END^{PAK} (90%
359 identity to END^{PaCF1}) into the pETDuet-1 vector in *E. coli* BL21 and were shown to co-
360 purify on a Ni-NTA column (Figure 5g). Additionally, we showed that both inhibitors co-
361 purified with END^{Pa4}, a distant homolog of END^{PaCF1} in *P. aeruginosa* (40% amino acid
362 identity), and inhibited END^{Pa4} *in vivo* in *E. coli* (Extended Data Figure 7). An
363 AlphaFold3[59] prediction of the defense system-inhibitor complex also suggests with
364 modest confidence that anti-END1 binds END^{PaCF1} (pTM = 0.62, ipTM=0.55) in the
365 iEndoIII region (Figure 5h).

366 Overall, we have identified a group of nucleases, the END nucleases, that is causal
367 for phage infection failure in a clinical isolate and target multiple DNA
368 hypermodifications. To counter the END nucleases, some phages with hypermodified
369 genomes encode anti-END inhibitors that can block multiple END homologs and bind
370 directly to these nucleases.

371

Discussion

The recent discovery of an unexpected, unprecedented diversity of antivirus defense systems in prokaryotes sparked much interest in prokaryotic immunity and stimulated the development of computational and experimental approaches and tools to identify immune systems and defense islands[5,6,12–14]. However, few tools exist to manipulate these islands in clinical strains, and over multiple genes. In this work, we adapted the recently developed Cascade-Cas3 technology for two purposes: first, to generate large deletions in defense islands, resulting in the identification of a distinct, single-gene defense system, and second, to probe phage genomes for inhibitors of defense systems. Given that the Cascade-Cas3 vector has been shown to be functional in other species, this approach can be widely exploited to interrogate defense islands and phages of other bacteria species in their native conditions.

The single-gene defense system identified in this work is an endonuclease homologous to type IIS endonucleases which belong to the PLD hydrolase superfamily. However, unlike previously studied type IIS endonucleases[1,25,29], END^{PaCF1} recognizes hypermodified DNA, without discriminating between different types of hypermodification including modifications with and without sugar moieties, and modifications on different nucleotide bases. We proposed that the promiscuous recognition of hypermodified bases is mediated by the C-terminal domain of END^{PaCF1}, an inactivated endonuclease III (iEndoIII). We found that iEndoIII domain can also combine with other nucleases that are predicted to target hypermodified DNA as well. Similarly to other domains involved in multiple defense systems, such as TIR and Sir2[66], the EndoIII domain is present across all three domains of life, eukaryotes[61], bacteria[62] and archaea[65]. The EndoIII domain is part of a larger helix-hairpin-helix (HhH) DNA glycosylase superfamily implicated in sensing damaged DNA for repair[65,67,68]. Mechanistic study of the HhH fold suggests that it can be easily mutated and evolved to recognize DNA adducts across all domains of life[69–71]. We show here that the inactivated EndoIII/Ogg derivative instead now serves as a sensor for multiple types of non-canonical bases in phage defense systems. We thus coined the term END nucleases for the defense enzyme with this type of modular organization.

The END nucleases are present in a broad variety of bacteria, from human pathogens such as *P. aeruginosa*, *E. coli*, *Klebsiella pneumoniae* to diverse environmental

bacteria such as *Rhizobium*, *Xanthomonas* sp. and probably many more as suggested by the identification of these genes in metagenomes. These findings imply that phages with hypermodified genomes are more common than currently known, and the vast array of DNA hypermodification mechanisms in phages and the defense systems targeting hypermodified DNA with varying degree of specificity remain to be characterized.

Additionally, using the END nucleases, we have identified the presence of hypermodified genomes on two phage families which were not known to be hypermodified previously. Currently, the identification of hypermodification in phage genomes is laborious, often relying on the presence of gene homologs previously implicated in modifying phage genomes, followed by confirmation with mass spectrometry analysis. However, as shown in this paper, this could miss out phages which do not carry such genes. One such example is the Pbunavirus family, which, despite being well studied due to its use in phage therapy, was not previously known to carry a hypermodified genome. Therefore, the END nucleases will provide a simple yet important tool to identify phages carrying a potentially hypermodified genome.

Our findings on END nucleases such as END^{PaCF1} support the modular model of modification dependent endonucleases[18], where the NTD is responsible for the catalytic activity, whereas the CTD is the sensor. In accord with this model, we found that PLD superfamily catalytic domains homologous to that of END^{PaCF1}'s also can combine with CTD sensors unrelated to iENDOIII, such as EVE, winged helix-turn-helix, pseudo barrel fold, helicases and helicases associated domains[1,22], which likely confer distinct substrate specificities. The modularity observed among END and other defense endonucleases is consistent with the previous observations on MDREs[24], antivirus STAND ATPases[18,44,72], and type III CRISPR systems[73] in all of which different sensors are combined with diverse effectors. This combinatorial modularity, driven by the Red Queen evolutionary dynamics, is emerging as a general principle of the evolution of immune systems. Another prominent aspect of the ubiquitous host-pathogen arms race is the evolution of defense system inhibitors by viruses, best explored in the case of anti-CRISPR proteins[74]. Here, we identified a family of anti-END inhibitors encoded by unrelated phages with distinct hypermodifications on DNA that protects against the END nuclease which *P. aeruginosa* phages are not able to

436 escape from. A lack of escape is consistent with some phages being unable to simply
437 lose their modified state, although this is possible for T4 and did induce END nuclease
438 escape. Undoubtedly, many more combinations of prokaryotic immune systems with
439 virus-encoded inhibitors or escape mechanisms remain to be discovered, and the
440 study of their interactions will enrich our understanding of the strikingly complex
441 landscape of microbial immunity.

442

Methods

Plasmid construction

To generate inserts in the pHERD30T (shortened to p30T) plasmid, the p30T backbone was amplified with primers WX_16/WX_17 (see Supplementary Table 1 for list of primers), before being subjected to DpnI (NEB) digest for 37 °C, 1 h. Inserts were amplified using primers with overhangs complementary to linearized p30T backbone, or ordered as gene fragments gblocks from Twist. For END nucleases from other bacteria, ATG was appended as the start codon if required. The inserts were then joined with the p30T backbone using Hi-Fi DNA Gibson Assembly (NEB) by incubating at 50 °C, 1 – 4 h. *E. coli* DH5α was then transformed with the resulting plasmids and verified by sequencing with Plasmidsaurus. *P. aeruginosa* cells were then electroporated with the p30T constructs and selected on 50 μg/ml gentamicin. p30T plasmids were also transformed into BW25113 and DH10B and selected on 15 μg/ml gentamicin. p30T sfCherry-EcoRI and sfCherry-EcoRI^{dead} have been previously constructed and tested for *in vivo* restriction function[57] before transforming into DH10B.

To generate all-in-one Cascade Cas3 plasmids carrying different guides, plasmids carry the Cascade-Cas3 system and targeting sequences in the 14-1 phage or in the Type I RM system in CF040 were constructed by first digesting the empty vector with BsaI at 37 °C for 4 h. Guides with complementary overhangs (in the form of oligos ordered from IDT) were subjected to T4 Polynucleotide Kinase (NEB) treatment for 3 h, 37 °C and inactivated at 65 °C, 10 min. Guides were then annealed with their complementary strands by heating to 95 °C in a thermocycler for 10 minutes, then gradually ramping down to room temperature Guides were then diluted and ligated into BsaI-digested Cascade Cas3 with T4 ligase (Thermofisher Scientific) at rtp for 1h. DH5α were then transformed with the plasmid before recovering for 2h and selected on 15 μg/ml gentamicin. Cascade Cas3 plasmids with the correct guides (verified using Sanger sequencing) were then electroporated into PAO1/CF040.

To clone END^{PaCF1} and homologs into pETDuet, pETDuet with/without the inhibitor was digested with NdeI and XhoI for 4 h. Gene fragments carrying the defense system were amplified with the appropriate primers. Fragments were the ligated into the backbone using Hi-Fi Assembly according to manufacturer's instructions, DH5α were

474 then transformed with the ligated plasmid and then selected in the presence of 100
475 μg/ml carbenicillin. Colony PCR was then carried out with primers WX_216/WX_217;
476 those with correct inserts were extracted and sent for sequencing with Plasmidsaurus
477 or Quintara BioSciences. Desired plasmids were then transformed into BL21 or
478 BW25113 using the heat shock chemical transformation protocol.

479 For cloning inhibitors into pETDuet, pETDuet with/without homologs of END^{PaCF1} was
480 digested with NcoI and SalI. Gene fragments carrying the inhibitors were amplified
481 with the appropriate fragments, ligated and selected as per previous. Colony PCR was
482 carried out with primers WX_257/WX_258.

483 To insert END^{PaCF1} into the PAO1 chromosome, the mini-Tn7 insertion method was
484 used[42]. Briefly, the defense system was first cloned into the lab's mini-Tn7 plasmid.
485 The plasmid backbone was first amplified with primers WX_129/WX_130; END^{PaCF1}
486 and the corresponding 60bp upstream were amplified using primers WX_167/WX_168,
487 before cloning into the plasmid. *P. aeruginosa* PAO1 cells were electroporated with the
488 resulting plasmid and helper plasmid pTNS3 and selected for using gentamicin. The
489 gentamicin marker was then excised with FLP recombinase as previously described[75].
490 Colonies were patched onto LB plates with/without gentamicin to identify colonies
491 which were gentamicin sensitive, i.e. the gentamicin cassette flipped out.

492 To delete END^{PaCF1} or END^{PAK} from CF040 or PAK respectively, the allelic exchange
493 method was used[75]. Briefly, PMQ30 plasmids were first digested with BamHI and
494 HindIII (NEB). Inserts with the required overhangs were ordered as gene fragments
495 from Twist to facilitate cloning. The backbone and inserts were ligated by HiFi DNA
496 Gibson Assembly (NEB), before being transformed into DH5α and selected on
497 gentamicin plates. Plasmids were then extracted using Zyppy miniprep kit (Zymo) as
498 per manufacturer's protocol and electroporated into SM10 cells. SM10 cells carrying
499 PMQ30 plasmids were then crossed with CF040 or PAK respectively, before selection
500 on VBMM + gentamicin. Gentamicin resistant cassette was then flipped out using FLP
501 recombinase, and negative selection was carried out on no salt LB + 15% sucrose at
502 30 °C. Colonies were patched onto LB plates with/without gentamicin to identify
503 colonies which were gentamicin sensitive, i.e. the gentamicin cassette flipped out.

504

505 **Plate reader assays**

506 Overnight cultures of PAO1 with/without the appropriate p30T constructs were diluted
507 in fresh media with 50 μg/ml gentamicin (if required) at a 1:1000 dilution. Phage lysates
508 were also serially diluted to the appropriate MOI. Diluted cultures (100 μl) were then
509 added to 96- well plates (Thermofisher Scientific) before 10 μl of phage lysates were
510 added and left to grown at 37 °C for 24h in the plate reader (Agilent Technologies),
511 shaking at 425 rpm.

512 **Plaque assays**

513 Plaque assays were carried out for *P. aeruginosa* and *E. coli*. For *P. aeruginosa*, strains
514 were first grown up overnight with/without antibiotics as required. The overnight
515 cultures (100-150 μl) were then mixed with 0.4% top agar and left to dry for at least 15
516 minutes. Phage lysates were diluted 10-fold, and 2 μl was spotted onto top agar,
517 before the plates were incubated at 37 °C. No inducers were added to the bottom agar
518 unless as indicated.

519 For *E. coli*, strains were first grown up overnight with/without antibiotics as required.
520 For experiments with BW25113 carrying p30T constructs, the overnight cultures were
521 then subcultured (1:10) and grown for another 2-3 h, before 100-150 μl were used for
522 mixing with 0.7% top agar and left to dry for at least 15 minutes. For all other
523 experiments with BW25113 and BL21, 100 μl of overnight cultures was mixed with 0.7%
524 top agar. For experiments with DH10B, 0.4% top agar was used to facilitate T4-C
525 plaque formation. Phage lysates were then similarly diluted and spotted before plates
526 were incubated at 37 °C.

527 **Generating competent *E. coli***

528 Chemically competent BW25113 and DH10B (from the Bushman lab) was generated
529 as per previous[76]. Briefly, cells were grown to OD_{600} = 0.3 at 37 °C and centrifuged at
530 4 °C. All remaining steps were done on ice. Cells were washed twice in 50 mM $CaCl_2$
531 with 10 mM Tris pH 7.5 and finally frozen in 50 mM $CaCl_2$ with 10 mM Tris pH 7.5 and
532 15% glycerol with liquid nitrogen before storing at -80 °C. For transformation, cells
533 were subjected to 42 °C heat shock, recovered in LB and plated on 100 μg/ml
534 carbenicillin or 15 μg/ml gentamicin as required.

535 **Generating recombinant phages**

536 To generate hybrid 14-1 and F8 phages, *P. aeruginosa* PAO1 was first grown overnight,
537 then subcultured and grown to OD_{600} 0.3 at 37 °C, 170 rpm. Phages 14-1 and F8 were
538 then mixed and added to the subculture at MOI = 10. After 10-20 minutes of growth,
539 bacterial cells were centrifuged and resuspended in fresh LB with 10 mM $MgSO_4$. After
540 2 h, a 1% volume of chloroform was added and mixed into the culture. The tubes were
541 then left at 37 °C for 15 min, followed by centrifugation at 10,000 x g, 2 min to remove
542 cell debris. Cells were lysed twice with chloroform, and the supernatant phage lysate
543 was stored at 4 °C.

544 The hybrid phages generated were then plated onto CF040 carrying Cascade-Cas3
545 plasmids with guides against 14-1. Large plaques suggestive of hybrid phages rather
546 than escape phages were collected; individual plaques were then propagated in the
547 same CF040-Cas3 background. Phages were then sequenced as per sequencing
548 protocol.

549 **Generating 14-1 with gp88 frameshifted.**

550 Full plate infection using phage 14-1 on PAO1 carrying Cascade-Cas3 plasmid
551 targeting gp88 was carried out in a full plate infection. Individual plaques were then
552 picked and plated on PAO1 with and without END^{PaCF1} to confirm if plaques were true
553 escapers. Phages were then checked for presence of gp88 with primers
554 WX_133/WX_128 and sent for sequencing with Quintara BioSciences.

555 **Propagation of T4 with different hypermodified/unmodified genomes**

556 Phages T4 wildtype carrying a glucose-hydroxymethyl-cytosine in its genome, T4
557 carrying hydroxymethyl-cytosine in the genome (T4-HMC) and T4 with unmodified
558 genome (T4-C) were received from the Bushman lab. Strains for phage propagation,
559 CR63 and DH10B, were also received in parallel from the Bushman lab. For wildtype
560 T4 and T4-HMC, phages were propagated on DH10B. For T4-C, phages were first
561 propagated on CR63, before propagating on DH10B to ensure that the resulting phage
562 for experiments have an unmodified genome. All experiments with these three phages
563 were carried out on a DH10B background as per previous[77].

564

Sequencing

CF040 with regions of immune islands deleted were sequenced with Plasmidsaurus with whole bacterial genome sequencing. Sequences were then manually compared to wildtype CF040.

Sequencing of phages was carried out in house as described previously. Briefly, genomic DNA was first extracted using a modified SDS/Proteinase K method. Briefly, 100 μl of phage lysate was mixed with lysis buffer to a final concentration of 10 mM Tris, 1 mM EDTA, 0.5% SDS before incubating at 37 °C for 30 min, followed by 55 °C for 30 min – 1 h. DNA was then purified using DNA Clean and Concentrator kit and quantified with Nanodrop. 20-100 ng genomic DNA was used to prepare WGS libraries using Illumina DNA Prep Kit. Subsequently, PCR indexing-amplification of tagmented DNA was performed using 2x Phusion Master Mix (NEB) and custom-ordered indexing primers, amplified for 12 cycles. Libraries were further purified by agarose gel electrophoresis and purified with Zymoclean Gel DNA Recovery Kit (Zymo) as per manufacturer's instructions, Trimmed reads were assembled de-novo with SPADES[78]. All hybrid phage genomes were then queried for 14-1/F8 SNPs in the genome using the command-line 'grep' function.

qPCR

PAO1 with/without END^{PaCF1} were grown overnight in LB + 10 mM $MgSO_4$. Cultures were then subcultured at a 1:50 dilution and grown to OD_{600} = 0.4. F8 were then added to cultures at MOI of approx. 1. Samples (500 μl) were then taken at t = 0, 15, 30 and 60 min post infection and washed once with cold 100 μl PBS (Gibco). Samples were then frozen on dry ice, before genomic DNA was extracted using the modified SDS/Proteinase K method as previously described. Genomic DNA (2 – 10 ng) was then used for qPCR with Luna Universal qPCR Master mix (NEB), with *rpoD* primers and primers WX_Q4F/WX_Q4R, on the CFX Connect Real-Time PCR Detection System (Bio-Rad) at UCSF.

Restriction digest

100 ng of F8/14-1 gDNA were incubated with 1 μl of restriction enzyme (NEB) as stated in 50 μl rCutSmart reaction at 37 °C, 1h. For BmrI, NEB buffer r2.1 was used. Reactions were then run on 0.8% agarose gel in TAE buffer.

Protein purification

BL21 carrying plasmids were first grown overnight at 37 °C in LB 150 μg/ml carbenicillin. Cells were then subcultured into 100 ml LB 150 μg/ml carbenicillin and grown to log phase for 1.5-2 h. Inducer (1mM IPTG) was then added and the cultures were left to grow at 18 °C overnight.

The next day, cultures were centrifuged at 4000 x g for 10 min. Cells were then resuspended in lysis buffer, before four rounds of sonication. Lysates were then put through a nickel column and washed four times, each with 4 volumes of resin. Proteins were then eluted with elution buffer three times, each with an equivalent volume of resin. The lysates, wash 4 and elution 1 were collected. SDS-PAGE gels were then run using Tris-Glycine SDS running buffer (Bio-Rad) at 170 V for 45 min and stained with Direct Blue. Gels were then imaged with the Gel Doc.

Characterization of phage-modified nucleosides

LC-MS/MS fragmentation analysis for identifying the chemical identity of the base modification in F8 and 14-1 genomes were carried out as per previous[48]. Briefly, F8 and 14-1 DNA were digested to nucleosides overnight with Nucleoside Digestion Mix in 1x Nucleoside Digestion Mix Reaction buffer (NEB). The resulting DNA nucleosides were then analyzed using Agilent 6490 Triple quadrupole LC/MS system. The chromatography was performed using a Waters XSelect HSS T3 C18 column (2.1 × 100 mm, 2.5 μm particle size) operated at a flow rate of 0.6 mL/min with a linear gradient of aqueous buffer (10 mM aqueous ammonium formate, pH 4.4) and methanol over 6 minutes. MS/MS fragmentation spectra were obtained by collision-induced dissociation in the positive product ion mode with the following parameters: gas temperature 200 °C, gas flow 14 L/min, nebulizer 45 psi, sheath gas temperature 350 °C, sheath gas flow 11 L/min, capillary voltage 2 kV, nozzle voltage 1.5 kV, fragmentor voltage 380 V and collision energy 5–65 V.

Bioinformatics analysis

Amino acid alignments were done using ClustalOmega[79] and webLOGO was used to identify the consensus sequence. To determine genomic neighborhoods, amino acid sequence of END^{PaCF1} were input into webFLAGS[38]. For determining the CTD domain, the amino acid sequence was input into HHpred[58]. For identifying potential genes

involved in synthesizing hypermodified genomes, Domainator[50] and Foldseek[51] were used. Alignment of phage genome sequences from NCBI was done with Clinker[80].

To determine total number of homologs, PSI-Blast was carried out as previously described[2]. Briefly, homologs of END^{PaCF1} were first identified using PSI-Blast with the maximum number of target sequence set at 5,000 and E-value cut-off of 0.005. Two more iterations were then carried out, and the retrieved sequences were aligned using MAFFT[81]. The resulting alignment was then visualized using Jalview[82], and entries which resulted in large gaps in the alignments were discarded. Due to the large difference in size, entries larger and smaller than 500 amino acid sequences were further differentiated.

For analysis of other homologs with shared domain, two iterations of PSI-BLAST was ran against complete genomes. The best 760 hits were analyzed (E-value $\leq 10^{-12}$). Only hits with domains aligning to either the N or the C terminal domain of END^{PaCF1} were selected. The iENDOIII CTD domain was further ran against the clustered NR and domain organizations of retrieved proteins with significant sequence similarity were examined.

Protein structures were predicted using AlphaFold3[59], and visualized using PyMoL[83].

Author contributions statement

W-X. Y., T. A. K., A. W., A. E. G. generated strains, conducted experiments and analyses. Y-J. Lee conducted nucleoside analysis. K. S. M. conducted bioinformatic analysis of homologs with shared domains. W-X. Y. drafted the initial manuscript, and T. A. K., B. C., K. S. M., E. V. K., P. R. W. and J. B. D. reviewed and edited the text. J. B. D. and P. R. W. provided overall supervision and secured funding.

Competing interests

J.B.D. is a scientific advisory board member of SNIPR Biome, Excision Biotherapeutics, and LeapFrog Bio, consults for BiomX, and is a scientific advisory board member and co-founder of Acrigen Biosciences and ePhective Therapeutics. The Bondy-Denomy lab received research support from Felix Biotechnology.

Acknowledgements

We would like to thank Sukrit Silas and Horia Todor for providing *E. coli* strain BW25113 and the Basel phages, the Brouns lab at TU Delft for the vB_PaeM_FBPa21 phage (annotated in this paper as vB-Pa21) and the Bushman lab at University of Pennsylvania for T4 phages with different genome hypermodifications.

References

1. Xu, S. Y., Corvaglia, A. R., Chan, S. H., Zheng, Y. & Linder, P. A type IV modification-dependent restriction enzyme SauUSI from Staphylococcus aureus subsp. aureus USA300. *Nucleic Acids Res* **39**, 5597–5610 (2011).

2. Huiting, E. *et al.* Bacteriophages inhibit and evade cGAS-like immune function in bacteria. *Cell* **186**, 864-876.e21 (2023).

3. Ofir, G. *et al.* Antiviral activity of bacterial TIR domains via immune signalling molecules. *Nature* **600**, 116–120 (2021).

4. LeRoux, M. *et al.* The DarTG toxin-antitoxin system provides phage defence by ADP-ribosylating viral DNA. *Nat Microbiol* **7**, 1028–1040 (2022).

5. Georjon, H. & Bernheim, A. The highly diverse antiphage defence systems of bacteria. *Nature Reviews Microbiology* vol. 21 686–700 Preprint at https://doi.org/10.1038/s41579-023-00934-x (2023).

6. Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science (1979)* **359**, (2018).

7. Wein, T. & Sorek, R. Bacterial origins of human cell-autonomous innate immune mechanisms. *Nature Reviews Immunology* vol. 22 629–638 Preprint at https://doi.org/10.1038/s41577-022-00705-4 (2022).

8. Bernheim, A., Cury, J. & Poirier, E. Z. The immune modules conserved across the tree of life: Towards a definition of ancestral immunity. *PLoS Biol* **22**, (2024).

9. Cury, J. *et al.* Conservation of antiviral systems across domains of life reveals immune genes in humans. *Cell Host Microbe* **32**, 1594-1607.e5 (2024).

10. Schwardmann, L. S., Nölle, V. & Elleuche, S. Bacterial non-specific nucleases of the phospholipase D superfamily and their biotechnological potential. *Applied Microbiology and Biotechnology* vol. 104 3293–3304 Preprint at https://doi.org/10.1007/s00253-020-10459-5 (2020).

11. Barrangou, R. & Doudna, J. A. Applications of CRISPR technologies in research and beyond. *Nat Biotechnol* **34**, 933–941 (2016).

12. Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. V. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. *J Bacteriol* **193**, 6039–6056 (2011).

13. Gao, L. *et al.* Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science (1979)* **369**, 1077–1084 (2020).

14. Johnson, M. C. *et al.* Core defense hotspots within *Pseudomonas aeruginosa* are a consistent and rich source of anti-phage defense systems. *Nucleic Acids Res* **51**, 4995–5005 (2023).

15. Hochhauser, D., Millman, A. & Sorek, R. The defense island repertoire of the Escherichia coli pan-genome. *PLoS Genet* **19**, (2023).

16. Vizzarro, G., Lemopoulos, A., Adams, D. W. & Blokesch, M. Vibrio cholerae pathogenicity island 2 encodes two distinct types of restriction systems . *J Bacteriol* (2024) doi:10.1128/jb.00145-24.

17. O'Hara, B. J., Alam, M. & Ng, W. L. The Vibrio cholerae Seventh Pandemic Islands act in tandem to defend against a circulating phage. *PLoS Genet* **18**, (2022).

18. Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A., Wilson, G. G. & Murrayy, N. E. Highlights of the DNA cutters: A short history of the restriction enzymes. *Nucleic Acids Res* **42**, 3–19 (2014).

19. Roberts, R. J. *et al.* A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* **31**, 1805–1812 (2003).

20. Oliveira, P. H., Touchon, M. & Rocha, E. P. C. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res* **42**, 10618–10631 (2014).

21. Pingoud, A. & Jeltsch, A. Structure and function of type II restriction endonucleases. *Nucleic Acids Res* **29**, 3705–3727 (2001).

22. Sapranauskas, R. *et al.* Novel subtype of type IIs restriction enzymes. BfiI endonuclease exhibits similarities to the EDTA-resistant nuclease nuc of Salmonella typhimurium. *Journal of Biological Chemistry* **275**, 30878–30885 (2000).

23. Weigele, P. & Raleigh, E. A. Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. *Chemical Reviews* vol. 116 12655–12687 Preprint at https://doi.org/10.1021/acs.chemrev.6b00114 (2016).

24. Bell, R. T., Sahakyan, H., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. CoCoNuTs are a diverse subclass of Type IV restriction systems predicted to target RNA. *Elife* **13**, (2024).

25. Bair, C. L. & Black, L. W. A Type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *J Mol Biol* **3**, 768–778 (2007).

26. Lutz, T. *et al.* A protein architecture guided screen for modification dependent restriction endonucleases. *Nucleic Acids Res* **47**, 9761–9776 (2019).

27. Pastor, M. *et al.* Crystal structures of the EVE-HNH endonuclease VcaM4I in the presence and absence of DNA. *Nucleic Acids Res* **49**, 1708–1723 (2021).

28. Hosford, C. J., Bui, A. Q. & Chappie, J. S. The structure of the Thermococcus gammatolerans McrB N-terminal domain reveals a new mode of substrate recognition and specificity among McrB homologs. *Journal of Biological Chemistry* **295**, 743–756 (2020).

29. Helbrecht, I. *et al.* Characterization of winged helix domain fusion endonucleases as N6-methyladenine-dependent type IV restriction systems. *Front Microbiol* **15**, (2024).

30. Selvy, P. E., Lavieri, R. R., Lindsley, C. W. & Brown, H. A. Phospholipase D: Enzymology, functionality, and chemical modulation. *Chemical Reviews* vol. 111 6064–6119 Preprint at https://doi.org/10.1021/cr200296t (2011).

31. Bullen, N. P. *et al.* An enterococcal phage protein inhibits type IV restriction enzymes involved in antiphage defense. *Nat Commun* **15**, (2024).

32. Ho, C. H., Wang, H. C., Ko, T. P., Chang, Y. C. & Wang, A. H. J. The T4 phage DNA mimic protein arn inhibits the DNA binding activity of the bacterial histone-like protein H-NS. *Journal of Biological Chemistry* **289**, 27046–27054 (2014).

33. Silas, S. *et al.* Activation of programmed cell death and counter-defense functions of phage accessory genes. Preprint at https://doi.org/10.1101/2023.04.06.535777 (2023).

34. Rifat, D., Wright, N. T., Varney, K. M., Weber, D. J. & Black, L. W. Restriction endonuclease inhibitor IPI* of bacteriophage T4: A novel structure for a dedicated target. *J Mol Biol* **3**, 720–734 (2008).

35. Pirnay, J. P. *et al.* Personalized bacteriophage therapy outcomes for 100 consecutive cases: a multicentre, multinational, retrospective observational study. *Nat Microbiol* **9**, 1434–1453 (2024).

36. Csörgő, B. *et al.* A compact Cascade–Cas3 system for targeted genome engineering. *Nat Methods* **17**, 1183–1190 (2020).

37. Shmidov, E. *et al.* Multigenerational Proteolytic Inactivation of Restriction Upon Subtle Genomic Hypomethylation. Preprint at https://doi.org/10.1101/2024.10.21.619520 (2024).

38. Saha, C. K., Pires, R. S., Brolin, H., Delannoy, M. & Atkinson, G. C. FlaGs and webFlaGs: Discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics* **37**, 1312–1314 (2021).

39. Tamulaitiene, G., Silanskas, A., Grazulis, S., Zaremba, M. & Siksnys, V. Crystal structure of the R-protein of the multisubunit ATP-dependent restriction endonuclease NgoAVII. *Nucleic Acids Res* **42**, 14022–14030 (2014).

769 40. Grazulis, S. *et al. Structure of the Metal-Independent Restriction Enzyme BfiI*
770 *Reveals Fusion of a Specific DNA-Binding Domain with a Nonspecific*
771 *Nuclease*. www.pnas.orgcgidoi10.1073pnas.0507949102 (2005).

772 41. Koonin, E. V. A duplicated catalytic motif in a new superfamily of
773 phosphohydrolases and phospholipid synthases that includes poxvirus
774 envelope proteins. *Trends Biochem Sci* **7**, 242–243 (1996).

775 42. Choi, K. H. & Schweizer, H. P. mini-Tn7 insertion in bacteria with single attTn7
776 sites: example Pseudomonas aeruginosa. *Nat Protoc* **1**, 153–161 (2006).

777 43. Rousset, F. & Sorek, R. The evolutionary success of regulated cell death in
778 bacterial immunity. *Curr Opin Microbiol* (2023).

779 44. Gao, L. A. *et al.* Prokaryotic innate immunity through pattern recognition of
780 conserved viral proteins. *Science (1979)* **377**, (2022).

781 45. Zhang, T. *et al.* Direct activation of a bacterial innate immune system by a viral
782 capsid protein. *Nature* **612**, 132–140 (2022).

783 46. Dy, R. L., Przybilski, R., Semeijn, K., Salmond, G. P. C. & Fineran, P. C. A
784 widespread bacteriophage abortive infection system functions through a Type
785 IV toxin-antitoxin mechanism. *Nucleic Acids Res* **42**, 4590–4605 (2014).

786 47. Lopatina, A., Tal, N. & Sorek, R. Abortive Infection: Bacterial Suicide as an
787 Antiviral Immune Strategy. (2025) doi:10.1146/annurev-virology-011620.

788 48. Lee, Y. J. *et al.* Identification and biosynthesis of thymidine hypermodifications
789 in the genomic DNA of widespread bacterial viruses. *Proc Natl Acad Sci U S A*
790 **115**, E3116–E3125 (2018).

791 49. Lood, C. *et al.* Integrative omics analysis of Pseudomonas aeruginosa virus
792 PA5oct highlights the molecular complexity of jumbo phages. *Environ Microbiol*
793 **22**, 2165–2181 (2020).

794 50. Johnson, S. R. *et al.* Domainator, a flexible software suite for domain-based
795 annotation and neighborhood analysis, identifies proteins involved in antiviral
796 systems. *Nucleic Acids Res* **53**, (2025).

797 51. van Kempen, M. *et al.* Fast and accurate protein structure search with
798 Foldseek. *Nat Biotechnol* **42**, 243–246 (2024).

799 52. Kokontis, C., Klein, T. A., Silas, S. & Bondy-Denomy, J. Conserved jumbo
800 phage factors required for protein import into a phage nucleus. Preprint at
801 https://doi.org/10.1101/2024.03.27.586873 (2024).

802 53. Hutinet, G. *et al.* 7-Deazaguanine modifications protect phage DNA from host
803 restriction systems. *Nat Commun* **10**, (2019).

54. Lehman, I. R. & Pratt, E. A. On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4, and T6. *J Biol Chem* **235**, 3254–3259 (1960).

55. Thomas, J. A., Orwenyo, J., Wang, L. X. & Black, L. W. The odd “RB” phage—identification of arabinosylation as a new epigenetic modification of DNA in T4-like phage RB69. *Viruses* **10**, (2018).

56. Hutinet, G., Lee, Y.-J., De Crécy-Lagard, V. & Weigele, P. R. Hypermodified DNA in Viruses of E. coli and Salmonella. *EcoSal Plus* **9**, (2020).

57. Kokontis, C., Klein, T. A., Silas, S. & Bondy-Denomy, J. Multi-interface licensing of protein import into a phage nucleus. *Nature* (2025) doi:10.1038/s41586-024-08547-x.

58. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, (2005).

59. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

60. Hossain, A. A. *et al.* DNA glycosylases provide antiviral defence in prokaryotes. *Nature* **629**, 410–416 (2024).

61. Bruner, S. D., Norman, D. P. G. & Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**, (2000).

62. Sarre, A. *et al.* Structural and functional characterization of two unusual endonuclease III enzymes from Deinococcus radiodurans. *J Struct Biol* **191**, 87–99 (2015).

63. Sarre, A. *et al.* The three Endonuclease III variants of Deinococcus radiodurans possess distinct and complementary DNA repair activities. *DNA Repair (Amst)* **78**, 45–59 (2019).

64. Mills, J. L. *et al.* Solution NMR structure of the helicase associated domain BVU-0683(627-691) from Bacteroides vulgatus provides first structural coverage for protein domain family PF03457 and indicates domain binding to DNA. *J Struct Funct Genomics* **14**, 19–24 (2013).

65. Faucher, F., Duclos, S., Bandaru, V., Wallace, S. S. & Doublié, S. Crystal Structures of Two Archaeal 8-Oxoguanine DNA Glycosylases Provide Structural Insight into Guanine/8-Oxoguanine Distinction. *Structure* **17**, 703–712 (2009).

66. Mariano, G. & Blower, T. R. Conserved domains can be found across distinct phage defence systems. *Mol Microbiol* **120**, 45–53 (2023).

67. Boiteux, S. & Radicella, J. P. The human OGG1 gene: Structure, functions, and its implication in the process of carcinogenesis. *Arch Biochem Biophys* **377**, 1–8 (2000).

68. Rollo, F. *et al.* Insights into the role of three Endonuclease III enzymes for oxidative stress resistance in the extremely radiation resistant bacterium Deinococcus radiodurans. *Front Microbiol* **14**, (2023).

69. Trasviña-Arenas, C. H., Demir, M., Lin, W. J. & David, S. S. Structure, function and evolution of the Helix-hairpin-Helix DNA glycosylase superfamily: Piecing together the evolutionary puzzle of DNA base damage repair mechanisms. *DNA Repair (Amst)* **108**, (2021).

70. Zhang, L. *et al.* Alkylated DNA repair by a novel HhH-GPD family protein from Crenarchaea. *Nucleic Acids Res* **53**, (2025).

71. Begley, T. J. & Cunningham, R. P. *Methanobacterium Thermoformicicum Thymine DNA Mismatch Glycosylase: Conversion of an N-Glycosylase to an AP Lyase*. *Protein Engineering vol* vol. 12 https://academic.oup.com/peds/article/12/4/333/1474858 (1999).

72. Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. Structure of the multimodular endonuclease FokI bound to DNA. *Nature* **388**, 97–100 (1997).

73. Hoikkala, V., Graham, S. & White, M. F. Bioinformatic analysis of type III CRISPR systems reveals key properties and new effector families. *Nucleic Acids Res* **52**, 7129–7141 (2024).

74. Stanley, S. Y. & Maxwell, K. L. Phage-Encoded Anti-CRISPR Defenses. *Annu Rev Genet* **11**, (2025).

75. Hmelo, L. R. *et al.* Precision-engineering the Pseudomonas aeruginosa genome with two-step allelic exchange. *Nat Protoc* **10**, 1820–1841 (2015).

76. Banta, A. B., Ward, R. D., Tran, J. S., Bacon, E. E. & Peters, J. M. Programmable Gene Knockdown in Diverse Bacteria Using Mobile-CRISPRi. *Curr Protoc Microbiol* **59**, (2020).

77. Bryson, A. L. *et al.* Covalent modification of bacteriophage T4 DNA inhibits CRISPRCas9. *mBio* **6**, (2015).

78. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**, 455–477 (2012).

79. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).

80. Gilchrist, C. L. M. & Chooi, Y. H. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473–2475 (2021).

81. Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059–3066 (2002).

82. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

83. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrodinger, LLC.

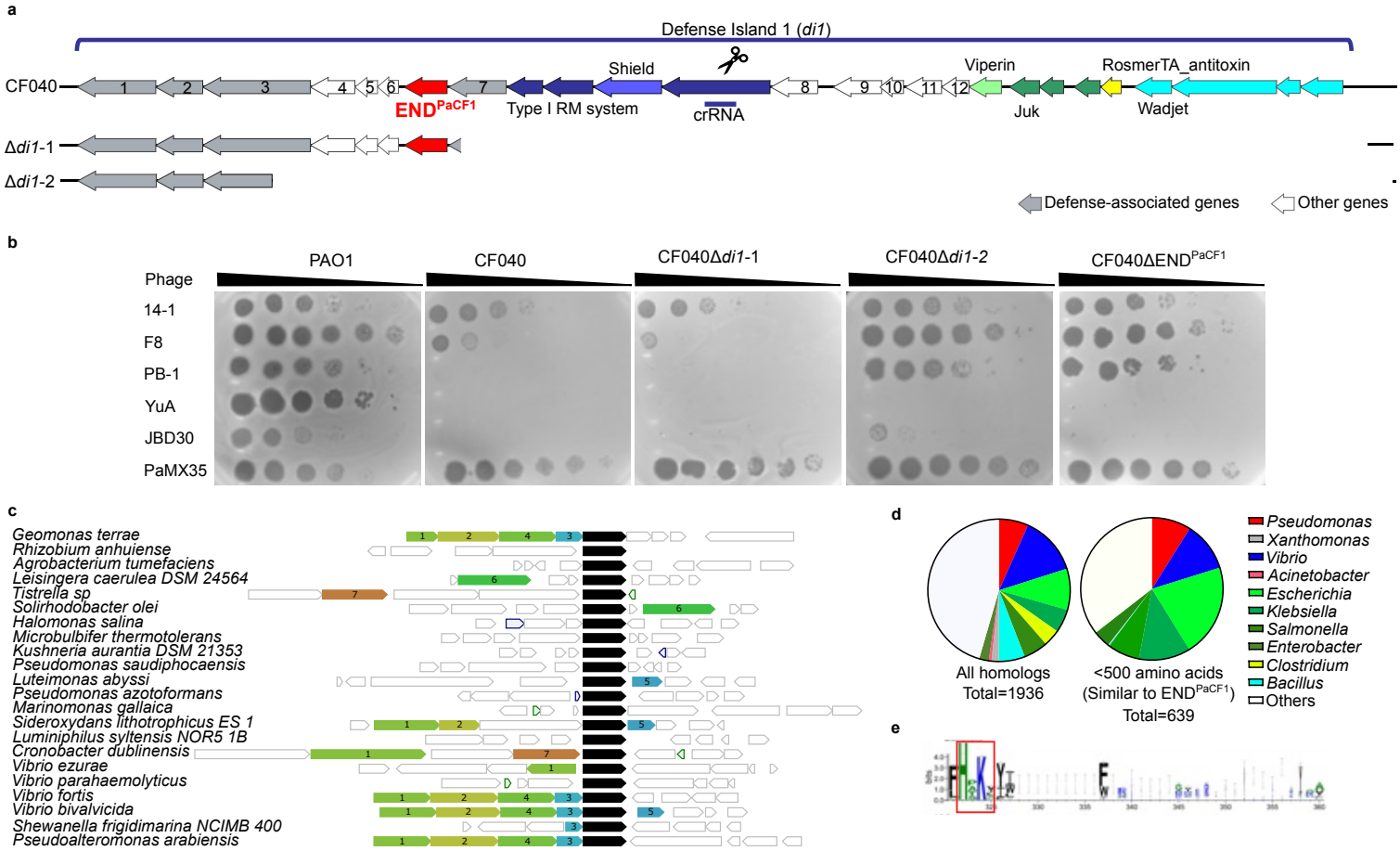


Figure 1: Identification of type IIS-like endonuclease END^{PaCF1} in clinical isolate CF040 using CASCADE Cas3 technology. **a**, Deletions generated in clinical isolate CF040 using CASCADE Cas3 against each type I restriction modification (RM) system in each of the two putative islands found in CF040. Grey arrows: Defense-related genes; white arrows: other genes (see Extended Data Table 1 for annotation of each numbered gene) The Type-IIS-like endonuclease was renamed as END^{PaCF1}. The Cascade Cas3 guide designed against the *hsdR* gene in this defense island is indicated as crRNA. **b**, Plaque assays with different phages onto CF040 with different genes within the immune island deleted (as seen in a). Phages are spotted in 10-fold serial dilutions onto bacterial lawns, and clearings represent phage replication. **c**, Amino acid sequence of END^{PaCF1} was input into web-FLAGS for neighbourhood analysis in other bacteria. 1: DNA methylase family; 2: Restriction endonuclease subunit S; 3: hypothetical protein; 4: AAA ATPase; 5: Restriction endonuclease; 6: VapE family protein; 7: BrxL. **d**, Genus of bacteria in which homologs of END^{PaCF1} are identified are shown in pie-chart. Homologs were shortlisted after three iterations of PSI BLAST, with a cut off of >70% coverage and E values < 0.0005. Left: all homologs; right: homologs with a similar size to END^{PaCF1} (i.e. < 500 amino acid). **e**, Consensus sequence of amino acid near the endonuclease active site (HxK, boxed in red). Sequences (n=639) aligned as per previous were input into WebLogo3. The height of the amino acid lettering corresponds to the conservation of this amino acid at the indicated position.

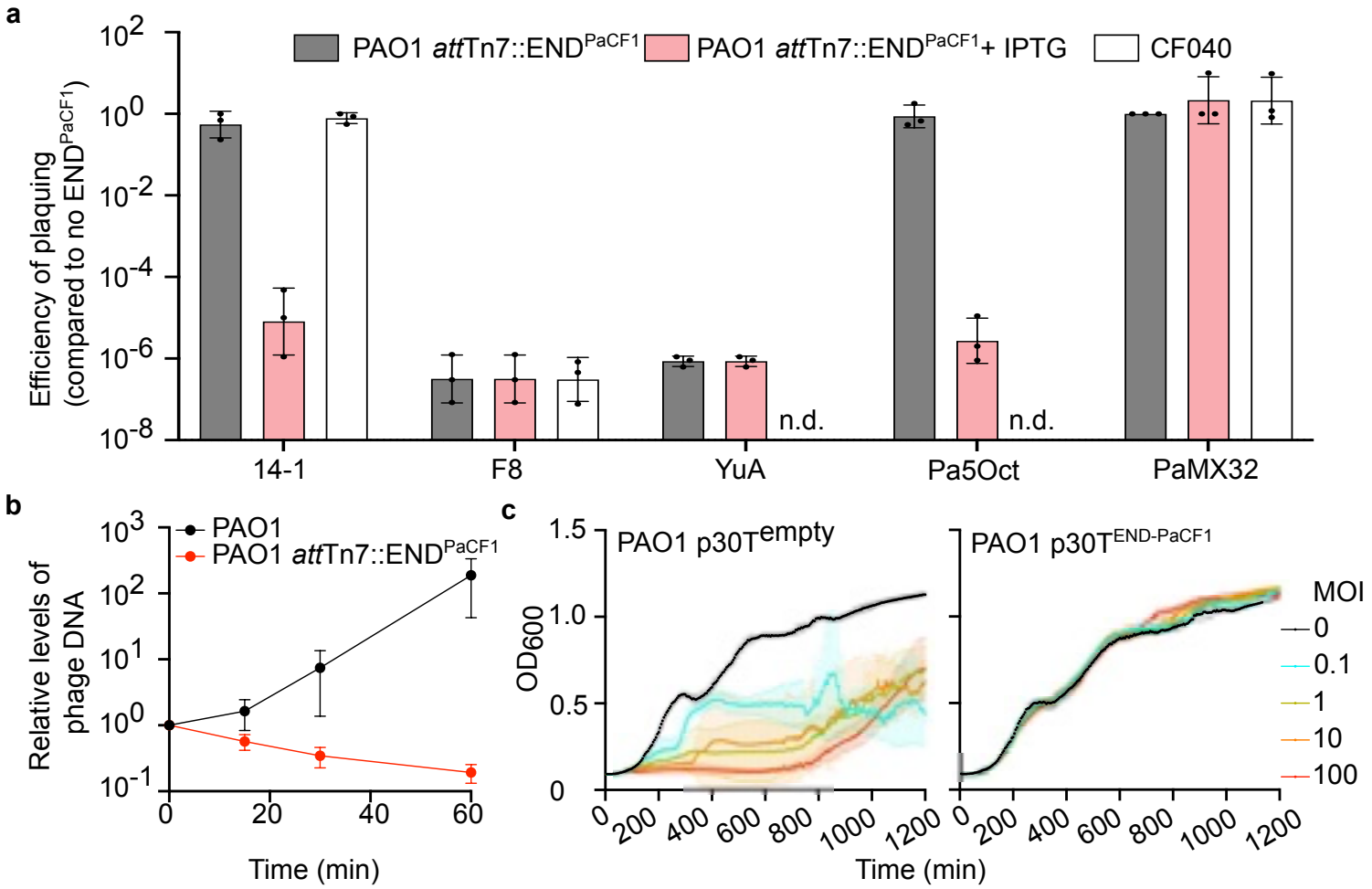


Figure 2: END^{PaCF1} functions within the half hour of infection and is not abortive. a, F8-like phages 14-1 and F8, YuA and Pa5Oct were plaqued onto PAO1 and onto PAO1 carrying END^{PaCF1} on the chromosome at the Tn7 integration site and CF040 with/without END^{PaCF1}. Efficiency of plaquing is calculated as percentage with respect to phage titer on PAO1/CF040ΔEND^{PaCF1}. PaMX32 is included as a plaque control, n.d. refers to not detected. 1mM IPTG was included to induce expression of integrated END^{PaCF1}. **b**, qPCR time course assay F8 infection in PAO1 and PAO1 *att*Tn7::END^{PaCF1} without induction. Samples (500 µl) were taken at t= 0, 15, 30, and 60 minutes post infection. Values were first normalized to PAO1 gDNA using primers binding to *rpoD*, before normalizing to t = 0. **c**, Growth curves of PAO1 with/without END^{PaCF1} infected with F8 at various MOIs.

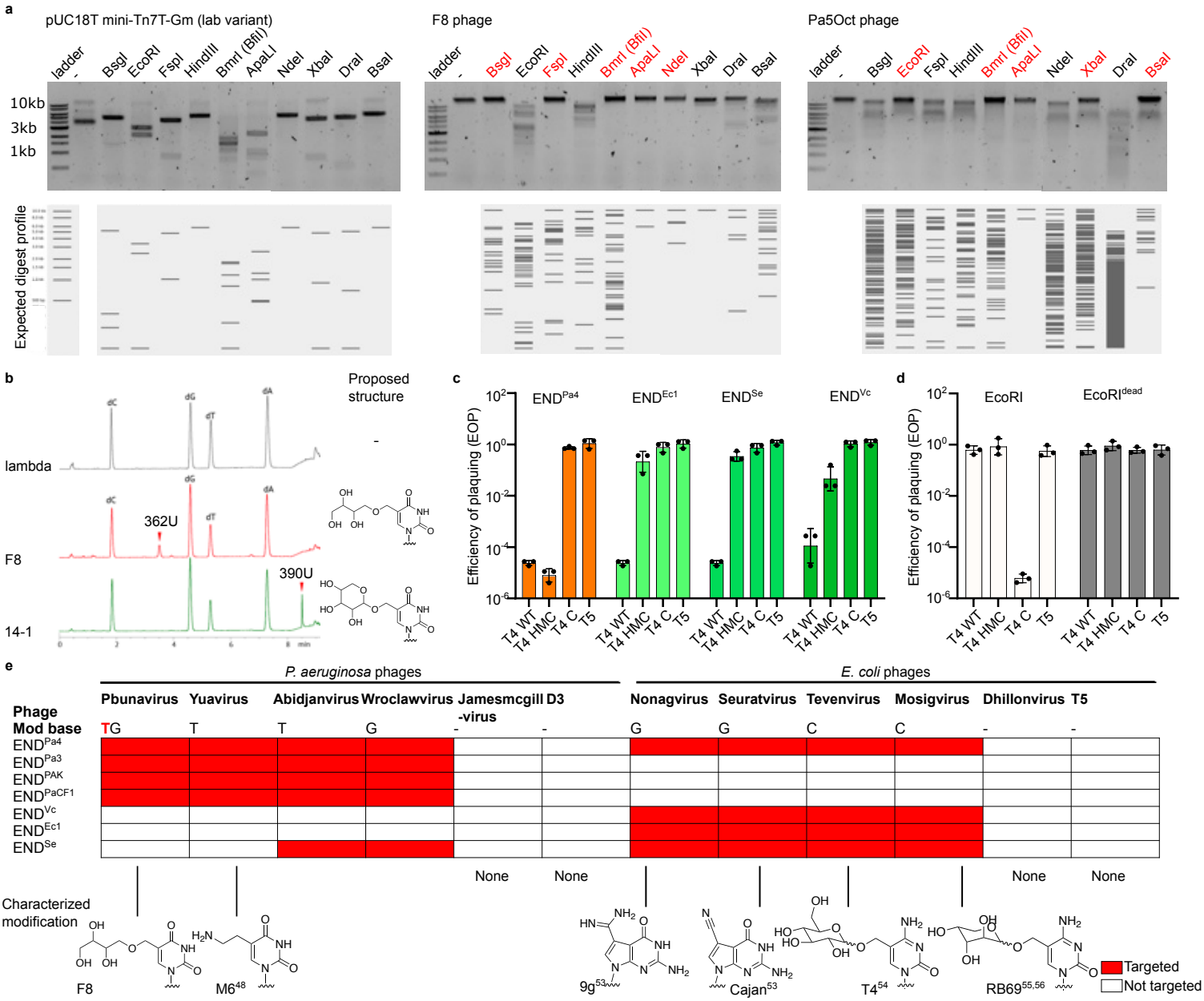


Figure 3: Pbunavirus F8 and 14-1 has a modified thymidine in its genome; END^{PaCF1} therefore targets only phages with modified genomes, but is agnostic to the exact modification. a, Restriction digest of pUC18T mini-Tn7T-Gm plasmid (lab variant), F8 genome and Pa5Oct genome with restriction enzymes. Restriction enzymes in red indicate that they did not digest F8/Pa5Oct phage shown. **b**, Mass spectrometry of F8 and 14-1 genomic DNA. Lambda phage, which does not have a modified genome, is included as a control. **c**, T4 phages with wildtype, hydroxylmethylcytosine (HMC) or unmodified cytosine (C) were plaqued against p30T carrying homologs of END^{PaCF1}, as indicated on the graph. T5 was included as a control. Efficiency of plaquing shown is relative to empty vector. **d**, T4 phages with different modifications were plaqued against DH10B with functional or dead EcoRI. Efficiency of plaquing shown is relative to empty vector; T5 was included as a control. **e**, Table of phages that are targeted by different homologs of END^{PaCF1}. An example of phage DNA hypermodifications is shown, including the phage from which the characterization was done. Only the F8 hypermodification is described here; the others were obtained from previously published papers.

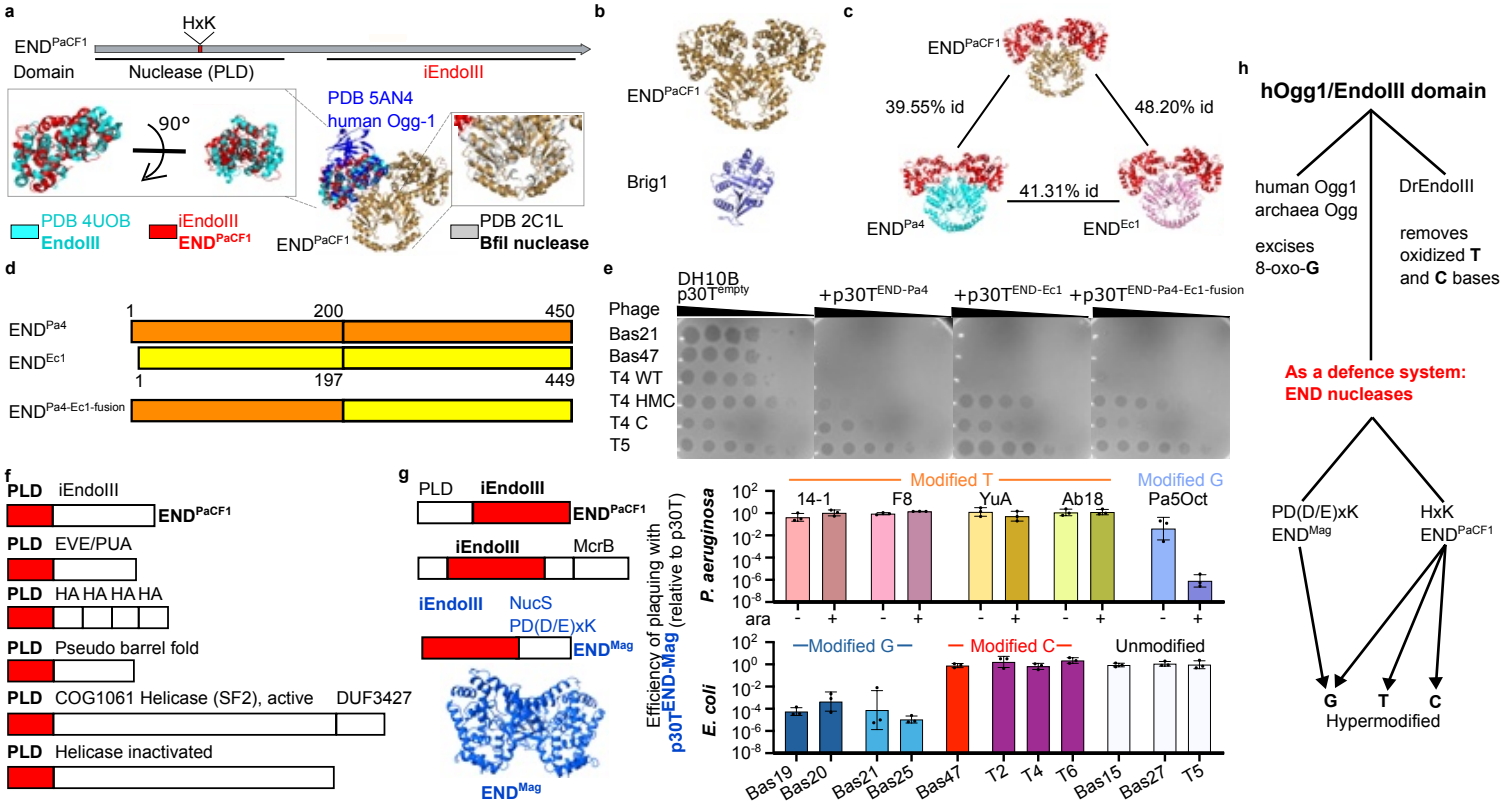


Figure 4: Domain organization and function of END nucleases. a, Domain organization and AlphaFold model for END^{PaCF1}. Approximately domain boundaries are based on HHpred results (Extended Data Figure 4). The N-terminal nuclease domain belongs to phospholipase D-like (PLD) family. Bottom: Overlay (cealign) of Alphafold3 prediction of END^{PaCF1} domains with known structures PDB 4UOB (endonuclease III, EndoIII), 5AN4 (human Ogg-1) and 2C1L (BfiI) **b**, AlphaFold model for END^{PaCF1} compared to the prediction for Brig1. **c**, AlphaFold model for END^{PaCF1} and distant homologs. The percent amino acid identity between each homolog is as shown. The C-terminal domain (CTD) is shown in red. **d**, Schematic of swap between CTD from END^{Pa4} homolog and END^{Ec1} homolog. The amino acids and the numbered position indicate position of swap. **e**, Plaque assays of Bas21, Bas 47 and T4 phages carrying different modifications on its genome with END^{Pa4}, END^{Ec1} and END$^{Pa4\text{-}Ec1\text{-}fusion}$ shown in d. HMC = hydroxymethylcytosine, C = unmodified cytosine. Plaque assays are representative of three repeats. **f**, Domain organization of other HxK/PLD nuclease fusions. The nuclease domain is colored red. Abbreviations: PUA - pseudouridine synthase and archaeosine transglycosylase domain, same family as EVE domain, HA - helicase associated domain, COG - cluster of orthologous genes, DUF - domain of unknown function. **g**, Domain organization of other of iENDO fusions. The iEndoIII domain is colored red. McrB is a GTPase associated with type IV restriction-modification systems and NucS belong to PD(D/E)xK nuclease superfamily. END^{Mag} (Alphafold structure prediction shown in blue), was tested in both *P. aeruginosa* (top) and *E. coli* (bottom). For *P. aeruginosa*, the system is further induced with arabinose (ara) **h**, Schematic of targets of known proteins carrying the hOgg1/EndoIII domain.

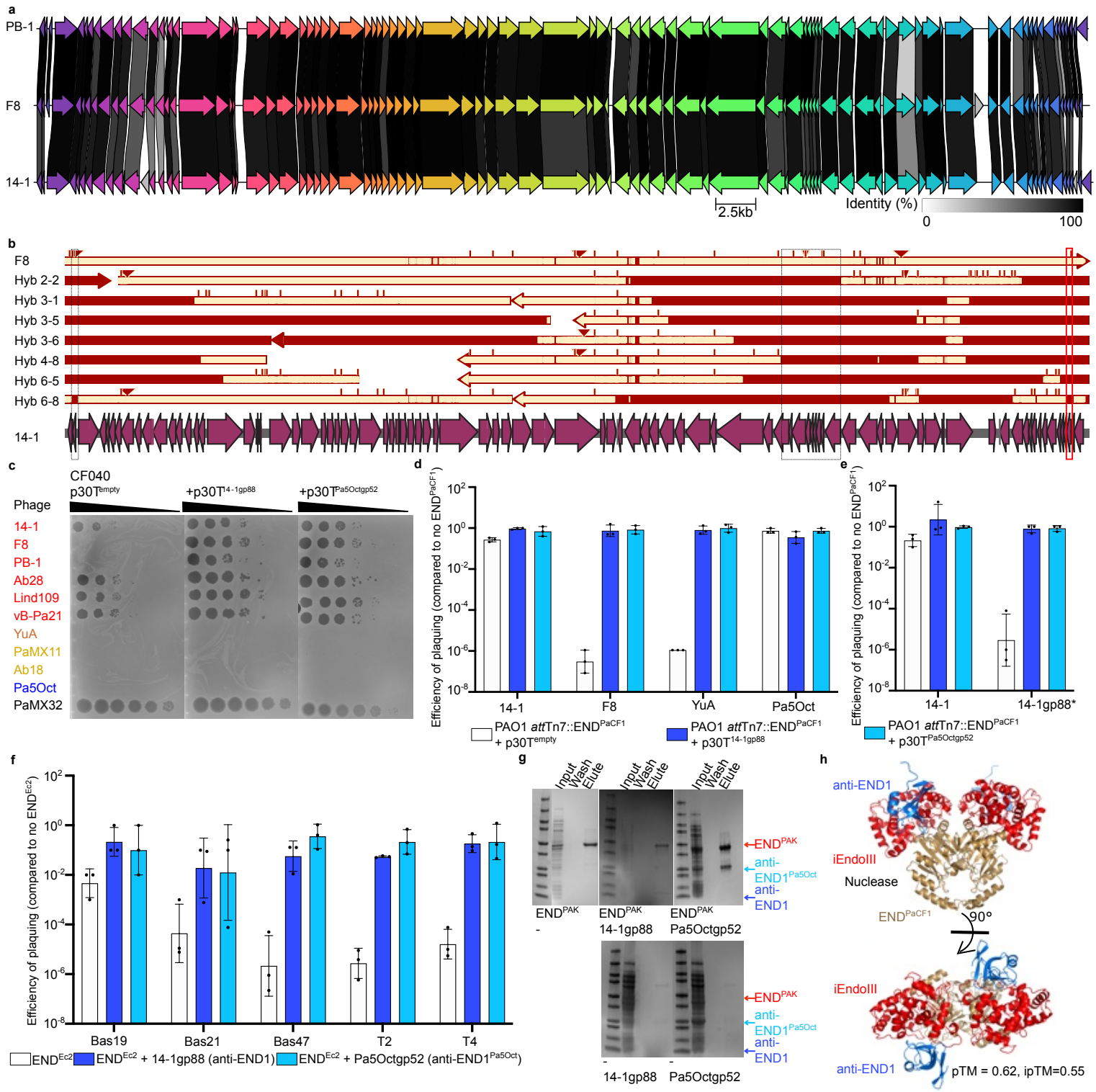
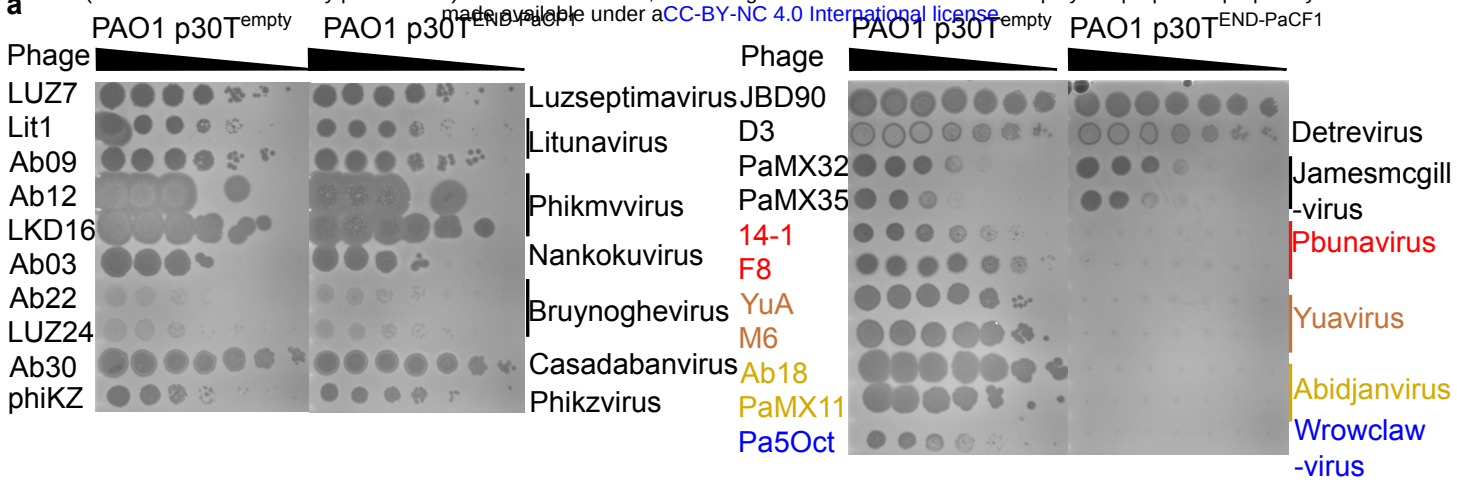
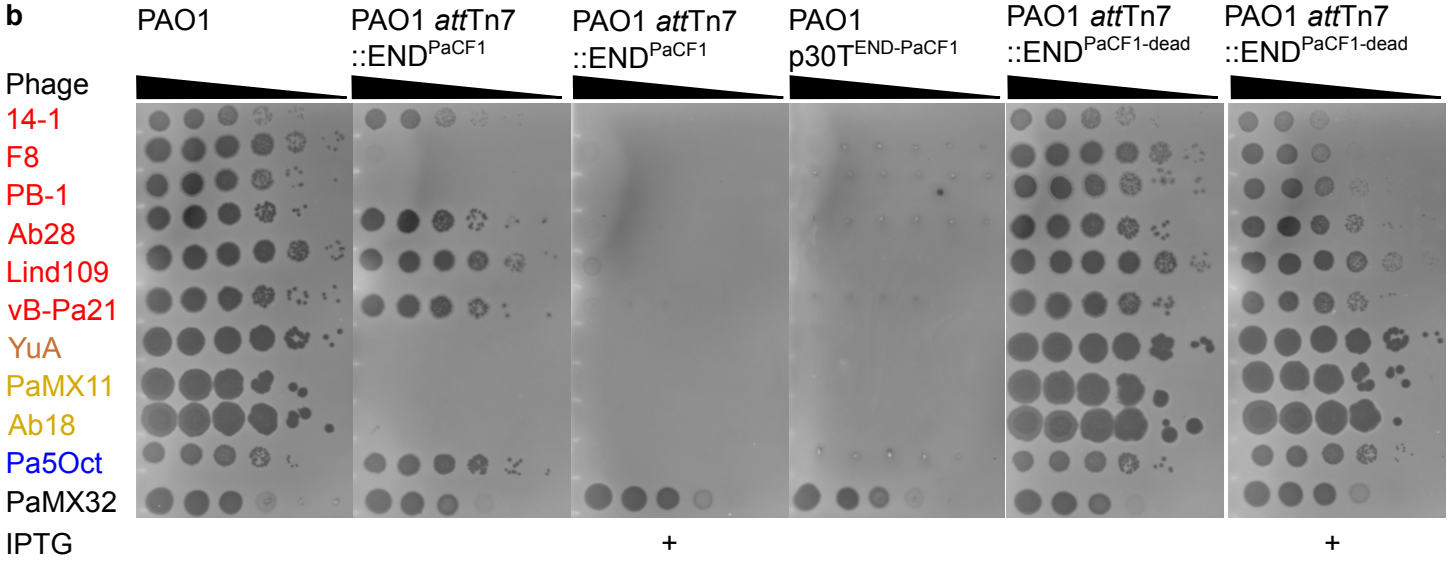
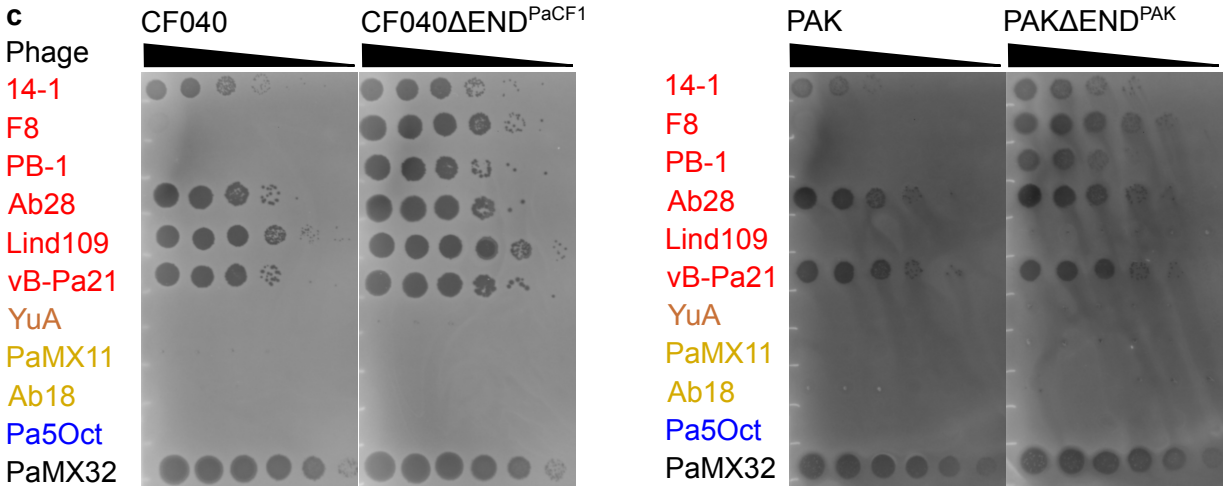
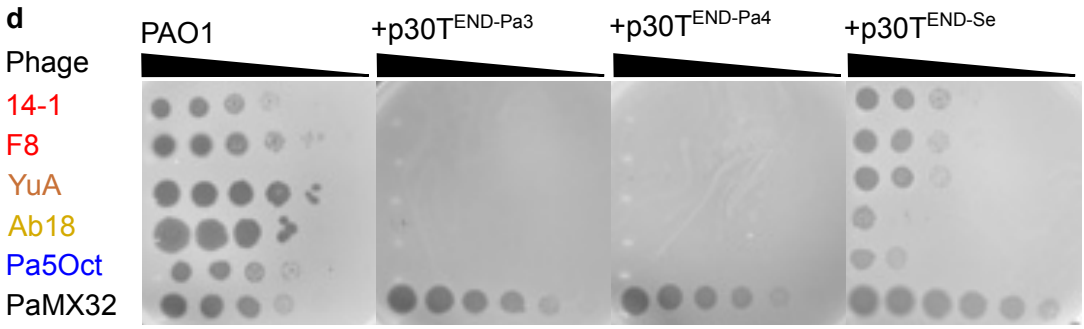


Figure 5: END^{PaCF1} inhibitors are encoded by some Pbunavirus phages including 14-1, as well as unrelated phage Pa5Oct and cross protect targeted phages from other families. **a**, Alignment of F8, 14-1 and PB-1. 14-1 is not inhibited by END^{PaCF1} in CF040, while F8 and PB-1 are. Alignment was done with clinker. **b**, Alignment of hybrid phages of 14-1 and F8. Hybrid phages of 14-1 (not targeted by endogenous END^{PaCF1} in CF040) and F8 (targeted by endogenous END^{PaCF1} in CF040) were constructed by co-infection, then sequenced with Illumina sequencing prior to assembly using SPADES. Alignment on Snapgene (shown here) and parsing with grep revealed all hybrid phages carried 14-1 gp88 (shown with a red box). **c**, Panel of *P. aeruginosa* phages were tested against END^{PaCF1} in the native strain CF040 with inhibitors cloned into p30T plasmid. Plaque assay shown is representative of three repeats. **d,** Plaque assay with PAO1 carrying END^{PaCF1} integrated into *att*Tn7 site, with 14-1 gp88 or Pa5Oct 0052 expressed in *trans*. Empty p30T plasmid was included as a control. **e**, Plaque assay with wildtype 14-1 or 14-1 with gp88 (14-1gp88*) with a frameshift mutation, onto PAO1 *att*Tn7::END^{PaCF1}. Presence of either inhibitor rescued plaquing of 14-1gp88*. **f,** Plaque assay with different hypermodified phages in *E. coli*, in the presence of END^{Ec2} with/without 14-1 gp88 (anti-END1) or Pa5Oct gp52 (anti-END1^{Pa5Oct}) inhibitors on pDUET. **g**, Pulldown of inhibitors in the presence of defense system. Only END^{PAK} is His-tagged. **h**, Alphafold3 prediction of END^{PaCF1}- anti-END1 interaction. The predicted structure shown is consistent across all five predicted outputs.

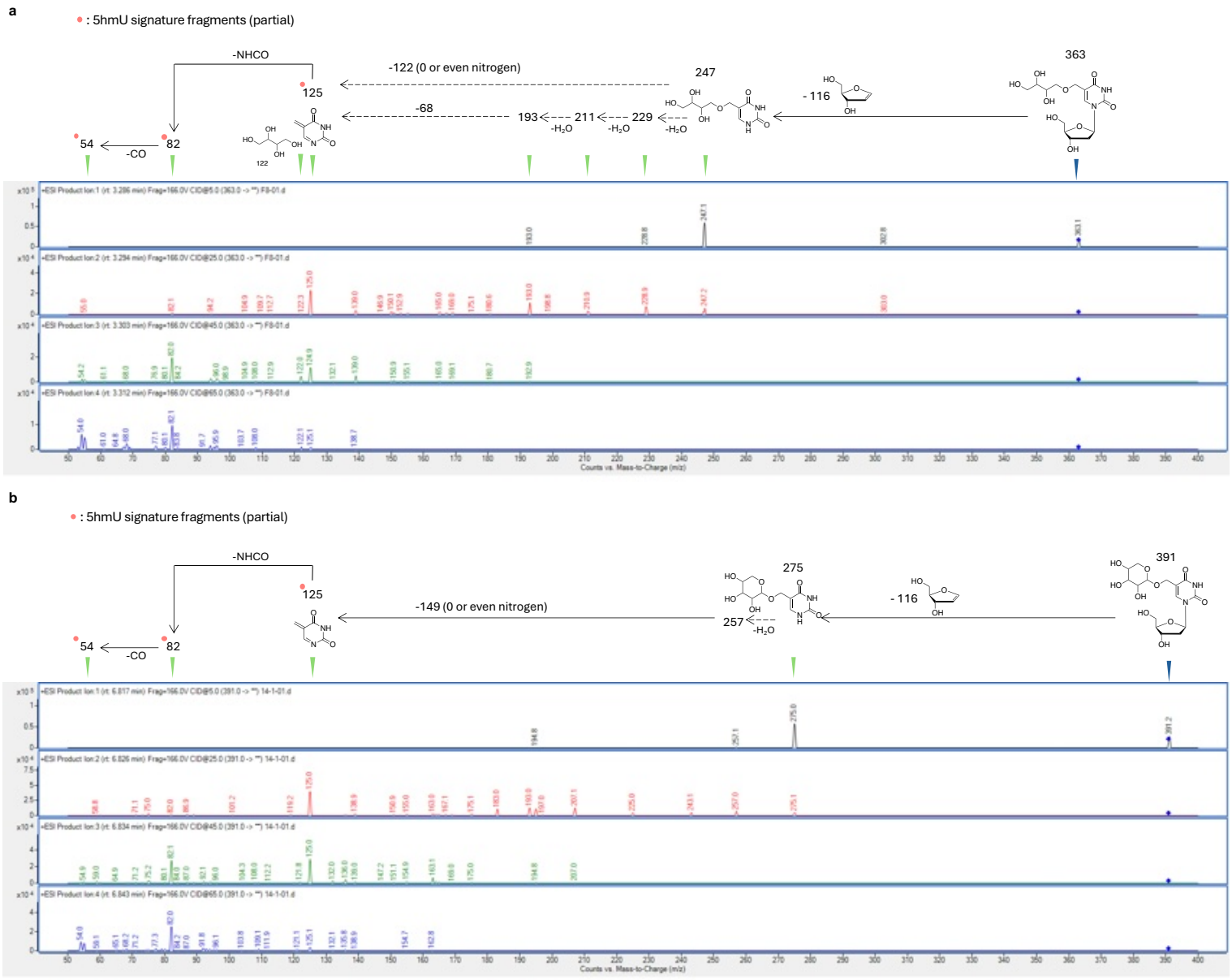




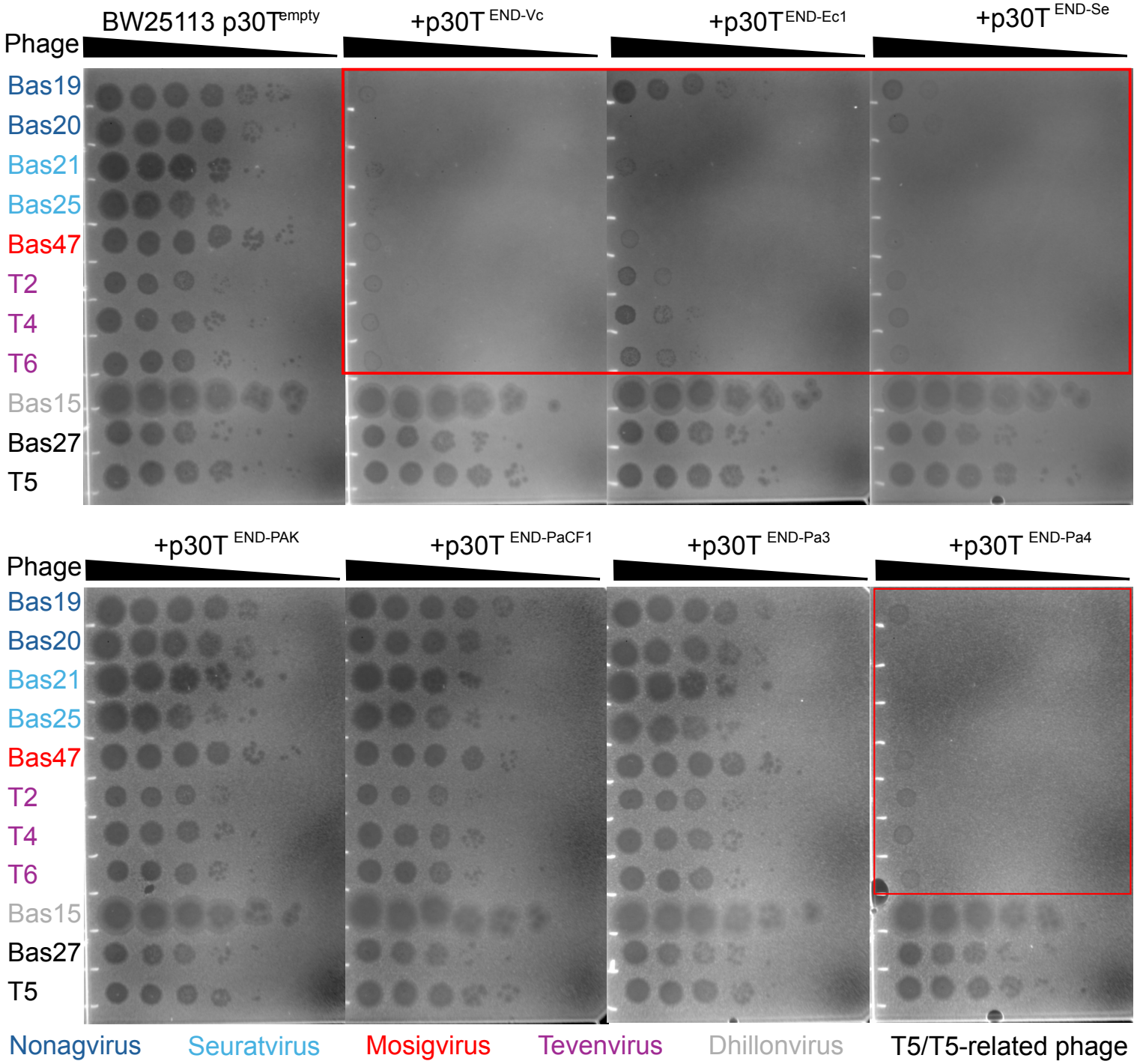




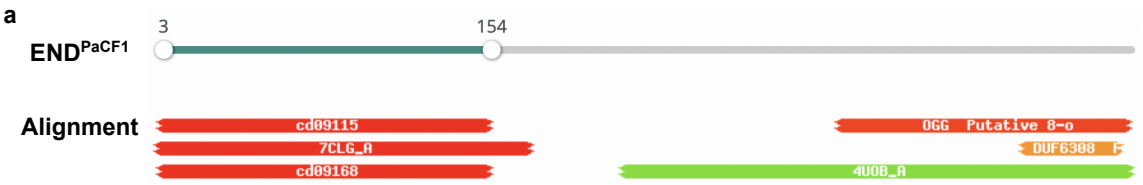
Extended Data Figure 1: END^{PaCF1} and other homologs targets F8-like and YuA-like phages, and is functional in native strains CF040 and PAK. a, Phages across different phage families were plaqued onto PAO1 carrying p30T$^{END-PaCF1}$. Empty p30T plasmid was included as a control; p30T plasmid was used to maximise expression of END^{PaCF1}. **b**, Plaque assays for all tested phages with increasing induction of END^{PaCF1}. 1mM IPTG was used to induce *att*Tn7::END^{PaCF1}. Plaque assays are representative of three repeats. **c**, Phages from F8 family, YuA family and Pa5Oct were plaqued onto CF040 and PAK with and without END^{PaCF1}/ END^{PAK} deleted. **d**, Plaque assays for PAO1 carrying p30T$^{END-PaCF1}$ homologs. PaMX32 was included as a phage control. Plaque assays shown are representative of at least three repeats.



Extended Data Figure 2: +ESI-MS/MS fragmentation spectra of nucleoside from phages F8 and 14-1. Product-ion spectra with collision-induced dissociation (CID) energies of 5, 25, 45 and 65 is as shown for **a,** F8 and **b,** 14.1



Extended Data Figure 3: Homologs of END^{PaCF1} inhibit *E. coli* phages with hypermodified genomes. Panel of *E. coli* phages from the Basel collection was tested against END^{PaCF1} homologs across different bacteria. Red boxes highlight phages which showed a decrease in titer in the presence of the defense system. T5 and T5 related phages, which are known to have non-modified genomes, are used as controls.



b

7CLG_B Putative helicase; Restriction endonuclease, ATPase, helicase, SRA, DNA BINDING PROTEIN; HET: SO4, MSE; 3.1A {Staphylococcus aureus subsp. aureus USA300}

Probability: 99.46%, E-value: 2.5e-12, Score: 144.64, Aligned cols: 155, Identities: 18%, Similarity: 0.223, Template Neff: 9.7

```
Q ss_pred            ccccCCCCcccHHHHHHHHHcccCCCEEEEEEccccCCcccc----hHHHHHHHHCCCcEEEEEecCCCCCCCCHHHHHHH
Q Q_7656688    2     KLIANRINGVHLRDLLPALSAEVQVDSVLAAVAVGSSASDIT----QDLVGHCVANKLRLDLWMRYDHTVPVSVELLKRL   77 (450)
Q Consensus    2     ~~~~~~~~~~l~~~l~~~a~~~i~ia~y~~~~g~~~~~~~~L~~~~~gv~v~ii~~~~~~~~~~~~l~~l   77 (450)
                       +++.|.+ +..+...+...+.  .+.++.|+++|++.   +.     ...|.++..+|++++|+++..... +.+.++..+
T Consensus   30    ~~~~n~~~~~~~~~l~~~l~~~a~~~~i~~~~~~~~~~~~~~~L~~~L~~~~~rgv~v~ii~~~~~~~~~~~~~~~l   102 (953)
T 7CLG_B      30    KLLVNNK-NEKVLSTIIDELQ--~KCEFFYFSVAFITE---SGLASLKAQLLDLSNKGVKGKILTSNYLGF-NSPKMYGEL   102 (953)
T ss_dssp           CCEECTT-TSCHHHHHHHHHH--TCSEEEEEESCBCH---HHHHHTHHHHHHHHHHTCCEEEEEECGGGC-SCHHHHHHH
T ss_pred           eeeeCCC-CccHHHHHHHHHH--hCCEEEEEEccccc---chHHHHHHHHHHHHHHcCCCeEEEEecCCCCC-CCHHHHHHH

Q ss_pred            HHcCCCcEEEEEcC-CCccccEEEEeCC---EEEEcCCCcCCchhhhcceEEEEEc---chhHhhcCchHHHHHHHHHH
Q Q_7656688   78    LNHQKDNVFTRFVP~DRFHPKVIWWKGY---~GTYIGSANHTDNGWLTNIEAGVFLT---EDELLANGMDSQLEDFFEYLR   150 (450)
Q Consensus   78    ~~~~~~~i~i~~~~~~H~K~~i~d~~~~~~~iGSaN~T~~gl~~N~E~~v~~~~~~~~~~~~~~~~~~~~f~~l~   150 (450)
                      .+..  ++.++.+. ..+|+|+|+|++.   .+++||+|+|.+|+|..|.|.++.+.   +.+     ....+...|+.+|
T Consensus  103    ~~~~~~i~v~~~~~~~H~K~~i~d~~~~~~~iGS~Nlt~~al~~n~E~~v~~~~~~~~~~~~~~~~~~~f~~~w   175 (953)
T 7CLG_B     103    LKLK--NVEVRLTDIAGFHAKGYIFEHKDYSSMVIGSSNLTSNALKVNYEHNVLLSTMKNGD-----LVDSVKSEFDLLW   175 (953)
T ss_dssp          HTCT--TEEEEEECCTTBCCEEEEEECSSEEEEEEESCCBSHHHHTSSBCCEEEEEEETTCH-----HHHHHHHHHHHH
T ss_pred          HHcC--CcEEEEecCCCCcceEEEEEcCCceEEEEeCCcCchhhhccCceEEEEEeecCCCh-----HHHHHHHHHHHH

Q ss_pred            hcCCCCCCCHHHHHHHHHHHH
Q Q_7656688   151   ELDVTIPISADYIAEMERLNAL   172 (450)
Q Consensus   151   ~~~~~~~~~~~l~~~~~~~~~   172 (450)
                      +.+.  +++.+.++.|...+..
T Consensus   176   ~~s~~~~l~~~~i~~y~~~~~~   195 (953)
T 7CLG_B      176   QKST--PLTEQWINSVKESFEY   195 (953)
T ss_dssp          HTSE--ECCHHHHHHHHHC---
T ss_pred          hhcC--cCCHHHHHHHHHHch
```

PF21790.2 ; OGG ; Putative 8-oxoguanine DNA glycosylase OGG-like protein

Probability: 95.54%, E-value: 0.054, Score: 50.23, Aligned cols: 109, Identities: 7%, Similarity: -0.067, Template Neff: 9.4

```
Q ss_pred            cCHHHHHhCCHHHHHHHHHHHHHHHHHHHHhCCHHHhCCcccCCCCHHHHHHHHHHHHHHHcccCCCCCCHHHHHHHHHhCCC
Q Q_7656688_1   311   LAQDKVQELTVDEFATLCRNTHATMDHIIKVPLAELGRPELKTLDREARVGLFAPLILRQRNRKGWDVRQLLHYVLYGGP   390 (450)
Q Consensus     311   l~~~~~~~l~~~~~~~~~~~~a~~~~~~~~~~~~~~g~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~i~~~l~~Ll~~~~   390 (450)
                        +++..+...+.++....+~.+..| +......+                .....+.+  ....~.+.+++++..+..+
T Consensus      39    ~i~R~~l~~~~~~~~~l~~v~~W~g~g~~~~~~~~~~~~~~~~~~~~~~~~r~~~~l~~~~~~~~~l~~a~~~~~~~~   97 (195)
T PF21790.2      39    IARSFIHARSNGDPVALFLMAMVW~GYRPKDYG----------------PHRTKAVL~~AQEGAADNIATIVDATRTE~~   97 (195)
T ss_pred            eeHHHHHhcCCCCHHHHHHHHHHHh~cCCCCCC~~~~~~~~~~~~~~~~~~HHHHHHHH~~hCCChHHHHHHHHHHHHHh~~

Q ss_pred            chHHHHHHHHHHhCcCcccCCCchHHHHHHHHccC-----cccCCCCCCchHHHHHHcC
Q Q_7656688_1   391   NDEIWQRLYHAGRDSDYSIARYGLNSLAEVVGWAR-----~PELVPPRNGRTSKALRALG   444 (450)
Q Consensus     391   ~~~~~~R~~~~~~~~~~~l~g~g~~~~~ell~~~~~~~~~Pd~~~~~N~r~~k~L~~LG   444 (450)
                        +..+.+..+...  .+|+|+|++++|++|.++.      |+.|++++.++.++|..++
T Consensus      98    ~~~~~a~~~l~~~~~~i~g~lG~a~~Sk~L~f~~~~~~~~p~~~~Ild~~v~~~l~~~~   152 (195)
T PF21790.2      98    ~~GAAAGWRALLRT~~HKIKGLNMSFGTKLLYFAGYTTSHRPRPLILDERVRSAIQNVS   152 (195)
T ss_pred            ~~hHHHHHHHHhcc~~CCCCCcccchHHHHHHHHhCCCCCCCCCchHHHHHHHHHHcC
```

4UOB_A ENDONUCLEASE III-3; LYASE, ENDONUCLEASE III, FES CLUSTER, BASE EXCISION REPAIR, DNA GLYCOSYLASE, DEINOCOCCUS RADIODURANS; HET: ACT, SO4; 1.31A {DEINOCOCCUS RADIODURANS}

Probability: 78.44%, E-value: 85, Score: 30.2, Aligned cols: 147, Identities: 11%, Similarity: 0.017, Template Neff: 10.2

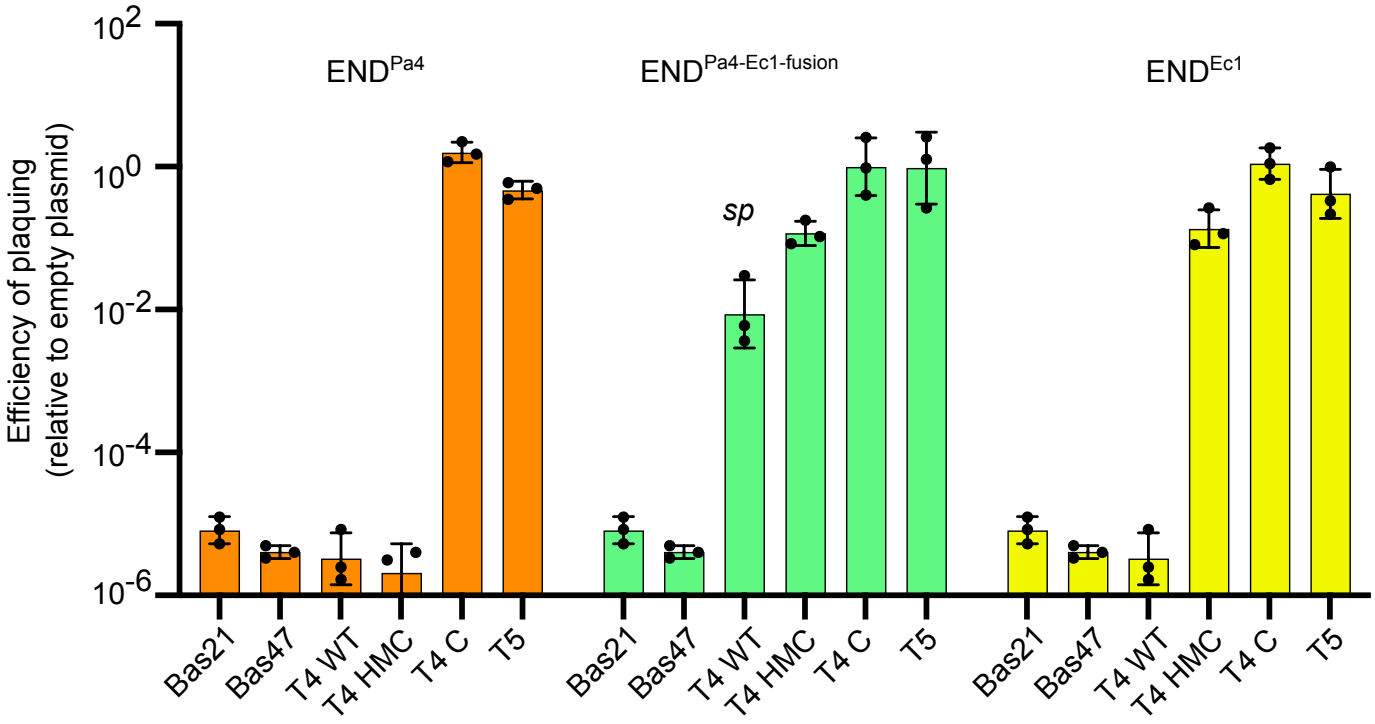
```
Q ss_pred            HHHHHHHHHHHHHHHhhCCCCcCCCCCCHHHHHHHHHHHHHHHHcccccCCCchHHHHHHHCCCHHHHHHHHHHHHCCCC
Q Q_7656688   213   LSTLGILQSIQQQLADYRPNWVGAEVPAAWQVDQFLHAYYYNHVGDSLHKPYEDYHRLNQGNPQVALETELKWWKSLSEA   292 (450)
Q Consensus   213   ~~~~~~~~~~~~~~~~~~~~~~~p~~~~~d~f~~~~~~~v~~~~~~~~~~~~~~~~~~~~~~~~~~l~~~~~~~~~~~~~   292 (450)
                      ...~..+..+.+.+....+~|.....~|....+..++.    ..+....
T Consensus      38    ~~~~~~~~~l~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~   81 (278)
T 4UOB_A         38    VWFGEYLSRLRDTYAPELPPPRQFPDPLGGLIRTILS----QQNTRRV-------------------------------   81 (278)
T ss_dssp            TTHHHHHHHHHHHHCSSCCCCCCCSSHHHHHHHHHHH----TTSCHHH-------------------------------
T ss_pred            hHHHHHHHHHHHHHCCCCCCCCCCCHHHHHHHHHHh----ccCCcHHH-------------------------------

Q ss_pred            CCHHHHHHHhHHHHHHHHHh~CHHHHhCCHHHHHHHHHHHH~~HHhHHHHHhCCHHHhCCcccCCCCHHHHHHHHHHHHH
Q Q_7656688   293   PSREDYALYESAPTVRALL~AQDKVQELTYDEFATLCRNT~~HATMDHIIKVPLAELGRPELKTLDREARVGLFAPLILR   369 (450)
Q Consensus   293   ~~~~~~~~~~~~~~~~~~l~~~~~~~~~~~~~~~~~~~~~a~~~~~~~~~~~~~~~~~~g~~~~~~~~~~~~~~~~~~~~   369 (450)
                            ....+..+...+ +++.+...+.+++..++...  ......+...
T Consensus      82    ~~~~~~~~~~~~~~~~~~l~~~~~~~l~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~   126 (278)
T 4UOB_A         82    --------AQRQWEVLTATYPQWEAALLDGPDGIEATLKSAGGGLSRMKADYI----------------------------   126 (278)
T ss_dssp            -------HHHHHHHHHHCSSHHHHHHHHHHHHHHHHHHHTSCSCHHHHHHHH----------------------------
T ss_pred            -------HHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHHHcCCCcHHHHHHHH----------------------------

Q ss_pred            hhccCCCCHHHHHHHHHHhCCCchHHH----------HHHHHHHHcCccccccCCccHHHHHHHHHHHHHHHhCcccCCCCCh
Q Q_7656688   370   QRNRKGWDVRQLLHYVLYGGPNDEIW----------QRLYHAGRDSDYSIARYGLNSLAEVVGWARPELVPPRNGRR   435 (450)
Q Consensus   370   ~~~~~~~~~~~l~~~~~~~~~~~~~~----------~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~   435 (450)
                              +...+.+...  .+.+.              +.+.+.+     ..++|+|.++..++.+.....++.+..
T Consensus     127   ~~~~~~~~~~a~~~~~~~~~g~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~   191 (278)
T 4UOB_A        127   ----------YGILAHLQEH--~HGGLSLRLRLREFPHTPEGHEQARQAL----AALPGVGNKTVALVLLFDLRRFAMPVDGN   191 (278)
T ss_dssp            ----------HHHHHHHHHH--~SSSSCCGGGGGSCSSHHHHHHHHHHH----HTSTTCCHHHHHHHHHHCSCCCCCCHHH
T ss_pred            ----------HHHHHHHHHH--~hCCCCHHHHhCCCCCCHHHHHHHHH----HCCCCCCHHHHHHHHHHcCCCCCCCHHH

Q ss_pred            hHHHHHHcCC
Q Q_7656688   436   TSKALRALGF   445 (450)
Q Consensus   436   ~~~~L~~LG~   445 (450)
                      ..+.+..+|.
T Consensus     192   ~~r~l~rlg~   201 (278)
T 4UOB_A        192   MERAAKRLEL   201 (278)
T ss_dssp            HHHHHHHTTS
T ss_pred            HHHHHHHhCC
```

Extended Data Figure 4: HHpred domain analysis of END^{PaCF1}. a, Alignment output from HHpred. Only the top 3 alignment of each domain is shown. **b**, Details of alignment of END^{PaCF1} with characterized proteins. First alignment is for END^{PaCF1} NTD and PLD nuclease domain from ATP-dependent restriction endonuclease SauUSI (PDB: 7CLG_A). Second alignment is for END^{PaCF1} CTD and 8-oxo-guanine DNA glycosylase OGG-like (PF21790). Third alignment is for END^{PaCF1} CTD and endonuclease III (PDB: 4UOB). For the third alignment, the catalytic residues of 4UOB is shown boxed in red.



Extended Data Figure 5: Histogram of plaque assays in *E. coli* DH10B with END^{Pa4}, END^{Ec1} and END$^{Pa4\text{-}Ec1\text{-}fusion}$. Panel of *E. coli* phages were tested against different END nuclease constructs. T5 is included as a negative control. HMC = hydroxymethylcytosine, C= unmodified cytosine. sp stands for small plaques (see Figure 4e).

a

b

3FHG_A N-glycosylase/DNA lyase; ogg, helix-hairpin-helix, glycosylase, 8-oxoguanine, 8-oxoG, SsOGG, DNA damage, DNA repair, Glycosidase, Hydrolase, Lyase, Multifunctional enzyme, Nuclease; HET: SO4, GOL; 1.9A {Sulfolobus solfataricus}

Probability: 96.86%, E-value: 0.14, Score: 44.07, Aligned cols: 152, Identities: 13%, Similarity: 0.095, Template Neff: 10.9

```
Q ss_pred           HHHHHHHHHHHHHhcHHHHHhCCHHHHHHHHHHHHHhcCCCCC--CcchhhHHHHHhcHHHHHHHHHHHHcCC--CCHHHH
Q Q_1190304    26   DQRKSDHFLWDQWVNQETVQKLSDEELKQRFLEYFHAGAGRH--PFNHIYRDRIIRDPKKFRDTLTFLLDET--ISLKER   101 (362)
Q Consensus    26   ~~~~~~~~~~~~l~~~~~~lt~~e~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~r~~l~~l~~~~~~~~~r   101 (362)
                    .........+...+      ...+.+++..++..         .  .+.+.....+......+...+..+....  .+....
T Consensus    49   ~~~~~~~~l~~~~~~~~~~~~~l~~~~~~~~~~~~~~~~~~~~~~~~ka~~i~~~~~~~~~l~~~~~~~~~~~~~   115 (207)
T 3FHG_A       49   ISAYQALNCLGQKI-----YVANEEEIRNILKS--------CKYRFYNLKAKYIIMAREKVYGRLKEEIKPLADEDQQLA   115 (207)
T ss_dssp           HHHHHHHHHHGGGG-----GTCCHHHHHHHHHH--------TTCTTHHHHHHHHHHHHHHHTTTHHHHHHHHHHHCHHHH
T ss_pred           HHHHHHHHHHhhhh-----hcCCHHHHHHHHHh--------CCCCcHHHHHHHHHHHHHHHhhhHHHHHhhhcCCCHHHH

Q ss_pred           HHHHHcccCCcccCCCCHHHHHHHHHHHCHHHcccccHhHHHHHHHHcCCCCCccCCCCHHHHHHHHHHHHHHHHcCCCCC
Q Q_1190304   102   INEVLNKNGKHHIEGLGRGLTTSLLMDLNPQQYVTWNNKTDKGLETLGCQPSFERGDDWGTKYEKILEAIREIQSLNPQS   181 (362)
Q Consensus   102   i~~~~~~~~~~i~G~g~~~~S~lL~~~~P~~y~i~n~~~~~l~~lg~~~~~~~~~~~~~~Y~~~~~~~~l~~~~l~~   181 (362)
                    ...+.      +++|+|...++.+|....+....+.|..+..++..+|..............|..+.+.+..+.. ..+.
T Consensus   116   ~~~l~~~~~~~~Gig~~~a~~~l~~~~~~~~p~D~~v~r~l~~~g~~~~~~~~~~~~~~~~~~l~~~~~~~~~   188 (207)
T 3FHG_A      116   RERLL------NIKGIGHQEASHFLRNVGYFDLAIIDRHIIDFMRRIGAIGETNVKQLSKSLYISFENILKSIAS-NLNM   188 (207)
T ss_dssp           HHHHT------TSTTCCHHHHHHHHHHHTTCCSSCCCHHHHHHHHHTTSSCCCCCSCCCHHHHHHHHHHHHHHHH-HTTS
T ss_pred           HHHHH------cCCCCCHHHHHHHHHHcCCccceecHHHHHHHHHHcCCCCCCCCccCCHHHHHHHHHHHHHHHH-HcCC

Q ss_pred           CHHHHHHHHHHHHcCC
Q Q_1190304   182   NFLEIDHFLHIVAVEP   197 (362)
Q Consensus   182   ~~~~~d~fl~~~~~~~   197 (362)
                    +..++|..+|.+....
T Consensus   189   ~~~~ld~~lw~~~~~~   204 (207)
T 3FHG_A      189   SVGILDLFIWYKETNT   204 (207)
T ss_dssp           CHHHHHHHHHHHHSS
T ss_pred           CHHHHHHHHHHHHcCc
```

3FHF_A N-glycosylase/DNA lyase; ogg, helix-hairpin-helix, glycosylase, 8-oxoguanine, 8-oxoG, MjOgg, DNA damage, DNA repair, Glycosidase, Hydrolase, Lyase, Multifunctional enzyme, Nuclease; HET: MSE; 1.995A {Methanocaldococcus jannaschii}

Probability: 95.87%, E-value: 1, Score: 39.22, Aligned cols: 148, Identities: 14%, Similarity: 0.151, Template Neff: 10.2

```
Q ss_pred           HHHHHHHHHHHHHHcCHHHHHhCCHHHHHHHHHHHHHhcCCCCC--CcchhhHHHHHhcHHHHHHHHHHHHcCCCCHHHHH
Q Q_1190304    25   RDQRKSDHFLWDQWVNQETVQKLSDEELKQRFLEYFHAGAGRH--PFNHIYRDRIIRDPKKFRDTLTFLLDETISLKERI   102 (362)
Q Consensus    25   ~~~~~~~~~~~~~~l~~~~~~lt~~e~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~r~~l~~l~~~~~~~~ri   102 (362)
                    ..........+...+     ...+.+++..++..         .  .+.+.......+..... +. .+...+....+.....
T Consensus    59   ~~~~~~~~~~l~~~~~~~~~~~~~~l~~~~~~~~~~~~~~~~~g~~~ka~~l~~~~~~~~~~~~~~~~~~~~~~~~   123 (214)
T 3FHF_A       59   AEGGIRIQKEIGDGF-----LTLPREELEEKLKN--------LGHRFYRKRAEYIVLARR-FK-NIKDIVESFENEKVAR   123 (214)
T ss_dssp           HHHHHHHHHHHTTHH-----HHSCHHHHHHHHHH--------TTCTTHHHHHHHHHHHHGG-GC-CHHHHHHHSSSHHHHH
T ss_pred           HHHHHHHHHHHHccH-----hcCCHHHHHHHHHH--------cCCCCchHHHHHHHHHHHH-Hh-cHHHHHhCCCCHHHHH

Q ss_pred           HHHHcccCCCc-cCCCCHHHHHHHHHHCH-HHccccchHHHHHHHHcCCCCCCccCCCCHHHHHHHHHHHHHHHHcCCCC
Q Q_1190304   103   NEVLNKNGKHH-IEGLGRGLTTSLLMDLNP-QQYVTWNNKTDKGLETLGCQPSFERGDDWGTKYEKILEAIREIQSLNPQ   180 (362)
Q Consensus   103   ~~~~~~~~~~i~G~g~~~~S~lL~~~~P~~y~i~n~~~~~l~~lg~~~~~~~~~~~~~~Y~~~~~~~~l~~~~l~~   180 (362)
                    ..+.      . ++|+|..+++.+| .... ....+.|..+..++..+|.....+.... ...|..+.+..... ..+
T Consensus   124   ~~l~~~~~~~~GiG~~~a~~~l~~~~~~~~~D~~v~r~~~~~g~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~   194 (214)
T 3FHF_A      124   EFLV------RNIKGIGYKEASHFL-RNVGYDDVAIIDRHILRELYENNYIDEIPKTLS-RRKYLEIENILRDIGE-EVN   194 (214)
T ss_dssp           HHHH------HHSTTCCHHHHHHHH-HHTTCCSCCCCCHHHHHHHHHTTSSSSCCSSCC-HHHHHHHHHHHHHHHH-HTT
T ss_pred           HHHH------HhCCCCCHHHHHHHH-HhcCcccccHhHHHHHHHHHHcCCCCCCCCCCC-HHHHHHHHHHHHHHHH-HHC

Q ss_pred           CCHHHHHHHHHHHcC
Q Q_1190304   181   SNFLEIDHFLHIVAVE   196 (362)
Q Consensus   181   ~~~~~d~fl~~~~~~~   196 (362)
                    +....+|..+|.....
T Consensus   195   ~~~~ld~~~W~~~~~   210 (214)
T 3FHF_A      195   LKLSELDLYIWYLRTG   210 (214)
T ss_dssp           CCHHHHHHHHHHHHS
T ss_pred           CCHHHHHhHHHHHHcC
```

3N0U_C Probable N-glycosylase/DNA lyase; structural genomics, ISFI, DNA REPAIR, 8-OXOGUANINE, base excision repair, PSI-2, Protein Structure Initiative, Integrated Center for; 1.5A {Thermotoga maritima}

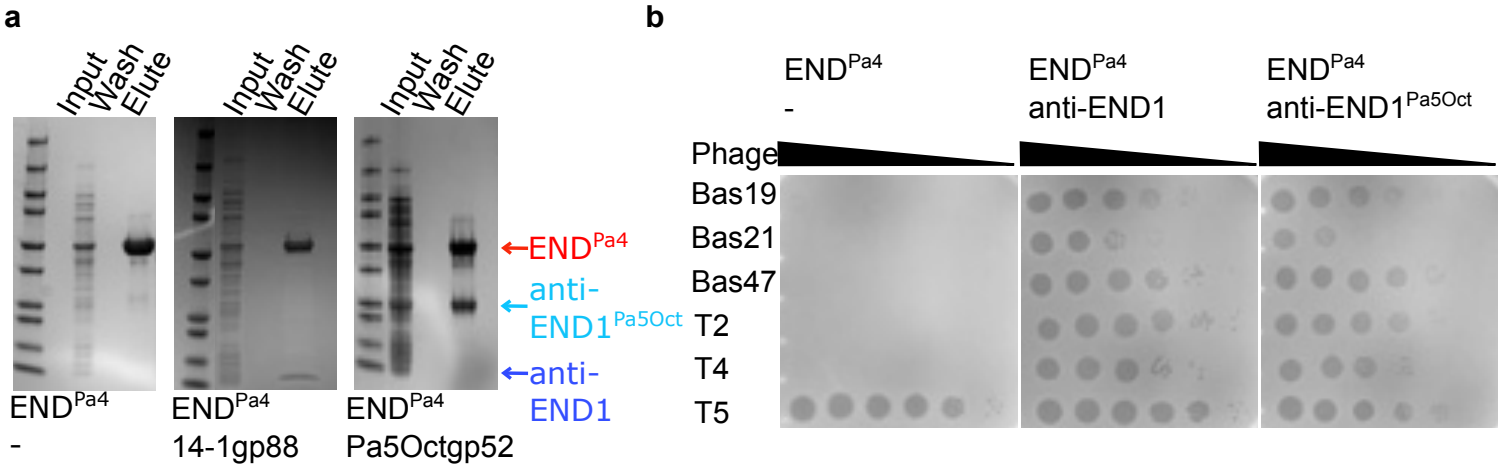
Probability: 95.24%, E-value: 2.4, Score: 37.07, Aligned cols: 151, Identities: 12%, Similarity: 0.108, Template Neff: 10

```
Q ss_pred           HHHHHHHHHHHHHcCHHHHHhCCHHHHHHHHHHHHHhcCCCCC--CcchhhHHHHHhcHHHHHHHHHHHHcCCCCHHHHH
Q Q_1190304    26   DQRKSDHFLWDQWVNQETVQKLSDEELKQRFLEYFHAGAGRH--PFNHIYRDRIIRDPKKFRDTLTFLLDETISLKERIN   103 (362)
Q Consensus    26   ~~~~~~~~~~~~~l~~~~~~lt~~e~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~r~~l~~l~~~~~~~~ri~   103 (362)
                    .........+.+.+      ...+.+++..++..         .  .+.+.....+.....+. .+..+...  +..+...
T Consensus    66   ~~a~~~~~~l~~~~~~~~~~~~~~~l~~~~~~~~~~~~~~~~~~ka~~i~~~a~~~~~~~~~~~~~~~~~~~~~~   129 (219)
T 3N0U_C       66   EGGIRAQKEIGKGF-----VHLPLEELAEKLRE--------VGHRYPQKRAEFIVENRKLLG-KLKNLVKG--DPFQSRE   129 (219)
T ss_dssp           HHHHHHHHHHTTHH-----HHSCHHHHHHHHHH--------TTCSSHHHHHHHHHHHGGGTT-THHHHHHS--CHHHHHH
T ss_pred           HHHHHHHHHHHccc-----cCCCHHHHHHHHHH--------cCCCCcHHHHHHHHHHHHHHh-HHHHHHhC--CHHHHHH

Q ss_pred           HHHcccCCcc-cCCCCHHHHHHHHHHHCHHHccccchHHHHHHHHcCCCCCCccCCCCHHHHHHHHHHHHHHHHcCCCCC
Q Q_1190304   104   EVLNKNGKHH-IEGLGRGLTTSLLMDLNPQQYVTWNNKTDKGLETLGCQPSFERGDDWGTKYEKILEAIREIQSLNPQSN   182 (362)
Q Consensus   104   ~~~~~~~~~~i~G~g~~~~S~lL~~~~P~~y~i~n~~~~~l~~lg~~~~~~~~~~~~~~Y~~~~~~~~l~~~~l~~~   182 (362)
                    .+.      . ++|+|..+++.+|....+....+.|..+..++..+|.....+ .......|..+...+.. .++.+
T Consensus   130   ~L~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~   201 (219)
T 3N0U_C      130   FLV------RNAKGIGWKEASHFLRNTGVEDLAILDKHVLRLMKRHGLIQEIP-KGWSKKRYLYVEEILRKVAE-AFGES   201 (219)
T ss_dssp           HHH------HHSTTCCHHHHHHHHTTTCCSCCCCCHHHHHHHHHTTSCSSCC-SSCCHHHHHHHHHHHHHHHH-HHTCC
T ss_pred           HHH------hhCCCCCHHHHHHHHHhCCcccceecHHHHHHHHHHcCCCCCCC-CCCCHHHHHHHHHHHHHHHH-HHCCC

Q ss_pred           HHHHHHHHHHHcCCCCc
Q Q_1190304   183   FLEIDHFLHIVAVEPGQ   200 (362)
Q Consensus   183   ~~~~d~fl~~~~~~~~~   200 (362)
                    ...+|..+|........+
T Consensus   202   ~~~~~~~W~~~~~~~~K   219 (219)
T 3N0U_C      202   PGKFDLYLWYLVKGKVDK   219 (219)
T ss_dssp           HHHHHHHHHHHHSCCCC
T ss_pred           HHHHHHHHHHHHHCCcC
```

Extended Data Figure 6: HHpred domain analysis of END^{Mag}. a, Alignment output from HHpred. Only the top 3 alignment of each domain is shown. **b**, Details of alignment of END^{Mag} sensor domains with characterized proteins. First alignment is for END^{Mag} with *Sulfolobus solfataricus* Ogg1 (PDB: 3FHG). Second alignment is for END^{Mag} with *Methanocaldococcus jannashii* Ogg1 (PDB: 3FH3). Third alignment is for END^{Mag} with *Thermotoga maritima* Ogg1



Extended Data Figure 7: END^{Pa4}, a distant homolog of END^{PaCF1}, also binds to anti-END inhibitors. **a,** Pulldown of inhibitors in the presence of defense system. Only END^{Pa4} is His-tagged; controls are shown in Figure 5. **b,** Plaque assay of different hypermodified phages in *E. coli*, in the presence of END^{Pa4} with/without anti-END inhibitors.

Notation	Coordinates	strand	DefenseFinder	HHpred
gene 1	6354280..6356735	-	Type IV RM (note no nuclease)	
gene 2	6356736..6358225	-	Mokosh	
gene 3	6358210..6361595	-	Mokosh (note no nuclease)	
gene 4	6361602..6363005	-		S/T kinase
gene 5	6363011..6363703	-		Protein serine/threonine phosphatase
gene 6	6363703..6364374	-		Von Willebrand factor type A (vWA) domain
gene 7	6365945..6367771	-	OLD/Gabija	
gene 8	6376092..6377573	-		RtcR, Regulator of RNA terminal phosphate cyclase
gene 9	6376092..6377573	-		RtcR, Regulator of RNA terminal phosphate cyclase
gene 10	6379523..6380257	-		N-terminal glycosyltransferase
gene 11	6380320..6381444	-		ComF-like phosphoribosyltransferase
gene 12	6381447..6382313	-		Histidinol phosphatase (PHP family)

Extended Data Table 1: Putative roles of defene-associated genes and other genes indicated in Figure 1b.

Notation	Protein Accession (if applicable)	Species	Strain	Source	Country source	Collection date
END^{PAK}		*P. aeruginosa*	PAK	Lab isolate		
END^{Pa3}	ELJ2661932.1	*P. aeruginosa*	2020QG-00001	Human, wound	USA	2020
END^{Pa4}	MBX6756561.1	*P. aeruginosa*	HBRM35	Human, sputum	China	2017
END^{Ec1}	HCP8353069.1	*E. coli*	07-06701	Human, stool	Germany	2007
END^{Ec2}	EFH3554530	*E. coli*	PSU-2108	Pig	USA	2000
END^{Se}	EHN1424678.1	*S. enterica*	FDA1160512-C001-001	Cheese	USA	2021
END^{Vc}	WP_267838393.1	*V. cholerae*	N/A	N/A	N/A	N/A
END^{Mag}	MDV2503288.1	Marine sediment metagenome	N/A	Pelagic sediments	Atlantic Ocean	2014

Extended Data Table 2: END nucleases tested experimentally in the paper.

Restriction enzyme	Recognition sequence
BsgI	GTGCAG(16/14)
EcoRI	G/AATTC
FspI	TGC/GCA
HindIII	A/AGCTT
BmrI	ACTGGG(5/4)
ApaLI	G/TGCAC
NdeI	CA/TATG
XbaI	T/CTAGA
DraI	CACNNN/GTG
BsaI	GGTCTC(1/5)

Extended Data Table 3: List of restriction enzymes and the corresponding recognition sequence as indicated on NEB website.

Primers	Sequence
WX_16	aagcttggcactggcc
WX_17	ggtatgtatatctccttcttaaagttaaac
WX_128	cgttgtaaaacgacggccagtgccaagctttcatttgacgtacttcccgaagtagc
WX_129	gctcgaattctgtttcctgtgtgaa
WX_130	tgcaggaattcctcgagaagcttgg
WX_133	gtttaactttaagaaggagatatacataccatgactagctccaaatggaccatcg
WX_167	acaatttcacacaggaaacagaattcgagcggttgcctgaaacggggc
WX_168	aggtaccaagcttctcgaggaattcctgcatcagtagaccttcacatcgaacccc
WX_216	aacagaaagtaatcgtattgtacacg
WX_217	ctaggttaattaagctgcgctag
WX_257	ttttgtttaactttaagaaggagatatacc
WX_258	tctgttcgacttaagcattatgc
WX_Q4F	gccaatagcgttcctgtgag
WX_Q4R	ggaagttatcgaggcgaaagc
WX_rpoD_F	gggcgaagaaggaaatggtc
WX_rpoD_R	caggtggcgtaggtggagaa

Supplementary Data Table 1: List of primers used.

Primers	Sequence
WX_16	aagcttggcactggcc
WX_17	ggtatgtatatctccttcttaaagttaaac
WX_128	cgttgtaaaacgacggccagtgccaagctttcatttgacgtacttcccgaagtagc
WX_129	gctcgaattctgtttcctgtgtgaa
WX_130	tgcaggaattcctcgagaagcttgg
WX_133	gtttaactttaagaaggagatatacataccatgactagctccaaatggaccatcg
WX_167	acaatttcacacaggaaacagaattcgagcggttgcctgaaacggggc
WX_168	aggtaccaagcttctcgaggaattcctgcatcagtagaccttcacatcgaacccc
WX_216	aacagaaagtaatcgtattgtacacg
WX_217	ctaggttaattaagctgcgctag
WX_257	ttttgtttaactttaagaaggagatatacc
WX_258	tctgttcgacttaagcattatgc
WX_Q4F	gccaatagcgttcctgtgag
WX_Q4R	ggaagttatcgaggcgaaagc
WX_rpoD_F	gggcgaagaaggaaatggtc
WX_rpoD_R	caggtggcgtaggtggagaa

Supplementary Data Table 1: List of primers used.