

Published in final edited form as:

Nature. 2021 January 01; 589(7840): 120–124. doi:10.1038/s41586-020-2762-2.

Prokaryotic viperins produce diverse antiviral molecules

Aude Bernheim¹, Adi Millman¹, Gal Ofir¹, Gilad Meitav¹, Carmel Avraham¹, Helena Shomar², Masha M. Rosenberg², Nir Tal², Sarah Melamed¹, Gil Amitai¹, Rotem Sorek[#]

¹Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel

²Pantheon Biosciences, Yavne, Israel

Abstract

Viperin is an interferon-induced cellular protein conserved in animals. It was shown to inhibit the replication of multiple viruses by producing a ribonucleotide called 3'-deoxy-3',4'-didehydro-CTP (ddhCTP), which acts as a chain terminator for the viral RNA polymerase. Here we show that the eukaryotic viperin has originated from a clade of bacterial and archaeal proteins that protect against phage infection. Prokaryotic viperins (pVips) produce a set of modified ribonucleotides that include ddhCTP, as well as ddhGTP and ddhUTP. We further provide evidence that pVips protect against T7 phage infection by inhibiting viral polymerase-dependent transcription, implying an anti-viral mechanism of action similar to the animal viperin. Our results unveil a potential repository of natural antiviral compounds produced by bacterial immune systems.

Viperin is an antiviral protein that becomes highly expressed in cells stimulated by interferons¹. In humans, this protein has broad antiviral activity against DNA and RNA viruses that include the human cytomegalovirus, West Nile virus, dengue, hepatitis C, and HIV^{1,2}. It was recently shown that viperin is an enzyme that catalyzes the conversion of CTP to 3'-deoxy-3',4'-didehydro-CTP (ddhCTP)³. This modified nucleotide lacks a hydroxyl group at the 3' carbon of the ribose and hence, when the viral polymerase incorporates it into the nascent chain of the viral RNA, it acts as a chain terminator that does not allow further polymerization of the RNA chain³. In accordance, it was shown that ddhCTP directly inhibits the RNA-dependent replication of RNA viruses such as Zika *in vivo*.

It was previously noted that some bacteria and archaea encode genes that have significant sequence similarity to vertebrate viperins, although their role was unknown^{13–15}. We set out

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

[#]Correspondence and requests for materials should be addressed to R.S. rotem.sorek@weizmann.ac.il.

Authors contribution

A.B. and R.S. led the study and A.B. performed all experiments unless otherwise indicated. A.M. and A.B. performed the computational analyses that appear in Figs. 1 and 2 and Extended Data Fig 9. H.S., M.R. and N.T. designed and performed purification of pVips and *in vitro* enzymatic assays that appear in Fig. 3 and Extended Data Fig 6. G.M., C.A., and S.M. assisted with the plaque assays that appear in Fig. 1 and Extended Data Fig. 1. C.A. assisted in the preparation of cell lysates that appear in Fig. 3 and Extended Data Figs 3, 4 and 5. G.O. and G.A. assisted in the design and analysis of GFP reporting studies presented in Fig 4 and Extended Data Fig 7. R.S. supervised the study. R.S. and A.B. wrote the paper together with the team.

Competing interests

R.S. is a scientific cofounder and advisor of BiomX, Pantheon Bioscience and Ecophage. A.B., A.M., and R.S. are inventors on patent application PCT/IL2020/050377 licensed to Pantheon Bioscience. H.S., M.R. and N.T. are employed by Pantheon Bioscience.

to examine whether prokaryotic homologs of the human viperin participate in defense against phages. To this end, we first performed a profile-based search for viperin homologs in a database of >38,000 bacterial and archaeal genomes. This search yielded 1,724 genes (1,112 non-redundant sequences) homologous to the human viperin, which were aggregated into 17 clusters based on sequence similarity (Methods) (Extended Data Table 1).

Viperin is a member of the radical S-adenosyl-methionine (SAM) family of enzymes, and shows both sequence and structural homology to other members of that family, particularly the housekeeping molybdenum cofactor biosynthesis enzyme MoaA¹³. To differentiate between viperin homologs with housekeeping properties and homologs that may participate in defense against phages, we took advantage of the fact that in prokaryotes, genes involved in antiviral activity tend to co-localize next to one another on the genome, forming “defense islands”^{12,16}. Most clusters of viperin homologs did not show a tendency to co-localize with defense genes (Extended Data Table 1). However, in one of the clusters, 60% of the genes were found in the vicinity of CRISPR-Cas systems, restriction-modification systems (RM), and other bacterial defense genes (Figure 1a). Such a high propensity for co-localization with defense systems is a strong predictor that the genes in the cluster play a role in phage resistance^{7,12}. We denoted the genes in the defensive cluster pVips (for prokaryotic viperin homologs). As pVips are relatively rare in prokaryotic genomes (164 genes in the cluster), we further performed an online homology search in additional genomes that were not included in our original database, retrieving 86 additional such genes and resulting in a total of 250 pVips (Extended Data Table 2).

To check whether pVips can defend against phages, we selected 59 genes that span the space of the pVip sequence diversity (Extended Data Table 2) and cloned them in *E. coli* under the control of an inducible promoter. GFP, as well as the MoaA gene from *E. coli*, were similarly cloned as negative controls. We then challenged the pVip-expressing bacteria with an array of phages that span several major phage families (*Myoviridae*: P1; *Siphoviridae*: Lambda-vir, SECphi6, SECphi18, SECphi27; *Podoviridae*: T7; *Leviviridae*: MS2, Qbeta) (Figure 1b-c, Extended Data Figure 1).

About half of the tested pVips conferred clearly identifiable defense against phages. Most of these protected against T7 as evidenced by plaque assays (up to ten fold reduction in T7 plaque sizes; Figure 1b, Extended Data Figure 1a), and by a delay or absence of culture collapse in T7 infection assays in liquid culture (Figure 1c; Extended Data Figure 2a). pVips mutated in cysteine residues in the CxxxCxxC motif predicted to coordinate the iron-sulfur cluster lost the defensive capacity against T7, suggesting that the catalytic activity of pVips is necessary for defense (Extended Data Figure 2b). A subset of the pVips also protected against phages P1, lambda, SECphi6, and SECphi18, reducing the observed number of plaques by between 10 to 10,000 fold (Extended Data Figure 1). Remarkably, when the human viperin gene was cloned and expressed in *E. coli* under the same conditions, it protected against T7 in a manner similar to that observed for many pVips (Figure 1, Extended Data Figures 1-2).

The pVips we found are present in phylogenetically very distant organisms, suggesting an ancient evolutionary origin, rampant horizontal gene transfer, or both. We found pVips in

176 species overall, belonging to 14 bacterial and archaeal phyla that include Proteobacteria, Firmicutes, Cyanobacteria, Actinobacteria, Bacteroidetes, Euryarchaeota, and others (Extended Data Table 2). To better understand their diversity and phylogenetic distribution, we generated a phylogenetic tree of the viperin family, including pVips, eukaryotic viperins, and MoaA genes from bacteria and eukaryotes as an outgroup (Figure 2, Extended Data Table 3). We found that pVips are grouped into seven major clades that partially follow the phyletic grouping of the encoding microbes (Figure 2). Remarkably, all eukaryotic viperins form a monophyletic clade within the tree, with the closest common ancestor predicted to localize to pVip clade 2, which is mostly composed of pVips from archaeal species. The clear monophyletic organization of the eukaryotic viperin clade and its position within the pVips tree suggest that a single event in the ancient history of the eukaryotic lineage resulted in the acquisition of viperin from prokaryotes.

In vertebrate genomes, the viperin gene is frequently encoded next to a cytidylate kinase gene that is co-expressed with the viperin during the interferon response^{3,18}. This kinase phosphorylates CMP to CTP, thus generating the substrate for viperin activity³. We found that 47 of the 250 pVips (19%) were encoded next to a gene annotated as a nucleotide kinase in their genome of origin (Figure 1a, Figure 2), and that in some cases the kinase was fused to the pVip gene (Figure 2, Extended Data Table 2). This further strengthens the hypothesis that the pVip substrate is a tri-phosphorylated nucleotide. While some pVip-associated kinases were annotated as cytidylate kinases, as in vertebrates, others were annotated as thymidylate or other kinases¹⁴, suggesting that the substrates of some pVips might be tri-phosphorylated nucleotides other than CTP.

The animal viperin catalyzes the production of ddhCTP³. We therefore sought to examine whether pVips produce ddhCTP and/or other types of modified nucleotides. For this, we expressed pVips in *E. coli* and then extracted the fraction of small molecules from the cell lysates, presuming that the pVip-produced molecules would be present in this fraction. We analyzed these lysates with liquid chromatography followed by mass spectrometry (LC-MS) using an untargeted approach. As a positive control, we similarly analyzed cell lysates from cells expressing the human viperin protein. As expected, a compound conforming to the mass of ddhCTP was readily detected in lysates from cells expressing the human viperin, but not in negative control lysates derived from MoaA-expressing cells (Extended Data Figure 3, Extended Data Figure 4a). Additional compounds found in the human viperin sample matched the masses of ddh-cytidine (ddhC) and ddh-cytidine monophosphate (ddhCMP), possibly derived from natural decay of ddhCTP as reported to occur for CTP at neutral or acidic pH²⁰. Analysis of fragment ions using MS/MS provided further support that the identified masses are likely derivatives of ddhCTP, which was additionally confirmed by subjecting a synthesized ddhC standard to MS/MS analysis (Extended Data Figure 3a, Extended Data Figure 4a, Extended Data Figure 5). These results confirm that the human viperin actively produces ddhCTP when expressed in *E. coli*, explaining its observed anti-phage activity (Figure 1).

We then analyzed the small molecule fraction from lysates of cells expressing 27 pVips that were found to have anti-phage activity. Derivatives of ddhCTP (including ddhC, as verified by LC-MS with the synthesized ddhC chemical standard) were detected by LC-MS in the

lysate of pVip50 (Figure 3a, Extended Data Figure 4), a protein derived from a methanogenic archaeon that is localized in clade 2 of the pVip tree, verifying that pVips are indeed functional homologs of the human viperin that produce similar anti-viral molecules. However, for most other pVips we could not detect ddhCTP or its derivatives in the cell lysates. We therefore searched for other masses that were markedly enriched in lysates of cells expressing pVips and absent from the negative control lysate. For ten of the pVips we found masses that conform with 3'-deoxy-3',4'-didehydro-guanosine-triphosphate (ddhGTP) and 3'-deoxy-3',4'-didehydro-guanosine-monophosphate (ddhGMP) (Figure 3, Extended Data Figure 4). In addition, for 15 pVips we found other molecules with masses matching 3'-deoxy-3',4'-didehydro-uridine triphosphate (ddhUTP) and 3'-deoxy-3',4'-didehydro-uridine monophosphate (ddhUMP) (Figure 3, Extended Data Figure 4). MS/MS analysis of fragment ions from the masses predicted as ddhGTP, ddhUTP and their monophosphorylated derivatives further supported that they most likely correspond to these molecules (Extended Data Figure 5).

To confirm that pVips convert the nucleotide substrates CTP, GTP, and UTP to their 3'-deoxy-3',4'-didehydro-variants, we performed *in vitro* biochemical assays with purified pVip enzymes. Three isolated recombinant pVips, pVip6, pVip8 and pVip56 - which were predicted to generate ddhCTP, ddhUTP and ddhGTP, respectively (Figure 3c) - were incubated with the SAM cofactor and a nucleotide substrate, in the presence of an artificial electron donor (dithionite). LC-MS analysis confirmed the appearance of the expected ddh-ribonucleotide products in the reaction samples, as compared to control reactions without nucleotide substrate (Figure 3c). Detailed analysis of the MS-MS fragmentation spectra of each compound further supported that these products correspond to ddh-ribonucleotides (Extended Data Figure 6). Together, these results suggest that pVips produce new types of anti-viral ribonucleotides which, to the best of our knowledge, were not observed before in nature.

For most pVips, predicted derivatives of a single modified nucleotide were observed in the lysate (either ddhCTP, ddhGTP or ddhUTP). However, eleven pVips were found to produce derivatives of multiple ddh-ribonucleotides. For example, in lysates derived from pVip46-expressing cells we found both ddhCTP and ddhUTP, and in lysates from pVip58 cells we detected ddhCTP, ddhUTP, ddhGTP and their derivatives (Figure 3). These results suggest that throughout evolution, some pVips may have become more promiscuous and can modify more than one ribonucleotide to its ddh antiviral form. Such pVips may have an advantage when encountering phages that can overcome one of these anti-viral molecules but not the other two.

For five of the tested pVips we did not detect any ddh nucleotide or its derivatives in the cell lysates, despite an anti-viral activity conferred by these pVips (Extended Data Figure 1, Figure 3). It is possible that these pVips produce a different antiviral molecule that could not be detected via our LC-MS protocol or, alternatively, that these pVips evolved to confer defense by another mechanism of action that does not involve the production of anti-viral molecules.

The identity of the molecules produced by the different pVips is largely consistent with their phylogenetic relatedness. All pVips from clades 4-7 seem to produce ddhUTP, with some of these also producing additional ddh-ribonucleotides. In clades 1 and 2, the latter of which resides on the same super-clade as the eukaryotic viperins, we found pVips that produce ddhCTP. Clade 3 includes pVips that appear to generate mostly ddhGTP but sometime ddhUTP (Figure 2, Figure 3).

While the human viperin and pVips produce ddh-ribonucleotides, we found that they protected against phages that have double-stranded DNA genomes (for example T7; Figure 1). We therefore hypothesized that in these cases, the products of pVips affect phage-dependent transcription rather than DNA replication. In support of this hypothesis, it was previously shown that in mammalian cells, T7 RNA polymerase-dependent transcription of GFP was impaired if the human viperin was co-expressed in the same cells²¹. We thus sought to examine whether T7 polymerase-dependent RNA synthesis is affected by pVips. For this we used a plasmid that encodes a GFP reporter gene under the control of a T7 promoter, and introduced it to *E. coli* BL21-DE3, a strain that encodes an inducible T7 RNA polymerase (Figure 4a). As expected, induction of the T7 RNA polymerase using IPTG led to accumulation of a fluorescent GFP signal (Figure 4b). But when pVips or the human viperin were co-expressed in the same cells, the GFP signal was fully repressed (Figure 4b). Repression of T7 RNA polymerase-mediated GFP expression was observed upon expression of pVips producing ddhGTP (pVip60), ddhUTP+ddhCTP (pVip8, pVip9), or ddhCTP (the human viperin) suggesting that the T7 RNA polymerase is sensitive to multiple types of modified ribonucleotides.

Notably, when the above experiment were conducted with pVips mutated to inactivate their active sites, no repression of GFP expression was observed, implying that the catalytic activity of pVips and the production of the ddh ribonucleotide products is required for expression inhibition (Extended Data Figure 7a).

To more directly confirm that the impact of pVips on T7 polymerase-dependent GFP expression was caused by reduced RNA synthesis, we examined GFP RNA levels using RNA-seq. RNA was extracted one hour after T7-mediated induction of GFP expression in cells that co-express pVips or the human viperin. We observed significant reduction in GFP RNA expression (as measured by RPKM, see Methods) when the pVips or the human viperin were expressed in the cell, as compared to control cells that expressed MoeA instead (Figure 4c). The expression levels of genes driven by induced endogenous promoters (specifically, the T7 RNA polymerase gene itself, Extended Data Figure 7b) did not show marked changes during pVip expression, further supporting that pVips specifically target transcription by the phage polymerase. Taken together, these results suggest that pVips can defend against phage T7 via suppression of transcription by the viral RNA polymerase, presumably because their products form RNA chain terminators. Notably, expression of pVips does not appear to be toxic to *E. coli* (Extended Data Figure 8), implying that the bacterial RNA polymerase may be less sensitive to ddh-ribonucleotides as compared to the T7 RNA polymerase. It was similarly shown that ddhCTP produced by the human viperin is not toxic to human cells³.

Bacterial anti-phage defense mechanisms are frequently encoded as multi-gene defense systems, with some genes in the system responsible for identifying the invading phage and others function in mitigating the infection^{7,22,23}. When examining the genomic context of pVips, we identified that most appear to be part of a conserved cassette of genes (Extended Data Figure 9). The most common configuration included, in addition to pVip, two other genes: a gene comprising an ankyrin repeats domain, and a gene encoding a predicted HicA-like RNase (Extended Data Figure 9). Ankyrin repeats domains are common biological recognition motifs involved in protein-protein interactions, and it is possible that the ankyrin repeats domain may serve as a sensor for phage infection. Under this hypothesis, following phage sensing, the ankyrin domain protein would activate expression of the pVip in a manner conceptually similar to interferon-mediated activation of human viperin expression. The associated RNase may be responsible for degradation of prematurely terminated phage RNAs, or for some other auxiliary function. As the pVip gene family we discovered is largely present in non-model organisms for which phages have not been isolated, it is not trivial to study them in their natural defensive settings.

Our data suggest that pVips protect against T7 infection by inhibiting transcription from the viral RNA polymerase. It is therefore puzzling that some pVips protect against phages lambda and P1, which do not encode their own RNA polymerase and rely on the host polymerase for their transcription²⁴. One possible explanation is that phage lambda transcribes its genome in very long operons that can reach 22kb of a continuous RNA molecule²⁵. Even if the host polymerase incorporates the ddh-nucleotide at a very low frequency, the chances to incorporate such a nucleotide in a very long polycistron are high and would affect phage transcription much more profoundly than host transcription. A second possible explanation may be that some proteins in phage lambda alter the properties of the host RNA polymerase to favor viral transcription²⁵, and that the altered polymerase may be more sensitive to the pVip-produced chain terminators. It was previously shown that a small change of one amino acid in a viral polymerase can dramatically affect its sensitivity to a synthetic chain terminator²⁶. Finally, it is also possible that pVips inhibit phage lambda in a manner that is independent from the production of ddh-nucleotides.

While phylogenetically widespread, we found pVips in less than 1% of all genomes that we analyzed. It is possible that pVips are much more abundant in nature, but due to the ongoing arms race with the infecting viruses the sequences of other pVip families have diverged and are no longer alignable to the vertebrate viperin. This hypothesis predicts that additional pVips may be identified in the future, and perhaps these pVips would catalyze the production of new types of antiviral molecules in addition to ddhCTP, ddhGTP, and ddhUTP.

It was recently shown that small molecules of the anthracycline family, produced by species of *Streptomyces*, have natural anti-phage properties and efficiently inhibit phage replication, presumably through intercalation into phage DNA²⁷. Our discovery of pVips reveals another strategy of chemical defense against phages and implies that fighting phages with small molecules may be a more common antiviral strategy than originally anticipated. We hypothesize that future mechanistic studies of bacterial defense systems may reveal additional genes involved in the synthesis of small-molecule antivirals that protect against infection.

Many of the most potent antiviral drugs used in the clinic are synthetic nucleoside chain terminators. These include aciclovir, a commonly used drug against herpes viruses²⁸; azidothymidine (AZT), an anti-HIV drug used clinically since the 1980s²⁸; and sofosbuvir, which in recent years is being used as part of a highly successful treatment for hepatitis C²⁹. It is possible that the new chain terminators that we have discovered could be adopted for clinical treatment of human viruses. Moreover, if it turns out that pVips are just one example of a widely used chemical defense strategy in bacteria, bacteria may prove to host a potent repository of anti-viral molecules that could be harvested and adopted for clinical use. If this would be the case, then environmental bacteria, after being used for many year as a repository for the discovery of novel antibiotics, may once again serve mankind in its battle against pathogens.

Methods

A search for viperin homologs in prokaryotic genomes

The human viperin protein sequence (NCBI accession NP_542388.2) was searched against the protein sequences of all genes in 38,167 bacterial and archaeal genomes downloaded from the Integrated Microbial Genomes (IMG) database¹⁷ in October 2017, using the ‘search’ option in the MMseqs2 package³⁰ (release 6-f5a1c) with default parameters (3 iterations), as previously described⁷. Hits with an e-value higher than 10^{-5} were discarded. The resulting set of proteins was clustered using the ‘cluster’ option of MMseqs2 release v6-f5a1c, with sensitivity parameter of ‘-s 7.5’, coverage parameter 60% and the remaining parameters being the default parameters (Extended Data Table 1). For each cluster, the fraction of genes associated with known defense genes was computed as previously described⁷. Additional candidate prokaryotic viperin homologs (pVips) were searched manually using the “top IMG homologs” function in IMG for the identified genes in the cluster of pVips.

To generate the phylogenetic tree, the protein sequence of prokaryotic viperins, eukaryotic viperins and MoaA sequences were aligned using mafft³¹ (version v7.402, default parameters). The sequences of the eukaryotic viperins and MoaA proteins used in the tree are provided in Extended Data Table 3. The tree was computed with IQ-TREE¹⁹ multicore version v.1.6.5 (option –m TESTNEW in IQ-TREE). The phylogenetic model LG+R6 was ultimately used because it gave the lowest Bayesian Information Criterion (BIC) among all models available for the tree. 1000 ultra-fast bootstraps were performed in order to evaluate node support (options –bb 1000 –wbtl in IQ-TREE). The online tool iTOL³² was used for tree visualization.

Eukaryotic viperins sequences used in the phylogenetic tree were chosen as follows. A homology based-search was performed on the non-redundant eukaryotic proteins database of NCBI using HMMER 3.2.1³³ in the MPI bioinformatics toolkit³⁴ with 205 non redundant pVips as a seed. This search yielded 4915 hits that were used to build an initial phylogenetic tree. The sequences of pVips, MoaA and these hits were aligned using mafft³¹ (version v7.402, default parameters). The tree was computed with IQ-TREE¹⁹ multicore version v.1.6.5 (option –m TESTNEW in IQ-TREE). On this tree, all the pVips were found in a monophyletic clade that also comprised 1298 eukaryotic sequences (Extended Data Figure

10). These 1298 eukaryotic protein sequences were then used to build a second phylogenetic tree. Sequences of these 1298 eukaryotic proteins, pVips and MoaA were aligned using mafft³¹ (version v7.402, default parameters). The tree was computed with IQ-TREE¹⁹ multicore version v.1.6.5 (option `-m TESTNEW` in IQ-TREE). All the eukaryotic viperin sequences represented a monophyletic clade that is internal to the pVips clades. Representative eukaryotic sequences for the tree in Figure 2 were then chosen to span the diversity of the eukaryotic viperin homologs, including sequences from mammals (human, dog, bat), lower animals such as mollusca, as well as fungi (Extended Data Figure 10).

Bacterial strains and growth conditions

Escherichia coli strains (MG1655, Keio *iscR*³⁵, Keio *iscR*-F+, DH5 α , BL21-DE3, BL21-*iscR*) were grown in LB or LB agar at 37 °C unless mentioned otherwise. Whenever applicable, media were supplemented with ampicillin (100 μgml^{-1}), kanamycin (50 μgml^{-1}), chloramphenicol (30 μgml^{-1}), or tetracycline (10 μgml^{-1}) to ensure the maintenance of plasmids.

Plasmids and strain construction

Primers used in this study are shown in Extended Data Table 4. pVip genes were codon optimized for expression in *E. coli* and synthesized by Twist Bioscience (pVips 1-14) or by Genscript (pVips 15-63, MoaA control, human viperin). Synthesized pVip sequences are indicated in Extended Data Table 2. Each candidate sequence was cloned in plasmid pBad/His A (ThermoFisher, Catalog number 43001). For pVips 1-14, flanking sequences were added to the synthesis for cloning purposes (before each gene “GGATCCTACCTGACGCTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTT TGGGCTAACAGGAGGAATTAACC”, after each gene “TAAGAATTCCCAGGCATCAAATAAAACGAAAGGCTCAGT CGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTGCGGTGAACGCTCTCCAGCTG GTACCATATGG”) the vector fragment was generated by PCR (template pBad/His A, primer AB1, AB2) and for each pVip, overhang sequences (matching primers AB3, AB4) overlapping the plasmid backbone were added to allow Gibson assembly. pVips 1-14 were amplified by PCR (using the template of synthesized DNA, primers AB3, AB4), and PCR fragments of pVip and plasmid backbone were joined using Gibson assembly. For pVips 15-63 and the MoaA, DNA synthesis and cloning into pBad/His A was performed by Genscript Corp. The codon-optimized protein-coding sequence of the human viperin was synthesized and cloned into pBad/His A by Genscript Corp. The sequence was then modified by PCR and Gibson assembly to remove the endoplasmic reticulum (ER)-targeting sequence of the human viperin protein (residues 1-50, primers AB86, AB87, AB88, AB89). All experiments involving the human viperin were performed with this shortened version.

All of the pVip plasmids were initially cloned and propagated in DH5 α , and then purified and transformed into the Keio *iscR*³⁵ strain. Because pVips are iron-sulfur cluster proteins, they necessitate active production of iron-sulfur clusters for their enzymatic activity³. We therefore conducted the experiments in *E. coli* strains deleted for *iscR*, a repressor of iron-sulfur cluster production in *E. coli*^{36,37}. For experiments involving phages Qbeta and MS2, the F plasmid is necessary for infection. Thus, the Keio *iscR* F+ strain was constructed

through conjugation of strain Keio *iscR* and strain top10 F+ and then later used as the relevant genetic background.

For protein purification, codon-optimized pVip genes were amplified by PCR (Extended Data Table 4) and cloned into the aTc-inducible expression vector pASG-IBA143 (IBA Lifesciences, pIBA143_vector_F, pIBA143_vector_R), for fusion of a Twin-Strep-tag to the C terminus of the pVips (pIBA143_VipX_F, pIBA143_VipX_R). To construct the *suf* operon expression vector “pSuf”, the complete *suf* operon (*sufABCDSE*) was amplified from *E. coli* MG1655 genomic DNA, and was cloned into the pACYC-184 (NEB) backbone (*Suf_operon F*, *Suf_operon R*) together with the arabinose expression system from pBad/His A vector (ThermoFisher, Catalog number 43001). Tagged pVip8 and pVip56 were transformed in BL21 *iscR* and tagged pVip6 was transformed in BL21 pSuf.

For experiments involving the GFP reporter assay, strains were constructed as follows. BL21-DE3 was knocked out for the *iscR* gene through P1 transduction, with P1- *iscR* phages propagated from strain Keio *iscR*³⁵ followed by kanamycin selection. The final reporter plasmid, pAB151, was constructed to encode: a. GFP under the control of T7 promoter; b. a gene cassette encoding the T7 lysozyme to limit the activity of T7 RNA polymerase; c. an insulator sequence between the chloramphenicol resistance gene and the GFP gene. pAB151 was constructed through three consecutive Gibson assemblies, each of which used two PCR fragments as described below. The first reaction used insert template pDR111 (provided by I. Kolodkin-Gal) with primers OG630, OG631, and vector template pACYC (obtained from Novagene), primers OG629, OG628, to generate plasmid pAB137. The second Gibson reaction used insert template pLysS (obtained from Novagene) with primers AB55, AB56, and vector template pAB137 with primers AB53, AB54, to generate pAB138. The third reaction used insert template pSG1 from ref¹² with primers AB121, AB122, and vector template pAB138 with primers AB119, AB1120, to generate pAB151 (Extended Data Table 4).

For the design of the inactive mutants of human viperin, pVip8, pVip9 and pVip60, the conserved closely spaced cysteine residues in the CxxxCxxC motif, which coordinate the iron-sulfur cluster³⁸, were mutated (Human viperin C32A, C36A, C39A ; pVip8: C22A, C26A, C29A ; pVip9: C17A, C21A, C24A ; pVip60: C192A, C196A, C199A). Mutants were built using Q5 Site directed Mutagenesis kit (NEB) using primers presented in Extended Data Table 4 (AB156-AB163).

Plaque assays

Phages were propagated on *E. coli* MG1655 using the plate lysate method as described in ref³⁹. Lysate titer was determined using the small drop plaque assay method as described in ref⁴⁰. Phages used in this study are presented in Extended Data Table 5.

Plaque assays were performed as previously described⁴⁰. Bacteria from overnight cultures were mixed with MMB agar (LB + 0.1 mM MnCl₂ + 5 mM MgCl₂ + 0.5% agar) supplemented with arabinose (final concentration 0.004%) for induction of pVip expression. Serial dilutions of phage lysate in MMB were dropped on top of the agar plates. After the drops dried up, plates were incubated at 37°C overnight.

Infection dynamics in liquid medium

Overnight cultures were diluted 1:100 in MMB medium and incubated at 37 °C while shaking at 250 r.p.m. for 45 minutes, at which point arabinose was added to a final concentration of 0.2%. Cells were then incubated at 37 °C while shaking at 250 r.p.m. for 45 minutes. 180ul of the diluted cultures were transferred into wells in a 96-well plate containing 20 µl of phage lysate for a final MOI of 0.001. Infections were performed with technical duplicates and OD₆₀₀ was followed using a TECAN Infinite 200 plate reader with measurement every 5 min.

CFU counts were measured using the same experimental setup and time points as above. 10 ul of cells were taken right after dilution (time 0), before induction (45min), and 45 and 90 minutes after induction of pVip expression, serially diluted and plated on selective agar plates. CFU were counted after overnight incubation at 37 °C.

Cell lysates preparation

Overnight cultures of Keio⁰ iscR encoding pVips, the human viperin, MoaA and GFP negative controls were diluted 1:100 in 100 ml LB medium and grown at 37 °C (250 r.p.m.) for 1 hour and 45 minutes. The expression of viperin or MoaA was induced by the addition of arabinose (final concentration 0.2%) and cells were further incubated at 37 °C (250 r.p.m.) for one hour. Cells were then centrifuged at 3,900 r.p.m. for 10 min at 4 °C and samples kept on ice throughout the cell lysate preparation. Pellets were resuspended in 600 µl PBS buffer containing 100 mM sodium phosphate (pH 7.4). The resuspended pellet was supplemented with 1 µl of hen-lysozyme (Merck) (final hen-lysozyme concentration of 10 µg/ml). The resuspended cells were then mixed with Lysing matrix B (MP) beads and cells were disrupted mechanically using a FastPrep-24 bead-beater device (MP) (2 cycles of 40 s, 6 m s⁻¹, at 4 °C). Cell lysates were then centrifuged at 12,000g for 10 min at 4 °C and the supernatant was loaded onto a 3-kDa filter Amicon Ultra-0.5 centrifugal filter unit (Merck) and centrifuged at 14,000g for 30 min at 4 °C. The resulting flow-through, containing substances smaller than 3 kDa, was used as the lysate sample for evaluating the presence of ddh nucleotides by LC-MS.

Detection of ddh-nucleotides in cell lysates

Sample analysis was carried out by MS-Omics (Vedbæk, Denmark) as follows. Samples were diluted 1:1 in 10 % ultra-pure water and 90% acetonitrile containing 10 mM ammonium acetate at pH 9 then filtered through a Costar® Spin-X® centrifuge tube filter 0.22 µm nylon membrane. The analysis was carried out using a UHPLC system (Vanquish, Thermo Fisher Scientific, US) coupled with a high-resolution quadrupole-orbitrap mass spectrometer (Q Exactive™ HF Hybrid Quadrupole-Orbitrap, Thermo Fisher Scientific, US) at a resolution of 120,000 (at 200 m/z). An electrospray ionization interface was used as ionization source. Analysis was performed in positive ionization mode from 200 to 1000 m/z at a scan rate of 3 Hz. The UPLC was performed using a slightly modified version of the protocol described in ref⁴¹. Peak areas were extracted using Compound Discoverer 3.1 (Thermo Fisher Scientific, US).

MS/MS of ddh-nucleotides was acquired using the same instrument with an inclusion list of the different ddh-nucleotide and ddh-nucleoside masses at a resolution of 30,000. Fragmentation was done through a higher-energy collisional dissociation cell using a normalized collision energy of 20, 40 and 60 eV where the spectrum is the sum of each collision energy. Intensity threshold was set to 2×10^4 , isolation window of 0.4 m/z and injection time of 100 ms. Analysis of ddhCTP and ddhGTP derivatives was performed in positive ionization mode, and for ddhUTP derivatives in negative ionization mode.

Raw data files were processed by Compound Discoverer™ 3.0 software. Unknown compounds were detected with a 3 ppm mass tolerance, signal to noise ratio of 3, 30% of relative intensity tolerance for isotope search, and minimum peak intensity of 5×10^5 . The compounds were grouped with a 5 ppm mass and 0.2 min retention time tolerances. Blank samples were used to remove background noise, and annotated peaks that were 5 times higher than the blanks were kept. Metabolites identified were searched against ChemSpider™ chemical structure database with 3 ppm mass tolerance, mzCloud spectral library with a precursor and fragment mass tolerance of 3 and 5 ppm respectively, and an internal MSMS library through mzVault with the same tolerance as mzCloud. Two data sources were searched in the ChemSpider database: Human Metabolome Database (HMDB) and *E. coli* Metabolome Database (ECMDB).

The raw data files for the MS and MS/MS data in this section, as well as additional technical details, are available for download on the Metabolights repository under study number MTBLS1750.

Quantification of 3'-deoxy-3',4'-didehydro cytidine (ddhC)

The 3'-deoxy-3',4'-didehydro cytidine molecule was synthesized by Jena Bioscience (Jena, Germany) at purity of 97.5% and was used as a standard for ddhC quantification in cell lysates using LC-MS. Sample analysis was carried out by MS-Omics (Vedbæk, Denmark) as follows. Samples were diluted 1:1 in 10 mM ammonium formate and 0.1% formic acid in ultra-pure water. The analysis was carried out using the LC-MS setup described above. An electrospray ionization interface was used as ionization source performed in positive ionization mode. The UHPLC method is based on Waters Application note 2011, 720004042en (Waters Corporation, Milford, US). Peak areas of 3'-deoxy-3',4'-didehydrocytidine (ddhC) were extracted using Trace Finder™ Version 4.1 (Thermo Fisher Scientific, US) and quantified using an external calibration with the standard.

pVips purification and in vitro enzymatic assays

Overnight cultures of BL21- *iscR* (pVip8 and pVip56) or BL21 pSuf (pVip6) cells freshly transformed with plasmids encoding the tagged pVip, were seeded at an initial OD₆₀₀ of ~0.06 in 1-2 L of selective LB medium. pSuf expression was induced at OD₆₀₀ 0.2-0.3 (0.2% arabinose, 100 μM FeCl₃, 100 μM L-cysteine). pVip expression was induced at OD₆₀₀ 0.6–0.8 (50 ng/mL aTc) and incubated at 37 °C with shaking for 3-4h. Pellets were then harvested by centrifugation and stored at -20 °C.

Frozen cell pellets were resuspended in cold lysis buffer [50 mM Tris-HCl, 500 mM NaCl, 5 mM dithiothreitol (DTT), 0.5 M arginine, and 20% glycerol], and sonicated with a Branson

Sonifier (15 sec ON, 45 sec OFF, 10 min total ON, 30% amplitude) on ice. Lysates were subjected to centrifugation for 30 min at 17,000 g and 4 °C. The lysate was loaded onto a StrepTactin Superflow High Capacity (IBA Lifesciences) column, previously equilibrated with 20 column volumes of Buffer W (100 mM Tris-HCl pH 8, 300 mM NaCl, 5 mM DTT, 10% glycerol). The column was washed twice with 10 column volumes of Buffer W and eluted with buffer E (50 mM Tris-HCl pH 8, 300 mM NaCl, 5 mM DTT, 2.5 mM desthiobiotin, 20% glycerol). The presence of the pVip proteins in the resulting fractions was confirmed by SDS-PAGE. Purified proteins were frozen in liquid nitrogen and stored at – 80 °C.

Protein reconstitution—Purified protein solutions were thawed on ice and introduced into an MBraun anaerobic chamber maintained at <0.1 ppm oxygen. All subsequent steps were performed in anaerobic conditions at 12 °C. Purified pVips were incubated for 1 hour with 50 mM DTT with gentle shaking. Protein solutions were supplemented with 8-fold molar excess $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$, incubated for 15 min with gentle shaking, followed by the addition 8-fold molar excess of Na₂S droplet by droplet. After incubation for 3-4h to overnight with slow shaking, the reconstituted pVips were transferred to the Reaction Buffer (50 mM HEPES pH 7.5, 150 mM KCl, 5 mM DTT, 20% Glycerol) using PD-10 desalting columns (GE Healthcare) and concentrated using an Amicon Ultra centrifugal 10 kDa filter to a final protein concentration of 20-50 μM . Proteins were then flash-frozen with liquid nitrogen and stored at – 80 °C.

In vitro enzymatic assays—For pVip6 and pVip8, reactions were performed in a total volume of 100 μL containing: 20-50 μM reconstituted enzyme in Reaction Buffer, 2 mM S-Adenosyl methionine (SAM), 1 mM of nucleotide substrate, and 5 mM sodium dithionite. Reactions were carried out inside the anaerobic chamber maintained at <0.1 ppm oxygen. A 10 μL aliquot was removed from the reaction mixture (sample before reaction). Reactions were then initiated with sodium dithionite and incubated at 37 °C for 1-2 h. After incubation, samples were taken out of the anaerobic chamber and stored at -80 °C until analysis.

For pVip56, to obtain sufficient amounts of ddhGTP for MS/MS analysis, an enzymatic reaction in a total volume of 1ml was performed, containing 113 mM pVip56, 2mM SAM, 2mM GTP and 5mM dithionite in Reaction Buffer. Reactions were carried out in anaerobic conditions as previously described and incubated at 37 °C for 3 hours. To remove the protein, 10K centrifugal filters were used. The flow through was diluted 2-fold into cold 10 mM ammonium bicarbonate buffer pH 9.0 (buffer A), then loaded onto Capto™ HiRes Q 5/50 (GE Healthcare) pre-equilibrated with buffer A. The column was washed with 25 mL of buffer A and elution was performed using linear elution gradient (100 mL) of 200 mM to 800 mM ammonium bicarbonate, pH 9. The purified product was lyophilized and resuspended in water prior to LC-MS analysis.

LC-MS analysis of *in vitro* assays—LC-MS measurements were performed with a Thermo Scientific Q Exactive Orbitrap mass spectrometry system equipped with a Dionex Ultimate 3000 UHPLC system. The software Thermo Xcalibur was used for instrument control and data processing. Prior analysis, 10 μL of sample from enzymatic assays were

mixed with 40 μ L of acetonitrile:methanol organic mixture (5:3 v/v ratio). The mixtures were vortexed, centrifuged at 17,000g for 2 min and 3 μ L of supernatant were injected onto an SeQuant® ZIC®-pHILIC 5 μ m polymeric 100 x 2.1 mm HPLC column. The mobile phase was composed of 20 mM ammonium carbonate pH 9.5 (solvent A) and 100% acetonitrile (solvent B). Samples were separated using a constant flow rate of 0.2 mL/min: 80% solvent B was held for 2 min, followed by a gradient from 80% to 20% of solvent B for 15 min, before immediately returning to 80% solvent B for equilibration for 9 min. Data analysis was performed using the Thermo Scientific FreeStyle software.

T7 dependent GFP expression assay

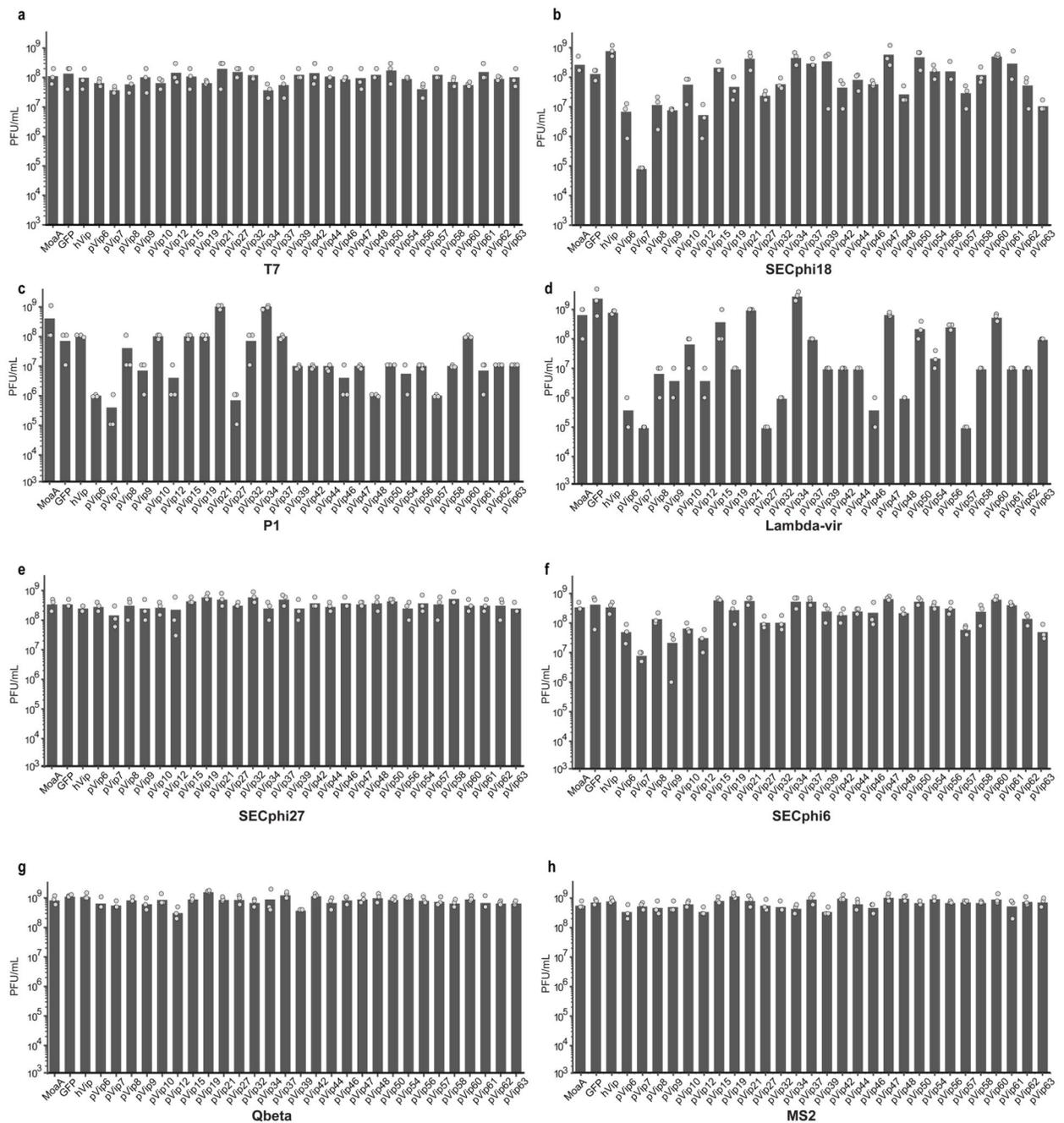
Overnight cultures of BL21-DE3 *iscR* cells containing pAB151 and pVip-encoding plasmids (or plasmids encoding MoaA or the human viperin) were diluted 1:100 in LB medium and incubated in a 96-well plate format at 37 °C with shaking of 250 r.p.m. until OD₆₀₀ reached 0.1. Arabinose was then added to a final concentration of 0.2%. After 45 minutes of incubation at 37 °C, 250 r.p.m., the expression of T7 RNA polymerase was induced by the addition of IPTG to a final concentration of 0.1 mM. Fluorescence levels (wavelength excitation 488nm, emission 520nm) and cell density (OD₆₀₀) were monitored using TECAN Infinite 200 plate reader with measurement every 15 min.

Quantification of GFP transcripts using RNA-seq

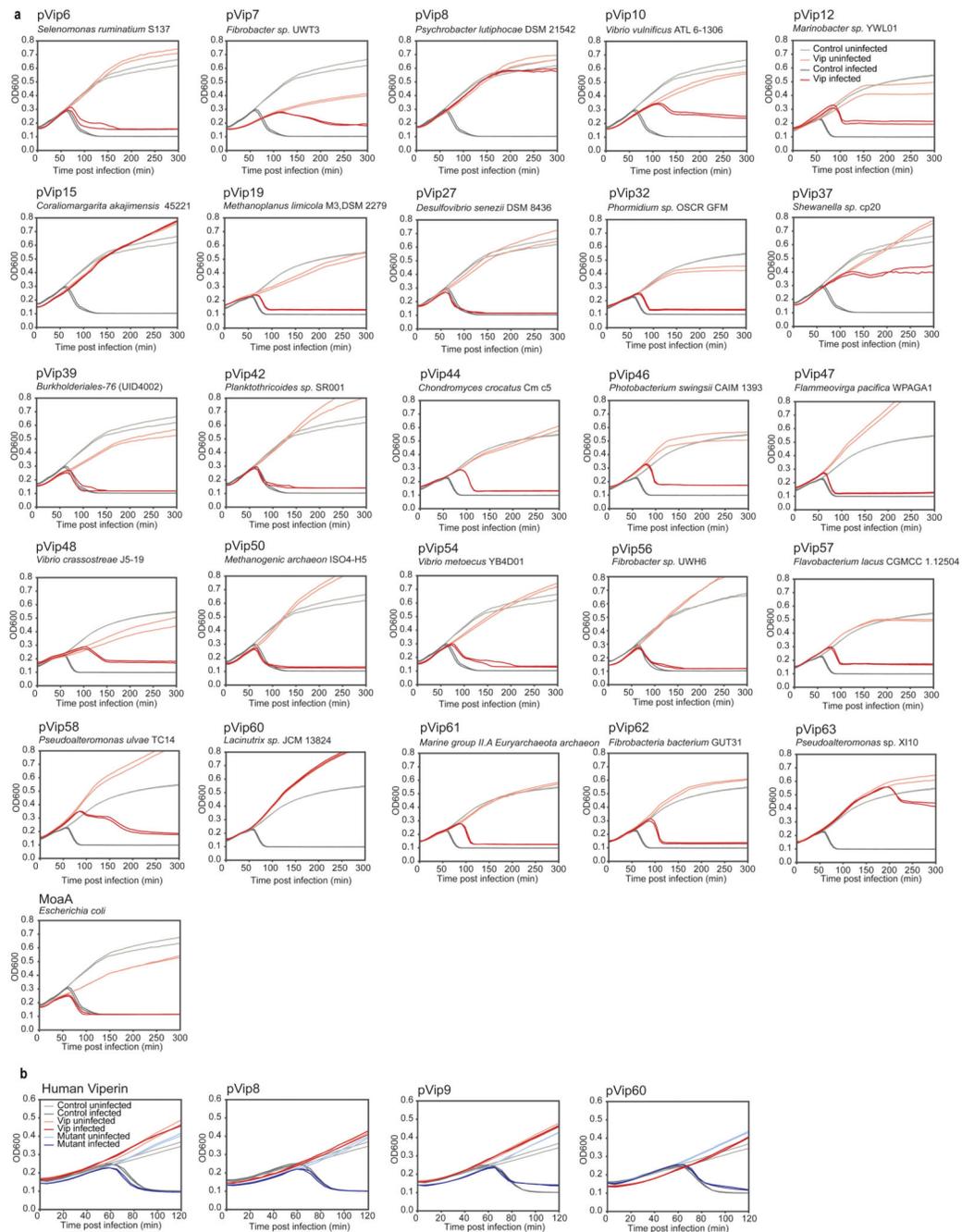
BL21-DE3 *iscR* cells containing pAB151 and pVip-encoding plasmids (or plasmids encoding MoaA or the human viperin) were diluted 1:100 in 5 ml LB medium supplemented with antibiotics (chloramphenicol, kanamycin and ampicillin). These cells were grown at 37 °C with shaking of 250 r.p.m. to OD₆₀₀ of 0.3 and expression of the viperin (or MoaA) protein was induced by the addition of arabinose (final concentration 0.2%). After 45 minutes of incubation at 37 °C, 250 r.p.m., the expression of T7 RNA polymerase was induced by the addition of IPTG to a final concentration of 0.1 mM. After one hour, samples were centrifuged for 10 minutes at 4000 r.p.m in 4 °C. The supernatant was discarded, and pellets were used for RNA extraction. Bacterial pellets were lysed using TRIzol and phenol-chloroform. Bacterial pellets were treated with 100 μ l of 2mg/ml lysozyme (in Tris 10mM EDTA 1mM pH 8.0) and incubated at 37 °C for 5 minutes. 1ml of TRI-reagent was added, samples were then vortexed for 10 seconds before addition of 200 μ l chloroform. Following another vortexing step, the samples were left at room temperature for 5 minutes to allow phase separation and then centrifuged at 12000g, 4 °C for 15 minutes. The upper phase was added to 500 μ l of isopropanol. Samples were then incubated overnight at -20 °C. Finally, following 30 minutes centrifugation at 12000g at 4 °C, samples were washed twice with ice cold 70% ethanol, and resuspended in 50 μ l water. RNA levels were measured using Nanodrop. All RNA samples were treated with TURBO™ DNase (Life technologies, AM2238). Ribosomal RNA depletion and RNA-seq libraries were prepared as described in ref⁴², except that all reaction volumes were reduced by a factor of 4.

RNA-seq libraries were sequenced using Illumina NextSeq platform, Reads were mapped as described in ref⁴² to the reference genome of *E. coli* BL21 DE3 (NC_012892) as well as the plasmids present in the relevant strain (pAB151 and plasmids encoding pVip/MoaA/human viperin). RNA-seq-mapped reads were used to generate reads-per-gene and RPKM counts.

Extended Data

**Extended Data Figure 1. pVips protect against phage infection.**

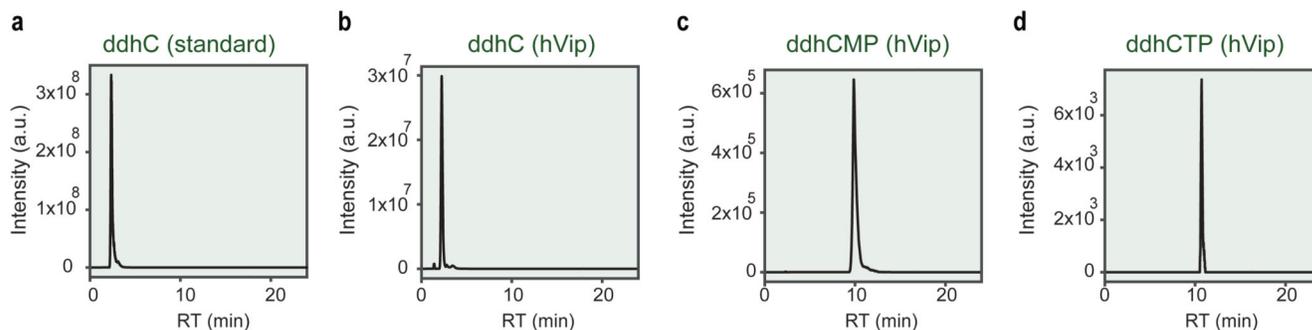
Bacteria expressing pVips, GFP or MoaA (negative controls), or the human viperin gene were grown on agar plates and tenfold serial dilutions of the phage lysate were dropped on the plates. a - h. Efficiency of plating (EOP) data, representing plaque-forming units per millilitre; each bar graph represents average of three replicates, with individual data points overlaid.



Extended Data Figure 2. T7 infection in liquid culture in the presence of pVips.

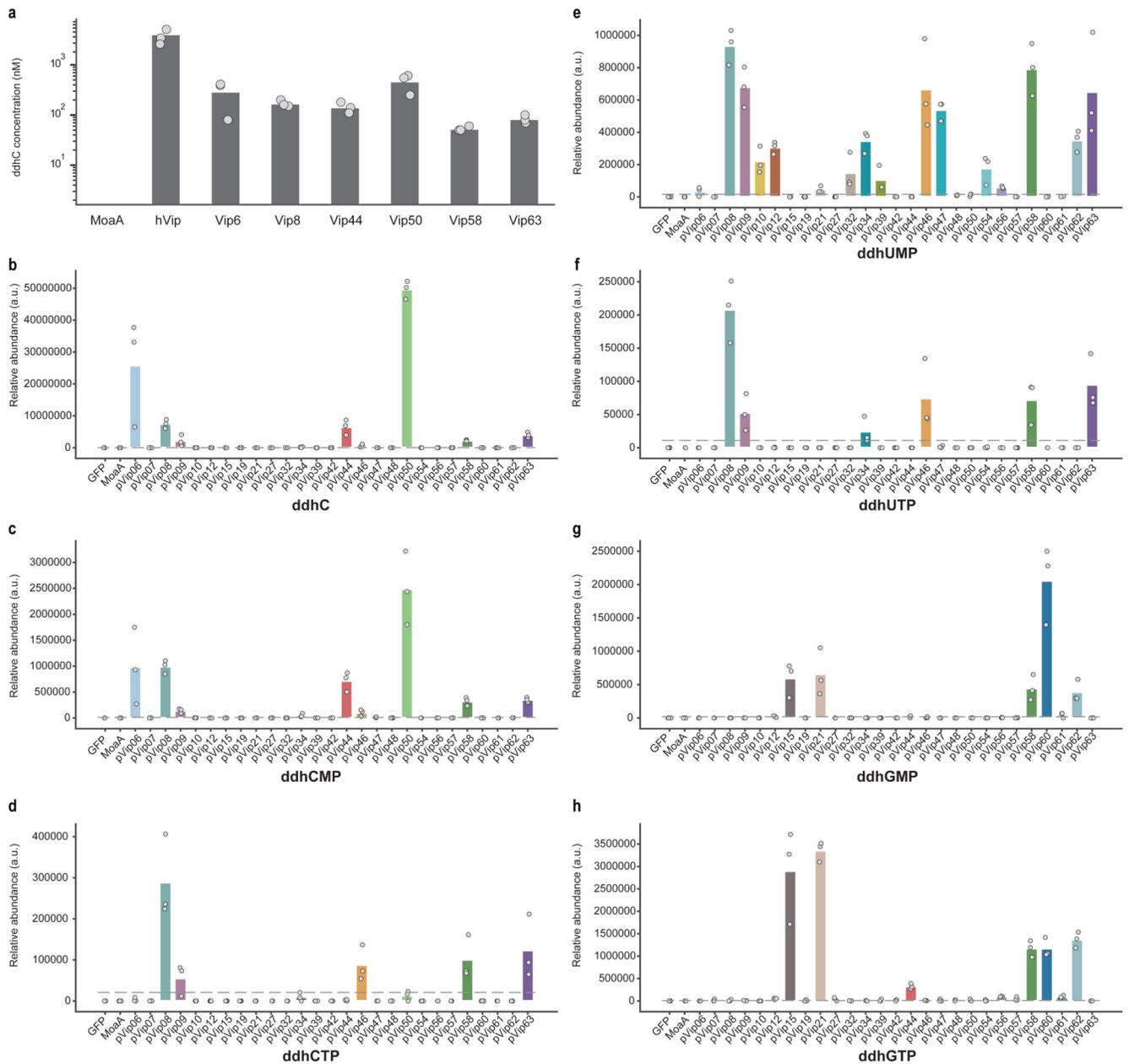
a. For each pVip, growth curves of liquid cultures infected by phage T7 (MOI 0.001) are shown. Light and dark grey are uninfected and infected controls (strain expressing GFP), respectively. Light and dark red are uninfected and infected strains expressing pVips, respectively. Two technical replicates are presented as individual curves; representative of three biological replicates. The negative controls (GFP uninfected, GFP infected) are the same for pVips 6, 7, 8, 10, 15, 27, 37, 39, 42, 50, 54, MoaA, and for pVip12, 19, 32, 44, 46, 47, 48, 57, 58, 60, 61, 62, 63. b. The catalytic activity of pVips is required for defense

against T7 phage. For each pVip and its respective mutant (mutation of three cysteines in the active site), growth curves of liquid cultures infected by phage T7 (MOI 0.001) are presented. Light and dark grey are uninfected and infected controls (strain expressing MoaA), respectively. Light and dark red are uninfected and infected strains expressing viperins, respectively. Light and dark blue are uninfected and infected strains expressing catalytically inactive mutants. Two technical replicates are presented as individual curves; representative of three biological replicates.



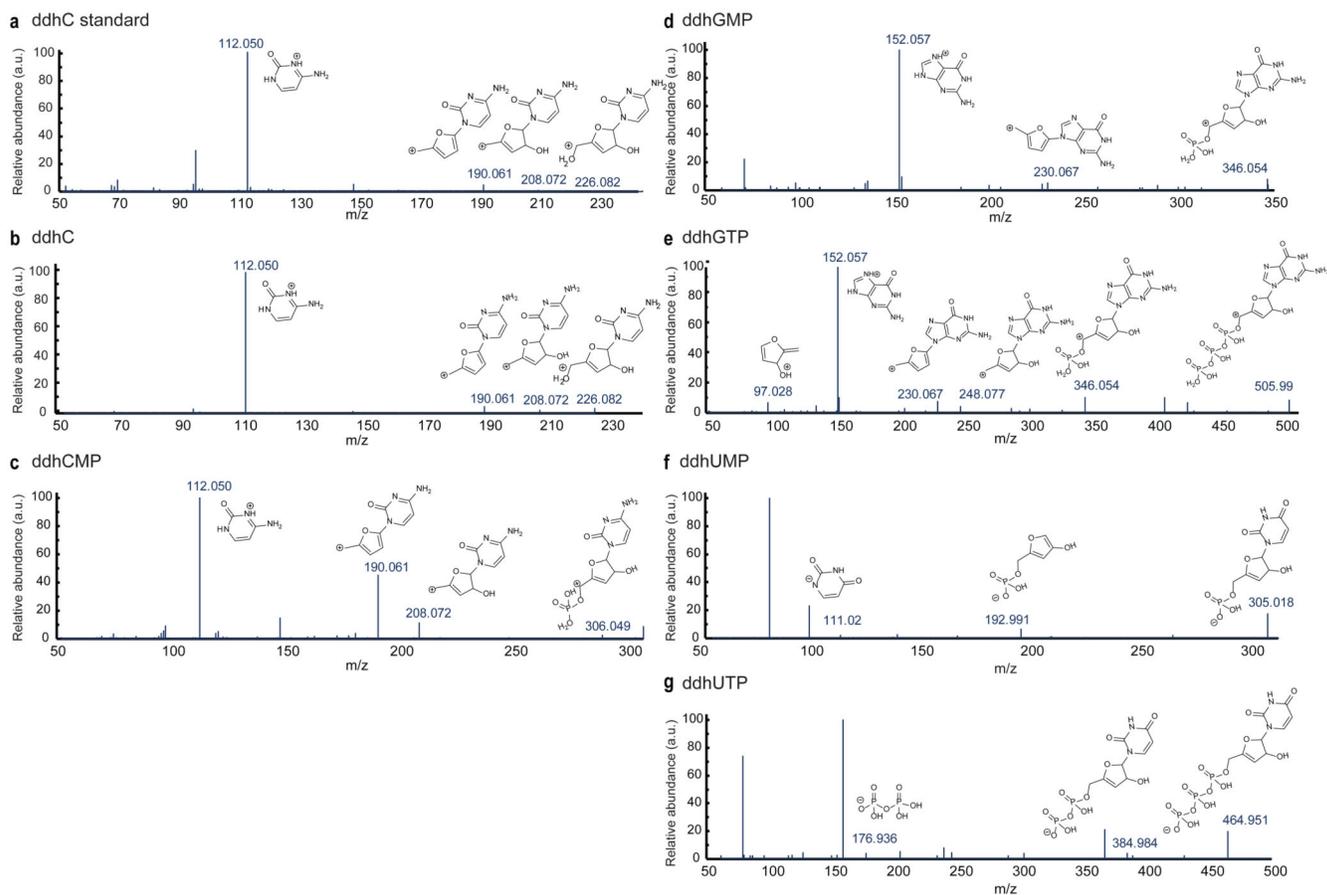
Extended Data Figure 3. Detection of ddhCTP and ddhCTP derivatives in cell lysates from an *E. coli* strain expressing the human viperin.

a. Extracted ion chromatogram of the ddhC standard. b-d. Extracted ion chromatogram for singly charged masses that are predicted to correspond to ddhC (m/z 226.08223, retention time (RT) of 2.2 minutes)(b), ddhCMP (m/z 306.04856, RT 9.7)(c), ddhCTP (m/z 465.98122, RT 10.7)(d) in cell lysates from an *E. coli* strain expressing the human viperin. Representative of three replicates.



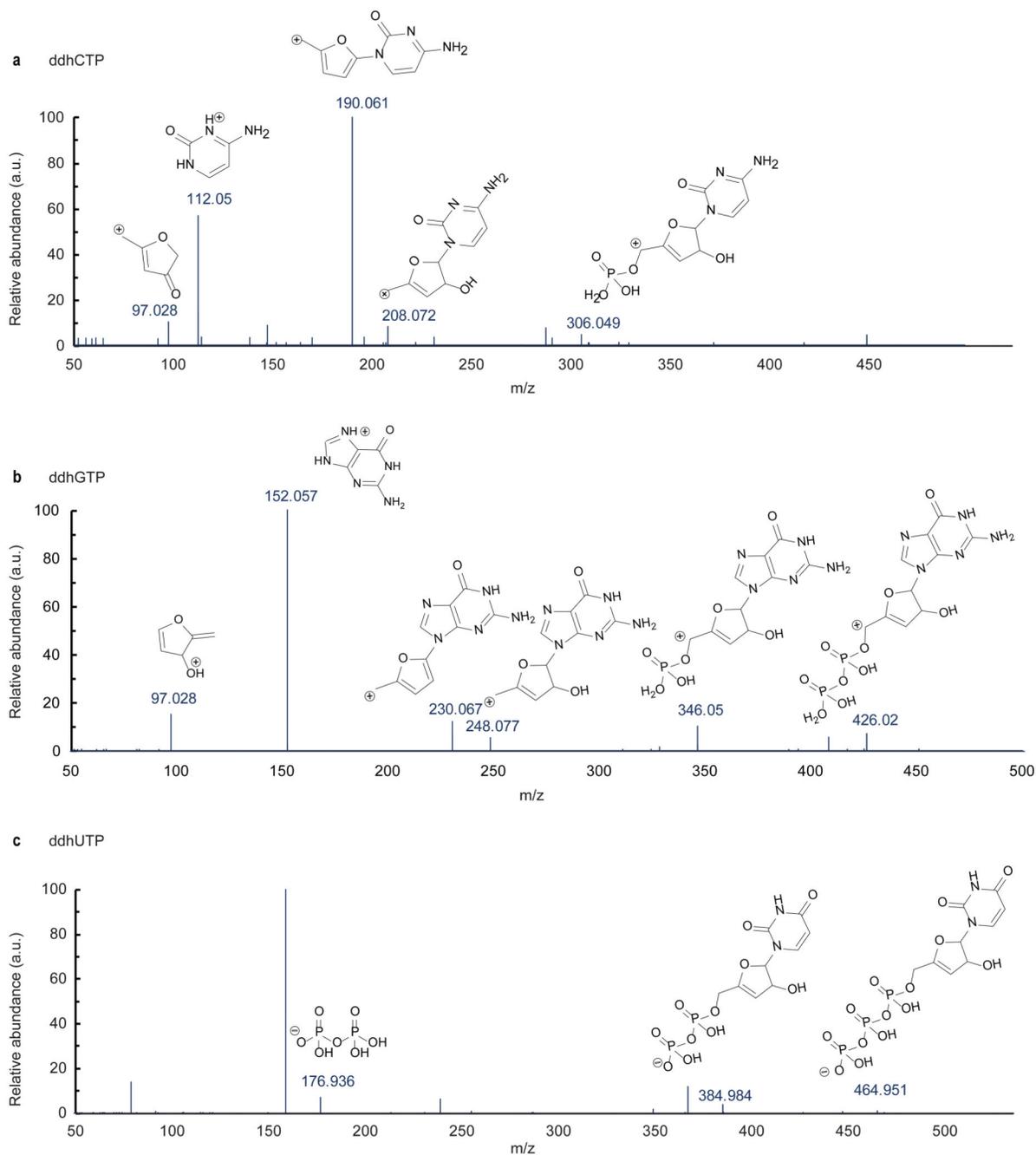
Extended Data Figure 4. Detection of ddh-ribonucleotides in lysates of cells that express pVips. a. Quantification of ddh-cytidine (ddhC) in lysates of cells expressing pVips. Detection and quantification of ddhC was performed using LC-MS with a synthesized chemical standard (Methods). For MoaA, the measurement was under the limit of detection (LOD 0.0003 μM). Bar graph represents average of three replicates, with individual data points overlaid. b-h. Relative abundance for singly charged masses that are predicted to correspond to ddhC (m/z 226.08223, retention time (RT) of 2.2 minutes) (b), ddhCMP (m/z 306.04856, RT 9.7) (c), ddhCTP (m/z 465.98122, RT 10.7) (d), ddhUMP (m/z 307.03258, RT 8.7) (e), ddhUTP (m/z 466.96524, RT 9.9) (f), ddhGMP (m/z 346.05471, RT 9.8) (g), and ddhGTP (m/z 505.98737, RT 10.7) (h). Average relative abundance is presented as bar graph, with

individual data points from three biological replicates overlaid. Limit of detection (LOD) is indicated by a dashed grey line. A compound was defined as present, in Figure 3, if all three replicated were above the LOD.



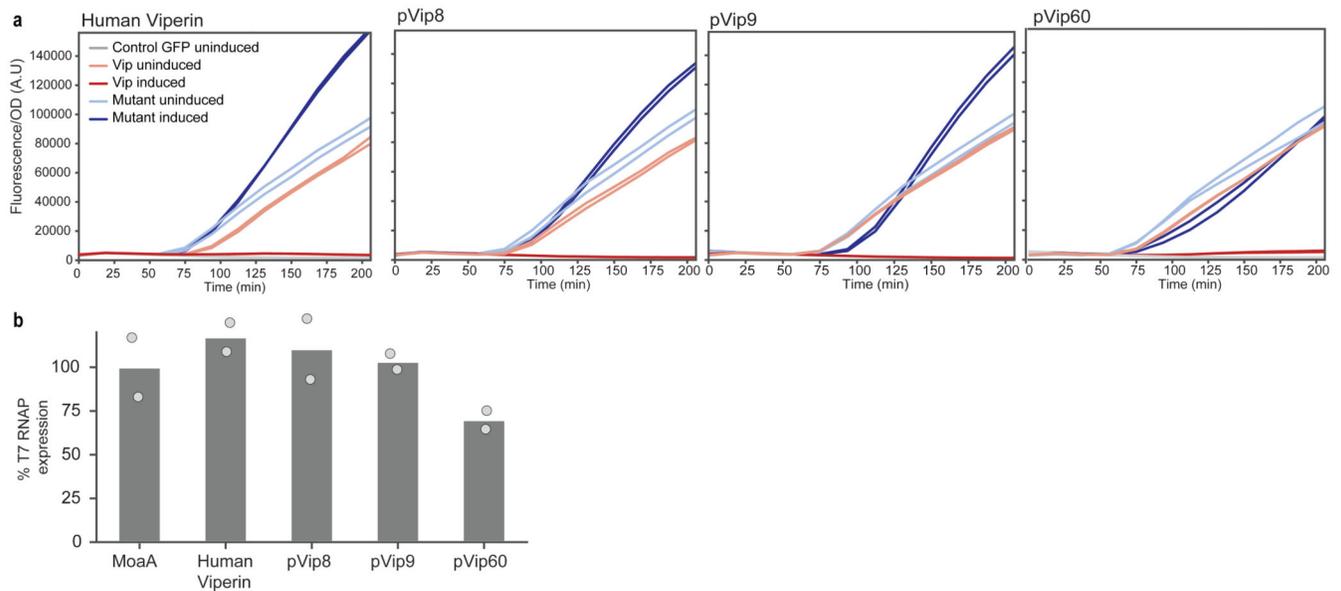
Extended Data Figure 5. MS/MS fragmentation spectra for predicted compounds.

MS/MS data were acquired in positive ionization mode for a synthesized chemical standard ddhC (a) as well as for masses from the human viperin cell lysate predicted to correspond to ddhC (b), and ddhCMP (c). Similar data were obtained for masses from the pVip21 cell lysate predicted to correspond to ddhGMP (d), and ddhGTP (e). MS/MS data were acquired, in negative ionization mode, from the pVip47 cell lysate for masses predicted to correspond to ddhUMP (f), and ddhUTP (g). In all panels, assignment of hypothetical structures is indicated for informative fragment ions. The ddhC molecule is annotated to level 1, and all other molecules are annotated to level 2b, per the Metabolomics Standards Initiative nomenclature.



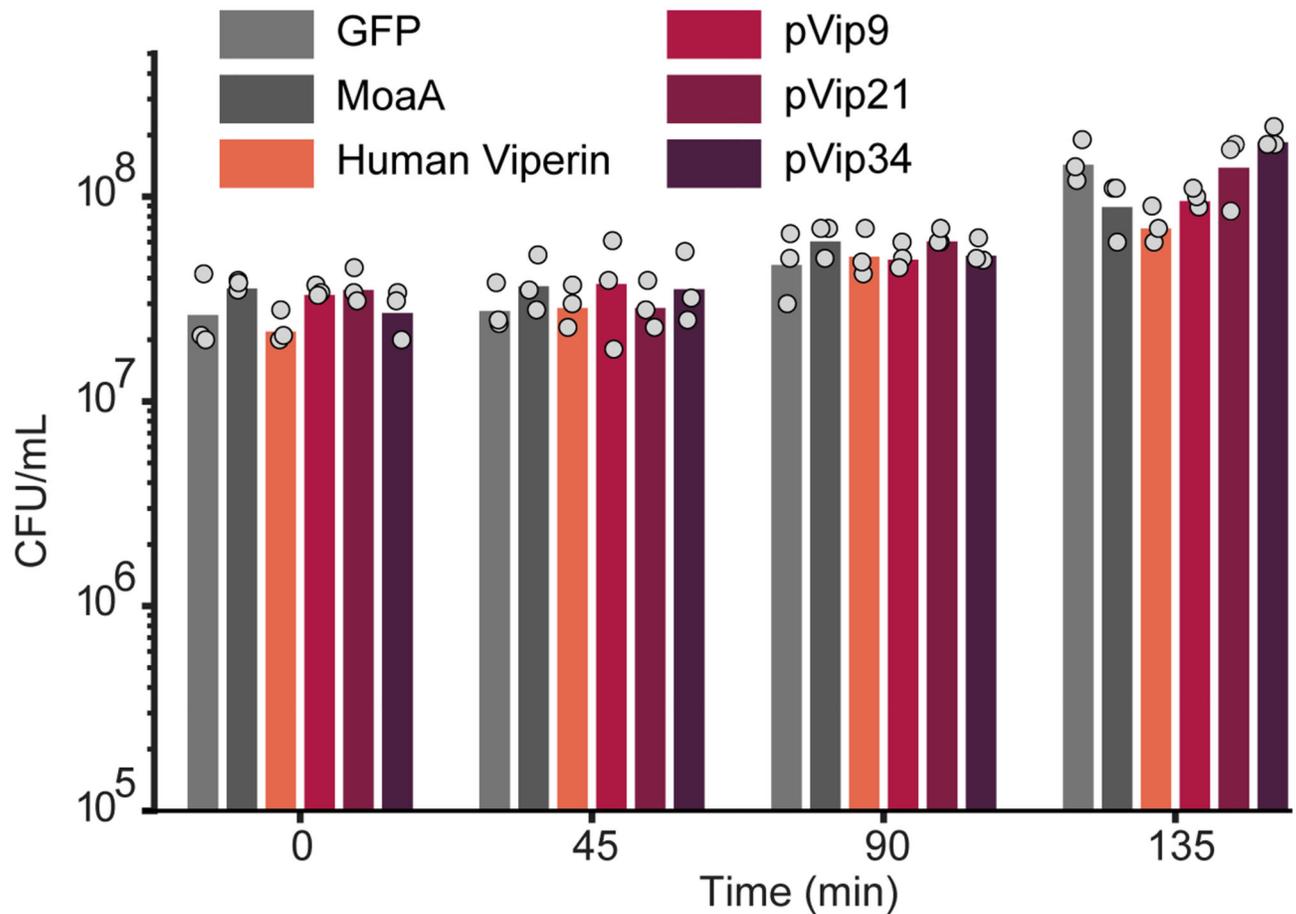
Extended Data Figure 6. MS/MS fragmentation spectra for predicted compounds from *in vitro* reactions with purified pVips.

(a-b) MS/MS data were acquired in positive ionization mode for the product detected in reaction samples using purified pVip6 or purified pVip56 and CTP and GTP as nucleotide substrates respectively; the resulting products are predicted to correspond to ddhCTP (a) and ddhGTP (b). (c) MS/MS data were acquired in negative ionization mode for product detected in reaction samples using purified pVip8 UTP as substrate; the resulting product is predicted to correspond to ddhUTP (c).



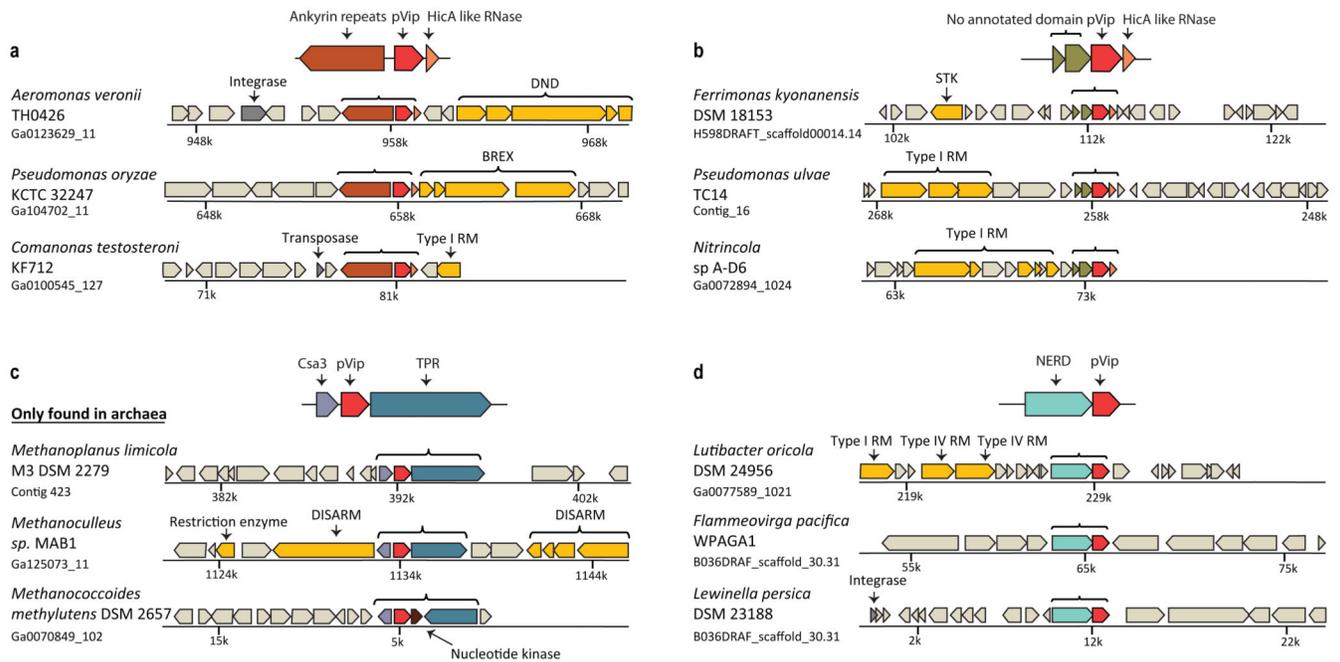
Extended Data Figure 7. Transcription during induction of WT and mutant pVips.

a. The catalytic activity of pVips is required for defense against T7 phage and repression of viral transcription. Application of the reporter assay (same as presented in Figure 4a) for strains expressing the human viperin, pVips and their cognate catalytically inactive mutants. Strains are first induced with arabinose for 45 minutes to express the pVip. At $t=0$, IPTG is added to express the GFP. Fluorescence/OD over time curves are presented for each strain. Dark and light red correspond to induced and non-induced wild type viperins, respectively; Dark and light blue correspond to induced and non-induced mutant viperins, respectively. Grey curve corresponds to negative control (WT viperin, no addition of IPTG). Two technical replicates are presented by individual curves. Representative of two biological replicates. b. T7 RNAP expression as measured by RNA-seq. The expression (RPKM) of T7 RNAP in cells expressing viperins was compared to that in cells expressing the MoaA negative control. Bar graphs represent average of two replicates, with individual data points overlaid.



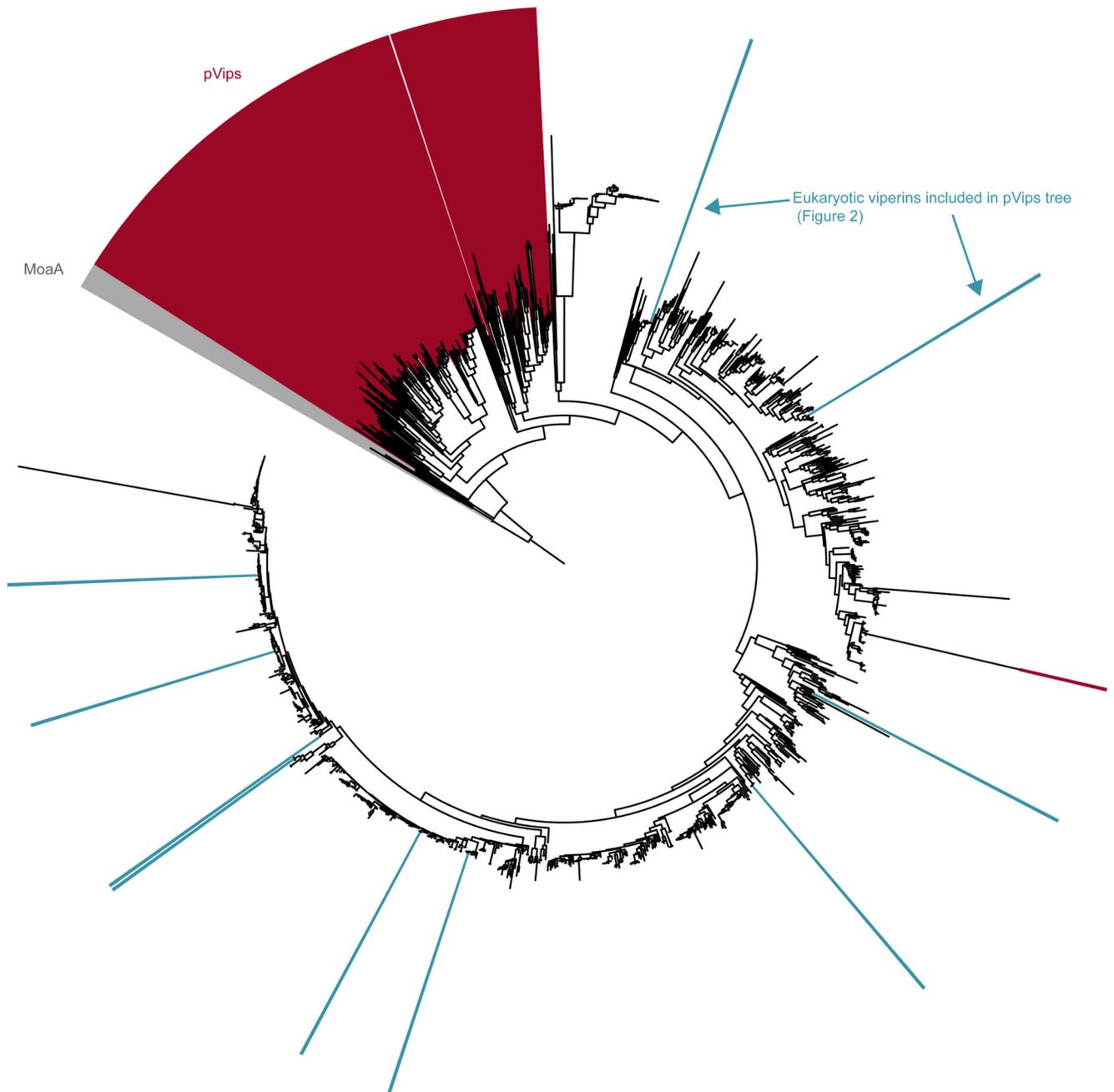
Extended Data Figure 8. Heterologous expression of pVips is not toxic in *E. coli*.

Expression of pVips, human viperin or negative controls (GFP, MoaA) was induced at 45min by addition of arabinose (final concentration 0.2%). CFU were measured right after dilution from overnight culture (t=0), before induction (t=45), and 45 and 90 minutes after induction (t=90, t=135).



Extended Data Figure 9. Putative multi-gene defense systems that include pVips.

Representative instances of pVips and their genomic neighborhood. Genes predicted to be part of the pVip-containing defense system are highlighted. Genes known to be involved in defense are in yellow. Genes of mobile genetic elements are in dark grey. RM, restriction-modification; TA, toxin-antitoxin. The name of bacterial species, and the accession of the relevant genomic scaffold in the IMG database¹⁷ are indicated on the left. Panels a-d represent four common configurations of putative pVip-containing systems found in bacterial and archaeal genomes.



Extended Data Figure 10. Phylogenetic tree of pVips and putative eukaryotic viperins. MoaA sequences were used as an outgroup (grey). pVips are depicted in red and putative eukaryotic viperins selected for the phylogenetic tree presented in Figure 2 are depicted in blue.

Extended Data Table 1
Clusters of genes retrieved by the homology-based
search of human viperin in prokaryotic genomes.

Genes used to calculate defense scores were those present on DNA scaffolds of sufficient size with at least ten genes from each side of the viperin homolog.

pVip number	IMG ID	IMG Genome ID	Genome Name	Domain	Phylum	Class	Order
1	2695043264	2693429896	<i>Lutibacter oricola</i> DSM 24956	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales
2	2684559953	2681813561	<i>Chryseobacterium gambrii</i> DSM 18014	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales
3	2507146842	2506783068	<i>Methanofollis liminatans</i> GKZPZ, DSM 4140	Archaea	Euryarchaeota	Methanomicrobia	Methanomicrobiales
6	2624749465	2623620517	<i>Selenomonas ruminantium</i> S137	Bacteria	Firmicutes	Negativicutes	Selenomonadales
7	2739066738	2738541339	<i>Fibrobacter</i> sp. UWT3	Bacteria	Fibrobacteres	Fibrobacteria	Fibrobacterales
8	2521798317	2521172648	<i>Psychrobacter lutiphocae</i> DSM 21542	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
9	2574301464	2574179732	<i>Vibrio porteresiae</i> DSM 19223	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
10	2720695169	2718218250	<i>Vibrio vulnificus</i> ATL 6-1306	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
11	2632766730	2630968672	<i>Shewanella baltica</i> OS678	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
12	2698137626	2695420938	<i>Ruegeria intermedia</i> DSM 29341	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales
13	2744653400	2744054531	<i>Marinobacter</i> sp. YWL01	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
14	2654783232	2654587543	<i>Pseudomonas nitroreducens</i> B	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
15	646713396	646564524	<i>Coraliomargarita akajimensis</i> DSM 45221	Bacteria	Verrucomicrobia	Opitutae	Puniciceoccales
17	2504625218	2504557017	<i>Marinomonas</i> sp. GOBB3-320	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales
18	2506474236	2506381025	<i>Methanoplanus limicola</i> M3, DSM 2279	Archaea	Euryarchaeota	Methanomicrobia	Methanomicrobiales
19	2506475787	2506381025	<i>Methanoplanus limicola</i> M3, DSM 2279	Archaea	Euryarchaeota	Methanomicrobia	Methanomicrobiales
20	2509664214	2509601008	<i>Methanomethylovorans hollandica</i> DSM 15978	Archaea	Euryarchaeota	Methanomicrobia	Methanosarcinales
21	2515428782	2515154070	<i>Lewinella persica</i> DSM 23188	Bacteria	Bacteroidetes	Saprosipria	Saprosipirales
22	2518436022	2518285547	<i>Pelobacter carbinolicus</i> Bd1, GraBd1	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfuromonadales
23	2522341593	2522125098	<i>Tolomonas lignilytica</i> BRL6-1	Bacteria	Proteobacteria	Gammaproteobacteria	Aeromonadales

pVip number	IMG ID	IMG Genome ID	Genome Name	Domain	Phylum	Class	Order
24	2524269675	2524023156	Conchiformibius kuhniae DSM 17694	Bacteria	Proteobacteria	Betaproteobacteria	Neisseriales
25	2525334630	2524614668	Methanocorpusculum bavaricum DSM 4179	Archaea	Euryarchaeota	Methanomicrobia	Methanomicrobiales
26	2557036911	2556921023	Pseudoalteromonas sp. H105 PacBio methylation	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
27	2574506394	2574179788	Desulfovibrio senezii DSM 8436	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfovibrionales
28	2574517928	2574179790	Endozoicomonas numazuensis DSM 25634	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales
29	2582805913	2582580599	Composite genome from Lake Mendota Epilimnion pan-assembly MEint.metabat.6813	Bacteria	Bacteroidetes	unclassified	unclassified
30	2582946381	2582580664	Composite genome from Trout Bog Hypolimnion pan-assembly TBhypo.metabat.2746	Bacteria	Verrucomicrobia	unclassified	unclassified
31	2596421479	2595698251	Kibdelosporangium aridum DSM 43828	Bacteria	Actinobacteria	Actinobacteria	Pseudonocardiales
32	2609132705	2608642208	Phormidium sp. OSCR GFM (version 2)	Bacteria	Cyanobacteria	unclassified	Oscillatoriales
33	2618018523	2617270916	Marinobacter zhejiangensis CGMCC 1.7061	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
34	2619892213	2619618891	Cryomorphaceae bacterium EBPR_Bin_135	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales
36	2631333032	2630968323	Nitrincola sp. A-D6	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales
37	2632937107	2630968711	Shewanella sp. cp20	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
38	2633985761	2630968972	Methanococcoides methylutens DSM 2657	Archaea	Euryarchaeota	Methanomicrobia	Methanosarcinales
39	2634960437	2634166261	Burkholderiales-76 (UID4002)	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales
40	2635314107	2634166348	Actinomadura echinospora DSM 43163	Bacteria	Actinobacteria	Actinobacteria	Streptosporangiales
41	2637497700	2636415666	Photobacterium leiognathi mandapamensis KNH6	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
42	2639213731	2636416084	Planktothricoides sp. SR001	Bacteria	Cyanobacteria	unclassified	Oscillatoriales
43	2641427518	2639762959	Actinobacteria bacterium OK074	Bacteria	Actinobacteria	Actinobacteria	unclassified
44	2648875132	2648501185	Chondromyces crocatus Cm c5	Bacteria	Proteobacteria	Deltaproteobacteria	Myxococcales
45	2649163162	2648501251	Moritella viscosa 06/09/139	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales

pVip number	IMG ID	IMG Genome ID	Genome Name	Domain	Phylum	Class	Order
46	2649993803	2648501459	Photobacterium swingsii CAIM 1393	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
47	2651203508	2648501771	Flammeovirga pacifica WPAGA1	Bacteria	Bacteroidetes	Cytophagia	Cytophagales
48	2651490945	2648501840	Vibrio crassostreae J5-19	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
49	2651585264	2648501863	Aeromonas caviae CECT 4221	Bacteria	Proteobacteria	Gammaproteobacteria	Aeromonadales
50	2661858798	2660238307	Methanogenic archaeon ISO4-H5	Archaea	Euryarchaeota	Thermoplasmata	unclassified
51	2665950188	2663763173	Legionella santicrucis SC-63-C7	Bacteria	Proteobacteria	Gammaproteobacteria	Legionellales
52	2674184607	2671180787	Pseudomonas stutzeri C2	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
53	2684813341	2684622550	Aquabacterium parvum B6	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales
54	2693697599	2693429564	Vibrio metoecus YB4D01	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
55	2694112273	2693429660	Helicobacter bilis Missouri	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacteriales
56	2701115162	2700988679	Fibrobacter sp. UWH6	Bacteria	Fibrobacteres	Fibrobacteria	Fibrobacteriales
57	2718503187	2718217692	Flavobacterium lacus CGMCC 1.12504	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales
58	2721736750	2718218507	Pseudoalteromonas ulvae TC14	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
59	2728147792	2724679805	Shimia sagamensis DSM 29734	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacteriales
60	2733913669	2731957952	Lacinutrix sp. JCM 13824	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales
61	2741341560	2740891962	Marine group II.A Euryarchaeota archaeon SCGC AG-487_M08 (contamination screened)	Archaea	Euryarchaeota	Candidatus Poseidoniia	Candidatus Poseidoniales
62	2743907592	2740892545	Fibrobacteria bacterium GUT31 IN01_31	Bacteria	Fibrobacteres	Fibrobacteria	unclassified
63	2744633848	2744054527	Pseudoalteromonas sp. XII0	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
64	2741409035	2740891993	Candidatus Heimdallarchaeota archaeon LC_3	Archaea	Candidatus Heimdallarchaeota	unclassified	unclassified
65	2504129180	2503982047	Anabaena cylindrica PCC 7122	Bacteria	Cyanobacteria	unclassified	Nostocales
66	637160692	637000327	Treponema denticola ATCC 35405	Bacteria	Spirochaetes	Spirochaetia	Spirochaetales
67	637364324	637000336	Vibrio vulnificus CMCP6	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
68	637468954	637000337	Vibrio vulnificus YJ016	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales

pVip number	IMG ID	IMG Genome ID	Genome Name	Domain	Phylum	Class	Order
69	637586319	637000206	Photobacterium profundum SS9	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
70	637752529	637000204	Pelobacter carbinolicus Bd1, GraBd1	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfuromonadales
71	639797708	639633052	Psychromonas ingrahamii 37	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
72	640805406	640753033	Marinomonas sp. MWYL1	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales
73	640830189	640753049	Shewanella baltica OS185	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
74	641096015	640963011	Beggiatoa sp. PS	Bacteria	Proteobacteria	Gammaproteobacteria	Thiotrichales
75	641147750	640963027	Marinobacter algicola DG893	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
76	641288534	641228507	Shewanella baltica OS195	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
77	643461066	643348574	Shewanella baltica OS223	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
78	646369858	646311927	Fibrobacter succinogenes S85	Bacteria	Fibrobacteres	Fibrobacteria	Fibrobacterales
79	646419713	646311963	Thermomonospora curvata DSM 43183	Bacteria	Actinobacteria	Actinobacteria	Streptosporangiales
80	647622404	647533121	Campylobacteriales sp. GD 1	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacterales
81	649804297	649633054	Helicobacter felis CS1, ATCC 49179	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacterales
82	650410387	650377991	Marinobacter adhaerens HP15	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
83	650419199	650377942	Fibrobacter succinogenes S85	Bacteria	Fibrobacteres	Fibrobacteria	Fibrobacterales
84	650463340	650377984	Vibrio furnissii 2510/74, NCTC 11218	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
85	650537321	650377925	Coprococcus catus GD/7	Bacteria	Firmicutes	Clostridia	Clostridiales
86	650742368	650716002	Acidiphilium multivorum AIU301	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales
87	650921542	650716044	Lacinutrix sp. 5H-3-7-4	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales
88	2501733929	2501651210	Photobacterium profundum 3TCK	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
89	2502233141	2502171154	Thermoplasmatales archaeon BRNA1	Archaea	Euryarchaeota	Thermoplasmata	Thermoplasmatales
90	2509552219	2509276055	Treponema saccharophilum PB, DSM 2985	Bacteria	Spirochaetes	Spirochaetia	Spirochaetales
91	2512440669	2512047059	Haemophilus haemolyticus M21621	Bacteria	Proteobacteria	Gammaproteobacteria	Pasteurellales
92	2519473577	2519103099	Methanobolus psychrophilus R15	Archaea	Euryarchaeota	Methanomicrobia	Methanosarcinales

pVip number	IMG ID	IMG Genome ID	Genome Name	Domain	Phylum	Class	Order
93	2519473579	2519103099	Methanobolus psychrophilus R15	Archaea	Euryarchaeota	Methanomicrobia	Methanosarcinales
94	2519484486	2519103103	Brachyspira pilosicoli B2904	Bacteria	Spirochaetes	Spirochaetia	Brachyspirales
95	2519815572	2519103180	Curvibacter lanceolatus ATCC 14669	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales
96	2521802859	2521172649	Rheinheimera perlucida DSM 18276	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales
97	2522303848	2522125086	Succinimonas amylyolytica DSM 2873	Bacteria	Proteobacteria	Gammaproteobacteria	Aeromonadales
98	2524107537	2524023060	Ferrimonas kyonanensis DSM 18153	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
99	2525610838	2524614740	Pseudomonas stutzeri MF28	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
100	2525930338	2524614816	Halodesulfobivrio aestuarii DSM 10141	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfovibrionales
101	2528325157	2528311002	Comamonas testosteroni ZNC0007	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales
102	2812941770	2529293096	Sulfurimonas gotlandica GD1	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacteriales
103	2532381218	2531839141	Kingella kingae PYKK081	Bacteria	Proteobacteria	Betaproteobacteria	Neisseriales
104	2532646932	2531839206	Thaueria sp. 63	Bacteria	Proteobacteria	Betaproteobacteria	Rhodocyclales
105	2538932271	2537561856	Brachyspira hamptonii 30446	Bacteria	Spirochaetes	Spirochaetia	Brachyspirales
106	2540642849	2540341105	Methanoculleus bourgensis MS2	Archaea	Euryarchaeota	Methanomicrobia	Methanomicrobiales
107	2540668036	2540341115	Candidatus Methanomethylophilus alvus Mx1201	Archaea	Euryarchaeota	Thermoplasmata	Methanomassiliococcales
108	2540825991	2540341170	Pseudodesulfobivrio piezophilus C1TLV30	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfovibrionales
109	2541039228	2540341248	Ruminococcus flavefaciens AE3010	Bacteria	Firmicutes	Clostridia	Clostridiales
110	2541315631	2541046975	Treponema medium ATCC 700293	Bacteria	Spirochaetes	Spirochaetia	Spirochaetales
111	2546450678	2545824694	Marinobacter santoriniensis NKSG1	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
112	2546738312	2545824767	Bacteriovorax sp. DB6_IX	Bacteria	Proteobacteria	Oligoflexia	Bacteriovoracales
113	2547718745	2547132187	Acinetobacter sp. MDS7A	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
114	2551476655	2551306039	Vibrio harveyi ZJ0603	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
115	2551491916	2551306042	Vibrio genomosp. F10 ZF-129	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
116	2551562099	2551306058	Vibrio splendidus 12E03	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales

pVip number	IMG ID	IMG Genome ID	Genome Name	Domain	Phylum	Class	Order
117	2551596444	2551306067	Vibrio rumoiensis IS-45	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
118	2553401559	2551306520	Aliivibrio logei ATCC 35077	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
119	2553886541	2551306646	Vibrio harveyi AOD131	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
120	2558097217	2556921621	Acinetobacter towneri DSM 14962	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
121	2559286049	2558860239	Spiroplasma culicicola AES-1	Bacteria	Tenericutes	Mollicutes	Entomoplasmatales
122	2559416375	2558860277	Treponema primitia ZAS-1	Bacteria	Spirochaetes	Spirochaetia	Spirochaetales
123	2562001279	2561511079	Selenomonas sp. FC4001	Bacteria	Firmicutes	Negativicutes	Selenomonadales
124	2563081558	2562617115	Myxococcus hansupus DSM 436	Bacteria	Proteobacteria	Deltaproteobacteria	Myxococcales
125	2563230595	2562617155	Helicobacter bilis ATCC 43879	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacteriales
126	2565569616	2563367142	Vibrio haliotocoli NBRC 102217	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
127	2565702223	2563367170	Helicobacter bilis WiWa	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacteriales
128	2566542256	2565956643	Acinetobacter parvus NIPH 1103	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
129	2566736970	2565956698	Acinetobacter towneri DSM 14962	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
130	2569938648	2568526421	Vibrio parahaemolyticus TUMSAT_H10_S6	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
131	2574423613	2574179766	Thiomonas sp. FB-Cd, DSM 25617	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales
132	2574578667	2574179802	Sulfitobacter mediterraneus KCTC 32188	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacteriales
133	2577747326	2576861245	Vibrio parahaemolyticus VIP4-0444	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
134	2577787495	2576861258	Pseudoalteromonas haloplanktis TB25	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
135	2580440151	2579778656	Pseudoalteromonas haloplanktis AC163	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
136	2581032418	2579778800	Vibrio metoecus PPCK-2014	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
137	2581542389	2579778918	Vibrio harveyi E385	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
138	2582293224	2579779100	Vibrio parahaemolyticus VIP4-0430	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
139	2582959978	2582580668	Composite genome from Trout Bog Hypolimnion pan-	Bacteria	Verrucomicrobia	unclassified	unclassified

pVip number	IMG ID	IMG Genome ID	Genome Name	Domain	Phylum	Class	Order
			assembly TBhypo.metabat.3004				
140	2583671671	2582580861	Pseudoalteromonas sp. TAE56	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
141	2584203718	2582580995	Vibrio parahaemolyticus TUMSAT_DE2_S2	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
142	2585240392	2582581301	Janthinobacterium sp. RA13	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales
143	2587265930	2585427937	Pseudoalteromonas sp. 520P1	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
144	2589217693	2588253911	Chondromyces apiculatus DSM 436	Bacteria	Proteobacteria	Deltaproteobacteria	Myxococcales
145	2597063350	2596583606	Fibrobacter succinogenes elongatus HM2	Bacteria	Fibrobacteres	Fibrobacteria	Fibrobacterales
146	2600497862	2600254970	Pseudomonas sp. 1-7	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
147	2600833866	2600255071	Vibrio ezurae NBRC 102218	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
148	2609594859	2609459643	Janthinobacterium sp. OK676	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales
149	2609930410	2609459764	Marinobacter sp. ES.048	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
150	2611345001	2609460080	Hyalangium minutum DSM 14724	Bacteria	Proteobacteria	Deltaproteobacteria	Myxococcales
151	2611749855	2609460164	Acidithiobacillus thiooxidans Licanantay	Bacteria	Proteobacteria	Acidithiobacillia	Acidithiobacillales
152	2612132826	2609460245	Delftia tsuruhatensis 391	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales
153	2617465221	2617270765	Marinobacter mobilis CGMCC 1.7059	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
154	2617538802	2617270789	Flavobacterium omnivorum CGMCC 1.2747	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales
155	2619647987	2619618818	Pseudidiomarina donghaiensis CGMCC 1.7284	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
156	2619760352	2619618853	Betaproteobacteria sp. genome_bin_13	Bacteria	Proteobacteria	Betaproteobacteria	unclassified
157	2620549291	2619619052	Unclassified Chloroflexi bacterium bin152	Bacteria	Chloroflexi	unclassified	unclassified
158	2621169600	2619619266	Photobacterium phosphoreum ANT220	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
159	2623278845	2622736530	Roseovarius lutimaris DSM 28463	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales
160	2632746825	2630968667	Nonlabens ulvanivorans JCM 19297	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales
161	2642232622	2639763156	Aeromonas sobria CECT 4245	Bacteria	Proteobacteria	Gammaproteobacteria	Aeromonadales

pVip number	IMG ID	IMG Genome ID	Genome Name	Domain	Phylum	Class	Order
162	2644760915	2643221740	Chryseobacterium sp. Leaf201	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales
163	2645912334	2645727543	Aeromonas tecta CECT 7082	Bacteria	Proteobacteria	Gammaproteobacteria	Aeromonadales
164	2647434260	2645727892	Comamonas testosteroni KF712	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales
165	2649993012	2648501459	Photobacterium swingsii CAIM 1393	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
166	2651793160	2648501913	Pseudomonas nitroreducens DPB	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
167	2652273697	2651869653	Rubrivivax sp. AAP121	Bacteria	Proteobacteria	Betaproteobacteria	unclassified
168	2654809173	2654587547	Achromobacter spanius CGMCC9173	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales
169	2658339966	2657245169	Methanoculleus sp. EBM-46	Archaea	Euryarchaeota	Methanomicrobia	Methanomicrobiales
170	2667505054	2663763602	Pseudomonas hussainii JCM 19513	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
171	2667963948	2667527390	Fabibacter pacificus CGMCC 1.12402	Bacteria	Bacteroidetes	Cytophagia	Cytophagales
172	2668144532	2667527434	Pseudomonas oryzae KCTC 32247	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
173	2668847476	2667527626	Vibrio parahaemolyticus S164	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
174	2672407511	2671180348	Vibrio tritonius AM2	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
175	2674782375	2671180928	Vibrio parahaemolyticus CFSAN007447	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
176	2677278474	2675903261	Anabaena sp. 4-3	Bacteria	Cyanobacteria	unclassified	Nostocales
177	2682061458	2681812894	Sphaerotilus natans ATCC 13338	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales
178	2684092807	2681813425	Methanoculleus sp. MAB1	Archaea	Euryarchaeota	Methanomicrobia	Methanomicrobiales
179	2688794699	2687453440	Aeromonas veronii TH0426	Bacteria	Proteobacteria	Gammaproteobacteria	Aeromonadales
180	2693209812	2690316327	Vibrio parahaemolyticus S165	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
181	2694949528	2693429874	Olleya namaensis DSM 28881	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales
182	2700499480	2698536835	Microgenomates bacterium JGI CrystG Apr02-3-G15 (contamination screened)	Bacteria	Candidatus Microgenomates	unclassified	unclassified
183	2701140257	2700988686	Fibrobacter sp. UWH9	Bacteria	Fibrobacteres	Fibrobacteria	Fibrobacterales
184	2701911183	2700989248	Vibrio parahaemolyticus CFSAN007448	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales

pVip number	IMG ID	IMG Genome ID	Genome Name	Domain	Phylum	Class	Order
185	2705695255	2703719122	unclassified Deltaproteobacteria bin 1	Bacteria	Proteobacteria	Deltaproteobacteria	unclassified
186	2706043000	2703719236	Fibrobacter sp. UWB7	Bacteria	Fibrobacteres	Fibrobacteria	Fibrobacterales
187	2712662546	2711768198	Arsukibacterium ikkense GCM72	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales
188	2714077658	2713896747	Vibrio alginolyticus V2	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
189	2719376594	2718217925	Alteromonas sp. Mex14	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
190	2719498267	2718217953	Marinobacter salinus Hb8	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
191	2719828580	2718218033	Lutibacter sp. LPB0138	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales
192	2722236530	2721755284	Gammaproteobacteria bacterium GWF2_41_13	Bacteria	Proteobacteria	Gammaproteobacteria	unclassified
193	2727845415	2724679709	Saccharicrinis carchari DSM 27040	Bacteria	Bacteroidetes	Bacteroidia	Marinilabiliales
194	2728971251	2728369061	Aliivibrio wodanis CL7	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
195	2729066335	2728369080	Dechloromonas denitrificans ATCC BAA-841	Bacteria	Proteobacteria	Betaproteobacteria	Rhodocyclales
196	2730169305	2728369366	Tenacibaculum sp. LPB0136	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales
197	2731232863	2728369654	Vibrio sp. JCM 19061	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
198	2735939253	2734482289	Sulfitobacter mediterraneus DSM 12244	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales
199	2740266671	2739367982	Oceanospirillales bacterium JGI 01_G13_750m (contamination screened)	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales
200	2741408272	2740891993	Candidatus Heimdallarchaeota archaeon LC_3	Archaea	Candidatus Heimdallarchaeota	unclassified	unclassified
201	2742412079	2740892189	Marinobacter sp. EN3	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
202	2742415354	2740892190	Acinetobacter sp. COS3	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
203	2743908240	2740892545	Fibrobacteria bacterium GUT31 IN01_31	Bacteria	Fibrobacteres	Fibrobacteria	unclassified
204	2751139676	2747843223	Janthinobacterium sp. 64	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales
205	2752652723	2751185612	Bacteroidales bacterium Bact_07	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales
206	2753090639	2751185737	Salinivibrio sp. DV	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
207	2753093587	2751185738	Salinivibrio sp. BNH	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
208	2753363234	2751185801	Aliivibrio sp. IS128	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales

pVip number	IMG ID	IMG Genome ID	Genome Name	Domain	Phylum	Class	Order
209	2753367132	2751185802	Aliivibrio sp. 1S165	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
210	2753371117	2751185803	Aliivibrio sp. 1S175	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
211	2753755176	2751185895	Haemophilus quentini MP1	Bacteria	Proteobacteria	Gammaproteobacteria	Pasteurellales
212	2758508848	2757320913	Diaphorobacter polyhydroxybutyrativovans SL-205	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales
213	2758538137	2757320982	Winogradskyella sp. PC-19	Bacteria	Bacteroidetes	Flavobacteriia	Flavobacteriales
214	2758668677	2758568024	Thermococcus siculi RG-20	Archaea	Euryarchaeota	Thermococci	Thermococcales
215	2766104288	2765235962	Neisseria sp. 10023	Bacteria	Proteobacteria	Betaproteobacteria	Neisseriales
216	2770832229	2767802753	Cystobacter ferrugineus Cbfe23	Bacteria	Proteobacteria	Deltaproteobacteria	Myxococcales
217	2558444101	2558309039	Megasphaera elsdenii T81	Bacteria	Firmicutes	Negativicutes	Veillonellales
218	2620552401	2619619052	Unclassified Chloroflexi bacterium bin152	Bacteria	Chloroflexi	unclassified	unclassified
219	2620553354	2619619052	Unclassified Chloroflexi bacterium bin152	Bacteria	Chloroflexi	unclassified	unclassified
220	2671326339	2671180039	Streptomyces rubidus CGMCC 4.2026	Bacteria	Actinobacteria	Actinobacteria	Streptomycetales
221	2722096198	2721755233	Nitrospirae bacterium GWD2_57_9	Bacteria	Nitrospirae	unclassified	unclassified
222	2725246328	2724679053	Photobacterium kishitanii 201212X	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
223	2518432012	2518285546	Pelobacter carbinolicus Bd1, GraBd1	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfuromonadales
224	2563551698	2563366541	Helicobacter bilis ATCC 51630	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacteriales
225	2587714015	2585428053	Phormidium sp. OSCR GFM	Bacteria	Cyanobacteria	unclassified	Oscillatoriales
226	2607643251	2606217509	Pelobacter carbinolicus Bd1, GraBd1	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfuromonadales
227	2621018896	2619619224	Psychromonas sp. SP041 (contamination screened)	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
228	2631330386	2630968323	Nitrincola sp. A-D6	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales
229	2635859667	2634166480	Vibrio parahaemolyticus T9109	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
230	2640547395	2639762741	Rubrivivax sp. AAP65	Bacteria	Proteobacteria	Betaproteobacteria	unclassified
231	2648083068	2645728046	Pseudomonas aeruginosa KF702	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
232	2652587282	2651869727	Vibrio parahaemolyticus 49	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales

pVip number	IMG ID	IMG Genome ID	Genome Name	Domain	Phylum	Class	Order
233	2659528208	2657245575	Vibrio parahaemolyticus CFSAN007439	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
234	2660786846	2660238041	Vibrio parahaemolyticus CFSAN007437	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
235	2661121363	2660238124	Vibrio parahaemolyticus S163	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
236	2661449018	2660238210	Vibrio parahaemolyticus S167	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
237	2669650778	2667527830	Vibrio parahaemolyticus ISF-29-3	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
238	2674100589	2671180763	Vibrio parahaemolyticus CFSAN007440	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
239	2684997840	2684622594	Vibrio crassostreae J5-4	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
240	2691769858	2690315984	Vibrio parahaemolyticus S162	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
241	2693670287	2693429558	Vibrio parahaemolyticus S160	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
242	2700498760	2698536834	Microgenomates bacterium JGI CrystG Apr02-3-G15 (unscreened)	Bacteria	Candidatus Microgenomates	unclassified	unclassified
243	2701137224	2700988685	Fibrobacter sp. UWH5	Bacteria	Fibrobacteres	Fibrobacteria	Fibrobacterales
244	2701611589	2700989176	Acinetobacter townieri KCTC 12419	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
245	2702436836	2700989395	Vibrio parahaemolyticus RM-13-3	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
246	2702853828	2700989481	Helicobacter bilis ATCC 51630	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacterales
247	2715393182	2713897062	Vibrio parahaemolyticus NCKU_TV_3HP	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
248	2715399811	2713897063	Vibrio parahaemolyticus NCKU_TV_5HP	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
249	2718509529	2718217694	Fibrobacter sp. UWH8	Bacteria	Fibrobacteres	Fibrobacteria	Fibrobacterales
250	2724146490	2721755831	Vibrio parahaemolyticus FORC_023_378	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
251	2725246629	2724679053	Photobacterium kishitanii 201212X	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales
252	2725300238	2724679064	Pseudoalteromonas sp. H105	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales
253	2729852792	2728369263	Fabibacter pacificus DSM 100771	Bacteria	Bacteroidetes	Cytophagia	Cytophagales
254	2747864490	2747842404	Vibrio vulnificus NV1	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales

Extended Data Table 2
List of pVips

Attached as an excel file. Gene and genome accessions in the IMG database¹⁷ are indicated.

Cluster name	# of genes	# of non-redundant genes (identical sequences removed)	# of genes used to calculate defense score	Defense score (% cases next to known defense genes)
2676290849	1394	855	735	6.10%
pVips	164	134	83	60.20%
2709749501	77	54	25	20.00%
2541272930	25	21	13	7.70%
2713134820	21	17	13	7.70%
2523876556	16	9	2	0.00%
2709564561	5	4	2	0.00%
2713748397	5	3	3	0.00%
2694949528	4	4	3	100.00%
2705785291	4	3	2	0.00%
2509529583	2	2	2	50.00%
2634981381	2	1	1	0.00%
2574215492	1	1	0	0.00%
2709307270	1	1	0	0.00%
2722379863	1	1	0	0.00%
2731410697	1	1	0	0.00%
2753755176	1	1	1	0.00%

Extended Data Table 3
MoaA genes (outgroup) and eukaryotic viperin sequences used in the phylogenetic tree in Figure 2.

Protein	Accession number	Species
MoaA	WP_011245749.1	<i>Bacillus clausii</i>
MoaA	WP_005712890.1	<i>Glaesserella parasuis</i>
MoaA	WP_011257906.1	<i>Xanthomonas oryzae</i>
MoaA	WP_009990662.1	<i>Sulfolobus solfataricus</i>
MoaA	KXZ35264.1	<i>Vibrio alginolyticus</i>
MoaA	PNG85565.1	<i>Pasteurella multocida</i>
MoaA	NP_005934.2	<i>Homo sapiens</i>
MoaA	XP_005761268.1	<i>Emiliana huxleyi CCMP1516</i>
MoaA	XP_009771542.1	<i>Nicotiana glauca</i>
MoaA	XP_015343178.1	<i>Marmota marmota marmota</i>
Viperin	XP_024064957.1	<i>Terrapene mexicana triunguis</i>
Viperin	PKC63257.1	<i>Rhizophagus irregularis</i>
Viperin	KIM76756.1	<i>Piloderma croceum F 1598</i>
Viperin	PNP59997.1	<i>Trichoderma harzianum</i>

Protein	Accession number	Species
Viperin	XP_851276.1	<i>Canis lupus familiaris</i>
Viperin	XP_001510936.1	<i>Ornithorhynchus anatinus</i>
Viperin	KFP16729.1	<i>Egretta garzetta</i>
Viperin	XP_006108914.2	<i>Myotis lucifugus</i>
Viperin	ALT07788.1	<i>Crassostrea gigas</i>
Viperin	NP_542388.2	<i>Homo sapiens</i>

Extended Data Table 4
Primers used in this study.

Primer name	Primer sequence
AB1	CTCCAGCTGGTACCATATGGCGGGCAGGACGC
AB2	AAAAGCGTCAGGTAGGATCCGCTAATCTTATG
AB3	TTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCG
AB4	TATGGCGGGCGTCTGCCCCCATATGGTACCAGCTGGAGAGC
AB53	TGGCTTCTGTTTCTATCAGCTGTCC
AB54	CATCATACTAAATCAGTAAGTTGGCAGCA
AB55	CAACTACTGATTTAGTGTATGATGCAAATGTAGCACCTGAAGTCAGC
AB56	ACAGCTGATAGAAACAGAAGCCAGCCGATCTTCCCAT
AB86	TGGTTAATTCCTCCTGTTAGCCCA
AB87	GAGCGAGGAAGCGGAAGAGC
AB88	CGCATCAGGCGCTCTCCGCTTCTCGTCACTGACTCGTGCCTC
AB89	GTTTTTTGGGCTAACAGGAGGAATTAACCATGTTACGCGGACCGGATGAAACAAAAG
AB119	CTTAAAAAATTACGCCCGCCCT
AB120	GTTATTGGTGCCCTTAAACGCCT
AB121	AGGCGTTAAGGGCACCAATAACCAAGAGTTTGTAGAAACGCAAAAGGCCA
AB122	GGGCGGGCGTAATTTTTTTAAGGAATTCGACTCTCTAGCTTGAGGC
OG628	CATGGTATATCTCCTTATTAAGTTAAAC
OG629	TAATTAACCTAGGCTGCTGC
OG630	TGTTAACTTTAATAAGGAGATATACCATGTCAAAGGAGAAGAAC
OG631	GGCAGCAGCCTAGGTTAATTAATAAAGTTCGTCCATACCGTG
pIBA143_vector_F	TCAGGGAGCGCTTGG
pIBA143_vector_R	CATTGTATATCTCCTTCTTAAAGTTAAACA
pIBA143_Vip6_F	TTAAGAAGGAGATATACAAATGGCTTACAAAGTAACTTAC
pIBA143_Vip6_R	TGGCTCCAAGCGCTCCCTGAACCTGCATAGCGACTTG
pIBA143_Vip8_F	GAAGGAGATATACAAATGCATAATCATAATAAGATTGCTAATAAAG
pIBA143_Vip8_R	TGGCTCCAAGCGCTCCCTGAATGAGTGACGCTCTTTTATAAC
pIBA143_Vip56_F	GAAGGAGATATACAAATGAATATCAAACAATTGTCATCAACTGGC
pIBA143_Vip56_R	TGGCTCCAAGCGCTCCCTGAATCTTTATCTCCAAACGGGCAA
Suf_operon F	ATGGACATGCATTAGGAACCT
Suf_operon R	TTAGCTAAGTGCAGCGGCTT

Primer name	Primer sequence
AB156- _hVIP_mut_F	GCCGGCTTTGCTTTTCATACAGCAAAAACATC
AB157- hVIP_mut_R	TTTGTAGTTAGCCTGTCTTGTAATAATGGTAG
AB158- VIP8_mut_F	GCTGGATATGCTTTTTCGCAAATGGGGTAAG
AB159-VIP8_mut-R	GCGATAGTTGGCTGCTTCGGTAATATGCCA
AB160-VIP9_mut_F	GCCCATTATGCTTATGCCAAGTGGGCAAAG
AB161-VIP9_mut_R	ATCGTAATTAGCCTTTTCGGTGATGTGATAGTTG
AB162- VIP60_mut_F	GCCAAATTTGCTTTTTCGACATTTTATAGATGTC
AB163- VIP60_mut	TCTCATGTTGGCCGGCTGCCAAAGATGAAA

Extended Data Table 5
Phages used in this study.

Phage	Host	Taxonomy	Accession number
SECphi27	<i>E. coli</i>	Siphoviridae	LT961732.1
Lambda_VIR	<i>E. coli</i>	Siphoviridae	NC_001416.1
SECphi6	<i>E. coli</i>	Siphoviridae	GCA_902807315
P1	<i>E. coli</i>	Myoviridae	AF234172.1
SECphi18	<i>E. coli</i>	Siphoviridae	LT960609.1
T7	<i>E. coli</i>	Podoviridae	NC_001604.1
Qbeta	<i>E. coli</i>	Leviviridae	NC_001890.1
MS2	<i>E. coli</i>	Leviviridae	NC_001417

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Morten Danielsen and Daniel Malheiro from MS-omics (Denmark) for conducting the MS experiments and for the extensive help with the data analysis. We also thank the Sorek laboratory members for comments on earlier versions of this manuscript. A.B. is the recipient of a European Molecular Biology Organization (EMBO) Long Term Fellowship (EMBO ALTF 186-2018). A.M. was supported by a fellowship from the Ariane de Rothschild Women Doctoral Program and, in part, by the Israeli Council for Higher Education via the Weizmann Data Science Research Center. G.O. was supported by the Weizmann Sustainability and Energy Research Initiative (SAERI) doctoral fellowship. R.S. was supported, in part, by the Israel Science Foundation (personal grant 1360/16), the European Research Council (grant ERC-CoG 681203), the Ernest and Bonnie Beutler Research Program of Excellence in Genomic Medicine, the Minerva Foundation with funding from the Federal German Ministry for Education and Research, the Ben B. and Joyce E. Eisenberg Foundation, and the Knell Family Center for Microbiology.

References

1. Mattijssen S, Pruijn GJM. Viperin, a key player in the antiviral response. *Microbes Infect.* 2012; 14:419–426. [PubMed: 22182524]

2. Helbig KJ, Beard MR. The role of viperin in the innate antiviral response. *J Mol Biol.* 2014; 426:1210–1219. [PubMed: 24157441]
3. Gizzi AS, et al. A naturally occurring antiviral ribonucleotide encoded by the human genome. *Nature.* 2018; 558:610–614. [PubMed: 29925952]
4. Bernheim A, Sorek R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat Rev Microbiol.* 2020; 18:113–119. [PubMed: 31695182]
5. Ablasser A, et al. CGAS produces a 2′-5′-linked cyclic dinucleotide second messenger that activates STING. *Nature.* 2013; 498:380–384. [PubMed: 23722158]
6. Sun L, Wu J, Du F, Chen X, Chen ZJ. Cyclic GMP-AMP Synthase Is a Cytosolic DNA Sensor That Activates the Type I Interferon Pathway. *Science.* 2013; 339:786–791. [PubMed: 23258413]
7. Cohen D, et al. Cyclic GMP-AMP signalling protects bacteria against viral infection. *Nature.* 2019; 574:691–695. [PubMed: 31533127]
8. Swarts DC, et al. DNA-guided DNA interference by a prokaryotic Argonaute. *Nature.* 2014; 507:258–61. [PubMed: 24531762]
9. Olovnikov I, Chan K, Sachidanandam R, Newman DK, Aravin AA. Bacterial Argonaute Samples the Transcriptome to Identify Foreign DNA. *Mol Cell.* 2013; 51:594–605. [PubMed: 24034694]
10. Joshua-Tor L, Hannon GJ. Ancestral roles of small RNAs: An ago-centric perspective. *Cold Spring Harb Perspect Biol.* 2011; 3:1–11.
11. Akira S, Takeda K. Toll-like receptor signalling. *Nat Rev Immunol.* 2004; 4:499–511. [PubMed: 15229469]
12. Doron S, et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science.* 2018; 359:eaar4120. [PubMed: 29371424]
13. Fenwick MK, Li Y, Cresswell P, Modis Y, Ealick SE. Structural studies of viperin, an antiviral radical SAM enzyme. *Proc Natl Acad Sci.* 2017; 114:6806–6811. [PubMed: 28607080]
14. Fenwick MK, Su D, Dong M, Lin H, Ealick SE. Structural Basis of the Substrate Selectivity of Viperin. *Biochemistry.* 2020; 59:652–662. [PubMed: 31917549]
15. Honarmand Ebrahimi K, Rowbotham JS, McCullagh J, James WS. Mechanism of diol dehydration by a promiscuous radical-SAM enzyme homologue of the antiviral enzyme viperin (RSAD2). *ChemBioChem.* 2020
16. Makarova KS, Wolf YI, Snir S, Koonin EV. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J Bacteriol.* 2011; 193:6039–56. [PubMed: 21908672]
17. Chen IMA, et al. IMG/M v.5.0: An integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* 2019; 47:D666–D677. [PubMed: 30289528]
18. Kambara H, et al. Negative regulation of the interferon response by an interferon-induced long non-coding RNA. *Nucleic Acids Res.* 2014; 42:10668–10681. [PubMed: 25122750]
19. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015; 32:268–274. [PubMed: 25371430]
20. Thiem J, Stangier P. Preparative-Enzymatic Formation of Cytidine 5′-Monophosphosialate by Integrated Cytidine 5′-Triphosphate Regeneration. *Liebigs Ann Chem.* 1990:1101–1105.
21. Dukhovny A, Shlomai A, Sklan EH. The antiviral protein Viperin suppresses T7 promoter dependent RNA synthesis-possible implications for its antiviral activity. *Sci Rep.* 2018; 8
22. Makarova KS, et al. An updated evolutionary classification of CRISPR Cas systems. *Nat Rev Microbiol.* 2015; 13:722–736. [PubMed: 26411297]
23. Tock MR, Dryden DTF. The biology of restriction and anti-restriction. *Curr Opin Microbiol.* 2005; 8:466–472. [PubMed: 15979932]
24. Calendar R, Abedon ST. *The Bacteriophages.* 2005
25. Hendrix RW, Casjens S. Bacteriophage lambda and its genetic neighborhood. *The Bacteriophages.* 2005:409–447.
26. Tran NQ, Rezende LF, Qimron U, Richardson CC, Tabor S. Gene 1.7 of bacteriophage T7 confers sensitivity of phage growth to dideoxythymidine. *Proc Natl Acad Sci U S A.* 2008; 105:9373–9378. [PubMed: 18599435]

27. Kronheim S, et al. A chemical defence against phage infection. *Nature*. 2018; 564:283–286. [PubMed: 30518855]
28. De Clercq E, Neyts J. Antiviral Agents Acting as DNA or RNA Chain Terminators. *Antiviral Strategies. Handbook of Experimental Pharmacology*. 2009
29. Feld JJ, et al. Sofosbuvir and velpatasvir for hcv genotype 1, 2, 4, 5, and 6 infection. *N Engl J Med*. 2015; 373:2599–2607. [PubMed: 26571066]
30. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol*. 2017; 35:1026–1028. [PubMed: 29035372]
31. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT : a novel method for rapid multiple sequence alignment based on fast Fourier transform. 2002; 30:3059–3066.
32. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 2016; 44:W242–W245. [PubMed: 27095192]
33. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol*. 2011; 7
34. Zimmermann L, et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol*. 2018; 430:2237–2243. [PubMed: 29258817]
35. Baba T, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*. 2006; 2
36. Roche B, et al. Iron/sulfur proteins biogenesis in prokaryotes: Formation, regulation and diversity. *Biochim Biophys Acta - Bioenerg*. 2013; 1827:923–937.
37. Schwartw C, et al. IscR, an Fe-S cluster-containing transcription factor, represses expression of Escherichia coli genes encoding Fe-S cluster assembly proteins. *Proc Natl Acad Sci United States Am Acad Sci*. 2001; 98:14895–14900.
38. Jiang D, et al. Identification of Three Interferon-Inducible Cellular Enzymes That Inhibit the Replication of Hepatitis C Virus. *J Virol*. 2008; 82:1665–1678. [PubMed: 18077728]
39. Fortier L-C, Moineau S. Phage Production and Maintenance of Stocks. *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions*. 2009:203–219.
40. Kropinski AM, Mazzocco A, Waddell TE, Lingohr E, Johnson RP. Enumeration of Bacteriophages by Double Agar Overlay Plaque Assay. *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions*. 2009; 69:76.
41. Hsiao JJ, Potter OG, Chu TW, Yin H. Improved LC/MS Methods for the Analysis of Metal-Sensitive Analytes Using Medronic Acid as a Mobile Phase Additive. *Anal Chem*. 2018; 90:9457–9464. [PubMed: 29976062]
42. Dar D, et al. Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science*. 2016; 352

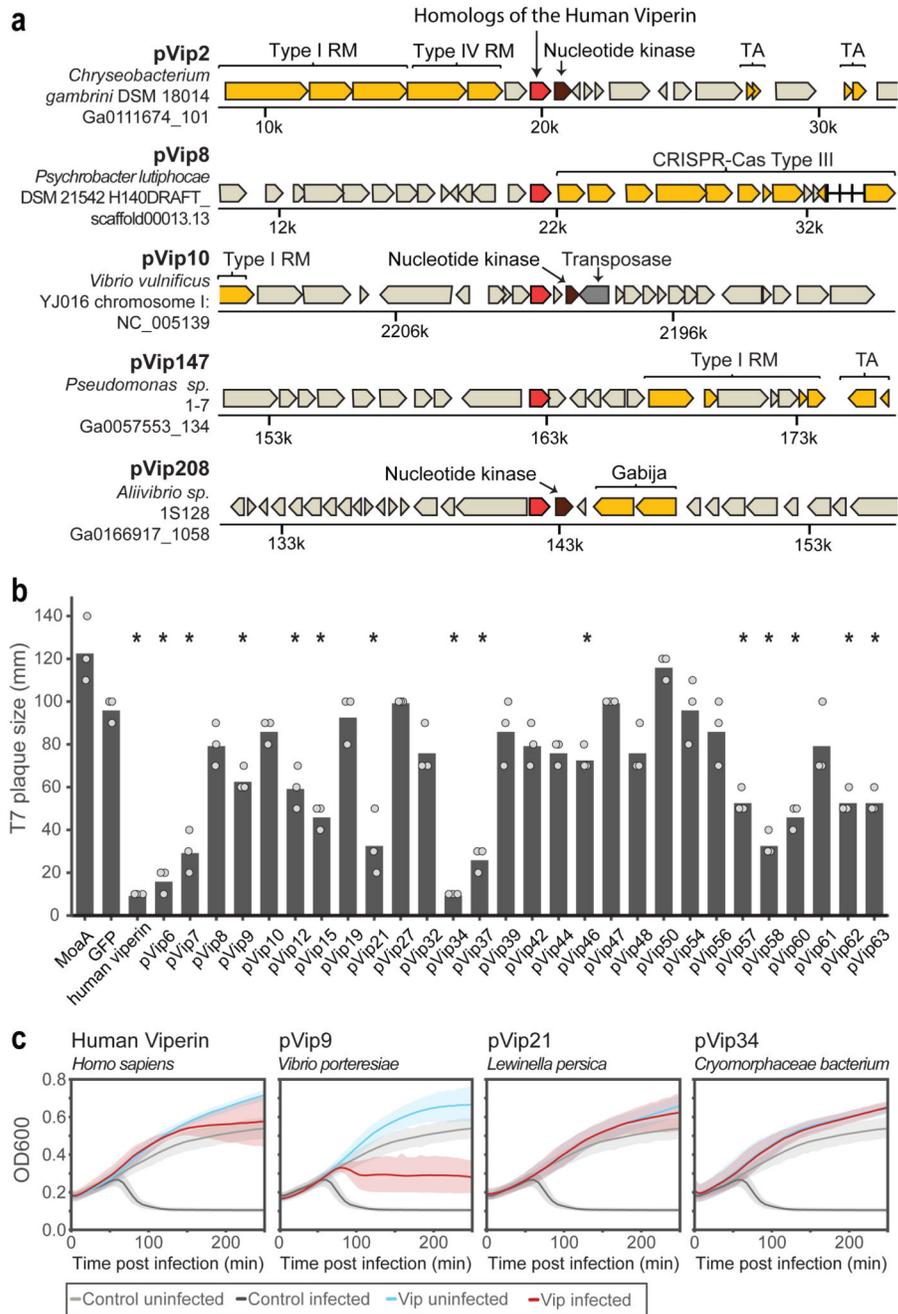


Figure 1. pVips and the human viperin have antiviral activity in bacteria.

a. Representative instances of pVips and their genomic neighborhood. Homologs of the human viperin are in red, genes annotated as nucleotide kinase in brown, genes known to be involved in defense in yellow, and genes of mobile genetic elements in dark grey. RM, restriction-modification; TA, toxin-antitoxin; Gabija is a recently described defense system¹². The name of the bacterial species, and the accession of the relevant genomic scaffold in the IMG database¹⁷ are indicated on the left. b. Plaque sizes of phage T7 infecting *E. coli* strains that express viperins. Bacteria expressing pVips, negative controls

(GFP, MoeA), or the human viperin gene were grown on agar plates and phage lysate was dropped on top of them. Bar graph represents average of three replicates, with individual data points overlaid. Star represents statistically significant difference compared to negative control (GFP) (two tailed t-test, p -value <0.01). c. Growth curves of *E. coli* strains expressing viperins that were infected by phage T7. Light and dark grey are uninfected and infected controls (strain expressing GFP), respectively. Blue and red are uninfected and infected strains expressing viperins, respectively. The negative control (GFP uninfected, GFP infected) is the same in all four graphs. Curve corresponds to the mean of three biological replicates, each with an average of two technical replicates, and the shade corresponds to a confidence interval (CI) of 95%.

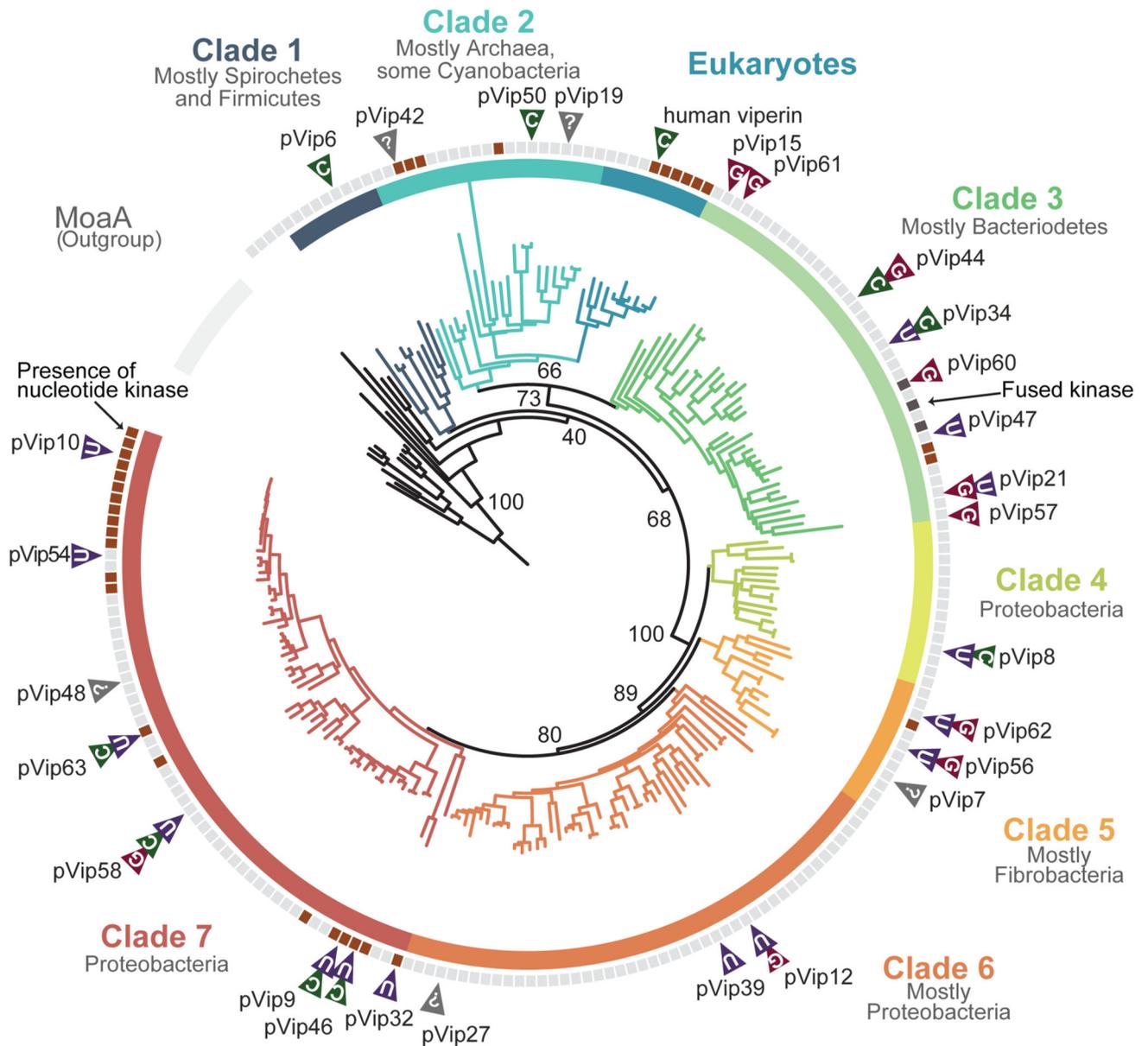


Figure 2. Phylogenetic tree of the viperin family.

Branches are colored according to major clades. Bootstrap values (derived from the ultrafast bootstrap function in the IQtree software¹⁹) are indicated for major nodes. The presence of a nucleotide kinase in the genomic vicinity of the pVip is shown by a brown rectangle in the surrounding ring (or a dark grey rectangle, in case the kinase is fused to the pVip gene).

Triangles correspond to the type of ddh-nucleotide derivatives produced by a specific pVip, as measured by mass spectrometry analysis. The phylogenetic tree was generated using a set of 205 non-redundant pVip sequences.

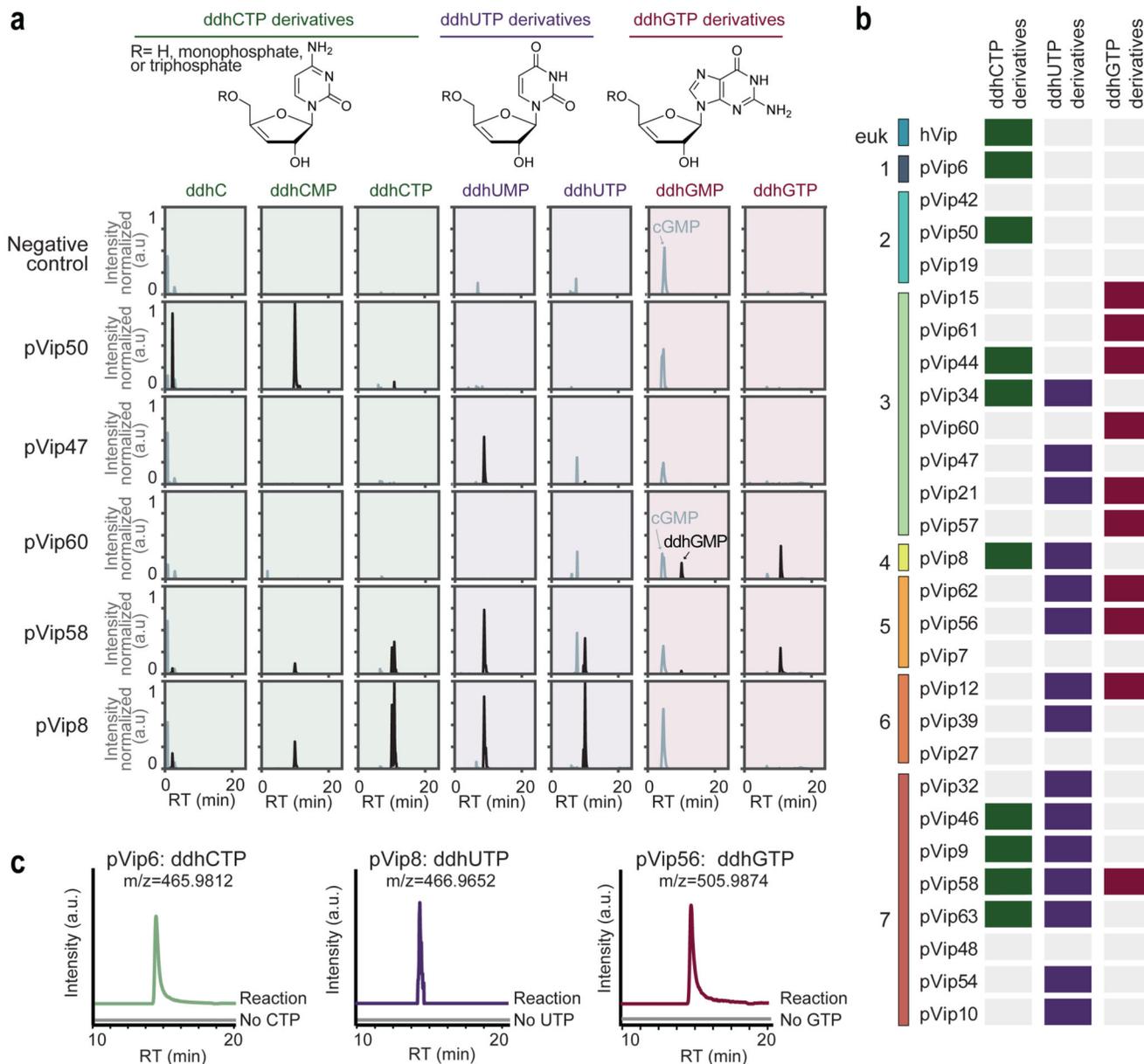


Figure 3. pVips produce a variety of modified ribonucleotides.

a. Extracted ion chromatograms for selected pVip lysates analyzed via LC-MS. Presented are chromatograms of singly charged masses with a precision +/- 5 ppm corresponding to ddhC (m/z 226.08223, retention time (RT) of 2.2 minutes), ddhCMP (m/z 306.04856, RT 9.7), ddhCTP (m/z 465.98122, RT 10.7), ddhUMP (m/z 307.03258, RT 8.7), ddhUTP (m/z 466.96524, RT 9.9), ddhGMP (m/z 346.05471, RT 9.8), and ddhGTP (m/z 505.98737, RT 10.7). X-axis depicts RT in minutes, y-axis depicts normalized ion intensity (A.U, arbitrary units). Normalization was performed on all pVips and MoaA (negative control) samples, with maximal values set to 1.0. In black, peak assigned to ddh nucleotides. In grey, peaks that appear in the negative controls and are not assigned to ddh nucleotides. Representative of three replicates. b. Production of ddh-nucleotide derivatives by pVips. Colored boxes

depict detected compounds. Colored rectangles on the left and associated numbers represent the clade of pVips as described in Figure 2. c. Chromatograms of ddh-nucleotides detected in reaction samples performed *in vitro* with purified pVips. The presence of a product corresponding to ddhCTP, ddhUTP, and ddhGTP is observed in samples where a pVip was incubated with SAM, dithionite, and the respective nucleotide substrate.

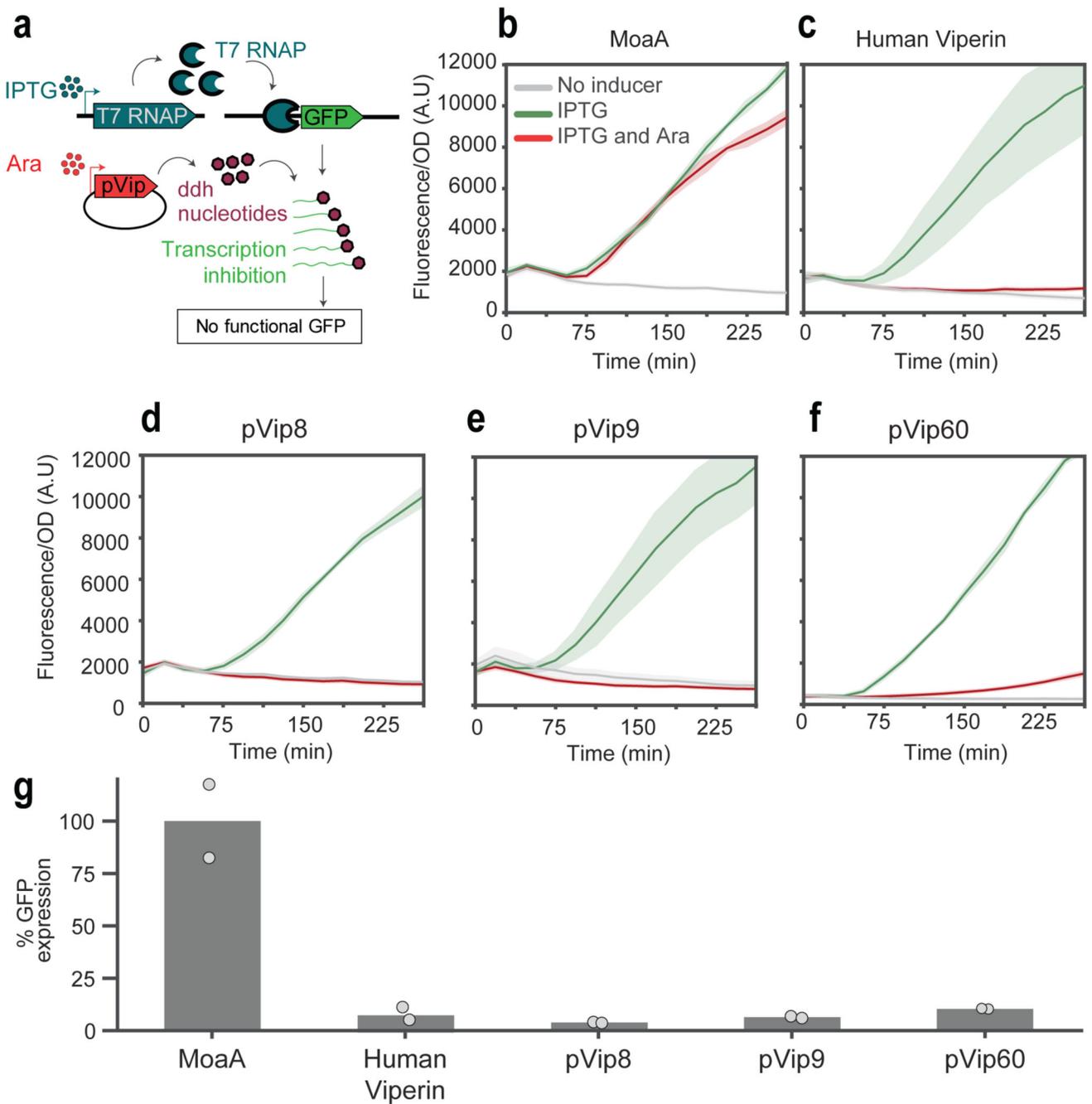


Figure 4. pVips inhibit T7 polymerase-dependent transcription.

a. Schematic representation of the reporter system for T7 polymerase-dependent transcription. *E. coli* BL21-DE3 encodes a chromosomal T7 RNA polymerase (T7 RNAP) under the control of an IPTG-inducible promoter. A reporter plasmid encodes GFP under the control of a T7 promoter. Upon IPTG induction, the T7 RNA polymerase is expressed and drives the expression of GFP. The pVip (or MoaA control) is encoded on a second plasmid under the control of an arabinose promoter. b-f. Application of the reporter assay for strains expressing MoaA (negative control), the human viperin, and pVips. Strains are first induced

with arabinose for 45 minutes to express the pVip. At $t=0$, IPTG is added to express the GFP. Fluorescence/OD over time curves are presented for each strain. Grey lines correspond to no induction (no arabinose, no IPTG), green to IPTG only (GFP expressed, viperin not expressed), and red to induction with both IPTG and arabinose. Curve corresponds to the mean of two technical replicates and the shade to a confidence interval (CI) of 95%. Representative of two biological replicates. b. Strain expressing MoaA (negative control). c. Strain expressing the human viperin. c-f. Strains expressing prokaryotic viperins. g. GFP expression as measured by RNA-seq. GFP expression (RPKM) in cells expressing viperins was compared to that in cells expressing the MoaA negative control. Bar graph represents average of two replicates, with individual data points overlaid.