A host of armor

Prokaryotic immune strategies against mobile genetic elements

Mayo-Muñoz, David; Pinilla-Redondo, Rafael; Birkholz, Nils; Fineran, Peter C.

Published in:
Cell Reports

DOI:
10.1016/j.celrep.2023.112672

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
A host of armor: Prokaryotic immune strategies against mobile genetic elements

David Mayo-Muñoz,1,2,3 Rafael Pinilla-Redondo,1,4,6 Nils Birkholz,1,2,3,5,6 and Peter C. Fineran1,2,3,5,*

1Department of Microbiology and Immunology, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand
2Genetics Otago, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand
3Maurice Wilkins Centre for Molecular Biodiscovery, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand
4Section of Microbiology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark
5Bioprotection Aotearoa, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand
6These authors contributed equally
*Correspondence: peter.fineran@otago.ac.nz
https://doi.org/10.1016/j.celrep.2023.112672

SUMMARY

Prokaryotic adaptation is strongly influenced by the horizontal acquisition of beneficial traits via mobile genetic elements (MGEs), such as viruses/bacteriophages and plasmids. However, MGEs can also impose a fitness cost due to their often parasitic nature and differing evolutionary trajectories. In response, prokaryotes have evolved diverse immune mechanisms against MGEs. Recently, our understanding of the abundance and diversity of prokaryotic immune systems has greatly expanded. These defense systems can degrade the invading genetic material, inhibit genome replication, or trigger abortive infection, leading to population protection. In this review, we highlight these strategies, focusing on the most recent discoveries. The study of prokaryotic defenses not only sheds light on microbial evolution but also uncovers novel enzymatic activities with promising biotechnological applications.

INTRODUCTION

Comparative genomics has demonstrated the dynamic nature of prokaryotic genomes.1 Such plasticity is due to horizontal gene transfer (HGT), which shapes microbial evolution by expanding and redistributing the gene pool for selection.2 The major mobile genetic elements (MGEs) facilitating HGT include viruses and (bacterio)phages, plasmids, transposons, and integrative conjugative elements (ICEs), and genetic material can transfer through transformation, transduction, and conjugation (Figure 1).3 Many MGEs promote their fitness by conferring host adaptive traits, like antibiotic resistance or virulence,4,5 and can enhance genomic rearrangements that promote evolutionary innovation.6

Although MGEs can benefit their hosts, interactions range from mutualistic to parasitic.6 For example, virulent phages are parasitic since they cause bacterial lysis.7 Other MGE-host interactions can be more complex. For example, temperate phages and plasmids can exist stably within bacteria and provide mutualistic traits,8 but there are costs of propagation,9 conjugation,0,11 and selfish maintenance through post-segregational killing.12,13 Finally, MGEs can limit acquisition of other MGEs with potentially beneficial traits14 through phage “superinfection exclusion”15 and other incompatibility or competition mechanisms.16–18

To control MGEs, prokaryotes present a wide variety of defense mechanisms that act at different stages during MGE invasion.19 Surface receptor mutation or modification disrupts phage adsorption and DNA injection and represents a first line of defense but typically incurs a fitness cost.20 Prokaryotes can also regulate receptor expression in response to specific environmental stimuli21,22 or mask receptors by producing capsules, exopolysaccharides, or alginate.23 Alternatively, some bacteria produce proteins that bind receptors and prevent phage attachment24 or outer membrane vesicles containing receptors, which bind phages and reduce productive infections.25 Other mechanisms confer resistance at the multicellular level such as the development of transient resistance or the formation of biofilm and spatial refuges where bacteria can hide from predators.26–29

Prokaryotes have also evolved dedicated defense systems, which are collectively considered the “prokaryotic immune system.”15,30,31 In light of the rapid expansion of this research field, we direct the reader to defense system identification tools that have compiled known defense systems.32,33 Individual cells frequently encode multiple defense systems that are exchanged over short evolutionary timescales.20,34 Most genomes (78%) encode at least two known defense systems (average of six), and these are unevenly distributed across taxa, with some genomes containing dozens and others none.31,33,35–37 Defense systems are often found co-located in genomic regions termed “defense islands,”33,37 although the reasons for their clustering remain unclear. Their frequent association with phages,38–43 phage satellites,43,47–49 ICEs,43,47–49 plasmids,18,50–53 and transposons15,54 suggests that genomic clustering may facilitate
horizontal dissemination of defense systems. Furthermore, defense islands may emerge due to synergy between defense systems and/or a need for coordinated regulation. However, certain defense systems may exhibit antagonistic relationships that prevent their stable co-existence, but incompatibility can be overcome by mechanisms such as epigenetic silencing.

In this review, we provide an overview of the prokaryotic immune landscape with a focus on recently discovered and characterized systems. We classify defense systems according to common modes of action, including the degradation of invading nucleic acids, inhibition of DNA or RNA synthesis, and induction of growth arrest or cell suicide for population-level protection.

DEGRADATION OF FOREIGN NUCLEIC ACIDS

The most abundant and widespread defense systems sense and degrade invading nucleic acids in a site-specific manner.

Restriction-modification systems

A large group of defenses distinguish self from non-self and destroy invading nucleic acids based on epigenetic modifications. These restriction-modification (RM) systems occur in more than 95% of sequenced bacterial and archaeal genomes and are well characterized due to the biotechnological utility of restriction enzymes. RM systems typically comprise two antagonistic components: (1) a methyltransferase that methylates adenines or cytosines within a specific sequence context and (2) a restriction endonuclease that recognizes and cleaves the same motif lacking methylation (Figure 2). Therefore, the host is protected by a distinct methylation pattern, while foreign DNA lacking this pattern is susceptible to endonucleolytic cleavage. There are four types of RM systems (I–IV), based on their subunit composition and biochemical characteristics (i.e., protein structure, restriction site recognition, co-factor requirements, and substrate specificity). Type IV systems cleave modified instead of unmodified DNA.

Several other defense systems contain RM components, such as methyltransferases, which work with other proteins, including proteases, phosphatases, and phospholipases, to provide defense. For instance, the phage growth limitation (Pgl) system methylates phages after a first infection cycle, but the modified progeny are cleaved during subsequent infections. Consequently, although the first infected cell dies, it marks phage progeny for subsequent destruction. A related system, termed bacteriophage exclusion (BREX), distinguishes self from non-self by methylation of a specific DNA site and provides resistance to several unrelated phages upon first exposure. In contrast, the defense island associated with RM (DISARM) system uses DNA topology, rather than specific sequences, to discriminate between self and non-self. Intriguingly, DNA methylation is necessary but not sufficient to avoid DISARM-mediated restriction of invaders by an unknown mechanism.

In addition to methylation, other DNA modifications provide defense. For example, modification of the host chromosome with 7-deazapurine in DNA (DPD) inhibits transformation with non-modified plasmids, although the protective mechanism remains unclear. Other systems involve site-specific phosphorothioate (PT) modifications of the DNA sugar-phosphate backbone in which the non-bridging oxygen is replaced by sulfur. Accordingly, the DNA degradation (Dnd) and the single-stranded DNA PT (Ssp) modification systems use distinct pathways to incorporate sulfur into the DNA and a restriction component to cleave and destroy unmodified invader DNA. Overall, defenses that rely on epigenetic modifications are the most abundant and widespread immune mechanisms in prokaryotes.

Adaptive CRISPR-Cas immune systems

Another sequence-specific means of targeting foreign nucleic acids involves clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated (Cas) proteins. For a more comprehensive overview of CRISPR-Cas biology, refer to recent reviews. CRISPR-Cas systems are one of the most widespread defenses, present in 42% of bacteria and 85% of archaea. They are the only known “adaptive” immune systems for their ability to generate and retain memories of past infections. Due to their programmability and
regions. Finally, the crRNAs assemble with one or more crRNAs consisting of a single spacer flanked by repeat-derived precursor crRNA (pre-crRNA), which is processed into mature crRNA. The crRNAs guide the interference complex to target sequences, indicating a common origin. Conversely, class 2 systems—types II, V, and VI—are mostly restricted to bacteria and archaea and possess an effector complex responsible for the acquisition of foreign DNA, and the other end encodes the interference protein or complex, which is involved in CRISPR-RNA (crRNA) processing, guidance, and cleavage of foreign genetic material. CRISPR-Cas immunity proceeds in three stages: adaptation of CRISPR loci, crRNA biogenesis, and target interference (Figure 2). First, short fragments of invading MGEs are processed and incorporated as spacers at one end of the CRISPR array. Thus, the CRISPR array constitutes a chronological record of past encounters with specific MGEs. Subsequently, the CRISPR array is transcribed as a precursor crRNA (pre-crRNA), which is processed into mature crRNAs consisting of a single spacer flanked by repeat-derived regions. Finally, the crRNAs assemble with one or more Cas proteins into interference complexes with the ability to detect and destroy invading nucleic acids complementary to the spacers.

The evolutionary arms race between prokaryotes and MGEs has led to CRISPR-Cas diversification. While the adaptation module is largely conserved in different CRISPR-Cas systems, interference complexes are diverse. On the basis of distinct architectures and organization of the interference complexes, CRISPR-Cas systems have been divided into class 1 and class 2, with each subdivided into three types and further subtypes. Class 1 systems—types I, III, and IV—are present in both bacteria and archaea and possess an effector complex composed of several Cas proteins. Even though the different types exhibit low sequence similarity, they share structural conservation, indicating a common origin. Conversely, class 2 systems—types II, V, and VI—are mostly restricted to bacteria and are defined by a single multidomain protein as an effector. The different types of systems recognize specific sequences of invaders at the DNA (types I, II, and V) or RNA (types III and VI) level, and upon invader detection, multiple systems initiate collateral RNA or DNA cleavage.

**Long-A prokaryotic Argonautes**

Argonaute proteins (Agos) use small oligonucleotide guides to recognize nucleic acid targets through complementary base pairing (Figure 3). Agos were first discovered in eukaryotes (eAgos) as RNA interference components for regulation of gene expression and anti-MGE defense, but prokaryotic Agos (pAgos) also exist in archaea and bacteria. pAgos are diverse and subdivided on the basis of their domain architecture and into long-A, long-B, and short pAgos. Long-A and long-B pAgos are similar to eAgos and form a bilobed scaffold comprised of the middle (MID), P-element-induced wimpy testis (PIWI), N-terminal (N), and PIWI-Ago-Zwille (PAZ) domains. In contrast, short pAgos (discussed later) contain only MID and PIWI domains.

Long-A pAgos are involved in host defense and use ssRNA or ssDNA guides to target and degrade invading nucleic acids, such as plasmids and viruses. How the guides are acquired from MGEs and loaded onto the pAgos is not completely understood. Upon entry, invader DNA and RNA are degraded into short fragments by pAgo endonuclease activity or other host factors. These fragments are loaded onto pAgos as small interfering nucleic acids (siDNA or siRNA), which guide pAgos to complementary nucleic acids (ssDNA, mRNA, or double-stranded DNA [dsDNA]), resulting in cleavage or blockage of transcription and replication. Because of their nucleic acid-guided cleavage properties, pAgos have potential for programmable gene editing.

**Gabija**

The Gabija defense system consists of the GajA and GajB components and is present in >8.5% of sequenced prokaryotic
mers.107 GajA acts as a site-specific endonuclease that nicks featuring a tetrameric core of GajA subunits flanked by GajB dimer.107 GajA acts as a site-specific endonuclease that nicks dsDNA, presumably triggering an abortive infection (Abi) response through the degradation of phage and host genomic DNA.105 Activation of GajA occurs in response to changes in nucleotide concentrations in the cell, which may result from phage DNA replication and transcription.105

Defenses without site-specific recognition
Some defense systems degrade nucleic acids in a non-sequence-specific manner. For instance, Wadjet senses DNA topology to cleave closed-circular DNA substrates, thus protecting bacteria from acquiring small plasmids.108–110 In contrast, linear and large plasmids (>100 kb), and the chromosome, evade restriction due to their length or topology. A different mechanism has been reported for the nuclease-helicase immunity (Nhi) system, which relies on a multifunctional enzyme that detects and degrades phage-specific DNA replication intermediates.111 Moreover, the RecBCD complex rapidly degrades linear dsDNA, including phage DNA, upon injection into the cell.112 In the Kiwa defense system, KwA detects inhibition of bacterial RNA polymerase by phage proteins and reduces its inhibition of KwB. KwB, in turn, reduces phage DNA replication in a RecBCD-dependent manner.113 In addition, other systems cleave nucleic acids non-specifically upon activation, ultimately inducing toxicity to the cell to halt the infection, as discussed later (see population-level protection section).

INHIBITION OF DNA AND RNA SYNTHESIS
Another group of defense systems protect against invading nucleic acids by interfering with essential processes, such as DNA replication or transcription.

Chemical-based defenses
Some small molecules provide defense by inhibiting phage nucleic acid synthesis. For instance, certain Streptomyces species synthesize anthracyclines (e.g., doxorubicin and daunorubicin), which are secondary metabolites that suppress infection by various dsDNA phages (Figure 3).114 These molecules intercalate DNA and block viral—but not host—DNA replication, although the mechanism for discrimination of self from non-self is unknown. Since these secondary metabolites are secreted, they might provide antiviral protection to microbial communities.115 In addition, aminoglycoside antibiotics produced by Streptomyces block an early step of the phage life cycle, inhibiting viral DNA replication and potentially impacting transcription (Figure 3).116 In an alternative mechanism, prokaryotic viperins (pVips), which are homologs of interferon-induced cellular proteins conserved in animals,117 inhibit phage replication by producing modified ribonucleotides.118 These act as RNA chain-terminator molecules to disrupt transcription and include 3’-deoxy-3’,4’-didehydro (ddh) cytidine triphosphate (ddhCTP), ddh-guanosine triphosphate (ddhGTP), and ddh-uridine triphosphate (ddhUTP) (Figure 3).

Depletion of DNA or RNA nucleotides
Since the replication of MGEs consumes large quantities of nucleotides, the nucleotide pool is a common target for immune systems in both prokaryotes and eukaryotes. For example, the human SAMHD1 is a triphosphohydrolase that depletes deoxynucleotides (dNTPs) to limit the replication of viruses, including retroviruses and herpesviruses.119 Analogously, bacteria also exploit defense enzymes to deplete specific dNTPs. In response to phage infection, deoxycytidine triphosphate (dCTP) deaminases convert dCTP into deoxyuracil nucleotides, and deoxycytidine deaminase (AvcID) depletes free deoxycytidine nucleotides and inhibits phage replication.120 RADAR (restriction by an adenosine deaminase acting on RNA) is a two-protein defense system formed by an adenosine triphosphatase (RdrA) and adenosine deaminase.
(RdrB). These assemble into a supramolecular complex with RdrA adopting a heptameric ring and RdrB a dodecameric cage with the catalytic pockets facing outward. While RdrA rings load and translocate nucleic acid substrates, RdrB cages exhibit potent deamination activity for RNA, ATP, and deoxy-ATP (dATP). The rapid accumulation of these inosine derivatives poisons the nucleotide pool and blocks phage replication. Overall, these depletion mechanisms starve the invading MGE of essential building blocks, halting their replication. Currently, the mechanisms that keep these systems inactivated in the absence of phage infection are unknown, and in the case of RADAR, its activity is toxic to bacterial cells. Therefore, this defense might involve cell suicide for the sake of the population (see below).

**POPULATION-LEVEL PROTECTION**

A common outcome during defense is Abi, whereby cells die upon viral or plasmid invasion, suppressing replication and the epidemic. Consequently, Abi processes are an “altruistic” trait that protects the population at a cost to the individual—in contrast to surface resistance, which advantages the individual but does not eliminate phages and is therefore “selfish.” In contrast to surface resistance and Abi, other intracellular defense systems enable both the survival of the infected individual and protect neighbors by suppressing the epidemic. Historically, defense systems that lead to growth arrest or cell death have been classified as “Abi systems,” despite their unrelated evolutionary origins. Most original research on Abi systems was performed in *Escherichia coli* and *Lactococcus lactis* (e.g., AbiA-AbiZ of lactococci). Recently, there has been a further expansion of systems shown to induce Abi. It is therefore important to highlight that Abi does not denote a defined class of immune systems but rather describes a defense outcome common to many immune systems.

While the Abi response often involves the cooperation of sensor and effector modules, how Abi is achieved varies widely. In cases involving sensor and effector components, the sensor detects an invasion signal or trigger and activates the effector (e.g., nucleases, proteases, phospholipases, toxic transmembrane proteins, NADases, etc.), which then initiates cell death or inhibits metabolism (Figure 4). Triggers include various infection cues, such as intermediates of phage genome replication or phage transcription, phage-mediated silencing of host gene expression, phage proteins or RNAs produced during infection, or even phage inhibition of other defense systems. Once systems are activated, effectors elicit immunity through various mechanisms, including degrading or depolarizing the membrane or indiscriminately degrading MGE and host DNA or RNA. Other effectors target transfer RNAs (tRNAs) or cleave essential host translation machinery. The Abi response is often bactericidal, causing cell death, but can be bacteriostatic, leading to growth arrest—a sometimes blurry distinction. However, an important difference is that bacteriostatic strategies can “buy time” for other defenses, such as RM and CRISPR-Cas systems, to clear the infection, allowing the cell to recover. Although prolonged bacteriostasis may eventually result in cell death, some defense systems appear to have evolved regulatory mechanisms to prevent Abi-induced death once the invader is cleared.

**Signaling-based defenses**

Many diverse defense pathways utilize intracellular signaling to trigger Abi. In these systems, MGE invasion is detected by a sensor protein that generates signals, which bind and activate effector proteins to induce cell death (Figure 4). Cyclic-oligonucleotide-based antiphage signaling systems (CBASS) share ancestry with the cyclic GMP-AMP synthase (cGAS)-stimulator of interferon genes (STING) innate immune response pathway. These systems function through abortive infection (Abi) trigger suicide or growth arrest of infected cells to avoid further viral propagation in the population. The Abi response pathway typically involves the recognition or sensing (inner box, light blue) of an infection cue such as the phage capsid or portal protein (inner box, yellow), which subsequently activates an immune effector (middle box, dark red) through signal molecule binding (CBASS, Pycsar, Thoeris and CRISPR-Cas type III), proteolytic activity and/or oligomerization of defense components (SEFIRs, bGSDMs, non-catalytic pAgos and Avs) or effector release (TA and retrons). The immune effector then carries out a cytotoxic function that induces cell death or dormancy before the virus can complete its replication cycle (outer box), e.g. membrane depolarization or degradation, NAD+ depletion, cellular mRNA or DNA degradation, tRNA sequestration, protein cleavage, etc.
pathway in animals.142,159 In bacterial CBASS, a cGAS/DncV-like nucleotidytransferase (CD-NTase) produces cyclic oligonucleotide signals upon detecting phage infection, including specific structural components (e.g., phage capsid).142,160 These signals activate an effector that promotes Abi through diverse mechanisms, including membrane impairment and DNA degradation (Figure 4).142,143,161–163 CBASS are diverse and are classified into types I–V according to operon composition, signaling molecules, and effector functions.147 CBASS contain diverse effectors and oligonucleotide cyclase enzymes that synthesize various cyclic di- and trinucleotide activators (e.g., cyclic GAMP, cyclic UMP-AMP, cyclic UMP-UMP, and cyclic AMP-AMP-GMPs).164 Similarly, the pyrimidine cyclase system for antiphage resistance (Pycsar) produces cyclic pyrimidine signals (cCMP and cUMP) upon phage infection, which activate effectors to execute Abi through membrane impairment or depletion of cellular nicotinamide adenine dinucleotide (NAD+) (Figure 4).120

A notable parallel to CBASS and Pycsar occurs in type III CRISPR-Cas systems. After crRNA-guided recognition of an invader transcript, the Cas10-Palm cyclase domain of the interference complex synthesizes cyclic oligoadenylate (cOA) messengers that activate separate non-specific accessory nucleases, which arrest growth and halt infection (Figure 4).133,134 Interestingly, the NucC family of endonucleases are effectors in both CBASSs and type III CRISPR-Cas systems and abort infection by degrading the host genome.134,145,165 This demonstrates the modular nature of effector sharing that appears widespread between defense systems.

The Thoeris defense system also utilizes intracellular signaling. In Thoeris, recognition of infection stimulates a Toll/interleukin-1 receptor (TIR) domain-containing protein (ThoB) to generate an isomer of cyclic adenosine 5’-diphosphate-ribose (cADPR) as a signal.166 This molecule binds and activates a second protein, ThsA, that contains a sirtuin (SIR2) domain responsible for depleting the cell of NAD+, thus triggering Abi (Figure 4).35 Importantly, TIR domains are also key components of immune receptors that identify pathogen invasion in plants and animals.167

**Defense-associated sirtuin and SEFIR proteins**

The SIR2 domain is also the key component of minimal, one-protein defense systems called defense-associated sirtuins (DSRs).122 Similar to Thoeris, the function of SIR2 in these systems is NAD+ depletion in infected cells (Figure 4) and, in the case of the DSR2 protein, Abi.168 Interestingly, NAD+ depletion by the protein DSR1 appears to be transient, inhibiting replication of the infecting phage but allowing the population to recover afterward, calling into question whether such metabolite depletion inadvertently leads to Abi.168

Another minimal defense system is represented by proteins with a SEFIR domain (named after the proteins SEF and interleukin-17 receptor [IL-17R]).169 SEFIRs bear structural similarity to TIRs and are another example of a domain originally discovered as part of immune pathways in higher organisms and later shown to be involved in prokaryotic defense.35 Immunity seems to be mediated through interaction between multiple SEFIR proteins and, similar to SIR2, includes NAD+ depletion, which subsequently leads to Abi (Figure 4).35

**Toxin-antitoxin systems**

Phage infection can also be aborted through toxin-antitoxin (TA) modules.170 These systems encode a toxin component that targets essential cellular processes/components and an antitoxin that counteracts the toxicity.171,172 TA systems are classified into eight types (I–VIII) based on the nature of their components and according to the interaction between them.172 Antitoxins are either RNAs or proteins and inhibit toxicity at different levels: translation, activity, or stability of the toxin. In contrast, the majority of toxins are proteins, except for the recently discovered type VIII RNA toxins.142,173

TA systems were originally shown to stabilize and maintain plasmids through post-segregational killing and are also proposed to have other diverse cellular roles in stress.172 Later studies have demonstrated that TA systems can provide defense against viruses, and some through Abi, e.g., type I (Hok/Sok),174 type II (MazEF),175 type III system (ToxIN),151 and type IV (AbiE/AbiEii).176,177 TA systems are proposed to be activated via one of two mechanisms:178 (1) “generalist,” where phage takeover of the host cell results in inhibition of transcription and/or promotes RNA/DNA degradation and leads to turnover of the labile antitoxin,179 or (2) “specific,” where the detection of infection (e.g., via phage hallmark proteins) deactivates the antitoxin.138 In both cases, the toxin is no longer suppressed by the antitoxin and elicits its function, leading to growth arrest or death (Figure 4).132 Generalist TA systems potentially provide defense against more diverse phages due to the lack of specificity required for activation.

**Bacterial gasdermins**

Another potent antiviral strategy shared by bacteria and animals involves gasdermins, whose bacterial homologs defend against phages by eliciting cell death.180,181 Upon phage infection, bacterial gasdermins (bGSDMs) are activated by caspase-like proteases that remove an auto-inhibitory C-terminal peptide. The mature proteins oligomerize into large membrane pores that disrupt membrane integrity and elicit cell death (Figure 4). Considering that eukaryotic GSDMs release proinflammatory cytokines to induce death of the neighboring cells upon lysis (pyroptosis), it has been speculated that bGSDMs also mediate the release of intracellular immune signaling molecules that regulate cell death at the population level.180,182

**Non-catalytic pAgos**

In contrast to long-A pAgos, which cleave target nucleic acids (see earlier), long-B and short pAgos contain a catalytically inactive PIWI domain.99,183 Nevertheless, these pAgos maintain the ability for guide-mediated nucleic acid binding and function with a variety of effector proteins that induce cell death upon target detection.168,184,185 For example, short pAgos associate with effectors that are commonly fusions of different domains involving analog-of-PAZ (APAZ) and either SIR2 or TIR domains. Short prokaryotic Agos associated with SIR2-APAZ or TIR-APAZ effectors constitute SPARSA184,186 and SPARTA184 systems, respectively, and act as heterodimeric Ago-effector complexes. In addition, short Sulfolobus islandicus (Si) Ago systems employ Aga1 with Aga2 effector proteins.150
Upon MGE entry, these systems acquire guide RNAs from invader transcripts to facilitate the recognition of invading DNA, resulting in catalytic activation of the effector domains. In SPARSA and SPARTA, SIR2 or TIR activation leads to NAD(P)+ depletion and cell death (Figure 4). In SiAgos, an Ago-Aga1 complex is recruited to the membrane protein Aga2, a toxic transmembrane effector that forms large oligomers and binds anionic phospholipids via a basic pocket upon activation, resulting in membrane depolarization and cell death (Figure 4). Long-B pAgo (BPA) systems have been subclassified in an analogous fashion, with nuclease (bAgaN), SIR2 (bAgaS), and trans-membrane (bAgaM) effectors involved in BPAN, BPAS, and BPAM systems, respectively. Therefore, these systems appear to elicit similar downstream effects following nucleic acid target recognition as short pAgos and other Abi-inducing systems.

**Defense systems associated with reverse transscriptases: Retrons and UG/Abi-RTs**

Although retrons were originally discovered in the 1980s, their biological function remained enigmatic until recent work revealed a role in phage defense. Retrons are tripartite systems composed of a reverse transcriptase (RT), a non-coding RNA (ncRNA), and accessory effector protein(s). The RT uses the ncRNA to produce a short, multicyclic ssDNA (msDNA) that forms a chimeric RNA-DNA molecule that interacts with the RT itself and the effector. While the mechanism of action is still poorly understood, it is proposed that phage infection disrupts the retron complex, likely through interactions with the msDNA component, triggering the effector (Figure 4). Retrons are widespread across taxa and are diverse in composition, with 13 types (I–XIII) and multiple subtypes. Future work is required to elucidate their immune mechanisms.

Other RT clades provide immunity, including Abi-RTs, CRISPR-Cas-associated RTs, and unknown group (UG)-RTs. CRISPR-Cas-associated RTs are associated with the adaptation modules of RNA-targeting CRISPR-Cas systems, such as types III and VI, and mediate spacer acquisition from RNA (i.e., from mRNAs or RNA viruses). The function of Abi/UG-RTs is less clear, yet they exhibit several parallels with retrons, including (1) the genetic associations of the RTs with different putative effector modules, many of which have predicted defensive or toxic functions, and (2) the identification of co-encoded structured ncRNAs. A number of Abi/UG-RTs have antiphage activity, but their mechanism(s) are unexplored.

**STAND NTPase-based immunity**

Antiviral ATPases/NTPases of the STAND superfamily (AVAST), also known as antiviral STAND (Avs), and NACH domain-containing proteins (named after the eukaryotic proteins NAIP, CIITA, HET-E, and TEP1, which contain this domain) are single multidomain proteins involved in antiviral defense. Avs and NACH proteins share a conserved tripartite organization, composed of (1) a C-terminal sensor domain that recognizes specific viral targets, (2) a core NTPase domain, and (3) an N-terminal effector domain. The effector domains in these proteins are highly diverse and can include DNases, RNases, SIRs, TIRs, and other enzymatic functions previously associated with antiviral defense.

Four Avs families (Avs1–4) were shown to function as specific sensors for conserved structural features in phage proteins, such as the large terminase subunit and phage portal protein (Figure 4). Upon target binding, Avs proteins tetramerize and activate an effector-mediated Abi-like response, which suppresses viral propagation. By recognizing structural protein folds, Avs can detect remarkably divergent homologs of the immune triggers (<5% pairwise amino acid sequence identity) and thus protect the cell against distantly related phages. While the molecular mechanisms underlying NACHT defense have not been resolved in detail, diverse NACHT proteins provide robust protection against dsDNA and ssRNA phages. The immune activity is associated with NACHT-dependent cell growth impairment, suggestive of an Abi-like mechanism. Notably, STAND NTPases belong to an extensive group of immune pattern recognition receptors found in many organisms (including plants, animals, and fungi) that also trigger an inflammatory or cell death response, indicating that recognition of pathogen-specific components is a common trigger across the Tree of Life. Finally, evidence indicates that eukaryotes have horizontally acquired NACHT modules from prokaryotes on multiple occasions, revealing that the exchange of immune components across domains of life may be more common than anticipated.

**Lamassu**

The Lamassu family of defense systems protects prokaryotes from both viral infection and plasmid replication through Abi. They are found in ~10% of all sequenced prokaryotic genomes and involve proteins of the structural maintenance of chromosomes (SMC) family. A split ATPase domain is found in both prokaryotes and eukaryotes as a part of high-order chromosome organization. In addition to the Lamassu-associated SMC gene (termed invuB), these systems encode the LmuA protein, whose N-terminal domain can be replaced with different effectors that execute Abi (Figure 4), including endonuclease, SIR2, hydrolase, protease, and monooxygenase domains. Many of these domains function as effectors in other defense systems, such as the Cap4 endonuclease cleaving both phage and host DNA in CBASSs or SIR2 causing NAD+ depletion in Thoeris, DSR, or SPARSA systems. While the mechanism of system activation is unknown, it was hypothesized that LmuB detects the DNA of the invading element by recognizing replication intermediates, which triggers ATP hydrolysis by the LmuB ATPase to activate the Abi-inducing effector LmuA.

**CONCLUSION**

MGEs aid microbial adaptation, but their selfish nature incurs fitness costs for host cells and has led to numerous defense strategies in prokaryotes. Recently, the field has gained a more complete understanding of the abundance and diversity of these defenses, but it is likely that many more remain to be discovered. Leveraging defense system co-occurrence in defense islands has facilitated the most recent discoveries. In contrast, identifying systems that tend to occur solitarily is less advanced and requires approaches agnostic to a genomic context.
Although many systems have been shown to provide defense against MGEs, especially phages, a complete mechanistic understanding of most is lacking, and entirely novel modes of action are likely to be uncovered. Furthermore, given that the cell-autonomous innate immune systems in higher organisms share ancestry with prokaryotic defense systems, the characterization of new defense systems in prokaryotes can contribute to the identification of defense systems in eukaryotes. The characterization of new defense systems will also inevitably lead to the unraveling of new strategies by MGEs to counteract prokaryotic immunity, as exemplified by the numerous and diverse defense inhibitors discovered in recent years. The resulting arms race is a major force driving the diversification and turnover of defenses and counter-defenses.

There is still limited knowledge of how defense systems are controlled. For some systems, such as CRISPR-Cas, we are gaining a better understanding of their control, which indicates that immune activation must be regulated to manage defense-associated costs. However, regulation of most defense systems is completely unknown. Since prokaryotes contain multiple defense systems, some of which may interact, co-regulation is likely widespread—as observed for different CRISPR-Cas systems. Induction of some Abi systems is also the result of coordinated activation of defenses, which is often interpreted as a protection of last resort when the first lines of defense have failed. Interestingly, certain Abi-inducing systems are proposed to follow a “guarding” strategy, only becoming activated when an invading element inhibits other, first-line defenses. Insights into the triggers that activate defense systems upon infection are also being gained, but more work is required to identify other possible routes of regulation and activation.

Most of the recently characterized defense systems have been studied heterologously in model organisms (e.g., E. coli or Bacillus), and their mechanisms were uncovered through a combination of bioinformatic, microbiological, genetic, biochemical, and structural techniques. This has been a powerful approach to fast track discovery and elucidate the molecular basis of defense. However, the genomics-led discovery of new systems has been biased by pathogen-focused sequencing efforts. Therefore, vast gaps remain in our understanding of the broader role and importance of these defense systems in nature and across microbes. Expanding our knowledge in these areas will require a collective effort on defense system characterization in their natural contexts and in a wider range of microbes, coupled with more ecological studies. Finally, considering the promising applications arising from defense systems (e.g., RM, CRISPR-Cas, retrons, Agos, etc.), it is expected that further defenses will provide useful tools for biotechnology and biomedicine.

ACKNOWLEDGMENTS

We thank Leah M. Smith for useful feedback on the manuscript. Research on phage defense systems in the Fineran laboratory is supported by the Bioprotection Aotearoa and the Marsden Fund, Royal Society of New Zealand (Te Pūtea Rangahau a Marsden, Te Aparangi). D.M.-M was supported by a University of Otago Doctoral Scholarship and a Postgraduate Publishing Bursary. R.P.-R. was supported by a Lundbeck Foundation grant (R347-2020-2346). We apologize to all authors whose work we were unable to cite due to referencing constraints.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES


