



# Identification of Uncharacterized Components of Prokaryotic Immune Systems and Their Diverse Eukaryotic Reformulations

A. Maxwell Burroughs,<sup>a</sup> L. Aravind<sup>a</sup>

<sup>a</sup>Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

**ABSTRACT** Nucleotide-activated effector deployment, prototyped by interferon-dependent immunity, is a common mechanistic theme shared by immune systems of several animals and prokaryotes. Prokaryotic versions include CRISPR-Cas with the CRISPR polymerase domain, their minimal variants, and systems with second messenger oligonucleotide or dinucleotide synthetase (SMODS). Cyclic or linear oligonucleotide signals in these systems help set a threshold for the activation of potentially deleterious downstream effectors in response to invader detection. We establish such a regulatory mechanism to be a more general principle of immune systems, which can also operate independently of such messengers. Using sensitive sequence analysis and comparative genomics, we identify 12 new prokaryotic immune systems, which we unify by this principle of threshold-dependent effector activation. These display regulatory mechanisms paralleling physiological signaling based on 3'-5' cyclic mononucleotides, NAD<sup>+</sup>-derived messengers, two- and one-component signaling that includes histidine kinase-based signaling, and proteolytic activation. Furthermore, these systems allowed the identification of multiple new sensory signal sensory components, such as a tetratricopeptide repeat (TPR) scaffold predicted to recognize NAD<sup>+</sup>-derived signals, unreported versions of the STING domain, prokaryotic YEATS domains, and a predicted nucleotide sensor related to receiver domains. We also identify previously unrecognized invader detection components and effector components, such as prokaryotic versions of the Wnt domain. Finally, we show that there have been multiple acquisitions of unidentified STING domains in eukaryotes, while the TPR scaffold was incorporated into the animal immunity/apoptosis signal-regulating kinase (ASK) signalosome.

**IMPORTANCE** Both prokaryotic and eukaryotic immune systems face the dangers of premature activation of effectors and degradation of self-molecules in the absence of an invader. To mitigate this, they have evolved threshold-setting regulatory mechanisms for the triggering of effectors only upon the detection of a sufficiently strong invader signal. This work defines general templates for such regulation in effector-based immune systems. Using this, we identify several previously uncharacterized prokaryotic immune mechanisms that accomplish the regulation of downstream effector deployment by using nucleotide, NAD<sup>+</sup>-derived, two-component, and one-component signals paralleling physiological homeostasis. This study has also helped identify several previously unknown sensor and effector modules in these systems. Our findings also augment the growing evidence for the emergence of key animal immunity and chromatin regulatory components from prokaryotic progenitors.

**KEYWORDS** ASK signalosome, NAD<sup>+</sup>, SMODS, TPR-S, Wnts, YEATS, biological conflict systems, cyclic nucleotides, histidine kinase, sensor domain

**R**eplicators compete with other replicators for raw materials to produce new copies, energy resources, and space (1, 2). Hence, under natural selection, this leads to the emergence of a variety of genetically specified systems that improve the competitive-

**Citation** Burroughs AM, Aravind L. 2020. Identification of uncharacterized components of prokaryotic immune systems and their diverse eukaryotic reformulations. *J Bacteriol* 202:e00365-20. <https://doi.org/10.1128/JB.00365-20>.

**Editor** Ann M. Stock, Rutgers University-Robert Wood Johnson Medical School

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply. Address correspondence to L. Aravind, aravind@ncbi.nlm.nih.gov.

For a commentary on this article, see <https://doi.org/10.1128/JB.00507-20>.

**Received** 24 June 2020

**Accepted** 25 August 2020

**Accepted manuscript posted online** 31 August 2020

**Published** 19 November 2020

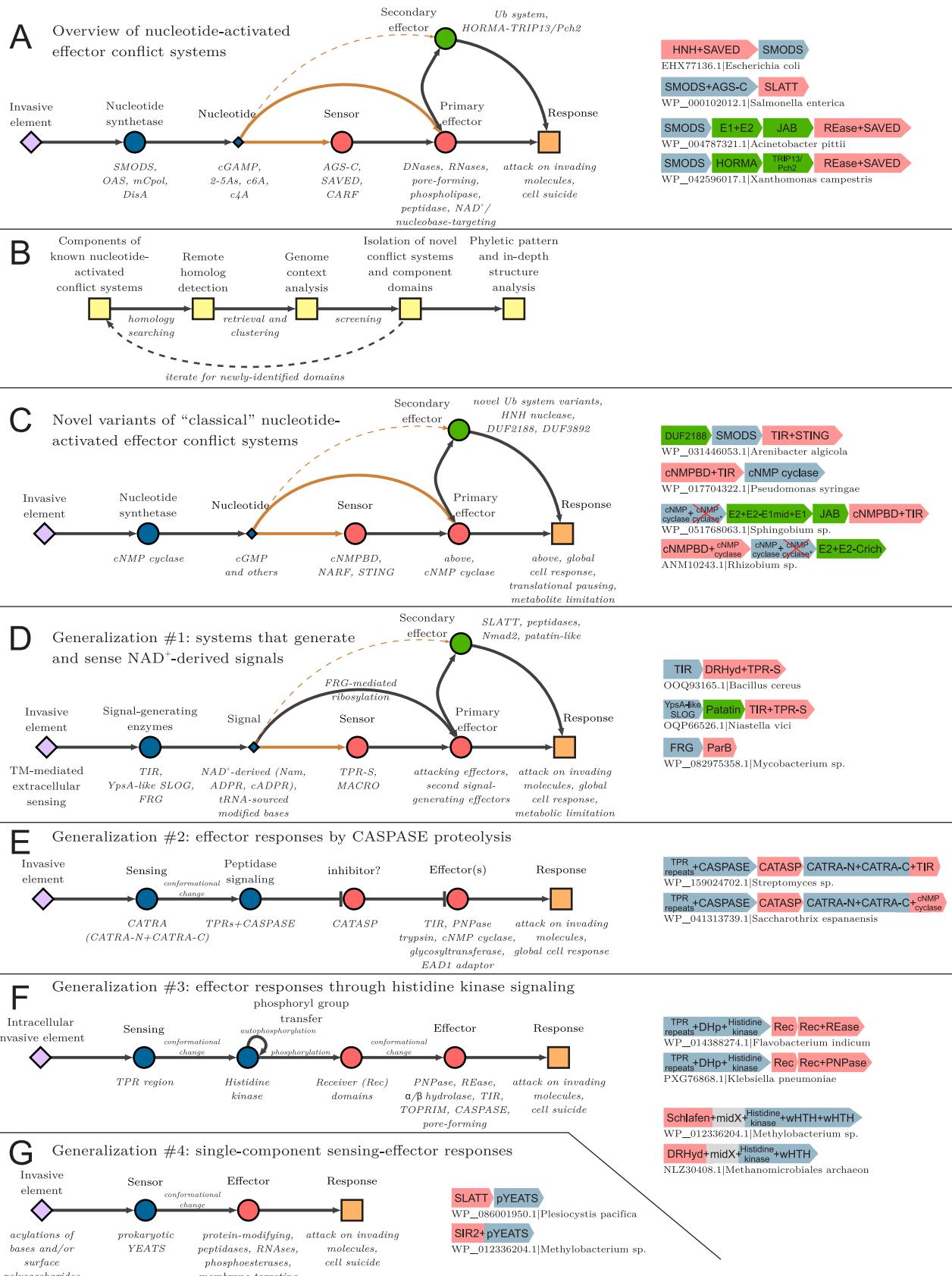
ness of replicators against rival replicators. Among these are systems that are directly deployed in antagonistic interactions with rival replicators and the counterdefenses against such (3–5). These might be collectively termed biological conflict systems and are under intense selective pressures as they have a decisive role in the fitness of replicators (1, 2, 6, 7). At the molecular level, the impact of this selection is seen in the form of an astonishing level of innovation and diversity in the organization, mechanisms, deployment, and regulation of such systems (6, 7).

A large fraction of these antagonistic interactions occur within the same cell, i.e., the conflict between viruses or plasmids and the cellular genome. Despite the great diversity of the systems involved in such conflicts, they can be described by a relatively simple “vocabulary” of constituents (6, 7) (Fig. 1): (i) components that recognize/sense the invasive entity, (ii) components that generate self-nonself markers facilitating the accurate discrimination of self versus invasive molecules, and (iii) effector components that limit the replication of the invasive entity. The first set of these usually functions through the recognition of unique features of invasive molecules (8, 9), while the second set typically tags self-molecules through modifications of nucleic acids, such as adenine and cytosine methylation in DNA (10, 11). Recognition of the invader molecules triggers the third set of players, which are most commonly enzymes targeting nucleic acids (endo-DNases and endo-RNases) or proteins (peptidases) (12, 13). However, these effectors are utilized in two distinct strategies to limit the invasive replicator: either by a direct attack on the invader macromolecules or by an attack on self-systems, such as ribosomes, which stalls invader replication by preventing the synthesis of new proteins or, in extreme cases, apoptosis (14, 15). The latter mechanism works via the principle of inclusive fitness, wherein by self-sacrifice, the organism limits infection of kin and accrues fitness through them (16).

Effector deployment has a cost for diverting limited resources toward defense (16). Furthermore, in the above-stated intracellular conflicts, there is always the potential for the accidental targeting of self-molecules and the untimely unleashing of suicidal effectors in the absence of a serious invader threat, which could nullify the organism's fitness from inadvertent cell death. As a result, the vocabulary of these systems has evolved to include a fourth set of components, namely, regulatory components, which can set a threshold for effector deployment (17, 18), thereby allowing the cell to respond to the strength of the invader threat rather than acting on simple detection. This paradigm first came to light in the 2'-5' oligoadenylate-based nucleotide signal used in the interferon response of jawed vertebrates to regulate the effector RNase L (19, 20).

Subsequent studies by us starting nearly 15 years ago broadened the scope of these nucleotide-regulated effector systems (21) and, along with a range of newer studies, have helped unify a variety of prokaryotic and eukaryotic immune responses such as the type III CRISPR-Cas systems, animal innate immunity pathways, and related bacterial systems that utilize nucleotide signals under a common umbrella (17). Thus, these systems can be recognized as having a unified grammar with two components: (i) a signal-generating enzyme that synthesizes either a linear or cyclic nucleotide (a nucleotidyltransferase) or a nucleotide-derived molecule (e.g., a nucleotide ribohydrolase) as a messenger and (ii) a recognition or “sensor” module that binds this messenger and activates the effector component, which is often directly combined with it in the same polypeptide (16, 17) (Fig. 1A). It has also become apparent that this core is augmented in some cases by a third component that might act as a “force multiplier” for the effector, such as a ubiquitin (Ub)-like protein (Ubl) conjugation system or an ATP-dependent system comprised of a HORMA and TRIP13/Pch2 ATPase dyad (17).

Following the description of these systems, several studies have begun exploring the mechanisms driving nucleotide recognition and thresholding in conflict (5, 22–31). Given the burgeoning interest in these systems and the continual expansion of genome data, we conducted a new search for these systems to better understand them and discover potential novel versions. We started our investigation by detecting novel versions of systems using the previously detailed domain repertoire found in these



**FIG 1** Discovery and generalized mechanisms of systems described in this study. (A) Generalized conceptual diagram of the interactions between core components of previously described nucleotide-activated effector conflict systems. Components are shown as nodes connected by arrows representing (Continued on next page)

systems (17). We then expanded our searches to detect thematically equivalent systems wherein the previously known components have been displaced by functionally equivalent but evolutionarily unrelated components, leading to the identification of novel domains involved in nucleotide generation, recognition, and effector activity. This also helped us identify hitherto unknown systems, which deploy distinct signaling mechanisms to activate their effector repertoire. Moreover, we show that there are several shared principles by which nucleotide-based and other signaling systems have been adapted parallelly for the regulation of conflict systems and cellular homeostatic signaling. Finally, we also identify sensor components that emerged in the context of such prokaryotic conflict systems, which were subsequently “institutionalized” for different chromatin-related and regulatory functions in eukaryotes.

## RESULTS AND DISCUSSION

**Search strategy to identify new conflict systems regulated by nucleotides, nucleotide derivatives, and analogous signals.** We first systematically identified nucleotide-activated effector conflict systems and their analogs by leveraging an extensive, previously described collection of domains found in such systems (17) as seeds for iterative sequence homology searches (Fig. 1B) (see Materials and Methods). We then extracted conserved gene neighborhood information for all newly identified candidate proteins and screened them for features characteristic of the mobility and variability of biological conflict systems: (i) high variability in domain architecture while retaining the same overall architectural “grammar” (7, 18, 32), (ii) rapid sequence divergence within predicted effector domains (this is one of the strongest indicators as it can be objectively measured using statistically significant differences in the mean column-wise Shannon entropy values for alignments of domains from conflict systems compared to their counterparts from nonconflict systems [see Materials and Methods and the supplemental material]), and (iii) extensive lateral transfer and dramatic differences in presence/absence between closely related organisms (13, 18, 33, 34) (Fig. 1B). Gene neighborhoods coding for conflict systems are characterized by frequent displacements of components by functionally equivalent but nonhomologous (analogous) components. Tracking the spread of these analogous components helps expand the horizon of systems; hence, we specifically flagged proteins/domains lacking prior information regarding roles as effectors or as nucleotide synthesis, modification, or binding components for further analysis. These were then used as seeds for further rounds of iterative exhaustive searches (Fig. 1B) (see Materials and Methods). The search for displacements of the signal-generating components allowed us to discover alternative mechanisms by which effectors previously identified in nucleotide-activated conflict systems might be activated.

Below, we describe these findings beginning with those that are closest to the previously known versions (Fig. 1A) and proceeding to the novel versions (Fig. 1C), distinct analogous systems (Fig. 1D), and, finally, examples involving the incorporation of components from such systems into distinct regulatory systems (Fig. 1E to G). Detailed breakdowns of the components and distributions of these systems are provided in Tables S1 and S2 in the supplemental material.

**Novel variants of nucleotide-activated effector conflict systems.** While these systems conform to the previously described basic archetype of nucleotide-derived signal-generating components linked to signal-sensing components, they may also

### FIG 1 Legend (Continued)

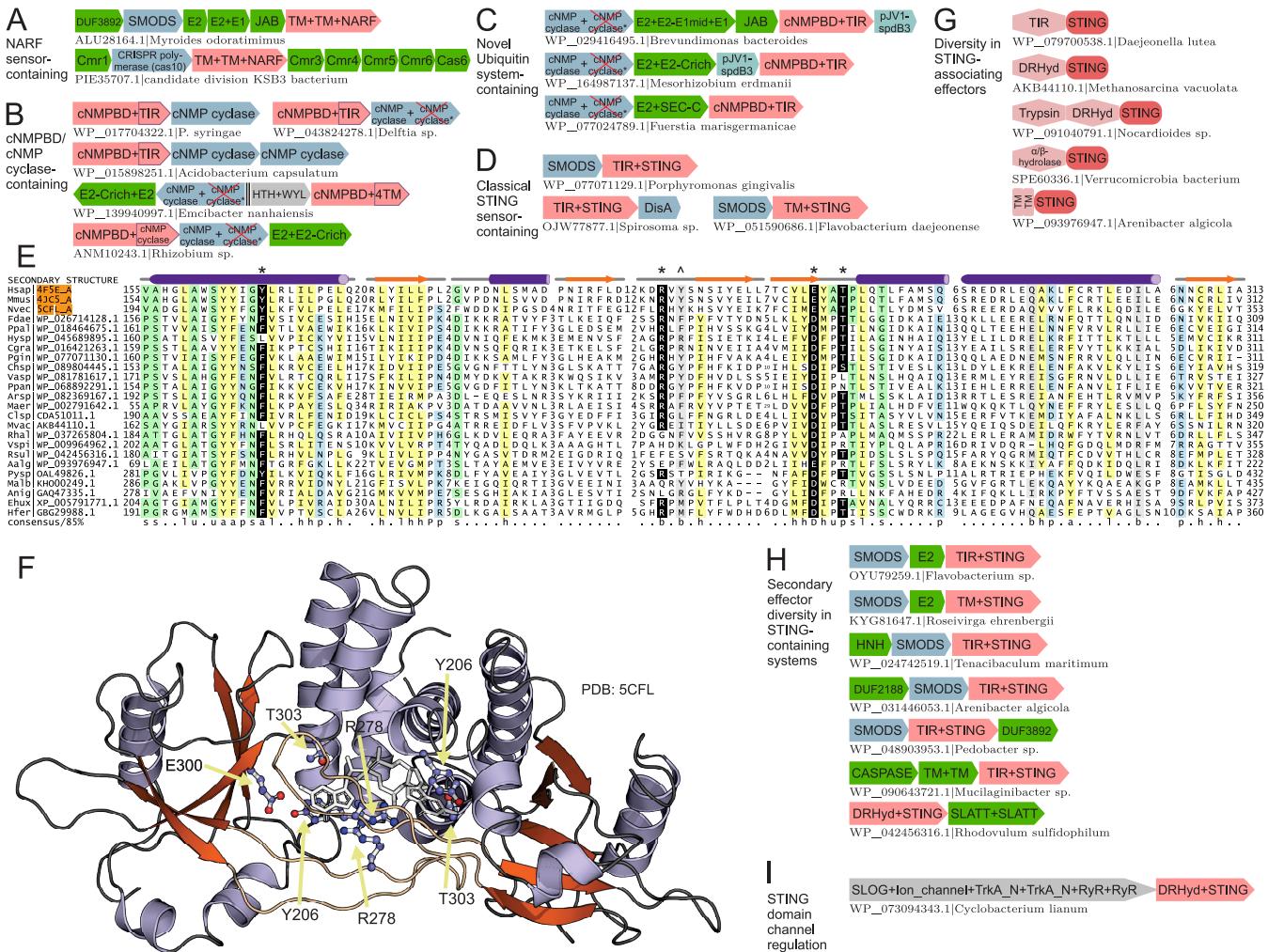
interactions. Orange lines are reserved for interactions mediated by second messenger diffusion. Dotted lines indicate predicted interactions. Examples of components are listed below or to the right of the nodes. Example gene neighborhoods are provided to the right of the diagram. Genes are depicted as boxed arrows, with the arrowhead pointing in the genome 3' direction. Gene colors match the coloring of diagram components. GenBank accession numbers and organism names are provided as labels below each neighborhood. Enzymatic domains predicted to be catalytically inactive are marked by a red “X.” (B) Flowchart showing the process used to identify novel conflict systems. (C to G) Generalized conceptual diagrams depicting novel discoveries in nucleotide-activated effector conflict systems (C), systems dependent on NAD<sup>+</sup>-derived signals (D), systems dependent on caspase-catalyzed proteolysis (E), systems dependent on a core histidine kinase signaling module (F), and single-component sensing and effector activation systems (G). Flat arrowheads depict repressing interactions in panel E. Otherwise, the depictions are as described above for panel A.

feature novel third components. These were predicted to act as “force-multiplying” or regulatory components (17) and subsequently experimentally shown to be required for phage protection in some conflicts (17, 27) or against a specific subset of infecting phages (28). They also share the same general pool of effector domains previously seen in such systems (Fig. 1A and C).

**(i) Cyclic nucleotide-regulated conflict systems: global signaling and distinct Ubl conjugation components.** Searches initiated with the constituent domains of the previously described systems featuring second messenger oligonucleotide or dinucleotide synthetases (SMODS) (see Materials and Methods) helped us discover several types of novel nucleotide-based signaling systems (Fig. 1C). The minority of these new systems, which served as our initial gateway into them, show nucleotide-generating, sensor, and effector components resembling the previously reported versions, in coupling a SMODS with either restriction endonuclease (REase) or HNH endonuclease effector domains fused to the SAVED nucleotide sensor domains or the stand-alone patatin-like lipase effectors (Fig. 1A). A novel sensor containing a Rossmannoid fold, most closely related to the receiver domain (Rec) superfamily, was observed in these systems, which we refer to as the NARF (nucleotidyltransferase-associated Rossmannoid fold) domain (Fig. 2A). Several conserved polar residues are observed in NARF domains at positions that typically house conserved catalytic residues in enzymatically active Rossmannoid domains, suggesting that the NARF domain possesses a nucleotide-binding pocket occupying a position equivalent to the Rec domain active site (see the supplemental material). Outside its association with the SMODS, NARF is observed in certain CRISPR polymerase-containing, so-called type III CRISPR-Cas systems, where they appear to displace the CARF nucleotide sensor/phosphoesterase domain (17, 29, 35) (Fig. 2A and Table S1). This suggests that the NARF domains are likely comparable sensors and also possibly signal-terminating phosphoesterase domains. NARF domains are always fused at the N terminus to two transmembrane (TM) helices. These TMs are positionally equivalent to the previously predicted membrane-perforating toxin SLATT domain (17) and show comparable conserved polar residues embedded in the TM segment (see the supplemental material), which are also seen in other multimerizing, TM channel-forming toxins. Specifically, the NARF parallels the previously identified SUa-2TM sensor (17), a domain distantly related to the SAVED domain which is also fused to a potential N-terminal 2TM membrane-perforating domain. Hence, it would be of interest to test if NARF-mediated nucleotide sensing could trigger a membrane-perforating cell suicide response.

Strikingly, in a subset of the newly identified systems, we observed a divergent version of the classical 3'-5' cyclic nucleotide-generating cyclase (cNMP cyclase) with the nucleic acid polymerase-like palm RRM fold catalytic domain (36) instead of the SMODS nucleotide-generating enzyme. This is coupled with a gene coding for a protein combining a nucleotide sensor domain of the cNMP-binding domain (cNMPBD) superfamily (37, 38) with different domains in the effector position (Fig. 1C and Fig. 2B). This contrasts with the formerly characterized SMODS-containing systems, which have a SAVED or an AGS-C domain as the nucleotide sensor (17). Hence, these new systems are likely to signal via cAMP or cGMP, which are the specific nucleotide ligands recognized by the cNMPBDs (Fig. 2B).

In these systems, the cNMPBD is most commonly fused to a TIR domain in the effector position (Fig. 1C and Fig. 2B), versions of which process NAD<sup>+</sup>. Notably, the TIR domain is sometimes replaced by a 4-transmembrane helix domain that might act as a pore-forming toxin comparable to the SLATT domain (14, 39) (Fig. 2B). In an interesting twist, the effector position may also feature a distinct version of the cNMP cyclase domain (Fig. 2B). This favors a hypothesis wherein the effector activity in some nucleotide-activated conflict systems triggers the production of a further nucleotide-derived messenger (17), with the second likely activating a more global cellular response (see below for further examples) apart from the localized activation of associated effector domains. Consistent with this, cNMP cyclases in the effector position appear most closely related to those found in more conventional (physiological)



**FIG 2** Novel components and examples in nucleotide-activated effector conflict systems. (A to D) Representative conserved gene neighborhoods, as labeled. Gene neighborhood depictions are as described in the legend to Fig. 1A. Conserved components of a neighborhood with no cognate functional node in Fig. 1 are shown in gray. (E) Multiple-sequence alignment of the STING domain. Sequences are labeled to the left by organism abbreviation (see the supplemental material) and GenBank accession number. The secondary structure is shown on the top line, and the consensus abbreviation/coloring on the bottom line are as follows: s, small/green; u, tiny/green; l, aliphatic/yellow; h, hydrophobic/yellow; a, aromatic/yellow; p, polar/blue; b, big/gray. Nucleotide-binding residues are shaded in black and colored in white, marked by an asterisk at the top of the alignment. A caret marks the position of importance in eukaryotic STING but of limited conservation in prokaryotes. (F) Structural rendering of the STING dimer. Conserved residues are rendered in ball-and-stick form and shown relative to the cyclic dinucleotide ligand. (G) Conserved domain architectures for the STING sensor fused to diverse effector domains. Domains in architectures are depicted as distinct shapes, with the organism name and GenBank accession number provided. (H and I) Conserved gene neighborhoods as labeled and described in the legend to Fig. 1A.

signaling systems (40, 41). These systems might additionally feature another copy of the cNMP cyclase domain, presumably with a regulatory role, either inactive and fused directly to the active version or encoded by a stand-alone gene (Fig. 2B and Table S1).

Previously, we described a set of SMODS-containing cyclic or linear nucleotide-activated effector conflict systems coupled alternatively to Ubl conjugation systems or HORMA-Trip13/Pch2 systems as third components (17, 27). The former versions of these systems display a direct fusion of the E1- and E2-like Ubl ligases in a single polypeptide and a JAB deubiquitinating peptidase in a separate polypeptide (Fig. 1A) (17). Notably, three of the types of these newly recovered systems also feature the incorporation of Ubl conjugation systems (21, 42) (Fig. 2C and Table S1). The first variant is characterized by an additional domain of unknown provenance inserted between the E1 and E2 ligase domains (Fig. 2C; see also the supplemental material). The second and third variants lack an E1 ligase, and their E2 ligase is respectively fused to a conserved, C-terminal cysteine-rich domain and the metal-binding C-terminal SEC-C motif, with

the latter restricted to planctomycetes (43, 44) (Fig. 2C and Table S1). Accretion of such additional elements that potentially act as force multipliers is a frequent feature in conflict systems supported by recent findings (27, 35).

Finally, across these newly recovered systems, we found a sporadic association with the pJV1-spdB3 TM domain (Fig. 2C and Table S1). These domains were previously implicated in plasmid DNA transfer among cells in multicellular bacterial communities (45–47), suggesting that this protein might act as an additional membrane-associated sensor that might help the activation of conflict systems via the direct detection of extrinsic DNA.

**(ii) Diversity of prokaryotic STING sensor domain-containing systems.** In parallel to a recent report (28), our searches identified a novel domain in the sensor position (Fig. 1C and Fig. 2D) of systems containing SMODS in the position of the nucleotide-generating enzyme. These are the long-sought prokaryotic cognate of the eukaryotic STING nucleotide receptor domain. While predominantly found in bacteriodetes, they also show a sporadic presence in lineages such as proteobacteria, firmicutes, verrucomicrobia, and euryarchaea (Table S1). Notably, the residues required for interaction with the cyclic GMP-AMP (cGAMP) ligand (48–50), which triggers animal innate immune response pathways, including the induction of type I interferon transcription (51–53), are almost uniformly conserved in the prokaryotic STING domains (Fig. 2E). These include the crucial tyrosine residue that mediates direct  $\pi$ - $\pi$  stacking interactions between the STING sensor and cGAMP and the entire complement of polar residues interacting with the nucleobase and the phosphate backbone (Fig. 2E and F) (48–50). Notably, we report for the first time a displacement of the SMODS domain with the structurally unrelated DisA-like cyclic di-AMP nucleotide synthase in some of the STING domain-encoding neighborhoods (Fig. 2D), suggesting a degree of flexibility in the dinucleotides recognized by the prokaryotic STING domains.

A previous report of prokaryotic STING describes only a single domain in the directly fused effector position: the TIR domain (54) (Fig. 1C and Fig. 2D). However, here, we report an extensive collection of effector domains fused to the STING sensor, comparable to the variability that is seen for the effectors fused to the SAVED and cNMPBD sensor domains (17) (Fig. 2G). Beyond the TIR domain, we observed fusions to the deoxyribohydrolase (DRHyd),  $\alpha/\beta$ -hydrolase, and trypsin peptidase domains. We also observed fusions to one or more N-terminal TM helices comparable to eukaryotic STING domains, which are fused to equivalently positioned TM helices (Fig. 2G). In prokaryotes, such TM fusions are occasionally found in neighborhoods with SMODS and also as stand-alone genes without any proximal synthetase gene (see the evolutionary discussion below) (Fig. 2D and Table S1).

We further note accretion of several distinct, additional components in these systems, including a stand-alone E2-like Ub ligase gene, which is located between those of the SMODS and the STING effector protein; an HNH endonuclease; small, mixed  $\alpha$ - and  $\beta$ -domains annotated as DUF2188 and DUF3982 (see below); caspase peptidases; and the previously predicted pore-forming toxin domain, SLATT (17) (Fig. 2H). These domains might again function as augmenters that back up the core STING-linked effector system. A sporadic gene neighborhood association was also observed with proteins combining ion channel domains with various known and predicted small-molecule ligand-binding domains, including RyR (55), TrkA-N (56), and the SLOG superfamily (17) domains (Fig. 2I). The latter associations could point to the regulation of these prokaryotic channels by cyclic dinucleotides sensed by the STING domain possibly in a role independent of biological conflicts.

Beyond these diverse prokaryotic systems, we also identified the first examples of the STING domain in distinct eukaryotic lineages outside the choanoflagellate-animal clade, such as ascomycete fungi, haptophytes, and stramenopiles (see below).

#### **Conflict systems that potentially generate and sense NAD<sup>+</sup>-derived signals.**

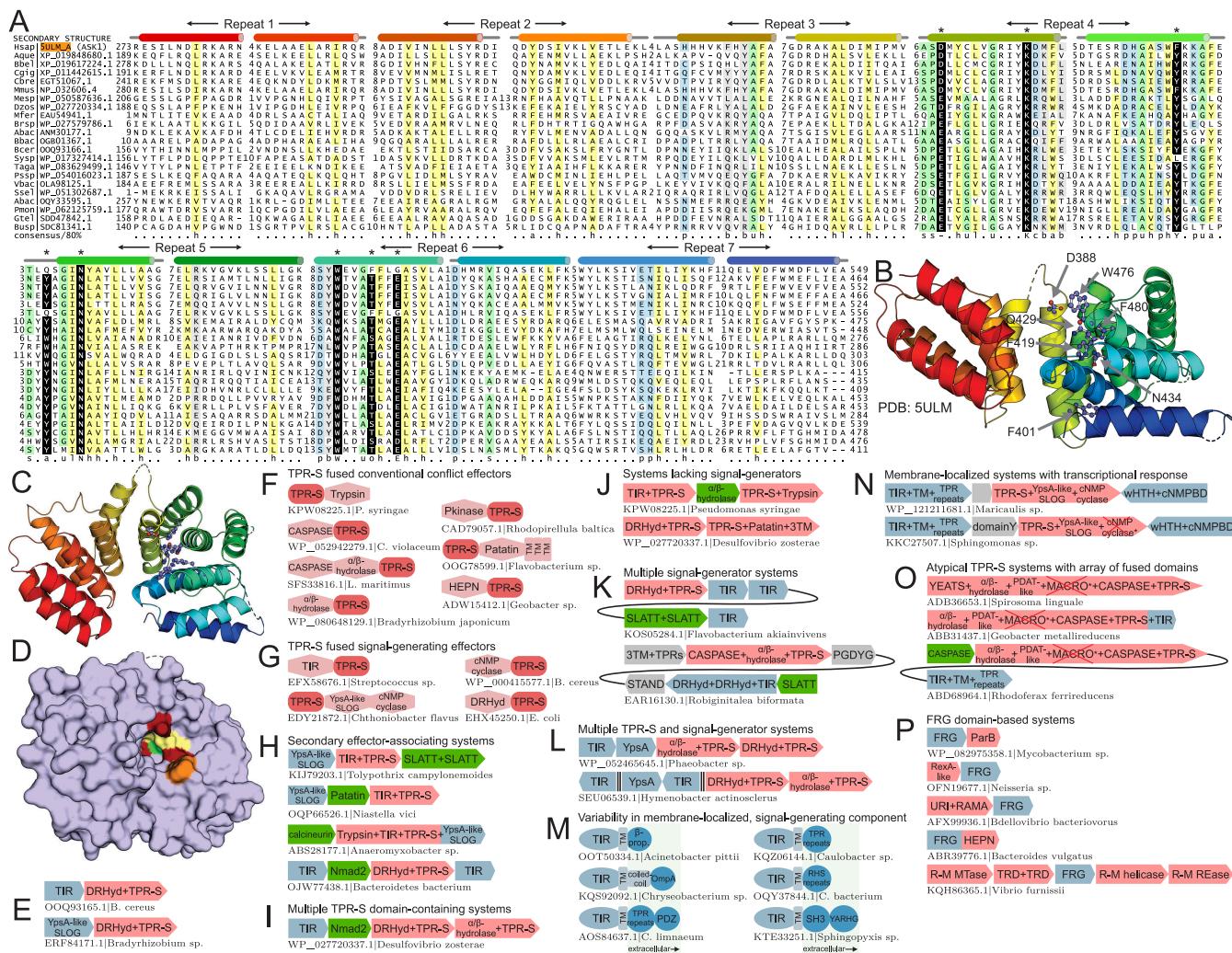
The above-described and previous analyses helped identify at least 5 distinct effector-fused nucleotide sensor domains, CARF, NARF, SAVED, cNMPBD, and STING, which are operonically linked to one or more signal nucleotide-generating synthetases of the

CRISPR polymerase, SMODS, DisA, and cNMP cyclase families (17) that are likely to sense the invasive entity indirectly or directly, as previously hypothesized (17) and now supported by experimental findings (27). This provided us with a consistent syntax shared by such systems (Fig. 1A and C) that could be used to identify novel versions wherein the conventional synthetase or any of these previously identified nucleotide sensor domains are replaced by other distinct functional equivalents (Fig. 1B) (see Materials and Methods). With regard to distinct functional replacements for synthetases, we had predicted previously that certain signals in conflict systems are likely derived from NAD (NAD<sup>+</sup>) or modified nucleobases in tRNA (57–59). Given the accumulating evidence in support of our previous prediction of TIR and SIR2 domains catalyzing the production of NAD<sup>+</sup>-derived signals (30, 57, 59, 60), we investigated if there are novel conflict systems that combined such signal-generating enzymes with novel sensors for them. We describe below such systems that we recovered in our screen.

**(i) The TPR-S domain, a novel sensor domain in conflict systems.** A diverse collection of systems contains a conserved region with tetratricopeptide repeats (TPRs) fused to a range of effector domains shared with other nucleotide-activated conflict systems (Fig. 1D and Fig. 3) (17, 60). Careful examination of this TPR-containing region led us to define a discrete module, which always contains a total of seven TPR repeats ( $\alpha$ -helical hairpin units) and several highly conserved aromatic and polar residues clustering in repeats 4 to 6 (Fig. 3A and B). These observations are reminiscent of other conserved modules formed from repeats under selective pressure to retain specific catalytic or ligand-binding functions, e.g., diverse enzymatic  $\beta$ -propeller units (61–63), the NIC domain (64), and the HEAT repeat deoxyhypusine hydroxylase (65). Remarkably, we also observed versions of this TPR module acquired via horizontal gene transfer (HGT) in the animal-choanoflagellate lineage (see the section on eukaryotic acquisitions in the evolutionary discussion below). A recent structure of this metazoan TPR module was published (66), revealing a tightly wound solenoid structure harboring a deep central pocket (Fig. 3C and D). By mapping the conserved residues onto this structure, we noted that they line the pocket, with one of them, an asparagine, at the very base of the pocket and an aromatic residue (mostly tryptophan) at the mouth of the pocket (Fig. 3D). These observations together with the fusions of this TPR module to diverse effector domains observed in other nucleotide-activated conflict systems are consistent with it binding a small-molecule ligand. Accordingly, we named this TPR module the TPR-S (TPR sensor) domain.

**(ii) Organization of simple systems with a TPR-S module.** When we investigated the neighborhoods of the genes coding for TPR-S domains fused to effectors, we found them to be almost invariably coupled in an operon with genes coding for TIR or the YpsA family of SLOG domains (60) (Fig. 3E). Catalytically active TIR domains generate NAD<sup>+</sup>-derived second messengers like nicotinamide (Nam), ADP ribose (ADPr), and cyclic ADPr (cADPr) (57–59), whereas certain SLOG domains catalyze comparable hydrolysis of modified nucleobase-ribose bonds (67, 68), potentially sourced from tRNA (69, 70), to generate signals (e.g., cytokinins in plants and certain bacteria). Since these enzymes take the place of the conventional nucleotide synthetase of nucleotide-activated effector conflict systems (Fig. 1A) (17) in the organizational syntax of these systems, we predict that they are signal-generating enzymes (Fig. 1D). They likely sense the invasive entity by analogy to conventional nucleotide synthetases such as 2'-5'-oligoadenylylate synthetase (OAS) or SMODS (17, 71). However, given the domains that they feature, we propose that they instead produce NAD<sup>+</sup> derivatives or cytokinin-like messengers, which are then sensed by the TPR-S domain to activate the associated effector domain.

The simplest versions of the TPR-S systems typically feature a gene coding for the signal-generating enzyme operonically linked to another gene encoding a polypeptide that combines a variable effector to the constant TPR-S domain (Fig. 1D and Fig. 3E). The effector domains are shared with conventional nucleotide-activated and other biological conflict systems, e.g., peptidases of the trypsin and caspase superfamilies,



**FIG 3** Identification and characterization of the TPR-S domain and its NAD<sup>+</sup>-derived nucleotide-activated effector conflict systems. (A) Multiple-sequence alignment of the TPR-S domain, as described in the legend to Fig. 2E. Conserved residues with a predicted role in substrate recognition are denoted with asterisks above the alignment and shaded in black.  $\alpha$ -Helical hairpin TPR repeats are labeled above the alignment. (B and C) Structural rendering of the TPR-S domain, in side view (B) and top view (C). Individual helices are color-coordinated with the secondary structure in panel A. (D) Surface rendering of the TPR-S domain, in top view as shown in panel C. Surfaces of absolutely conserved tryptophan and asparagine residues are shown in orange and green, respectively. Other conserved polar and aromatic residues are shown in red and yellow, respectively. (E to P) Representative conserved gene neighborhoods and domain architectures of NAD<sup>+</sup>-derived nucleotide-activated conflict systems, as labeled. Depictions are as described in the legends to Fig. 1A and Fig. 2G. C. violaceum, *Chromobacterium violaceum*; L. maritimus, *Lutibacter maritimus*; C. bacterium, “*Candidatus Cloacimonetes bacterium*”; C. limnaeum, *Chlorobaculum limnaeum*.

protein kinases, potential pore-forming TM domains, lipid-targeting  $\alpha/\beta$ -hydrolases, and HEPN RNases (17, 35, 72) (Fig. 3F and Table S1). Such effectors might be termed “attacking effectors” as they typically directly degrade self or nonself target macromolecules to cripple invader proliferation. In contrast to these, a variant theme in the TPR-S systems features a second signal-generating enzymatic domain in the effector position, namely, versions of the trio of related Rossmannoid fold domains, TIR, DRHyd, and SLOG (17), or cNMP cyclases (Fig. 3G). Especially in the case of the last of these effectors, as in the above-described nucleotide-activated systems (Fig. 1C and Fig. 2C), they might help propagate more global signals within the cell by generating a second messenger molecule. Another possible functional role for the first three of these effectors, supported by certain previous studies (13, 73, 74), could involve metabolic limitation or cell death through the degradation of NAD<sup>+</sup> or a related metabolite.

Paralleling the nucleotide-activated conflict systems, two-gene systems are sometimes augmented by a third gene coding for a protein with a stand-alone effector domain. These include membrane-attacking effectors such as SLATT, which is a potential membrane

perforator, and patatin-like phospholipases (14, 17, 75, 76) or calcineurin-like phosphoesterases, which might help modulate or terminate the secondary nucleotide signal, stand-alone caspase peptidases, and the Nmad2 effector found in certain DNA-modifying restriction-modification (R-M)-like systems (9) (Fig. 3H). These could be directly activated by the nucleotide-derived signal and might function as either a second line of defense or a backup that effects cell suicide upon the failure of the primary signal-activated system (16, 17).

A subset of systems (Table S1) contains multiple genes coding for distinct TPR-S-effector-fused proteins (Fig. 3I). Infrequently, two TPR-S-effector-fused genes also colocalize on the genome with no proximal candidate genes for signal-generating enzymes (Fig. 3J). These organizations suggest that multiple alternative effectors could be activated by the same signal generated by an enzyme encoded either in the proximity or elsewhere in the genome. We also observe the presence of multiple signal-generating enzymes in a single system that otherwise resembles the usual organization of the TPR-S systems (Fig. 3K) or clustered groups of multiple signal generator and TPR-S-effector-fused pairs (Fig. 3L). On one hand, the multiplicity of effectors could point to multiple parallel lines of attack being opened against the invasive entities. On the other, the presence of systems with multiple signal-generating enzymes might imply that they are responsive to different invasive entities or a chained relay of signals through the multiple signal generators.

**(iii) TPR-S domain systems with variable invader-sensing domains fused to the signal-generating component.** A specialized subclass of TPR-S-based systems is defined by a variable N-terminal region connected to the signal-producing enzymatic domain via a TM helix (Fig. 3M and N). This variable region features domains predicted to localize to the extracellular periplasmic region that are either superstructure-forming repeats such as the  $\beta$ -propeller, RHS, or TPRs (distinct and not closely related to TPR-S) or potential peptidoglycan-binding domains like OmpA, YARHG, and SH3 (Fig. 3M) (77, 78). Thus, this variant class of TPR-S systems is predicted to function at the cell surface, with the above-mentioned extracellular domains likely sensing the presence of the invasive entity via either direct binding or recognition of disrupted peptidoglycan, e.g., by lysozymes used by phages during invasion (79–81). The intracellular signal-generating domain, usually a TIR domain, is then predicted to produce the signal proximal to the inner face of the membrane.

In terms of their effectors, those found in these variant systems overlap with those from the above-described basic systems (Table S1). However, one distinct theme, most prevalent in alphaproteobacteria, features a pair of domains in the effector position, a YpsA-like SLOG and a cNMP cyclase domain, and a further gene coding for a Crp-like transcription factor that combines a cNMPBD with a winged helix-turn-helix (wHTH) domain (82) (Fig. 3N). On some occasions, the cNMP cyclase domain is predicted to be inactive, and the operon might additionally feature a previously uncharacterized gene that we predict to potentially encode a novel nucleotide synthetase (Fig. 3N; see also the supplemental material). Both the SLOG and cNMP cyclase domains of these systems are likely activated through the TPR-S-dependent sensing of a primary NAD<sup>+</sup>-derived signal generated by the TIR domain. Thus, paralleling the systems mentioned above (Fig. 2B), we propose that they then generate further, more global secondary messengers like cNMP, which might activate a transcription response via the associated Crp-like transcription factor (Fig. 3N).

**(iv) Atypical TPR-S systems with fusions to tandem arrays of domains.** An interesting variant of the TPR-S systems, observed primarily in proteobacteria and bacteroidetes, combines a signal-generating component in the form of a TIR domain with a second component featuring several N-terminal effector domains fused to the TPR-S domain. From the N terminus to the TPR-S domain, these usually include two  $\alpha/\beta$ -hydrolase domains, an inactive Macro domain, and caspase domains (Fig. 3O). The first  $\alpha/\beta$ -hydrolase domain is a distinct version of the superfamily with general affinities for families of  $\alpha/\beta$ -hydrolases showing lipase activity. The second  $\alpha/\beta$ -hydrolase domain is specifically related to the eukaryotic phospholipid:diacylglycerol acyltransferase

(PDAT) enzymes to the exclusion of all other families. In eukaryotes, PDAT catalyzes the production of triacylglycerols (TAGs) (83, 84), and, to our knowledge, this represents the first report of PDAT orthologs in bacteria. These TPR-S-fused versions are a subset of the total complement of bacterial PDATs, which are otherwise encoded as stand-alone genes in bacteria (our unpublished observations). While some bacteria have known TAG biosynthetic pathways, including the well-studied Kennedy pathway (85, 86), an alternative TAG synthesis pathway centered on a PDAT-like enzyme has long been postulated based on biochemical observations but has so far evaded characterization (85, 87). Further investigation of these bacterial representatives of the PDAT-like family might help resolve this long-standing mystery. However, in the context of these TPR-S systems, we propose that the  $\alpha/\beta$ -hydrolase domains might operate in hydrolyzing and altering the composition of membrane lipids.

The third domain in these polypeptides, the inactive Macro domain (Fig. 3O), potentially functions as a secondary sensor domain, as versions of this superfamily specifically bind ADPr, cADPr, and their derivatives generated from NAD<sup>+</sup> (57, 58, 88). Hence, its presence here strengthens the proposal that the TIR signal-generating enzymes function in producing NAD<sup>+</sup>-derived signals. The caspase domain, a further version of which occasionally occurs as a stand-alone gene in these neighborhoods, could function either in cleaving and releasing the other fused effector domains or as an “attacking” peptidase in its own right. Of note, variants of this system display the first detected prokaryotic version of a YEATS domain (89) fused to the extreme N terminus (Fig. 3O) (see below).

**(v) FRG domain-based systems.** The FRG domain (Pfam identifier PF08867) was previously recognized as a divergent member of the ADP-ribosyltransferase (ART) superfamily of enzymes (90). We observed that FRG domains often occur as the constant domain of proteins where they are combined with a range of other domains, several of which are previously described effectors in biological conflict systems (Fig. 3P). These include the RexA-like abortive infection system activator (10), an endo-DNase combining the modified DNA-binding RAMA domain with a URI endonuclease domain, and the HEPN RNase domains (91) (Fig. 3P). In certain cases, stand-alone FRG domains are integrated into conventional R-M systems (Fig. 3P). Members of the ART superfamily utilize NAD<sup>+</sup> either as NADases that degrade NAD<sup>+</sup> to release ADPr or other derivatives or as transferases that link ADPr to substrate molecules such as proteins and nucleic acids (57, 92). Indeed, other members of the ART superfamily proteins from toxin-antitoxin (T-A) systems are known to modify target proteins/nucleic acids and are also incorporated into R-M systems where they are predicted to function as accessory effector enzymes that might back up the core restriction components (74). Accordingly, we propose that these FRG domain proteins generate an NAD<sup>+</sup>-derived signal upon the direct sensing of invader molecules and activate the associated effector domains through either protein ADP-ribosylation or the release of a soluble ADPr derivative (Fig. 1D).

Thus, the above-described systems extend the examples of distinct sensor domains combined with effector domains in the same polypeptide beyond cyclic (oligo)nucleotides and (oligo)nucleotides to encompass potential NAD<sup>+</sup>-derived signals (Fig. 1A and D).

**Conflict systems utilizing diverse activating inputs other than soluble messengers.** An even more abstract version of the above-defined syntax can be conceived wherein an operon couples a gene coding for a polypeptide combining a signal (not necessarily nucleotide- or NAD<sup>+</sup>-derived)-sensing domain and effector domains with another coding for the corresponding signal-generating enzyme. Using such a generalized syntactical template, below we identify the conflict systems that might leverage signal sensors and signal-generating enzymes beyond those operating on nucleotides or NAD<sup>+</sup> derivatives to activate associated effectors (Fig. 1E to G).

**(i) Counterinvader responses predicted to be activated by caspase proteolysis.** One of the systems that we uncovered based on the above-described, generalized template is defined by three genes: (i) a gene that codes for a previously uncharacter-

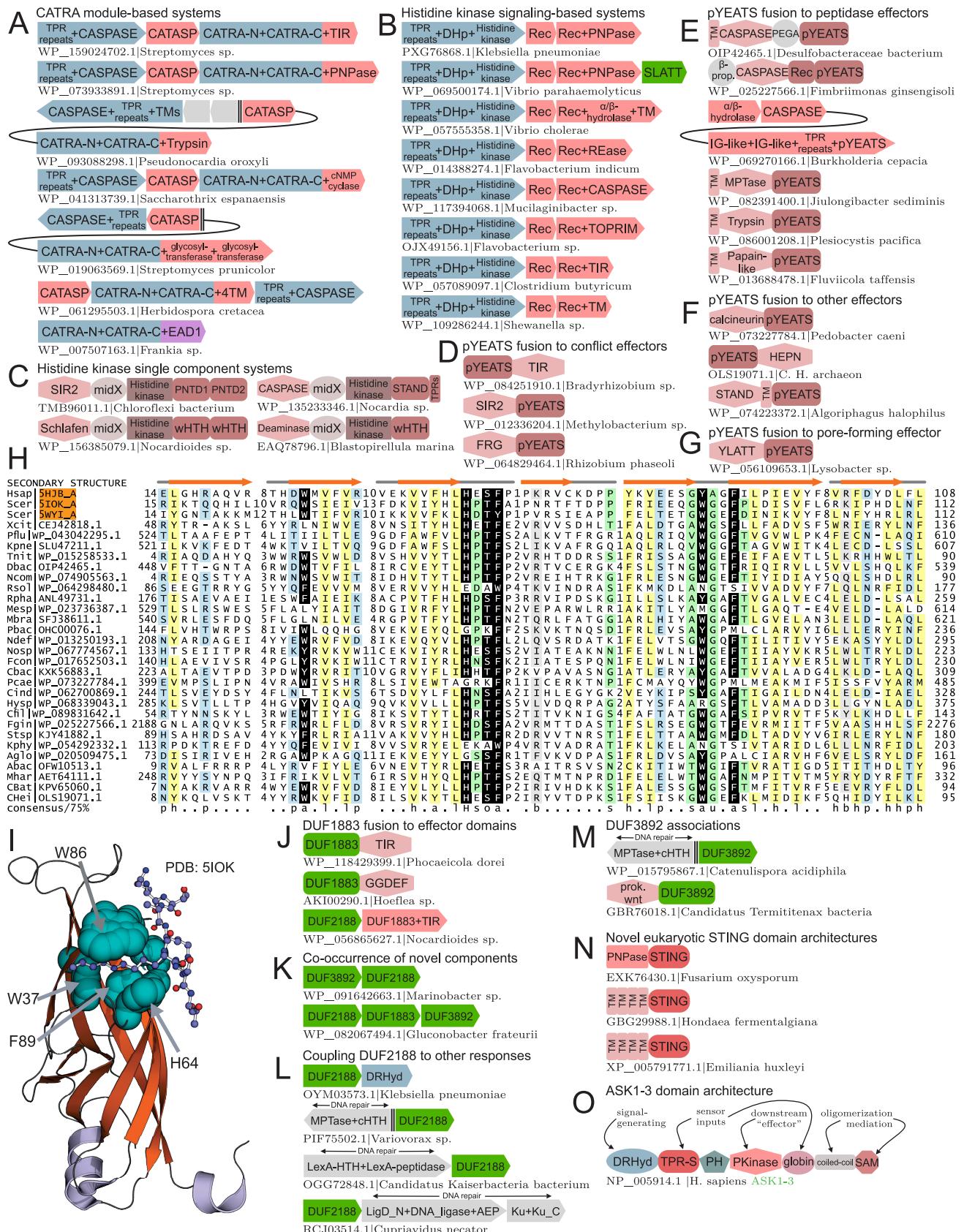
ized protein showing a conserved constant region fused to a C-terminal hypervariable region, which contains one or more of a diverse collection of predicted effector domains; (ii) a gene encoding a protein with a series of N-terminal TPR repeats not closely related to TPR-S fused to a C-terminal caspase domain; and (iii) a gene encoding a small protein predicted to possess a compact, uncharacterized four-helical domain. Hence, we name the conserved constant region in the first protein of the triad the CATRA (caspase and TPR repeat-associated) module (Fig. 1E and Fig. 4A). Analysis of the CATRA module showed that it is constituted of two conserved domains, the first of mixed  $\alpha/\beta$  character and the second entirely  $\alpha$ -helical, here referred to as CATRA-N and CATRA-C, respectively, with CATRA-C containing a nearly absolutely conserved aspartate (see alignments in the supplemental material).

Effector domains fused to the CATRA module include TIR, purine nucleotide phosphorylase (PNPase), trypsin-like peptidase, and cNMP cyclase domains shared with other systems. A pair of Rossmannoid fold glycosyltransferase domains and a novel four-TM domain are uniquely found in the effector position of these systems (Fig. 4A). A further domain observed in the effector position was EAD1, previously described as having functional parallels to the Death-like domains from animal apoptosis and immunity systems (18). Similar to the Death-like domains, EAD1 is predicted to mediate homotypic interactions with other proteins that contain EAD1. Thus, proteins containing effector domains fused to EAD1 can be coupled to the above-described core 3-component system (18).

The overall organization and domain content of this system are highly suggestive of the previously described “ternary” conflict systems (18), wherein the three different components (in this case CATRA-effector, TPR-caspase, and the small protein) are predicted to form a ternary complex. We posit that in the resting state, the effector domains in this ternary complex are in an inactive state. Upon the detection of an invasive threat, the caspase domain likely cleaves some target in the complex, either directly freeing the effectors or acting on another component like the small protein, which could play an inhibitory role. This could then result in a conformational change necessary for the effectors to assume an active state. Therefore, it seems likely that, akin to other ternary systems (18), CATRA-N and/or CATRA-C and the TPRs could function as direct sensors of invasive molecules and as transmitters of a signal through conformational change followed by proteolysis rather than as generators of a soluble signaling messenger.

**(ii) Histidine kinase signaling in invader sensing and defense.** A further group of systems uncovered in this analysis was a remarkable class of two-component systems, i.e., a histidine kinase (HisKin)-receiver domain couple, with a predicted role in counterinvader defense (Fig. 1F). These systems are defined by a three-gene operon (Fig. 4B) with a strictly conserved gene order coding for (i) a large protein composed of two distinct modules, an N-terminal fast-evolving TPR region with a distinct conservation pattern (not closely related to TPR-S) and a divergent C-terminal HisKin module (93, 94) (see the supplemental material); (ii) a protein with a stand-alone receiver (Rec) domain; and (iii) a protein with a Rec domain combined with a hypervariable region that features one of a diverse array of effector domains similar to those from the above-noted systems. The hypervariable effector position is most frequently occupied by a PNPase domain, with the others being  $\alpha/\beta$ -hydrolase, REase, caspase, TIR, Toprim, and potential membrane pore-forming TM domains (Fig. 4B and Table S1). The REase domain in these systems specifically unifies with the equivalent effector domain observed in previously described prokaryotic PIWI-containing conflict systems (95).

While the HisKin domains in these systems are divergent relative to their counterparts in conventional (physiological) signal transduction (93, 96), we observed absolutely conserved and positionally equivalent residues of the HisKin active site, such as the metal ion-coordinating asparagine residue, indicative of an active enzyme (97) (see the supplemental material). Furthermore, active HisKin domains are associated with an upstream elongated  $\alpha$ -hairpin, known as the DHP (98, 99), containing the invariable histidine residue which is autophosphorylated in the initial step of the signaling



**FIG 4** (A to G) Representative conserved gene neighborhoods and domain architectures of CATRA (A)- and histidine kinase (B and C)-mediated and pYEATS-centered (D to G) effector conflict systems. Depictions are as described in the legends to Fig. 1A and Fig. 2G. (H) Multiple-sequence alignment of

(Continued on next page)

cascade (100). Consistent with this, we detected an equivalent of the DHp with an absolutely conserved RH motif upstream of the core HisKin domain of current systems. However, secondary structure predictions and an unusually long separation from the catalytic domain suggest the presence of an extended insert not previously observed in the HisKin proteins (see the supplemental material). Both Rec domains from the current system retain the conservation patterns typical of the superfamily, including the absolutely conserved aspartate residue that receives the phosphate group from the autophosphorylated histidine residue of the DHp domain (Fig. 4B; see also the supplemental material).

In conventional signaling systems, HisKins are linked to other signaling modules such as the HAMP domain (101, 102); the S-helix, PAS, and GAF domains; or associated extracellular or TM ligand-binding domains (93, 103). Similarly, the Rec domains from such signaling systems are strongly associated with DNA-binding HTH domains, which mediate a downstream transcriptional response, or secondary signaling enzymes such as cNMP cyclases and the corresponding phosphodiesterases (104, 105). In stark contrast, the fast-evolving N-terminal TPR module of the HisKin component and the hypervariable effector domains fused to the second Rec component are not seen in conventional signaling systems. Thus, while these systems recapitulate the HisKin-Rec domain association typical of classical signaling systems (93, 106), they appear to be distinct variants, which are likely to mediate a response to invaders rather than relaying homeostatic signals. We accordingly reconstruct the following mechanism for these systems: the fast-evolving N-terminal TPR region of the HisKin component probably directly senses an invasive element inside the cell, initiating a conformational change and autophosphorylation of the dimerized HisKin domain. As in typical two-component signaling (107–111), this phosphate is transferred to the Rec domains, spurring their dimerization and activation of the Rec-fused effector domain via a conformational change. The activated effector domain then likely contains the invader by cell suicide or by a direct attack on invader molecules (Fig. 1F).

Aside from these systems, we had previously noted an unusual HisKin domain coupled to a Sir2 effector in prokaryotic Piwi-dependent conflict systems (95). Here, we observed that these HisKin domains can occur independently of the Piwi systems coupled with a variety of N-terminal effector domains, namely, endo-RNases (Schlafen [112, 113] and HEPN [35]), caspase, toxin deaminases (114, 115), and NAD<sup>+</sup>-processing enzymes (SIR2, TIR, and DRHyd) (Fig. 1F and Fig. 4C). These HisKin-containing proteins usually feature a variable C-terminal domain that is a DNA-binding wHTH domain or infrequently a STAND nucleoside triphosphatase (NTPase) or actin-like domain. Genes coding for these proteins are often found embedded in R-M systems or associated with ABC ATPases implicated in conflicts or in cell surface polysaccharide modification operons that might protect against specific phages by altering capsular composition (116). These HisKin domains are predicted to be active as they have all the above-mentioned features of catalytically active HisKin ATP-binding domains and are further typified by a universally conserved histidine close to the active site. However, they remarkably lack associated Rec or DHp domains. Hence, we posit that these might regulate the associated effector domains by merely binding and hydrolyzing ATP or autophosphorylation on the conserved internal histidine upon sensing invasive nucleic acids via the C-terminal wHTH or other stimuli if coupled to alternative C-terminal domains.

### (iii) The prokaryotic YEATS domains in membrane-localized conflict systems.

Given the above-mentioned examples, we wondered if in conflict systems, just as in

#### **FIG 4 Legend (Continued)**

The pYEATS domain, as described in the legend to Fig. 2E. Conserved residues contributing to the “aromatic cage” are shaded in black. (I) Structural rendering of the eukaryotic YEATS domain bound to crotonylated lysine. Conserved residues shared with the pYEATS domain are rendered as space filled and labeled. (J to M) Representative conserved gene neighborhoods for new components of prokaryotic counterinvader systems, as described in the legend to Fig. 1A. (N and O) Representative domain architectures for eukaryotic STING (N) and TPR-S (O) domain acquisitions, as described in the legend to Fig. 2G. C. H. archaeon, *Candidatus Heimdallarchaeota archaeon*.“

physiological signaling systems, the roles played by two-component systems can be mirrored by one-component systems that directly sense the invasive entity. We were alerted to such a possibility by the YEATS domain first encountered in the above-described TPR-S domain systems (Fig. 3O). As these are the first YEATS domains observed outside eukaryotic chromatin proteins, we investigated their presence in prokaryotes more thoroughly. Thus, we uncovered a sporadic yet widely distributed group of prokaryotic YEATS (pYEATS) domains in diverse bacterial lineages (Table S1).

These pYEATS proteins are linked to a hypervariable region displaying a remarkably diverse array of domains that often occur as effectors in other conflict systems (Fig. 1G). These include (i) NAD<sup>+</sup>-degrading or NAD<sup>+</sup>-dependent protein-modifying enzymes like SIR2, FRG-type ART (see above), and TIR domains (57, 117) (Fig. 4D); (ii) diverse peptidase domains, such as those of the caspase, metallopeptidase, trypsin-like, and papain-like superfamilies (36) (Fig. 4E) (these peptidase associations are often accompanied by further TM segments frequently coupling them to extracellular adhesion or peptidoglycan-binding domains like PEGA, immunoglobulin, TPRs, and  $\beta$ -propellers [118] [Fig. 4E]); (iii) a calcineurin-like phosphoesterase domain that might target cyclic nucleotides; (iv) the HEPN RNase domains; (v) STAND NTPase domains (Fig. 4F); and (vi) membrane-targeting domains such as an  $\alpha/\beta$ -hydrolase domain and a hitherto uncharacterized three-TM domain, which we refer to as the YLATT (YEATS-like-associating three-TM) domain, that might function as a membrane-perforating toxin (Fig. 4E and G). Notably, these pYEATS proteins tend to be encoded by stand-alone genes typically lacking gene neighborhood associations with any signal-generating enzymes (Fig. 4D to G). This, together with the striking hypervariability of the pYEATS proteins along with a phyletic pattern indicative of extensive dissemination by HGT (Table S1), suggests that the pYEATS proteins constitute a novel class of single-component conflict systems.

In eukaryotes, the YEATS domain has been shown to function as a “reader” domain that recognizes specific crotonylated lysine residues in histone H3 (119) by caging it through conserved aromatic residues (120). Crucially, pYEATS domains conserve aromatic residues at these positions, indicating a capacity for an interaction with a comparable chemical group (Fig. 4H and I). Through structure similarity searches with the DALI program, we established that the YEATS domains are most closely related to the eukaryote-specific lipid-binding C2 domains (Fig. S1A). Indeed, the binding of crotonyl lysine by the YEATS domain closely parallels the Ca<sup>2+</sup>-dependent binding of lipids by the C2 domains (Fig. S1B). Over half of the identified pYEATS proteins tend to have TM segments or have other associations indicating a membrane-proximal function (Fig. 4E to G and Table S1). These observations raise the possibility that the pYEATS domains bind membrane lipid moieties, e.g., modified basic head groups similar to the C2 domains. Hence, one prediction is that they specifically recognize invader-induced modifications of macromolecules by acyl moieties (e.g., phage-encoded acylations of surface polysaccharides [121]) to unleash the fused effectors through a conformational change (Fig. 1G).

**(iv) DUF1883 systems and other new components of prokaryotic counterinvader systems: DUF2188, DUF3982, and bacterial Wnt domains.** Continuing with the theme of the above-described pYEATS systems, we found a comparable group of predicted one-component conflict systems centered on a domain of unknown function, labeled DUF1883 in the Pfam database (Fig. 4J). DUF1883 typically occurs in proteins with a variable region, which contains one of several effector modules, including TIR and GGDEF domains (Fig. 4J and Table S1). Structural analysis of DUF1883 indicates that the domain forms an obligately dimeric structure related to the PPC-like  $\beta$ -sandwich fold (122, 123), with a deep cleft that could bind a ligand (see the supplemental material). Therefore, DUF1883 might function as a direct sensor of invader (or invader-induced) molecules, which then activates the fused effector to trigger a secondary signal or an invader restriction attack.

In certain operons, the gene for a stand-alone DUF1883 protein might be combined with two others encoding “domains of unknown function,” labeled DUF2188 and DUF3982 in Pfam (Fig. 4K). We accordingly predict that these two domains are further

effector domains. Consistent with this proposal, DUF2188 is also observed in (i) certain SMODS-containing operons in a position typical of secondary effectors previously observed in some of these operons (17) (Fig. 1A and Fig. 2H) and (ii) two-gene operons comparable to those described above (Fig. 2D and Fig. 3P), with a gene coding for a DRHyd enzyme (Fig. 4L and Table S1), which could generate an inducing signal from NAD<sup>+</sup> or another nucleotide. Furthermore, a previous experimental study showed a DUF2188 protein to associate with ribosomes (124). A considerable subset of DUF2188 domains contain a conserved histidine residue (see the supplemental material), raising the possibility of an enzymatic role for at least this subset either as an RNase or in RNA end processing; accordingly, we predict that this domain might evince its action by translation inhibition. We found that beyond these conflict-related gene neighborhoods, DUF2188 is also found in certain DNA-processing-related operons coupled with genes for SOS response peptidase-HTH proteins, e.g., LexA, the DNA ligase-Ku DNA end repair complex, and type I R-M systems (11, 125, 126) (Fig. 4L and Table S1). These associations again suggest that DUF2188 could mediate translation inhibition alongside the repair of DNA damage or restriction.

In a similar vein, we found via profile-profile comparisons that DUF3892 is related to the truncated RNase H fold-containing ribosome hibernation factors (see the supplemental material) (4, 127). This supports a role for these domains as possible effectors acting at the ribosome like DUF2188, also consistent with its comparable associations (Fig. 2H, Fig. 4M, and Table S1). Notably, DUF3892 is also fused to hitherto unknown minimal prokaryotic versions of the Wnt signaling domain of animals (128, 129) (Fig. 4M). Based on their conservation pattern and frequent occurrence as a toxin in the C-terminal “tip” position in a class of previously unreported polymorphic-toxin-related systems (13), we predict that these minimal Wnt domains are enzymatic effectors (our unpublished data) that might act along with the above-described domains.

**Eukaryotic reuse of components from the above-described prokaryotic conflict systems.** Our initial study of nucleotide-centric conflict systems in prokaryotes (17) revealed the considerable reuse of components from these systems for various conflict and nonconflict roles in eukaryotes. Beyond the direct use of animal nucleotide cGAS and OAS synthetases sharing a close ancestry with the prokaryotic SMODS (8, 17) in activating animal immune responses, other proteins derived from such systems also gave rise to components of diverse eukaryotic cellular systems. A striking example is the emergence of the pan-eukaryotic chromosome “coating” HORMA proteins and their AAA<sup>+</sup> ATPase partner TRIP13/Pch2 from such bacterial conflict systems. Another case is the recruitment of the SLOG domain as a nucleotide-binding gating module for the animal TRPM class of ion channels and other eukaryotic membrane channels (17, 130, 131). These previous findings prompted us to look more closely for comparable eukaryotic acquisitions from the above-mentioned systems.

**(i) Multiple eukaryotic acquisitions of the STING domain.** As noted above, we identified previously undetected eukaryotic STING domains suggesting multiple independent acquisitions. The cGAS-STING pair shows a strongly congruent phyletic distribution in *Monosiga brevicollis* and animals (Fig. S2A). cGAS has undergone extensive radiation with sequence and architectural diversification across animals (Fig. S2B). Concomitant lineage-specific expansions (LSEs) are observed for both cGAS and STING (Fig. S2C) in marine invertebrates like cnidarians and molluscs. Notably, certain molluscan LSEs of cGAS catalytic domains are fused to superstructure-forming TPR or ankyrin repeats that might play the role of invader detection modules. These observations, with the association of the cGAS-like SMODS and its STING sensor domain partner in prokaryotes, suggest that they were acquired from a bacterial precursor in a single lateral transfer event at the base of the choanoflagellate-metazoan lineage.

The presence of a direct fusion to a TIR domain in several molluscs (Fig. S2B and C) (132) led to the suggestion that the TIR-fused versions observed in certain bacteroïdetes (Fig. 2E, G, and H) are the direct precursors of the equivalent animal domains (28). Closer analysis suggests that this is unlikely: the TIR domains in these molluscs do not show a specific relationship to prokaryotic versions and instead are most closely

related to eukaryotic TIR domains. Our analyses indicate that the eukaryotic STING domains are most closely related to prokaryotic versions fused to N-terminal TM helices (Fig. 2G, Fig. S2D, and Table S1), consistent with the presence of at least two conserved N-terminal TM helices in animal-choanoflagellate STING domains which anchor the domain to the endoplasmic reticulum (ER) membrane (133, 134). In prokaryotic proteins combining the STING with TM segments, the latter could function as an effector that forms pores in the membrane (Fig. 2G). Hence, it may be worth evaluating a comparable role for the STING-fused TM helices at the ER membrane. Such a function might be connected to the observed translocation of the STING domain to the Golgi apparatus following its activation through cGAS-generated nucleotide recognition (135, 136). These observations suggest that the STING and TIR domain association has convergently emerged in certain lineages. More generally, this proposal is consistent with the previously reported, cross-superkingdom convergent origin of domain architectural themes in immunity and apoptosis systems (4, 18).

The STING domain has been independently acquired in eukaryotes on at least two other occasions. The first is in a subset of filamentous ascomycete fungi, where it is invariably fused to an N-terminal PNPase superfamily domain (137) (Fig. 4N and Table S1) also found as an effector in diverse bacterial conflict systems (17). Hence, these fungal PNPase domains might be deployed against invasive threats upon the recognition of a nucleotide or nucleotide-derived signal. The STING domain is also detected in multiple copies in the haptophyte *Emiliania huxleyi* and in a single copy in the stramenopile *Hondaea fermentalgiana*. These versions are fused to four N-terminal TM helices, again suggesting a possible membrane-linked role regulated by nucleotide detection (Fig. 4N). Interestingly, no corresponding SMODS-like domain has been detected in any of these lineages, suggesting that (i) STING is capable of recognizing signals generated independently of SMODS, (ii) these organisms possess cryptic SMODS-like enzymes that continue to evade detection, and (iii) it responds to nucleotide signals generated by an invasive or symbiotic entity.

**(ii) The TPR-S domain in the ASK signalosome.** Remarkably, we also found homologs of the above-described TPR-S domain in choanoflagellates and animals (Table S1). These map to a portion of Pfam DUF4071 (domain of unknown function) (138) and an overlapping portion of the previously described “central regulatory region” of the apoptosis signal-regulating kinases (ASKs) (66). In vertebrates, these form the oligomeric ASK signalosome complex incorporating three paralogous components (ASK1-3; also known as MAP3K5-MAP3K6-MAP3K15) during initiation of the apoptotic mitogen-activated protein kinase (MAPK) signaling cascade (66, 139, 140). These TPR-S domains of ASK1-3 retain the same configuration of conserved polar and aromatic residues as in their prokaryotic homologs (Fig. 3A), suggesting that they too might bind NAD<sup>+</sup> derivatives similar to the ones predicted for their prokaryotic counterparts.

To glean further insight into the function of these eukaryotic TPR-S proteins, we analyzed the domain architectures of the ASKs, which in addition to TPR-S contain the previously identified pleckstrin homology (PH) (66) and protein kinase domains. We observed that the region C terminal to the protein kinase domain contains previously unrecognized globin and sterile alpha motif (SAM) superfamily domains (141–143) connected via a disordered linker bracketing a central coiled-coil (CC) region (Fig. 4O and Table S1). Notably, this identification of the SAM domain supersedes the previous misidentification of a “ubiquitin-like sequence” in the same region (144). The novel globin domain is a non-heme-binding version with the strongest affinities for the HisK-N family of sensor domains, which inhibit histidine kinase activation required for sporulation in bacteria of the firmicutes lineage through the possible binding of a fatty acid- or lipid-derived ligand (145). Finally, we identified a conserved globular  $\alpha/\beta$ -domain at the N terminus of ASK1-3. Strikingly, we were able to unify this domain with the DRHyd superfamily (Fig. 4O and Fig. S3), one of the three superfamilies of enzymes linked to NAD<sup>+</sup> processing/NAD<sup>+</sup>-derived signal generation in prokaryotic conflict systems (17) (see above). This DRHyd domain features highly conserved residues at positions where active-site residues are typically observed in this superfamily (Fig. S3).

Notably, ASK1/MAP3K5 and potentially ASK2 have been characterized in the activation of innate antiviral immunity pathways in conjunction with initiating apoptotic pathways (146–149) reminiscent of the function that we have predicted for the prokaryotic TPR-S domains. Specifically, the multidomain architecture of the TPR-S domain parallels bacterial systems where the TPR-S domain is fused to tandem arrays of domains. Thus, we propose the following functional interpretation for the above-reported domain architecture of the ASK proteins (Fig. 4O). Experiments have implicated the region encompassing the TPR-S domain and the upstream N-terminal region of ASK1-3 in attenuating the activity of the C-terminal protein kinase domain (66, 150). Hence, we suggest that the TPR-S domain serves as a sensor whose inhibitory effect on the kinase domain is likely relieved when it binds a messenger (likely NAD<sup>+</sup> derived), echoing its prokaryotic counterparts where the associated effector domains are likely unleashed upon binding the messenger. If the N-terminal DRHyd domain is catalytically active, then it is likely to function as a signal-generating component of the ASK signalosome. Alternatively, if the DRHyd domain is enzymatically inactive, then it might simply function as a further ligand-binding domain for an NAD<sup>+</sup> or nucleotide derivative generated by a distinct process (151). The region corresponding to the DRHyd domain has been implicated in a negative regulatory interaction with the active cysteine of thioredoxin and activation by reactive oxygen species (150, 152–154). These could be the proximal signals that control the activity of the DRHyd domain. Finally, the ASK proteins are likely to integrate this NAD<sup>+</sup>-derived messenger-based regulation with other interactions and sensory inputs. On one hand, the regions mapping to the SAM domain and CC region have been respectively implicated in mediating homooligomerization (150) and binding USP9X and 14-3-3 proteins (144, 155, 156). On the other, the HisK-N-like globin domain is likely to represent an independent sensory element that recognizes a fatty acid or a related membrane-derived molecule. Thus, these observations bring the ASK signalosome under the same mechanistic umbrella as other conflict systems that initiate an effector response via the sensing of a threshold-setting nucleotide or nucleotide-derived messenger (17).

**(iii) The YEATS domain: from prokaryotic conflict systems to eukaryotic chromatin.** The eukaryotic YEATS domain has been characterized as a “reader” of acylated marks on histones, specifically preferring crotonylated H3K18Cr marks generated by the P300/CBP-like divergent GNAT superfamily enzymes over acetylated lysine (89, 119). Their domain architectures strongly suggest that all eukaryotic YEATS domain proteins are likely involved in chromatin-related roles as opposed to biological conflicts (89). Furthermore, the domain can be confidently traced to the last eukaryotic common ancestor (LECA) by merit of its nearly pan-eukaryotic phyletic pattern suggesting that its epigenetic role is ancestral to eukaryotes (89). The characterization of the pYEATS family suggests that the domain had considerably diversified in bacterial conflict systems prior to its acquisition by the stem eukaryote. Multiple associations of the pYEATS domain with TM segments point to a role in conflict in proximity to the membrane, which might include binding of macromolecules modified by acyl groups or particular lipid head group moieties (Fig. 4E to G). It is conceivable that within the pYEATS radiation there were versions specifically recognizing crotonyl moieties. In the stem eukaryote, this role was likely exapted to recognize crotonyl groups of modified histones in chromatin. This adaptation is reminiscent of another widespread histone modification reader domain, the chromodomain, that appears to have been acquired from bacterial secreted proteins associated with the peptidoglycan (157).

**Mechanistic and evolutionary considerations.** The systematic survey of the effector domains identified here and in previous studies indicates that the majority of them have the potential for a negative consequence on the deploying cell, in the worst case eliciting cell death. On one side, the inclusive fitness gained from kin favors cell suicide when a virus cannot be contained by other defensive systems in the cell; such a response can avert infection of kin (16). On the other side, “misfiring” or deployment of the effector in the absence of a sufficiently strong threat could nullify the fitness of

the deploying cell without benefitting kin. These opposing pressures have selected for finely balanced systems that set a suitable threshold for the deployment of potentially deleterious effectors (17, 18). The original characterization of the nucleotide-centric conflict systems indicated that such “thresholding” is a general principle unifying otherwise disparate versions of such systems. This was proposed to be the role of the nucleotide-derived messenger generated upon detecting the invasive entity through direct or indirect means (17). The effector is unleashed only after this messenger is sensed by an associated sensor. More recently, we have expanded this principle to include the so-called “ternary conflict systems” with more elaborate regulatory steps (18). In these systems, different convergent but mechanistically equivalent regulatory components bridge the invader detection and effector components. They are proposed to set a threshold for the activation of the effector components via NTPase or peptidase switches that likely ensure that the inducing signal is of sufficient strength (18).

The present study shows that this principle of thresholding has allowed the incorporation of a whole spectrum of regulatory mechanisms and parallel pathways into conflict systems, the latter of which include a diversified collection of mobile Ub systems (Fig. 1A and C). We observe that beyond conventional nucleotide-dependent systems with diverse synthetases, the signal generator components encompass a rich array of representatives drawn from the three related superfamilies of Rossmannoid domains, namely, TIR, DRHyd, and SLOG (17). These are predicted to generate NAD<sup>+</sup> derivatives or modified nucleobases that are frequently sensed by their cognate sensor, the TPR-S domain (Fig. 1D). These systems are also unique in their propensity to utilize these Rossmannoid domains as both signal generators and effectors in different systems. While this appears paradoxical at first sight, it is supported by their distinct positional occurrences in different systems. We suggest that this situation arises from the utility of NAD<sup>+</sup> as both a signal substrate and a potential effector target, being the “redox currency” in the cell (158). Finally, our discovery of histidine kinase-activated, FRG-dependent, and one-component-like (pYEATS and DUF1883) conflict systems reveals that the principle of thresholding in conflict need not depend on just nucleotide or NAD<sup>+</sup> derivatives (Fig. 1E to G). These illustrate that signals that are otherwise common in physiological signaling systems, such as histidine phosphorylation, diverse cyclic nucleotides, and small molecular ligands, have also been recruited in threshold setting in biological conflicts. Indeed, while the detected inputs and generated responses in conflict and classical signaling systems that deploy nucleotide and derived second messengers are predominantly distinct, the intervening steps thematically match each other (40, 41, 88) (Fig. 5). However, these two “classes” of signaling systems, conflict and physiological, barring a few exceptions, draw from mostly distinct pools of domains to carry out these steps, most prominently at the signal generation and signal sensor step (17, 103) (Fig. 5, red). The overall tendency toward exclusivity in domain recruitment suggests that there are selective pressures active in the respective signaling classes, which requires further investigation; for example, it may be that the synthetases in conflict systems are more suitable for the direct detection of invader inputs. Nevertheless, conflict and physiological signaling systems have converged to certain specific commonalities, such as the deployment of global cyclic nucleotide signals, as indicated by the cNMP cyclases seen in the effector proteins of conflict systems described in this study (40, 41, 57, 159) (Fig. 3G and N and Fig. 5).

We also show that the recruitment of components from prokaryotic conflict systems to diverse eukaryotic systems is not a limited phenomenon. Paralleling our previous findings regarding the HORMA-PCH2 AAA<sup>+</sup> ATPase (17), we show that some of these events, such as the recruitment of the YEATS domain as a reader of epigenetic marks, likely happened early in eukaryotic evolution. Thus, these observations are consistent with previous documentation of prokaryotic conflict systems acting as “nurseries” for molecular innovation that provided the evolutionary grist for several major eukaryotic systems (7). In other cases, they were incorporated into comparable conflict systems in eukaryotes. This was particularly common close to the stem of the metazoan lineage. Notably, both the SMODS and STING receptor and the DRHyd-

## Characteristics of conflict vs. physiological nucleotide-based signaling

nucleotide signals	<i>SMODS, OAS, mCpol, DisA, cNMP cyclase, GrpB?, DUF2204?</i>	<i>cGAMP, cA4, cA6, c-di-AMP, linear nucleotides, cGMP/cAMP?</i>	<i>AGS-C, SAVED, CARF, cNMPBD, SUa-2TM, NARF, STING</i>	<i>DNase, RNase, pore-forming, phospholipase, peptidase, PNPase, cNMP cyclases</i>	<i>attack on invading molecules, cell-suicide, global cell response</i>
NAD <sup>+</sup> -derived signals	<i>above + sensing at membrane</i>	<i>TIR, YpsA-like SLOG</i>	<i>ADPR, cADPR, nam?</i>	<i>TPR-S</i>	<i>above + TIR, DRHyd, SLOG</i> <i>above, translational pausing, metabolite (NAD<sup>+</sup>) limitation</i>
↑ CONFLICT SIGNALING					
↓ PHYSIOLOGICAL SIGNALING					
nucleotide signals	<i>detection of environmental cues, metabolite concentration or gradient trigger, detection of upstream signal</i>	<i>GGDEF, cNMP cyclase, CyaA-like, RelA/SpoT, DisA</i>	<i>cAMP, cGMP, c-di-GMP, c-di-AMP, alarmone</i>	<i>GAF, cNMPBD, riboswitches, PilZ</i>	<i>signal propagation: kinases, GTPases, transcription factors, ion channels signal degradation: HD, EAL, calcineurin-like</i>
NAD <sup>+</sup> -derived signals	<i>environmental cues through cell surface receptor signal transduction, intracellular stimuli</i>	<i>ADP-ribosyl cyclase</i>	<i>cADPr, NAADP</i>	<i>?</i>	<i>ryanodine receptor and two-pore calcium channels membrane depolarization</i>

**FIG 5** Generalized conceptual diagrams comparing nucleotide-activated effector conflict and physiological signaling systems. Diagrams are as described in the legend to Fig. 1. Example domains and functional categories for both classical nucleotide and NAD<sup>+</sup>-derived second messenger systems are provided. Domains shared across the two classes are shown in red.

TPR-S pairs were acquired from prokaryotic messenger-activated conflict systems in the choanoflagellate-metazoan common ancestor. Furthermore, after their initial acquisition, TIR domains have been independently expanded and utilized in animal immunity and apoptosis systems, particularly in the basal branches of metazoa (160, 161). This supports the proposal that the origin of colonial and multicellular forms in the “greater” animal lineage and the associated aspects of immunity probably drew heavily from preadaptations provided by the horizontal acquisition of components from prokaryotic conflict systems (18).

**Conclusions.** Effector domain sharing is a defining characteristic across diverse classes of conflict systems in prokaryotes, with thematically similar classes of conflict

systems demonstrating stronger overlap in their effectors (3, 4, 13, 17, 18, 72). Leveraging this knowledge and the general syntax associated with threshold-dependent regulation, we were able to identify several novel biological conflict systems. The new systems described in this work run the entire gamut from versions that are relatively straightforward variants of previously described systems to those that utilize entirely different signals and signaling mechanisms. Furthermore, our current work is likely to have consequences for understanding eukaryotic immunity and apoptosis systems such as the ASK signalosome (Fig. 3A to D and Fig. 4N). Our previous work has inspired several directed experimental studies by several other groups. These studies have confirmed the roles of at least a subset of these systems in mediating conflict with invasive entities (22, 27, 28). However, more experimental work remains to be done in terms of further understanding the biochemical, mechanistic, and ecological roles of threshold-dependent conflict systems. Chief among these is the need for better structural characterization of the sensor-ligand interactions of domains such as SAVED, AGS-C, and TPR-S as well as exploration of the identity of the range of messengers generated by the various signal generators. We expect that the current work will thus provide additional avenues for the study of nucleotide-activated effector and other threshold-based conflict systems.

## MATERIALS AND METHODS

**Sequence analysis.** Sequence profile searches were performed using the PSI-BLAST (RRID SCR\_001010) and JACKHMMER (RRID SCR\_005305) programs, with the profile being built at each iteration (162, 163). Clustering for both classification and culling of nearly identical sequences was performed with the BLASTCLUST program (<ftp://ftp.ncbi.nih.gov/blast/documents/blastclust.html>) (RRID SCR\_016641). It uses both the length of the pairwise alignment ( $L$ ) and measure of similarity, i.e., bit score ( $S$ ), and these were adjusted depending on the degree of clustering required. For example, the  $L$  and  $S$  parameters for clustering nearly identical proteins were set at an  $L$  value of 0.9 and an  $S$  value of 1.89.

Hidden Markov model (HMM) searches were run using either HMMsearch initiated with an HMM built from an alignment or iteratively using JACKHMMER from single seeds. The sequence searches were run against either the nonredundant (nr) protein database frozen at 10 October 2019 of the National Center for Biotechnology Information (NCBI) or the same database clustered down to 50% similarity using the MMseqs program (164) (RRID SCR\_008184). Profile-profile searches conducted using HHpred (RRID SCR\_010276) (165) were used for and run against (i) HMMs derived from the PDB, (ii) Pfam models (166), and (iii) a custom database of alignments of diverse domains curated by the Aravind group. Previously known domains in Pfam were corrected for boundaries where required and augmented using divergent members that were not detected by the original Pfam models. All novel alignments are provided in the supplemental material. Multiple-sequence alignments were built using Kalign (167) (RRID SCR\_011810) and Muscle (168), followed by manual adjustments based on profile-profile and structural alignments. Secondary structures were predicted using the JPred program (169) (RRID SCR\_016504). To assess significant patterns in sequence divergence that help distinguish proteins in conflict systems, we used statistically significant differences in the mean column-wise Shannon entropy values of alignments as a measure (170). Entropy comparisons (see examples provided in Fig. S4 in the supplemental material) were performed after initiating single-iteration BLAST searches against the above-described, clustered nr database and then clustering and aligning the recovered sequences. The significance of the mean entropy difference was tested using Wilcoxon signed-rank test scores.

**Structure analysis.** Structure similarity searches were performed using the DALI Lite program (171) (RRID SCR\_003047) run against the PDB database clustered at 75% sequence similarity. Structure similarity trees were constructed based on Z-scores obtained from an all-versus-all search of the compared structures using average linkage clustering. Structural visualization and manipulations were performed using the PyMOL (<http://www.pymol.org>) (RRID SCR\_000305) and MOL\* (<http://molstar.org>) programs.

**Comparative genomics.** Taxonomic lineages were obtained from the NCBI Taxonomy database. Contextual information from prokaryotic gene neighborhoods was retrieved using a Perl script to extract upstream and downstream genes of the query gene from the GenBank genome file. Their products were then clustered with BLASTCLUST to identify conserved gene neighborhoods based on conservation between different taxa. Several additional filters were then applied to recognize valid neighborhoods for further analysis: (i) nucleotide distance constraints (generally 50 nucleotides), (ii) conservation of gene directionality within the neighborhood, and (iii) presence in more than one phylum. Phylogenetic trees were constructed using an approximate maximum likelihood method implemented in the FastTree 2.1 (172) (RRID SCR\_015501) program under default parameters. Data processing (knitr and dplyr libraries), network analysis (igraph and circlize libraries), and visualization were performed using the R language. Delimited, downloadable data sets are available at [ftp://ftp.ncbi.nih.gov/pub/aravind/New\\_ConfSystems/](ftp://ftp.ncbi.nih.gov/pub/aravind/New_ConfSystems/).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 3.3 MB.

## ACKNOWLEDGMENT

Work by A.M.B. and L.A. is supported by the intramural funds of the National Library of Medicine, NIH.

## REFERENCES

- Hurst LD, Atlan A, Bengtsson BO. 1996. Genetic conflicts. *Q Rev Biol* 71:317–364. <https://doi.org/10.1086/419442>.
- Smith JM, Price GR. 1973. The logic of animal conflict. *Nature* 246: 15–18. <https://doi.org/10.1038/246015a0>.
- Iyer LM, Burroughs AM, Anand S, de Souza RF, Aravind L. 2017. Polyvalent proteins, a pervasive theme in the intergenomic biological conflicts of bacteriophages and conjugative elements. *J Bacteriol* 199: e00245-17. <https://doi.org/10.1128/JB.00245-17>.
- Burroughs AM, Aravind L. 2016. RNA damage in biological conflicts and the diversity of responding RNA repair systems. *Nucleic Acids Res* 44:8525–8555. <https://doi.org/10.1093/nar/gkw722>.
- Athukoralege JS, Graham S, Rouillon C, Gruschow S, Czekster CM, White MF. 2020. The dynamic interplay of host and viral enzymes in type III CRISPR-mediated cyclic nucleotide signalling. *Elife* 9:e55852. <https://doi.org/10.7554/elife.55852>.
- Werren JH. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci U S A* 108(Suppl 2):10863–10870. <https://doi.org/10.1073/pnas.1102343108>.
- Aravind L, Anantharaman V, Zhang D, de Souza RF, Iyer LM. 2012. Gene flow and biological conflict systems in the origin and evolution of eukaryotes. *Front Cell Infect Microbiol* 2:89. <https://doi.org/10.3389/fcimb.2012.00089>.
- Kato K, Ishii R, Hirano S, Ishitani R, Nureki O. 2015. Structural basis for the catalytic mechanism of DncV, bacterial homolog of cyclic GMP-AMP synthase. *Structure* 23:843–850. <https://doi.org/10.1016/j.str.2015.01.023>.
- Iyer LM, Zhang D, Burroughs AM, Aravind L. 2013. Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. *Nucleic Acids Res* 41:7635–7655. <https://doi.org/10.1093/nar/gkt573>.
- Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. *Nat Rev Microbiol* 8:317–327. <https://doi.org/10.1038/nrmicro2315>.
- Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, Blumenthal RM, Degtyarev S, Dryden DT, Dybvig K, Firman K, Gromova ES, Gumpert RL, Halford SE, Hattman S, Heitman J, Hornby DP, Janulaitis A, Jeltsch A, Josephsen J, Kiss A, Klaenhammer TR, Kobayashi I, Kong H, Kruger DH, Lacks S, Marinus MG, Miyahara M, Morgan RD, Murray NE, Nagaraja V, Piekarowicz A, Pingoud A, Raleigh E, Rao DN, Reich N, Repin VE, Selker EU, Shaw PC, Stein DC, Stoddard BL, Szybalski W, Trautner TA, Van Etten JL, Vitor JM, Wilson GG, Xu SY. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* 31:1805–1812. <https://doi.org/10.1093/nar/gkg274>.
- Russell AB, Peterson SB, Mougous JD. 2014. Type VI secretion system effectors: poisons with a purpose. *Nat Rev Microbiol* 12:137–148. <https://doi.org/10.1038/nrmicro3185>.
- Zhang D, de Souza RF, Anantharaman V, Iyer LM, Aravind L. 2012. Polymorphic toxin systems: comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol Direct* 7:18. <https://doi.org/10.1186/1745-6150-7-18>.
- Gilbert RJ. 2002. Pore-forming toxins. *Cell Mol Life Sci* 59:832–844. <https://doi.org/10.1007/s00018-002-8471-1>.
- Cook GM, Robson JR, Frampton RA, McKenzie J, Przybilski R, Fineran PC, Arcus VL. 2013. Ribonucleases in bacterial toxin-antitoxin systems. *Biochim Biophys Acta* 1829:523–531. <https://doi.org/10.1016/j.bbagen.2013.02.007>.
- Makarova KS, Anantharaman V, Aravind L, Koonin EV. 2012. Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. *Biol Direct* 7:40. <https://doi.org/10.1186/1745-6150-7-40>.
- Burroughs AM, Zhang D, Schaffer DE, Iyer LM, Aravind L. 2015. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res* 43:10633–10654. <https://doi.org/10.1093/nar/gkv1267>.
- Kaur G, Burroughs AM, Iyer LM, Aravind L. 2020. Highly regulated, diversifying NTP-dependent biological conflict systems with implications for the emergence of multicellularity. *Elife* 9:e52696. <https://doi.org/10.7554/elife.52696>.
- Clemens MJ, Williams BR. 1978. Inhibition of cell-free protein synthesis by pppA'p5'A'p5'A: a novel oligonucleotide synthesized by interferon-treated L cell extracts. *Cell* 13:565–572. [https://doi.org/10.1016/0092-8674\(78\)90329-X](https://doi.org/10.1016/0092-8674(78)90329-X).
- Sadler AJ, Williams BR. 2008. Interferon-inducible antiviral effectors. *Nat Rev Immunol* 8:559–568. <https://doi.org/10.1038/nri2314>.
- Iyer LM, Burroughs AM, Aravind L. 2006. The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like beta-grasp domains. *Genome Biol* 7:R60. <https://doi.org/10.1186/gb-2006-7-7-r60>.
- Severin GB, Ramliden MS, Hawver LA, Wang K, Pell ME, Kieninger AK, Khataokar A, O'Hara BJ, Behrmann LV, Neiditch MB, Benning C, Waters CM, Ng WL. 2018. Direct activation of a phospholipase by cyclic GMP-AMP in El Tor *Vibrio cholerae*. *Proc Natl Acad Sci U S A* 115: E6048–E6055. <https://doi.org/10.1073/pnas.1801233115>.
- Niewohner O, Garcia-Doval C, Rostol JT, Berk C, Schwede F, Bigler L, Hall J, Marraffini LA, Jinek M. 2017. Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* 548:543–548. <https://doi.org/10.1038/nature23467>.
- Kazlauskienė M, Kostiuk G, Venclovas C, Tamulaitis G, Siksnys V. 2017. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* 357:605–609. <https://doi.org/10.1126/science.aao0100>.
- Li F, Cimdins A, Rohde M, Jansch L, Kaever V, Nimtz M, Romling U. 2019. DncV synthesizes cyclic GMP-AMP and regulates biofilm formation and motility in *Escherichia coli* ECOR31. *mBio* 10:e02492-18. <https://doi.org/10.1128/mBio.02492-18>.
- Lau RK, Ye Q, Birkholz EA, Berg KR, Patel L, Mathews IT, Watrous JD, Ego K, Whiteley AT, Lowey B, Mekalanos JJ, Kranzusch PJ, Jain M, Pogliano J, Corbett KD. 2020. Structure and mechanism of a cyclic tri nucleotide-activated bacterial endonuclease mediating bacteriophage immunity. *Mol Cell* 77:723–733.e6. <https://doi.org/10.1016/j.molcel.2019.12.010>.
- Ye Q, Lau RK, Mathews IT, Birkholz EA, Watrous JD, Azimi CS, Pogliano J, Jain M, Corbett KD. 2020. HORMA domain proteins and a Trip13-like ATPase regulate bacterial cGAS-like enzymes to mediate bacteriophage immunity. *Mol Cell* 77:709–722.e7. <https://doi.org/10.1016/j.molcel.2019.12.009>.
- Cohen D, Melamed S, Millman A, Shulman G, Oppenheimer-Shaanan Y, Kacen A, Doron S, Amitai G, Sorek R. 2019. Cyclic GMP-AMP signalling protects bacteria against viral infection. *Nature* 574:691–695. <https://doi.org/10.1038/s41586-019-1605-5>.
- Athukoralege JS, Rouillon C, Graham S, Gruschow S, White MF. 2018. Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate. *Nature* 562:277–280. <https://doi.org/10.1038/s41586-018-0557-5>.
- Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R. 2018. Systematic discovery of antiphage defense systems in the microbial pan genome. *Science* 359:eaar4120. <https://doi.org/10.1126/science.aar4120>.
- Whiteley AT, Eaglesham JB, de Oliveira Mann CC, Morehouse BR, Lowey B, Nieminen EA, Danilchanka O, King DS, Lee ASY, Mekalanos JJ, Kranzusch PJ. 2019. Bacterial cGAS-like enzymes synthesize di-

- verse nucleotide signals. *Nature* 567:194–199. <https://doi.org/10.1038/s41586-019-0953-5>.
32. Leplae R, Geeraerts D, Hallez R, Guglielmini J, Dreze P, Van Melderen L. 2011. Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Res* 39:5513–5525. <https://doi.org/10.1093/nar/gkr131>.
  33. Van Melderen L. 2010. Toxin-antitoxin systems: why so many, what for? *Curr Opin Microbiol* 13:781–785. <https://doi.org/10.1016/j.mib.2010.10.006>.
  34. Bork P, Dandekar T, Diaz-Lazcoz Y, Eisenhaber F, Huynen M, Yuan Y. 1998. Predicting function: from genes to genomes and back. *J Mol Biol* 283:707–725. <https://doi.org/10.1006/jmbi.1998.2144>.
  35. Anantharaman V, Makarova KS, Burroughs AM, Koonin EV, Aravind L. 2013. Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biol Direct* 8:15. <https://doi.org/10.1186/1745-6150-8-15>.
  36. Murzin AG. 1998. How far divergent evolution goes in proteins. *Curr Opin Struct Biol* 8:380–387. [https://doi.org/10.1016/s0959-440x\(98\)80073-0](https://doi.org/10.1016/s0959-440x(98)80073-0).
  37. Yau KW. 1994. Cyclic nucleotide-gated channels: an expanding new family of ion channels. *Proc Natl Acad Sci U S A* 91:3481–3483. <https://doi.org/10.1073/pnas.91.9.3481>.
  38. Schultz SC, Shields GC, Steitz TA. 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science* 253:1001–1007. <https://doi.org/10.1126/science.1653449>.
  39. Menestrina G, Dalla Serra M, Comai M, Coraiola M, Viero G, Werner S, Colin DA, Monteil H, Prevost G. 2003. Ion channels and bacterial infection: the case of beta-barrel pore-forming protein toxins of *Staphylococcus aureus*. *FEBS Lett* 552:54–60. [https://doi.org/10.1016/s0014-5793\(03\)00850-0](https://doi.org/10.1016/s0014-5793(03)00850-0).
  40. Gancedo JM. 2013. Biological roles of cAMP: variations on a theme in the different kingdoms of life. *Biol Rev Camb Philos Soc* 88:645–668. <https://doi.org/10.1111/brv.12020>.
  41. Kalia D, Merey G, Nakayama S, Zheng Y, Zhou J, Luo Y, Guo M, Roembke BT, Sintim HO. 2013. Nucleotide, c-di-GMP, c-di-AMP, cGMP, cAMP, (p)ppGpp signaling in bacteria and implications in pathogenesis. *Chem Soc Rev* 42:305–341. <https://doi.org/10.1039/c2cs35206k>.
  42. Burroughs AM, Iyer LM, Aravind L. 2012. The natural history of ubiquitin and ubiquitin-related domains. *Front Biosci (Landmark Ed)* 17: 1433–1460. <https://doi.org/10.2741/3996>.
  43. Anantharaman V, Aravind L. 2002. MOSC domains: ancient, predicted sulfur-carrier domains, present in diverse metal-sulfur cluster biosynthesis proteins including molybdenum cofactor sulfurases. *FEMS Microbiol Lett* 207:55–61. <https://doi.org/10.1111/j.1574-6968.2002.tb11028.x>.
  44. Fekkes P, de Wit JG, Boorsma A, Friesen RH, Driessens AJ. 1999. Zinc stabilizes the SecB binding site of SecA. *Biochemistry* 38:5111–5116. <https://doi.org/10.1021/bi982818r>.
  45. Servin-Gonzalez L, Sampieri AI, Cabello J, Galvan L, Juarez V, Castro C. 1995. Sequence and functional analysis of the *Streptomyces* phaeochromogenes plasmid pJV1 reveals a modular organization of *Streptomyces* plasmids that replicate by rolling circle. *Microbiology* 141(Part 10):2499–2510. <https://doi.org/10.1099/13500872-141-10-2499>.
  46. Huang CH, Chen CY, Tsai HH, Chen C, Lin YS, Chen CW. 2003. Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon. *Mol Microbiol* 47:1563–1576. <https://doi.org/10.1046/j.1365-2958.2003.03403.x>.
  47. Yang Y, Kurokawa T, Takahama Y, Nindita Y, Mochizuki S, Arakawa K, Endo S, Kinashi H. 2011. pSLA2-M of *Streptomyces rochei* is a composite linear plasmid characterized by self-defense genes and homology with pSLA2-L. *Biosci Biotechnol Biochem* 75:1147–1153. <https://doi.org/10.1271/bbb.110054>.
  48. Zhang X, Shi H, Wu J, Zhang X, Sun L, Chen C, Chen ZJ. 2013. Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol Cell* 51:226–235. <https://doi.org/10.1016/j.molcel.2013.05.022>.
  49. Gao P, Ascano M, Zillinger T, Wang W, Dai P, Serganov AA, Gaffney BL, Shuman S, Jones RA, Deng L, Hartmann G, Barchet W, Tuschl T, Patel DJ. 2013. Structure-function analysis of STING activation by c[G(2',5')pA(3',5')p] and targeting by antiviral DMXAA. *Cell* 154: 748–762. <https://doi.org/10.1016/j.cell.2013.07.023>.
  50. Wu X, Wu FH, Wang X, Wang L, Siedlow JN, Zhang W, Pei ZM. 2014. Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic Acids Res* 42:8243–8257. <https://doi.org/10.1093/nar/gku569>.
  51. Ishikawa H, Ma Z, Barber GN. 2009. STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature* 461: 788–792. <https://doi.org/10.1038/nature08476>.
  52. Liu S, Cai X, Wu J, Cong Q, Chen X, Li T, Du F, Ren J, Wu YT, Grishin NV, Chen ZJ. 2015. Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation. *Science* 347:aaa2630. <https://doi.org/10.1126/science.aaa2630>.
  53. Tanaka Y, Chen ZJ. 2012. STING specifies IRF3 phosphorylation by TBK1 in the cytosolic DNA signaling pathway. *Sci Signal* 5:ra20. <https://doi.org/10.1126/scisignal.2002521>.
  54. Novatchkova M, Leibbrandt A, Werzowa J, Neubuser A, Eisenhaber F. 2003. The STIR-domain superfamily in signal transduction, development and immunity. *Trends Biochem Sci* 28:226–229. [https://doi.org/10.1016/S0968-0004\(03\)00067-7](https://doi.org/10.1016/S0968-0004(03)00067-7).
  55. Ponting CP. 2000. Novel repeats in ryanodine and IP3 receptors and protein O-mannosyltransferases. *Trends Biochem Sci* 25:48–50. [https://doi.org/10.1016/s0968-0004\(99\)01513-3](https://doi.org/10.1016/s0968-0004(99)01513-3).
  56. Anantharaman V, Koonin EV, Aravind L. 2001. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. *J Mol Biol* 307:1271–1292. <https://doi.org/10.1006/jmbi.2001.4508>.
  57. Essuman K, Summers DW, Sasaki Y, Mao X, Yim AKY, DiAntonio A, Milbradt J. 2018. TIR domain proteins are an ancient family of NAD(+)-consuming enzymes. *Curr Biol* 28:421–430.e4. <https://doi.org/10.1016/j.cub.2017.12.024>.
  58. Essuman K, Summers DW, Sasaki Y, Mao X, DiAntonio A, Milbradt J. 2017. The SARM1 Toll/interleukin-1 receptor domain possesses intrinsic NAD(+) cleavage activity that promotes pathological axonal degeneration. *Neuron* 93:1334–1343.e5. <https://doi.org/10.1016/j.neuron.2017.02.022>.
  59. Wan L, Essuman K, Anderson RG, Sasaki Y, Monteiro F, Chung EH, Osborne Nishimura E, DiAntonio A, Milbradt J, Dangl JL, Nishimura MT. 2019. TIR domains of plant immune receptors are NAD(+) -cleaving enzymes that promote cell death. *Science* 365:799–803. <https://doi.org/10.1126/science.aax1771>.
  60. Brzozowski RS, Huber M, Burroughs AM, Graham G, Walker M, Alva SS, Aravind L, Eswara PJ. 2019. Deciphering the role of a SLOG superfamily protein YpsA in Gram-positive bacteria. *Front Microbiol* 10:623. <https://doi.org/10.3389/fmicb.2019.00623>.
  61. Jawad Z, Paoli M. 2002. Novel sequences propel familiar folds. *Structure* 10:447–454. [https://doi.org/10.1016/s0969-2126\(02\)00750-5](https://doi.org/10.1016/s0969-2126(02)00750-5).
  62. Anantharaman V, Aravind L, Koonin EV. 2003. Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr Opin Chem Biol* 7:12–20. [https://doi.org/10.1016/s1367-5931\(02\)00018-2](https://doi.org/10.1016/s1367-5931(02)00018-2).
  63. Nurizzo D, Turkenburg JP, Charnock SJ, Roberts SM, Dodson EJ, McKie VA, Taylor EJ, Gilbert HJ, Davies GJ. 2002. *Cellvibrio japonicus* alpha-L-arabinanase 43A has a novel five-blade beta-propeller fold. *Nat Struct Biol* 9:665–668. <https://doi.org/10.1038/nsb835>.
  64. Aravind L, Koonin EV. 2000. Eukaryote-specific domains in translation initiation factors: implications for translation regulation and evolution of the translation system. *Genome Res* 10:1172–1184. <https://doi.org/10.1101/gr.10.8.1172>.
  65. Park JH, Aravind L, Wolff EC, Kaevel J, Kim YS, Park MH. 2006. Molecular cloning, expression, and structural prediction of deoxyhypusine hydroxylase: a HEAT-repeat-containing metalloenzyme. *Proc Natl Acad Sci U S A* 103:51–56. <https://doi.org/10.1073/pnas.0509348102>.
  66. Weijman JF, Kumar A, Jamieson SA, King CM, Caradoc-Davies TT, Ledgerwood EC, Murphy JM, Mace PD. 2017. Structural basis of auto-regulatory scaffolding by apoptosis signal-regulating kinase 1. *Proc Natl Acad Sci U S A* 114:E2096–E2105. <https://doi.org/10.1073/pnas.1620813114>.
  67. Samanovic MI, Tu S, Novak O, Iyer LM, McAllister FE, Aravind L, Gygi SP, Hubbard SR, Strnad M, Darwin KH. 2015. Proteasomal control of cytokinin synthesis protects *Mycobacterium tuberculosis* against nitric oxide. *Mol Cell* 57:984–994. <https://doi.org/10.1016/j.molcel.2015.01.024>.
  68. Tokunaga H, Kojima M, Kuroha T, Ishida T, Sugimoto K, Kiba T, Sakakibara H. 2012. *Arabidopsis lonely guy* (LOG) multiple mutants reveal a central role of the LOG-dependent pathway in cytokinin activation. *Plant J* 69:355–365. <https://doi.org/10.1111/j.1365-313X.2011.04795.x>.
  69. Schweizer U, Bohleber S, Frajedas-Villar N. 2017. The modified base isopentenyladenosine and its derivatives in tRNA. *RNA Biol* 14: 1197–1208. <https://doi.org/10.1080/15476286.2017.1294309>.
  70. Hrtyan M, Šliková E, Hejátko J, Růžička K. 2015. RNA processing in auxin

- and cytokinin pathways. *J Exp Bot* 66:4897–4912. <https://doi.org/10.1093/jxb/erv189>.
71. Hornung V, Hartmann R, Ablasser A, Hopfner KP. 2014. OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids. *Nat Rev Immunol* 14:521–528. <https://doi.org/10.1038/nri3719>.
  72. Zhang D, Burroughs AM, Vidal ND, Iyer LM, Aravind L. 2016. Transposons to toxins: the provenance, architecture and diversification of a widespread class of eukaryotic effectors. *Nucleic Acids Res* 44: 3513–3533. <https://doi.org/10.1093/nar/gkw221>.
  73. Skjerning RB, Senissar M, Winther KS, Gerdes K, Brodersen DE. 2019. The RES domain toxins of RES-Xre toxin-antitoxin modules induce cell stasis by degrading NAD<sup>+</sup>. *Mol Microbiol* 111:221–236. <https://doi.org/10.1111/mmi.14150>.
  74. de Souza RF, Aravind L. 2012. Identification of novel components of NAD-utilizing metabolic pathways and prediction of their biochemical functions. *Mol Biosyst* 8:1661–1677. <https://doi.org/10.1039/c2mb05487f>.
  75. Doeblemann G, Reissmann S, Assmann D, Fleckenstein M, Kahmann R. 2011. Two linked genes encoding a secreted effector and a membrane protein are essential for *Ustilago maydis*-induced tumour formation. *Mol Microbiol* 81:751–766. <https://doi.org/10.1111/j.1365-2958.2011.07728.x>.
  76. Gonzalez-Juarbe N, Gilley RP, Hinojosa CA, Bradley KM, Kamei A, Gao G, Dube PH, Bergman MA, Orihuela CJ. 2015. Pore-forming toxins induce macrophage necroptosis during acute bacterial pneumonia. *PLoS Pathog* 11:e1005337. <https://doi.org/10.1371/journal.ppat.1005337>.
  77. Coggill P, Bateman A. 2012. The YARHG domain: an extracellular domain in search of a function. *PLoS One* 7:e35575. <https://doi.org/10.1371/journal.pone.0035575>.
  78. Stoll R, Bosserhoff A. 2008. Extracellular SH3 domain containing proteins—features of a new protein family. *Curr Protein Pept Sci* 9:221–226. <https://doi.org/10.2174/138920308784534014>.
  79. Fastrez J. 1996. Phage lysozymes. *EXS* 75:35–64. [https://doi.org/10.1097/978-3-0348-9225-4\\_3](https://doi.org/10.1097/978-3-0348-9225-4_3).
  80. Hermoso JA, Monterroso B, Albert A, Galan B, Ahrazem O, Garcia P, Martinez-Ripoll M, Garcia JL, Menendez M. 2003. Structural basis for selective recognition of pneumococcal cell wall by modular endolysins from phage Cp-1. *Structure* 11:1239–1249. <https://doi.org/10.1016/j.str.2003.09.005>.
  81. Piuri M, Hatfull GF. 2006. A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. *Mol Microbiol* 62:1569–1585. <https://doi.org/10.1111/j.1365-2958.2006.05473.x>.
  82. Korner H, Sofia HJ, Zumft WG. 2003. Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol Rev* 27:559–592. [https://doi.org/10.1016/S0168-6445\(03\)00066-4](https://doi.org/10.1016/S0168-6445(03)00066-4).
  83. Dahlqvist A, Stahl U, Lenman M, Banas A, Lee M, Sandager L, Ronne H, Stymne S. 2000. Phospholipiddiacylglycerol acyltransferase: an enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proc Natl Acad Sci U S A* 97:6487–6492. <https://doi.org/10.1073/pnas.120067297>.
  84. Tall AR. 1990. Plasma high density lipoproteins. Metabolism and relationship to atherosclerosis. *J Clin Invest* 86:379–384. <https://doi.org/10.1172/JCI114722>.
  85. Alvarez HM. 2016. Triacylglycerol and wax ester-accumulating machinery in prokaryotes. *Biochimie* 120:28–39. <https://doi.org/10.1016/j.biochi.2015.08.016>.
  86. Kennedy EP. 1961. Biosynthesis of complex lipids. *Fed Proc* 20:934–940.
  87. Arabolaza A, Rodriguez E, Altabe S, Alvarez H, Gramajo H. 2008. Multiple pathways for triacylglycerol biosynthesis in *Streptomyces coelicolor*. *Appl Environ Microbiol* 74:2573–2582. <https://doi.org/10.1128/AEM.02638-07>.
  88. Peterson FC, Chen D, Lytle BL, Rossi MN, Ahel I, Denu JM, Volkman BF. 2011. Orphan macrodomain protein (human C6orf130) is an O-acetyl-ADP-ribose deacetylase: solution structure and catalytic properties. *J Biol Chem* 286:35955–35965. <https://doi.org/10.1074/jbc.M111.276238>.
  89. Schulze JM, Wang AY, Kobor MS. 2009. YEATS domain proteins: a diverse family with many links to chromatin modification and transcription. *Biochem Cell Biol* 87:65–75. <https://doi.org/10.1139/O08-111>.
  90. Aravind L, Zhang D, de Souza RF, Anand S, Iyer LM. 2015. The natural history of ADP-ribosyltransferases and the ADP-ribosylation system. *Curr Top Microbiol Immunol* 384:3–32. [https://doi.org/10.1007/82\\_2014\\_414](https://doi.org/10.1007/82_2014_414).
  91. Iyer LM, Zhang D, Aravind L. 2016. Adenine methylation in eukaryotes: apprehending the complex evolutionary history and functional potential of an epigenetic modification. *Bioessays* 38:27–40. <https://doi.org/10.1002/bies.201500104>.
  92. Corda D, Di Girolamo M. 2003. Functional aspects of protein mono-ADP-ribosylation. *EMBO J* 22:1953–1958. <https://doi.org/10.1093/emboj/cdg209>.
  93. West AH, Stock AM. 2001. Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* 26:369–376. [https://doi.org/10.1016/s0968-0004\(01\)01852-7](https://doi.org/10.1016/s0968-0004(01)01852-7).
  94. Marina A, Waldburger CD, Hendrickson WA. 2005. Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *EMBO J* 24:4247–4259. <https://doi.org/10.1038/sj.emboj.7600886>.
  95. Burroughs AM, Ando Y, Aravind L. 2014. New perspectives on the diversification of the RNA interference system: insights from comparative genomics and small RNA sequencing. *Wiley Interdiscip Rev RNA* 5:141–181. <https://doi.org/10.1002/wrna.1210>.
  96. Grebe TW, Stock JB. 1999. The histidine protein kinase superfamily. *Adv Microb Physiol* 41:139–227. [https://doi.org/10.1016/s0065-2911\(08\)60167-8](https://doi.org/10.1016/s0065-2911(08)60167-8).
  97. Kim DJ, Forst S. 2001. Genomic analysis of the histidine kinase family in bacteria and archaea. *Microbiology* 147:1197–1212. <https://doi.org/10.1099/00221287-147-5-1197>.
  98. Singh M, Berger B, Kim PS, Berger JM, Cochran AG. 1998. Computational learning reveals coiled coil-like motifs in histidine kinase linker domains. *Proc Natl Acad Sci U S A* 95:2738–2743. <https://doi.org/10.1073/pnas.95.6.2738>.
  99. Anantharaman V, Balaji S, Aravind L. 2006. The signaling helix: a common functional theme in diverse signaling proteins. *Biol Direct* 1:25. <https://doi.org/10.1186/1745-6150-1-25>.
  100. Varughese KI, Madhusudan, Zhou XZ, Whiteley JM, Hoch JA. 1998. Formation of a novel four-helix bundle and molecular recognition sites by dimerization of a response regulator phosphotransferase. *Mol Cell* 2:485–493. [https://doi.org/10.1016/s1097-2765\(00\)80148-3](https://doi.org/10.1016/s1097-2765(00)80148-3).
  101. Aravind L, Ponting CP. 1999. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiol Lett* 176:111–116. <https://doi.org/10.1111/j.1574-6968.1999.tb13650.x>.
  102. Hulko M, Berndt F, Gruber M, Linder JU, Truffault V, Schultz A, Martin J, Schultz JE, Lupas AN, Coles M. 2006. The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell* 126:929–940. <https://doi.org/10.1016/j.cell.2006.06.058>.
  103. Aravind L, Iyer LM, Anantharaman V. 2010. Natural history of sensor domains in bacterial signaling systems, p 1–38. In Spiro S, Dixon R (ed), *Sensory mechanisms in bacteria: molecular aspects of signal recognition*. Caister Academic Press, Norfolk, United Kingdom.
  104. Aravind L, Koonin EV. 1998. The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem Sci* 23: 469–472. [https://doi.org/10.1016/S0968-0004\(98\)01293-6](https://doi.org/10.1016/S0968-0004(98)01293-6).
  105. Ryan RP, Dow JM. 2010. Intermolecular interactions between HD-GYP and GGDEF domain proteins mediate virulence-related signal transduction in *Xanthomonas campestris*. *Virulence* 1:404–408. <https://doi.org/10.4161/viru.1.5.12704>.
  106. Hoch JA. 2000. Two-component and phosphorelay signal transduction. *Curr Opin Microbiol* 3:165–170. [https://doi.org/10.1016/s1369-5274\(00\)00070-9](https://doi.org/10.1016/s1369-5274(00)00070-9).
  107. Lewis RJ, Scott DJ, Brannigan JA, Ladds JC, Cervin MA, Spiegelman GB, Hoggett JG, Barak I, Wilkinson AJ. 2002. Dimer formation and transcription activation in the sporulation response regulator SpoOA. *J Mol Biol* 316:235–245. <https://doi.org/10.1006/jmbi.2001.5331>.
  108. Hwang I, Thorgeirsson T, Lee J, Kustu S, Shin YK. 1999. Physical evidence for a phosphorylation-dependent conformational change in the enhancer-binding protein NtrC. *Proc Natl Acad Sci U S A* 96:4880–4885. <https://doi.org/10.1073/pnas.96.9.4880>.
  109. Paul R, Weiser S, Amiot NC, Chan C, Schirmer T, Giese B, Jenal U. 2004. Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev* 18:715–727. <https://doi.org/10.1101/gad.289504>.
  110. O'Connor TJ, Nodwell JR. 2005. Pivotal roles for the receiver domain in the mechanism of action of the response regulator RamR of *Streptomyces coelicolor*. *J Mol Biol* 351:1030–1047. <https://doi.org/10.1016/j.jmb.2005.06.053>.
  111. Menon S, Wang S. 2011. Structure of the response regulator PhoP from

- Mycobacterium tuberculosis reveals a dimer through the receiver domain. *Biochemistry* 50:5948–5957. <https://doi.org/10.1021/bi2005575>.
112. Pisareva VP, Muslimov IA, Tcherepanov A, Pisarev AV. 2015. Characterization of novel ribosome-associated endoribonuclease SLFN14 from rabbit reticulocytes. *Biochemistry* 54:3286–3301. <https://doi.org/10.1021/acs.biochem.5b00302>.
  113. Yang JY, Deng XY, Li YS, Ma XC, Feng JX, Yu B, Chen Y, Luo YL, Wang X, Chen ML, Fang ZX, Zheng FX, Li YP, Zhong Q, Kang TB, Song LB, Xu RH, Zeng MS, Chen W, Zhang H, Xie W, Gao S. 2018. Structure of Schlafen13 reveals a new class of tRNA/rRNA-targeting RNase engaged in translational control. *Nat Commun* 9:1165. <https://doi.org/10.1038/s41467-018-03544-x>.
  114. Iyer LM, Zhang D, Rogozin IB, Aravind L. 2011. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res* 39:9473–9497. <https://doi.org/10.1093/nar/gkr691>.
  115. Mok BY, de Moraes MH, Zeng J, Bosch DE, Kotrys AV, Raguram A, Hsu F, Radey MC, Peterson SB, Mootha VK, Mougous JD, Liu DR. 2020. A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* 583:631–637. <https://doi.org/10.1038/s41586-020-2477-4>.
  116. Mangalea MR, Duerkop BA. 2020. Fitness trade-offs resulting from bacteriophage resistance potentiate synergistic antibacterial strategies. *Infect Immun* 88:e00926-19. <https://doi.org/10.1128/IAI.00926-19>.
  117. Hawse WF, Wolberger C. 2009. Structure-based mechanism of ADP-ribosylation by sirtuins. *J Biol Chem* 284:33654–33661. <https://doi.org/10.1074/jbc.M109.024521>.
  118. Adindla S, Inampudi KK, Guruprasad K, Guruprasad L. 2004. Identification and analysis of novel tandem repeats in the cell surface proteins of archaeal and bacterial genomes using computational tools. *Comp Funct Genomics* 5:2–16. <https://doi.org/10.1002/cfg.358>.
  119. Zhao D, Li Y, Xiong X, Chen Z, Li H. 2017. YEATS domain—a histone acylation reader in health and disease. *J Mol Biol* 429:1994–2002. <https://doi.org/10.1016/j.jmb.2017.03.010>.
  120. Li Y, Wen H, Xi Y, Tanaka K, Wang H, Peng D, Ren Y, Jin Q, Dent SY, Li W, Li H, Shi X. 2014. AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell* 159:558–571. <https://doi.org/10.1016/j.cell.2014.09.049>.
  121. Allison GE, Verma NK. 2000. Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *Trends Microbiol* 8:17–23. [https://doi.org/10.1016/s0966-842x\(99\)01646-7](https://doi.org/10.1016/s0966-842x(99)01646-7).
  122. Aravind L, Iyer LM, Koonin EV. 2006. Comparative genomics and structural biology of the molecular innovations of eukaryotes. *Curr Opin Struct Biol* 16:409–419. <https://doi.org/10.1016/j.sbi.2006.04.006>.
  123. Anantharaman V, Aravind L. 2002. The GOLD domain, a novel protein module involved in Golgi function and secretion. *Genome Biol* 3:research0023. <https://doi.org/10.1186/gb-2002-3-5-research0023>.
  124. Catazaro J, Lowe AJ, Cerny RL, Powers R. 2017. The NMR solution structure and function of RPA313: a putative ribosomal transport protein from Rhodopseudomonas palustris. *Proteins* 85:93–102. <https://doi.org/10.1002/prot.25201>.
  125. Michel B. 2005. After 30 years of study, the bacterial SOS response still surprises us. *PLoS Biol* 3:e255. <https://doi.org/10.1371/journal.pbio.0030255>.
  126. Blier PR, Griffith AJ, Craft J, Hardin JA. 1993. Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. *J Biol Chem* 268:7594–7601.
  127. Ueta M, Ohniwa RL, Yoshida H, Maki Y, Wada C, Wada A. 2008. Role of HPF (hibernation promoting factor) in translational activity in *Escherichia coli*. *J Biochem* 143:425–433. <https://doi.org/10.1093/jb/mvm243>.
  128. Richards GS, Degnan BM. 2009. The dawn of developmental signaling in the metazoa. *Cold Spring Harb Symp Quant Biol* 74:81–90. <https://doi.org/10.1101/sqb.2009.74.028>.
  129. Jessen S, Gu B, Dai X. 2008. Pygopus and the Wnt signaling pathway: a diverse set of connections. *Bioessays* 30:448–456. <https://doi.org/10.1002/bies.20757>.
  130. Huang Y, Fliegert R, Guse AH, Lu W, Du J. 2020. A structural overview of the ion channels of the TRPM family. *Cell Calcium* 85:102111. <https://doi.org/10.1016/j.ceca.2019.102111>.
  131. Himmel NJ, Gray TR, Cox DN. 2020. Phylogenetics identifies two eu-metazoan TRPM clades and an eighth TRP family, TRP soromelastatin (TRPS). *Mol Biol Evol* 37:2034–2044. <https://doi.org/10.1093/molbev/msaa065>.
  132. Kranzusch PJ, Wilson SC, Lee AS, Berger JM, Doudna JA, Vance RE. 2015. Ancient origin of cGAS-STING reveals mechanism of universal 2'3' cGAMP signaling. *Mol Cell* 59:891–903. <https://doi.org/10.1016/j.molcel.2015.07.022>.
  133. Ishikawa H, Barber GN. 2008. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature* 455:674–678. <https://doi.org/10.1038/nature07317>.
  134. Sun W, Li Y, Chen L, Chen H, You F, Zhou X, Zhou Y, Zhai Z, Chen D, Jiang Z. 2009. ERIS, an endoplasmic reticulum IFN stimulator, activates innate immune signaling through dimerization. *Proc Natl Acad Sci U S A* 106:8653–8658. <https://doi.org/10.1073/pnas.0900850106>.
  135. Barber GN. 2015. STING: infection, inflammation and cancer. *Nat Rev Immunol* 15:760–770. <https://doi.org/10.1038/nri3921>.
  136. Li T, Chen ZJ. 2018. The cGAS-cGAMP-STING pathway connects DNA damage to inflammation, senescence, and cancer. *J Exp Med* 215:1287–1299. <https://doi.org/10.1084/jem.20180139>.
  137. Mao C, Cook WJ, Zhou M, Koszalka GW, Krentsky TA, Ealick SE. 1997. The crystal structure of *Escherichia coli* purine nucleoside phosphorylase: a comparison with the human enzyme reveals a conserved topology. *Structure* 5:1373–1383. [https://doi.org/10.1016/s0969-2126\(97\)00287-6](https://doi.org/10.1016/s0969-2126(97)00287-6).
  138. Bateman A, Coggill P, Finn RD. 2010. DUFs: families in search of function. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 66:1148–1152. <https://doi.org/10.1107/S1744309110001685>.
  139. Obsil T, Obsilova V. 2017. Structural aspects of protein kinase ASK1 regulation. *Adv Biol Regul* 66:31–36. <https://doi.org/10.1016/j.jbbior.2017.10.002>.
  140. Nishida T, Hattori K, Watanabe K. 2017. The regulatory and signaling mechanisms of the ASK family. *Adv Biol Regul* 66:2–22. <https://doi.org/10.1016/j.jbbior.2017.05.004>.
  141. Schultz J, Ponting CP, Hofmann K, Bork P. 1997. SAM as a protein interaction domain involved in developmental regulation. *Protein Sci* 6:249–253. <https://doi.org/10.1002/pro.5560060128>.
  142. Kim CA, Bowie JU. 2003. SAM domains: uniform structure, diversity of function. *Trends Biochem Sci* 28:625–628. <https://doi.org/10.1016/j.tibs.2003.11.001>.
  143. Trevelyan SJ, Brewster JL, Burgess AE, Crowther JM, Cadell AL, Parker BL, Croucher DR, Dobson RCJ, Murphy JM, Mace PD. 2020. Structure-based mechanism of preferential complex formation by apoptosis signal-regulating kinases. *Sci Signal* 13:eaay6318. <https://doi.org/10.1126/scisignal.aay6318>.
  144. Nagai H, Noguchi T, Homma K, Katagiri K, Takeda K, Matsuzawa A, Ichijo H. 2009. Ubiquitin-like sequence in ASK1 plays critical roles in the recognition and stabilization by USP9X and oxidative stress-induced cell death. *Mol Cell* 36:805–818. <https://doi.org/10.1016/j.molcel.2009.10.016>.
  145. Stranzl GR, Santelli E, Bankston LA, La Clair C, Bobkov A, Schwarzenbacher R, Godzik A, Perego M, Grynberg M, Liddington RC. 2011. Structural insights into inhibition of *Bacillus anthracis* sporulation by a novel class of non-heme globin sensor domains. *J Biol Chem* 286:8448–8458. <https://doi.org/10.1074/jbc.M110.207126>.
  146. Miyakawa K, Matsunaga S, Kanou K, Matsuzawa A, Morishita R, Kudo A, Shindo K, Yokoyama M, Sato H, Kimura H, Tamura T, Yamamoto N, Ichijo H, Takaori-Kondo A, Ryo A. 2015. ASK1 restores the antiviral activity of APOBEC3G by disrupting HIV-1 Vif-mediated counteraction. *Nat Commun* 6:6945. <https://doi.org/10.1038/ncomms7945>.
  147. Okazaki T, Higuchi M, Takeda K, Iwatsuki-Horimoto K, Kiso M, Miyagishi M, Yanai H, Kato A, Yoneyama M, Fujita T, Taniguchi T, Kawaoka Y, Ichijo H, Gotoh Y. 2015. The ASK family kinases differentially mediate induction of type I interferon and apoptosis during the antiviral response. *Sci Signal* 8:ra78. <https://doi.org/10.1126/scisignal.lab1883>.
  148. Matsuzawa A, Saegusa K, Noguchi T, Sadamitsu C, Nishitoh H, Nagai S, Koyasu S, Matsumoto K, Takeda K, Ichijo H. 2005. ROS-dependent activation of the TRAF6-ASK1-p38 pathway is selectively required for TLR4-mediated innate immunity. *Nat Immunol* 6:587–592. <https://doi.org/10.1038/ni1200>.
  149. Gelezunas R, Xu W, Takeda K, Ichijo H, Greene WC. 2001. HIV-1 Nef inhibits ASK1-dependent death signalling providing a potential mechanism for protecting the infected host cell. *Nature* 410:834–838. <https://doi.org/10.1038/35071111>.
  150. Fujino G, Noguchi T, Matsuzawa A, Yamauchi S, Saitoh M, Takeda K, Ichijo H. 2007. Thioredoxin and TRAF family proteins regulate reactive oxygen species-dependent activation of ASK1 through reciprocal modulation of the N-terminal homophilic interaction of ASK1. *Mol Cell Biol* 27:8152–8163. <https://doi.org/10.1128/MCB.00227-07>.

151. del Sol A, Fujihashi H, Amoros D, Nussinov R. 2006. Residue centrality, functionally important residues, and active site shape: analysis of enzyme and non-enzyme families. *Protein Sci* 15:2120–2128. <https://doi.org/10.1110/ps.062249106>.
152. Saitoh M, Nishitoh H, Fujii M, Takeda K, Tobiume K, Sawada Y, Kawabata M, Miyazono K, Ichijo H. 1998. Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ASK) 1. *EMBO J* 17: 2596–2606. <https://doi.org/10.1093/emboj/17.9.2596>.
153. Kosek D, Kylarova S, Psenakova K, Rezabkova L, Herman P, Vecer J, Obsilova V, Obsil T. 2014. Biophysical and structural characterization of the thioredoxin-binding domain of protein kinase ASK1 and its interaction with reduced thioredoxin. *J Biol Chem* 289:24463–24474. <https://doi.org/10.1074/jbc.M114.583807>.
154. Nadeau PJ, Charette SJ, Toledano MB, Landry J. 2007. Disulfide bond-mediated multimerization of Ask1 and its reduction by thioredoxin-1 regulate H(2)O(2)-induced c-Jun NH(2)-terminal kinase activation and apoptosis. *Mol Biol Cell* 18:3903–3913. <https://doi.org/10.1091/mbc.e07-05-0491>.
155. Cockrell LM, Puckett MC, Goldman EH, Khuri FR, Fu H. 2010. Dual engagement of 14-3-3 proteins controls signal relay from ASK2 to the ASK1 signalosome. *Oncogene* 29:822–830. <https://doi.org/10.1038/onc.2009.382>.
156. Petrvalska O, Kosek D, Kukacka Z, Tosner Z, Man P, Vecer J, Herman P, Obsilova V, Obsil T. 2016. Structural insight into the 14-3-3 protein-dependent inhibition of protein kinase ASK1 (apoptosis signal-regulating kinase 1). *J Biol Chem* 291:20753–20765. <https://doi.org/10.1074/jbc.M116.724310>.
157. Kaur G, Iyer LM, Subramanian S, Aravind L. 2018. Evolutionary convergence and divergence in archaeal chromosomal proteins and chromo-like domains from bacteria and eukaryotes. *Sci Rep* 8:6196. <https://doi.org/10.1038/s41598-018-24467-z>.
158. Gazzaniga F, Stebbins R, Chang SZ, McPeek MA, Brenner C. 2009. Microbial NAD metabolism: lessons from comparative genomics. *Microbiol Mol Biol Rev* 73:529–541. <https://doi.org/10.1128/MMBR.00042-08>.
159. Galione A. 2006. NAADP, a new intracellular messenger that mobilizes Ca<sup>2+</sup> from acidic stores. *Biochem Soc Trans* 34:922–926. <https://doi.org/10.1042/BST0340922>.
160. Poole AZ, Weis VM. 2014. TIR-domain-containing protein repertoire of nine anthozoan species reveals coral-specific expansions and uncharacterized proteins. *Dev Comp Immunol* 46:480–488. <https://doi.org/10.1016/j.dci.2014.06.002>.
161. van der Burg CA, Prentis PJ, Surm JM, Pavasovic A. 2016. Insights into the innate immunome of actinarians using a comparative genomic approach. *BMC Genomics* 17:850. <https://doi.org/10.1186/s12864-016-3204-2>.
162. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
163. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR. 2015. HHMMER Web server: 2015 update. *Nucleic Acids Res* 43:W30–W38. <https://doi.org/10.1093/nar/gkv397>.
164. Hauser M, Steinegger M, Soding J. 2016. MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics* 32:1323–1330. <https://doi.org/10.1093/bioinformatics/btw006>.
165. Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol* 430:2237–2243. <https://doi.org/10.1016/j.jmb.2017.12.007>.
166. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427–D432. <https://doi.org/10.1093/nar/gky995>.
167. Lassmann T. 2020. Kalign 3: multiple sequence alignment of large data sets. *Bioinformatics* 36:1928–1929. <https://doi.org/10.1093/bioinformatics/btz795>.
168. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. <https://doi.org/10.1186/1471-2105-5-113>.
169. Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* 43: W389–W394. <https://doi.org/10.1093/nar/gkv332>.
170. Krishnan A, Iyer LM, Holland SJ, Boehm T, Aravind L. 2018. Diversification of AID/APOBEC-like deaminases in metazoa: multiplicity of clades and widespread roles in immunity. *Proc Natl Acad Sci U S A* 115: E3201–E3210. <https://doi.org/10.1073/pnas.1720897115>.
171. Holm L. 2019. Benchmarking fold detection by DALI Lite v.5. *Bioinformatics* 35:5326–5327. <https://doi.org/10.1093/bioinformatics/btz536>.
172. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>.