Title

A eukaryotic-like ubiquitination system in bacterial antiviral defence.

Journal

Nature: New biology, 631(8022)

Authors

Chambers, Lydia

Ye, Qiaozhen

Cai, Jiaxi

et al.

Publication Date

2024-07-01

DOI

10.1038/s41586-024-07730-4

Peer reviewed

Published in final edited form as:

Nature. 2024 July ; 631(8022): 843–849. doi:10.1038/s41586-024-07730-4.

A eukaryotic-like ubiquitination system in bacterial antiviral defense

Lydia R. Chambers[1,*], **Qiaozhen Ye**[2,*], **Jiaxi Cai**[3], **Minheng Gong**[2], **Hannah E. Ledvina**[4], **Huilin Zhou**[2], **Aaron T. Whiteley**[4], **Raymond T. Suhandynata**[5], **Kevin D. Corbett**[2,6,‡]

[1]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla CA, USA

[2]Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla CA, USA

[3]Department of Bioengineering, University of California San Diego, La Jolla CA, USA

[4]Department of Biochemistry, University of Colorado Boulder, Boulder CO, USA

[5]School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla CA, USA

[6]Department of Molecular Biology, University of California San Diego, La Jolla CA, USA

Summary

Ubiquitination pathways play crucial roles in protein homeostasis, signaling, and innate immunity[1–3]. In these pathways, an enzymatic cascade of E1, E2, and E3 proteins conjugates ubiquitin or a ubiquitin-like protein (Ubl) to target-protein lysine residues[4]. Bacteria encode ancient relatives of E1 and Ubl proteins involved in sulfur metabolism[5,6] but these proteins do not mediate Ubl-target conjugation, leaving open the question of whether bacteria can perform ubiquitination-like protein conjugation. Here, we demonstrate that a bacterial operon associated with phage defense islands encodes a complete ubiquitination pathway. Two structures of a bacterial E1:E2:Ubl complex reveal striking architectural parallels with canonical eukaryotic ubiquitination machinery. The bacterial E1 encodes an N-terminal inactive adenylation domain (IAD) and a C-terminal active adenylation domain (AAD) with a mobile α-helical insertion containing the catalytic cysteine (CYS domain). One structure reveals a pre-reaction state with the bacterial Ubl C-terminus positioned for adenylation, and a second structure mimics an E1-to-E2 transthioesterification state with the E1 CYS domain adjacent to the bound E2. We show

‡Correspondence: Correspondence and requests for materials should be addressed to K.D.C. (kcorbett@ucsd.edu).
*These authors contributed equally

Author Contributions

L.R.C. designed experiments, cloned expression vectors, purified proteins, determined crystal structures of $E1^{BilD}$:$E2^{BilB}$:Ubl^{BilA}, performed $E1^{BilD}$:$E2^{BilB}$:Ubl^{BilA} coexpression/activity assays, and wrote the paper with K.D.C.. **Q.Y.** designed experiments, cloned expression vectors, purified proteins, determined crystal structures of DUB^{BilC}:Ubl^{BilA}, performed DUB^{BilC} activity assays, and wrote the paper with K.D.C.. **J.C.** designed experiments and performed mass spectrometry analysis of Ubl^{BilA}-associated proteins. **M.G.** cloned expression vectors, purified proteins, and assisted with protein crystallization. **H.E.L.** performed bioinformatics analysis. **H.Z.** oversaw mass spectrometry analysis. **A.T.W.** designed experiments and oversaw bioinformatics analysis. **R.T.S.** designed experiments and oversaw mass spectrometry analysis. **K.D.C.** designed experiments, oversaw structural and biochemical assays, and wrote the paper with L.R.C. and Q.Y..

Competing Interests

The authors declare no competing interests.

that a deubiquitinase (DUB) in the same pathway pre-processes the bacterial Ubl, exposing its C-terminal glycine for adenylation. Finally, we show that the bacterial E1 and E2 collaborate to conjugate Ubl to target-protein lysine residues. Together, these data reveal that bacteria possess *bona fide* ubiquitination systems with strong mechanistic and architectural parallels to canonical eukaryotic ubiquitination pathways, suggesting that these pathways arose first in bacteria.

The conjugation of ubiquitin and Ubls to target proteins involves three enzymes termed E1, E2, and E3 (Extended Data Figure 1a). An E1 "activating protein" catalyzes the formation of a high-energy acyl adenylate intermediate through the reaction of ATP with the Ubl C-terminus, then a cysteine on the E1 attacks that intermediate to form a thioester link to the Ubl[4]. Next, a cysteine on an E2 "carrier protein" attacks the E1~Ubl thioester to form a second E2~Ubl thioester. Finally, conjugation of the Ubl to a lysine side chain on a target protein is mediated by an E3 "ligase," which may simply act as an E2-target adapter (as in RING-family E3s) or form a third cysteine~Ubl thioester intermediate before transfer to a target (as in HECT-family E3s). Some ubiquitination pathways circumvent the need for an E3, with their E2 proteins directly recognizing target proteins[7–9]. In all Ubl conjugation pathways, specific peptidases termed deubiquitinases or DUBs cleave Ubl-lysine isopeptide linkages in addition to pre-processing many Ubls to expose their reactive C-terminal glycine residue for E1-mediated catalysis[10].

Bacteria encode widespread E1-like adenylation enzymes (ThiF, MoeB) that act on ubiquitin-like proteins (ThiS, MoaD), but these pathways do not encode E2-like proteins and their primary biological roles are in sulfur metabolism rather than Ubl-target conjugation[5,6]. Mycobacteria and other Actinobacteria conjugate Pup (prokaryotic ubiquitin-like protein) to targets to mediate degradation, but the mechanism of Pup-target ligation is distinct from the canonical E1-E2-E3 ubiquitination pathway[11–15]. Many pathogenic bacteria encode E3 ligases and DUBs that modulate ubiquitin signaling by their eukaryotic host, but these do not constitute complete ubiquitination pathways[16–18].

Over the past two decades, sparsely distributed bacterial operons have been reported that encode distinct combinations of putative E1, E2, Ubl, and DUB proteins[19–22], plus rare examples of RING-type E3 proteins[23]. One such operon family, termed Bil (Bacterial ISG15-like gene) encodes putative E1, E2, Ubl, and DUB proteins and protects its host against bacteriophage (phage) infection by specifically modifying a virion structural protein[24,25]. These findings suggest that counter to prevailing models in which ubiquitination pathways arose in archaea[26,27], bacteria encode ancient ubiquitination-like pathways that participate in antiviral defense.

Identification of Type II BilABCD operons

We recently showed that bacterial Type II CBASS immune systems encode an E1-E2 fusion protein (Cap2) that resembles the noncanonical eukaryotic E1 Atg7 and its partner E2s Atg3 and Atg10 [28]. Cap2 conjugates a cGAS (cyclic GMP-AMP synthase)-like protein to target proteins using a ubiquitination-like mechanism[28,29]. We additionally described four bacterial operon families that encode different combinations of E1, E2, JAB-family DUB, and Ubl proteins, one of which encoded a multidomain E1 protein and an uncharacterized protein

termed CEHH after four conserved residues (cysteine, glutamate, histidine, histidine)[28]. In an earlier bioinformatic analysis of prokaryotic ubiquitination-related proteins, this operon family was denoted “6C” and described as encoding three proteins: multidomain E1, CEHH, and a JAB-family DUB[19]. AlphaFold structure predictions[30] and sequence alignments suggested that CEHH likely represents a diverged E2-like protein, and we further found that most CEHH-containing operons also encode a previously unidentified Ubl (Supplementary Table 1). Together, these observations suggested that the CEHH/6C operons may represent a complete ancestral ubiquitination system.

We performed comprehensive sequence searches using a multidomain E1 protein from *Ensifer aridi* TW10 and identified 136 complete nonredundant CEHH/6C operons, with most instances in the plant-associated Rhizobiaceae family (Figure 1a, Supplementary Table 1). Forty-one CEHH/6C operons are located in immune-enriched loci identified by the PADLOC server[31] (Extended Data Figure 1b, Supplementary Table 1), suggesting that they occur in so-called “defense islands” [31–34]. Further, twenty-five CEHH/6C operons are associated with CapH/CapP transcriptional regulators, which are associated with diverse bacterial immune systems and which activate these systems’ expression in response to DNA damage[35]. Together, these data show that CEHH/6C operons encode a complete ubiquitination-like pathway with predicted E1, E2, Ubl, and DUB proteins, and further suggest that they participate in antiviral defense.

We observed that the gene order in CEHH/6C operons (Ubl–CEHH/E2–DUB–E1) matches that of BilABCD operons (Ubl–E2–DUB–E1; Extended Data Figure 1c)[24]. Sequence alignments of the E1, E2, and DUB-like proteins in CEHH/6C and BilABCD operons show that in each case, these proteins share identifiable sequence homology but segregate into distinct groups in unrooted evolutionary trees (Extended Data Figure 2a–c). The broad parallels between CEHH and BilABCD operons, combined with these specific differences, prompted us to rename the CEHH/6C operons as Type II BilABCD operons.

Bacterial E1-E2-Ubl complex structure

We cloned and expressed the four proteins from a Type II BilABCD operon from *Ensifer aridi* TW10 (Ubl^{BilA}, $E2^{BilB}$, DUB^{BilC}, and $E1^{BilD}$) for biochemical and structural analysis. While both $E1^{BilD}$ and $E2^{BilB}$ were insoluble when overexpressed in *E. coli*, they formed a soluble complex when coexpressed with Ubl^{BilA}. We crystallized and determined two independent X-ray crystal structures of the $E1^{BilD}$:$E2^{BilB}$:Ubl^{BilA} complex to 2.5 Å resolution (Form 1) and 2.7 Å resolution (Form 2) (Figure 1b, Extended Data Table 1).

$E1^{BilD}$ encodes an N-terminal domain of unknown function (Pfam14459) annotated as “prokaryotic E2 family C” and a C-terminal E1-like adenylation domain (Extended Data Figure 1c), and was originally speculated to comprise an E2-E1 fusion protein[19]. Our structure shows that $E1^{BilD}$ adopts an overall architecture strikingly similar to that of canonical eukaryotic E1 proteins (Figure 1b, Figure 2a–b, Extended Data Figure 3). Whereas all known bacterial E1 and E1-like proteins including ThiS, MoeB, and Cap2 form homodimers through their adenylation domains[5,6,28], $E1^{BilD}$ possesses an N-terminal inactive adenylation domain (IAD) and a C-terminal active adenylation domain (AAD) that

form a structurally related pseudo-dimer. Inserted into the "crossover loop" of the E1^{BilD} AAD is an ~80 amino acid α-helical domain containing a conserved cysteine residue (CYS domain). This domain architecture – IAD, AAD, and CYS – is a hallmark of canonical eukaryotic E1 proteins. The closest structural relatives of E1^{BilD} are the human ubiquitin/FAT10 E1 UBA6 [36] and the heterodimeric human NEDD8 E1 NAE1-UBA3[37] (Figure 2a–b, Table 1, Extended Data Figure 3). Like these eukaryotic E1s, E1^{BilD} possesses a single active adenylation site in the AAD, with one key arginine residue (Arg9) provided by the IAD (Figure 1c, Extended Data Figure 3c, i). The E1^{BilD} CYS domain is linked to the AAD via flexible linkers and is highly mobile, showing high B-factors compared to the IAD/AAD and adopting different positions in Form 1 versus Form 2 structures. In the Form 1 structure, E1^{BilD} adopts an open-like conformation with the CYS domain positioned directly above the adenylation active site and the conserved cysteine residue (Cys417) 14 Å away from the C-terminus of the bound Ubl^{BilC} (Figure 1c). In Form 2, the CYS domain is rotated upward to contact the bound E2^{BilB} (see below). E1^{BilD} coordinates a zinc ion through four cysteine residues, two in the AAD crossover loop and two near the protein's C-terminus (Figure 1d); similar structural zinc ions are observed in SUMO and NEDD8 E1 proteins (Extended Data Figure 3d,j)[4,37,38]. Overall, while E1^{BilD} lacks accessory domains found in many eukaryotic E1 proteins –including the first catalytic cysteine half-domain (FCCH) or coiled-coil domain (CC) that is often inserted within the IAD, and the C-terminal ubiquitin fold domain (UFD) that mediates E2 binding – the overall structural parallels between E1^{BilD} and canonical eukaryotic E1 proteins suggest a common ancestry.

In contrast to identified Type I BilABCD operons whose Ubl (BilA) contains two predicted ubiquitin-like β-grasp domains[24], *E. aridi* Ubl^{BilA} contains a single β-grasp domain (Figure 2c). Ubl^{BilA} is structurally more similar to eukaryotic ubiquitin and SUMO than it is to the bacterial β-grasp proteins ThisS and MoaD (1.7 Å Cα r.m.s.d. for eukaryotic homologs, 3.3 Å for bacterial homologs; Table 1). While *E. aridi* Ubl^{BilA} encodes a single β-grasp domain, Ubl^{BilA} proteins across Type II BilABCD operons show remarkable structural diversity: we identified examples of Ubls predicted to encode up to three tandem β-grasp domains, or containing N-terminal domains with predicted coiled-coil or disordered regions (Extended Data Figure 2d) This variety in Ubl architecture suggests that ubiquitination by BilABCD operons could disrupt target-protein function through diverse mechanisms.

In our structures of the E1^{BilD}:E2^{BilB}:Ubl^{BilA} complex, Ubl^{BilA} is bound to E1^{BilD} equivalently to known Ubl-E1 complexes, with its C-terminus positioned in the E1^{BilD} adenylation active site. We do not observe electron density for a bound ATP or AMP molecule in the active site. Indeed, the C-terminal alanine residue of Ubl^{BilA} (Ala98) is positioned in a manner that would physically prevent binding of ATP. In a sequence alignment of all Ubl^{BilA} homologs from Type II BilABCD operons, we found that these proteins encode a universally-conserved glycine followed by up to nine residues of non-conserved sequence (Extended Data Figure 2d). In *E. aridi* Ubl^{BilA}, the conserved glycine (Gly97) is positioned one residue from the C-terminus. Close inspection of our structures reveals that Ubl^{BilA} residue Gly97 is positioned properly for adenylation, but cannot undergo this reaction because of the presence of the C-terminal residue Ala98 (Extended Data Figure 3c, i). Thus, in this system Ubl^{BilA} likely requires proteolytic processing by

its cognate DUB^{BilC} to expose the reactive Gly97 residue for catalysis by E1^{BilD} (see next section).

In our structures, E2^{BilB} adopts a fold similar to canonical E2 proteins from *S. cerevisiae* and *H. sapiens* (Figure 2d, Table 1). The conserved CEHH motif in this protein includes residues Cys138, Glu144, His146, and His151 [19]. Of these, Cys138 is positioned equivalently to canonical E2 proteins' catalytic cysteine. His151 is adjacent to Cys138, and may participate directly in catalysis. The other two conserved residues (Glu144 and His146) appear to play purely structural roles in E2^{BilB}, with their side chains hydrogen-bonding to nearby backbone amide and carbonyl groups (Extended Data Figure 4a). In keeping with their likely functional importance, the putative catalytic cysteine (Cys138) and its adjacent histidine (His151) are highly conserved in E2^{BilB} proteins across both Type I and Type II BilABCD operons, while Glu144 and His146 are not conserved in Type I BilABCD (Extended Data Figure 4a).

In both structures of the E1^{BilD}:E2^{BilB}:Ubl^{BilA} complex, E2^{BilB} is bound to the E1^{BilD} AAD near the adenylation active site and CYS domain (Figure 1b). While E1^{BilD} lacks the C-terminal UFD that participates in E2 binding in many canonical E1 proteins, the binding of E2^{BilB} to E1^{BilD} nonetheless closely resembles canonical E1-E2 binding. The interaction involves a loop and α-helix near the C-terminus of E1^{BilD} (residues 490-504), which is rigidified by the bound Zn^{2+} ion coordinated by Cys340 and Cys343 in the AAD crossover loop, plus Cys491 and Cys493 near the E1^{BilD} C-terminus (Figure 1d, Extended Data Figure 3e,j). In the E1^{BilD}:E2^{BilB}:Ubl^{BilA} Form 2 structure, the E1^{BilD} CYS domain is rotated away from the adenylation active site, and its putative catalytic cysteine is positioned adjacent to the E2^{BilB} putative catalytic cysteine (Cys138; Figure 1e, Extended Data Figure 3f). In the structure, the two residues are apparently linked via a disulfide bond. This state closely mimics the structural state of an E1-to-E2 transthioesterification intermediate, in which the E1 and E2 catalytic cysteine residues approach to within ~2 Å of one another[39]. Thus, our two structures reveal how the mobile E1^{BilD} CYS domain likely reacts first with a Ubl^{BilA}-adenylate intermediate to generate the E1^{BilD}~Ubl^{BilA} thioester, then rotates upward to mediate Ubl^{BilA} handoff to E2^{BilB}.

In both structures of the E1^{BilD}:E2^{BilB}:Ubl^{BilA} complex, E2^{BilB} forms a symmetric homodimer either through non-crystallographic symmetry (Form 1) or crystallographic symmetry (Form 2; Extended Data Figure 4b–d). Supporting the idea that E2^{BilB} forms a homodimer, size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) shows that a purified E1^{BilD} (C417A):E2^{BilB}:Ubl^{BilA} complex forms a 2:2:2 complex in solution, while complexes with E2^{BilB} mutated to disrupt the dimer interface form 1:1:1 complexes (Extended Data Figure 4e–f).

Ubl preprocessing by a bacterial DUB

In eukaryotic Ubl pathways, deubiquitinases play two roles: they pre-process Ubls to expose their reactive C-terminal glycine (peptide bond cleavage), and they also cleave Ubl-target conjugates (isopeptide bond cleavage)[10]. Our structures and sequence alignments suggest that in many BilABCD systems including that of *E. aridi*, DUB^{BilC} is required to

pre-process Ubl^{BilA} to expose the reactive C-terminal glycine. To test the activity of *E. aridi* DUB^{BilC}, we generated a model substrate comprising Ubl^{BilA} fused at its C-terminus to green fluorescent protein (GFP). We found that wild type DUB^{BilC}, but not two different active-site mutants (Glu33 to Ala (E33A) and Asp106 to Ala(D106A)), showed robust cleavage of Ubl^{BilA}-GFP (Figure 3a). We isolated the C-terminal product of Ubl^{BilA}-GFP cleavage and subjected it to N-terminal sequencing by Edman degradation, and found that DUB^{BilC} cleaves Ubl^{BilA} one residue upstream of the C-terminus, between Gly97 and Ala98 (Figure 3b–c, Extended Data Figure 5a). Thus, *E. aridi* DUB^{BilC} likely pre-processes Ubl^{BilA} to expose the reactive Gly97 for catalysis.

We next purified a complex of *E. aridi* DUB^{BilC}(E33A) and full-length Ubl^{BilA}, and determined two X-ray crystal structures at 1.36 Å resolution (Form 1) and 1.68 Å resolution (Form 2) (Figure 3d–e, Extended Data Figure 5b–c, Extended Data Table 1). The structures reveal DUB^{BilC} as a JAB/JAMM family peptidase, with the closest structural homologs being an archaeal JAMM peptidase and several eukaryotic AMSH enzymes (Extended Data Figure 5d). In both structures, Ubl^{BilA} is bound to DUB^{BilC} with its C-terminus positioned in the active site near the catalytic Zn^{2+} ion (Figure 3d–e). Despite using the catalytic mutant DUB^{BilC}(E33A) for crystallization, we observed that in both structures Ubl^{BilA} was cleaved at residue Gly97. This observation is likely explained by residual low-level cleavage activity in this mutant (Figure 3a). These structural data support a model in which *E. aridi* DUB^{BilC} pre-processes Ubl^{BilA} by cleaving between Gly97 and Ala98.

BilABCD mediates Ubl-lysine conjugation

Our structural data on E1^{BilD}, E2^{BilB}, and Ubl^{BilA} suggest that these proteins comprise a complete ubiquitination system equivalent to canonical eukaryotic ubiquitination pathways. To test this idea, we coexpressed E1^{BilD} and E2^{BilB} with Ubl^{BilA} truncated at Gly97 (Ubl^{BilA}Δ97) in *E. coli*. Using an N-terminal His_6-tag on Ubl^{BilA}, we purified Ubl^{BilA} and associated proteins. In native buffer conditions that maintain both covalent and noncovalent complexes with Ubl^{BilA}, we observed three major bands representing E1^{BilD}, E2^{BilB}, and Ubl^{BilA}, plus minor bands with a wide range of molecular weights (Figure 4a). These minor bands were not present when Ubl^{BilA} contained its native C-terminal Ala98 residue; when E1 was mutated to eliminate Ubl^{BilA} adenylation (Arg246 to Ala); when the E1^{BilD} putative catalytic cysteine (Cys417) was mutated to alanine; or when the E2^{BilB} putative catalytic cysteine (Cys138) was mutated to alanine (Figure 4a). These data suggest that the observed minor bands represent covalent Ubl^{BilA}-target protein conjugates that depend on Ubl^{BilA} adenylation by E1^{BilD}, followed by sequential thioester formation with E1^{BilD} and E2^{BilB}. A similar experiment with E2^{BilB} dimer mutants shows that these mutants also generate Ubl^{BilA}-target conjugates (Extended Data Figure 6a), albeit at a lower level than wild type E2^{BilB}; thus, E2^{BilB} dimerization is not strictly required for its catalytic activity.

We next purified His_6-tagged Ubl^{BilA}Δ97 in denaturing conditions to eliminate copurification of non-covalently associated proteins. When coexpressed with wild-type E1^{BilD} and E2^{BilB}, we again observed that Ubl^{BilA} copurified with a range of minor bands, which were not present when either the E1 or E2 putative catalytic cysteine residues were mutated to alanine (Figure 4b). These data confirm that the minor bands are likely covalent

complexes between Ubl^{BilA} and *E. coli* proteins, that are generated by E1^{BilD} and E2^{BilB}. We incubated the purified mixture of Ubl^{BilA}-target conjugates with purified DUB^{BilC}, then visualized the reactions by SDS-PAGE and western blotting with an antibody that recognizes the His$_6$-tag on Ubl^{BilA}. Untreated samples showed a large number of bands, suggesting that each band represents an *E. coli* protein conjugated to His$_6$-tagged Ubl^{BilA} (Figure 4c). After incubation with wild-type DUB^{BilC}, the minor bands were mostly lost and western blotting showed a single major band corresponding to monomeric Ubl^{BilA} (Figure 4c). The minor bands were not lost upon incubation with the DUB^{BilC} catalytic mutants E33A or D106A, confirming that this loss is due to proteolysis by DUB^{BilC}. Thus, E1^{BilD} and E2^{BilB} can generate covalent Ubl^{BilA}-target conjugates, which DUB^{BilC} can cleave.

Next, we coexpressed all four proteins (E1^{BilD}, E2^{BilB}, His$_6$-Ubl^{BilA}, and DUB^{BilC}) in *E. coli* and purified Ubl^{BilA} and associated proteins. When Ubl^{BilA} was full-length (ending at Ala98), we observed the formation of Ubl^{BilA}-target conjugates in the presence of wild-type DUB^{BilC}, but not the DUB^{BilC} E33A active-site mutant (Extended Data Figure 6b–c). These data demonstrate that DUB^{BilC} can pre-process Ubl^{BilA} to expose the reactive Gly97 residue and enable catalysis by E1^{BilD} and E2^{BilB}.

Finally, we used mass spectrometry (MS) to verify that Ubl^{BilA} is conjugated to target proteins through an isopeptide linkage. We generated a His$_6$-tagged Ubl^{BilA}Δ97 construct with a Val95 to Lys (V95K) mutation, such that trypsin digestion of Ubl^{BilA}-target conjugates would leave an alanine-glycine dipeptide linked to a target lysine residue (Figure 4d). We verified that the V95K mutant did not disrupt Ubl^{BilA}-target conjugate formation (Extended Data Figure 6d), then purified His$_6$-Ubl^{BilA}(Δ97, V95K) in denaturing conditions after coexpression with E1^{BilD} and E2^{BilB}. We separated the purified proteins by SDS-PAGE, then excised a region of the gel containing ~20-40 kDa proteins (importantly, excluding monomeric Ubl^{BilA}). We extracted proteins from the gel, cleaved with trypsin, and performed LC-MS/MS with data-dependent acquisition to search for peptides that contained a lysine residue modified with an alanine-glycine dipeptide. We identified a peptide corresponding to residues 76-95 of Ubl^{BilA}(Δ97, V95K) modified at Lys92, then modified the MS parameters to selectively detect its daughter ions (Figure 4e). All detected daughter ions showed identical chromatographic elution profiles by reverse-phase liquid chromatography (Figure 4f), supporting the identification of this modified peptide. We similarly identified the unmodified Ubl^{BilA}(Δ97, V95K) 76-95 peptide with an essentially identical chromatographic elution profile (Figure 4g), and the Ubl^{BilA}(Δ97, V95K) 76-92 peptide that is properly cleaved by trypsin at Lys92 (Extended Data Figure 6e). These data confirm that E1^{BilD} and E2^{BilB} can catalyze formation of a *bona fide* Ubl^{BilA}-target isopeptide linkage.

Discussion

Bacteria encode widespread E1- and Ubl-like proteins involved in sulfur metabolism, but true ubiquitin-like protein conjugation was not demonstrated in bacteria until the CBASS-associated protein Cap2 was shown to conjugate a cGAS-like protein to target proteins as part of an antiviral immune response[28,29]. In a variation on this theme, other CBASS systems encode an E2-like regulator that mediates covalent oligomerization of

its cognate cGAS-like protein through a ubiquitination-like mechanism[40]. Here, we reveal that bacterial BilABCD operons encode full ubiquitination systems with E1, E2, DUB, and Ubl proteins that architecturally and mechanistically resemble canonical eukaryotic ubiquitination machinery. We demonstrate that $E1^{BilD}$ encodes tandem IAD and AAD domains, with a mobile CYS domain capable of shuttling Ubl^{BilA} from the $E1^{BilD}$ adenylation site to $E2^{BilB}$. We show that all BilABCD proteins more closely resemble canonical eukaryotic ubiquitination machinery than functionally related bacterial proteins. When coexpressed in *E. coli*, $E1^{BilD}$ and $E2^{BilB}$ mediate Ubl^{BilA} conjugation to both cellular proteins and Ubl^{BilA} itself. Finally, the DUB protein BilC pre-processes Ubl^{BilA} to expose its reactive C-terminal glycine, and can also cleave Ubl^{BilA}-target conjugates. Overall, these data show that bacteria encode ancient relatives of canonical ubiquitination pathways. Remarkably, several other uncharacterized bacterial operon families are predicted to encode distinct combinations of Ubl, E1, E2, RING E3, and DUB proteins[19–22,28]. As with Type II CBASS and BilABCD systems, these operon families are predicted to function in antiviral defense[19–22]. Thus, the evolutionary pressure of the phage-bacteria arms race has given rise to a profusion of ubiquitination-like protein conjugation pathways that likely interrupt phage infections through a variety of mechanisms.

Our data reveal that when expressed in a heterologous host, *E. aridi* $E1^{BilD}$ and $E2^{BilB}$ can mediate nonspecific Ubl^{BilA}-target conjugation. In the context of bacteriophage infection, BilABCD could protect its host by non-specifically ubiquitinating abundant phage proteins, potentially inhibiting their self-assembly into functional phage progeny. A more likely scenario is that BilABCD specifically ubiquitinates one or a few phage proteins to interrupt key life-cycle events like DNA replication, virion assembly, or host-cell lysis. Alternatively, BilABCD could cause production of progeny phage that are non-infectious due to ubiquitination of key structural proteins. Indeed, recent work on a Type I BilABCD operon from *Collimonas sp.* OK412 shows that this system specifically ubiquitinates the central tail fiber protein of phages Secphi27 and Secphi4, and that this activity results in both defective virion assembly and impaired infectivity of progeny phage[25]. We propose that $E2^{BilB}$ may directly recognize phage target proteins, providing specificity to this system in the absence of a dedicated E3 protein.

BilABCD operons were originally dubbed “Bacterial ISG15-like” because of shared innate-immune function and structural similarity between Ubl^{BilA} and the eukaryotic Ubl ISG15 [24]. Our data solidify the functional parallels between bacterial BilABCD pathways and canonical eukaryotic ubiquitination pathways, including the innate-immune ISG15 pathway. ISG15 and its E1 UBA7 are found only in vertebrates and likely evolved from UBB/UBC and UBA1, respectively[20,41]; this pathway is therefore not a direct descendent of bacterial BilABCD. Nonetheless, these pathways are likely to function similarly; in bacterial immunity, Ubl conjugation in different BilABCD systems likely has diverse effects on infecting phages, including the known effects on virion assembly and infectivity through modification of a key tail fiber protein[25], plus potentially aberrant oligomerization or aggregation of modified target proteins mediated by diverse Ubls. The ISG15 pathway has been proposed to function similarly, by modifying viral proteins to interrupt key life-cycle events[3,42,43]. While the ISG15 pathway appears to have emerged after the establishment of eukaryotes rather than being inherited from a bacterial or archaeal ancestor[20], the functional

parallels between BilABCD and ISG15 pathways strongly point to the effectiveness of ubiquitination-like protein conjugation pathways in mediating innate immunity across kingdoms.

Methods

Bioinformatics

To identify bacterial CEHH operons, the *Ensifer aridi* TW10 E1^{BilD} protein (IMG accession 2509379665) was used as a query for BLAST searches in the IMG database of bacterial genomes (https://img.jgi.doe.gov/). Hit sequences were filtered for redundancy, aligned using the MAFFT[44] version 7 web server (https://mafft.cbrc.jp/alignment/server/index.html) and visualized in Jalview[45]. For each non-redundant hit, the genomic neighborhood of the E1^{BilD} gene was visually inspected in IMG for neighboring genes similar to *E. aridi* Ubl^{BilA}, E2^{BilB}, and DUB^{BilC}, plus genes similar to CapH (usually annotated as Pfam01381; XRE-family HTH domain) and CapP (usually annotated as Pfam06114; Peptidase_M78)[35]. In all cases, the *bilA-bilB-bilC-bilD* gene order was consistent, and in all but one case *capH* and *capP* were positioned upstream of *bilABCD* and oriented in the opposite coding direction (i.e. sharing a common promoter region with *bilABCD*). Average linkage (UPGMA) trees were calculated by the MAFFT web server and visualized using the Interactive Tree of Life server (https://itol.embl.de/). Protein structure predictions (monomer and multimer) were performed using the AlphaFold_MMseqs2 implementation of ColabFold[46], which implements AlphaFold 2 [30,47] predictions using sequence alignments generated by MMseqs2 [48].

Protein expression and purification

Ensifer aridi Ubl^{BilA} (Joint Genome Institute Integrated Microbial Genomes (JGI-IMG) accession 2509379668), E2^{BilB} (IMG accession 2509379667), DUB^{BilC} (IMG accession 2509379666), and E1^{BilD} (IMG accession 2509379665) were individually cloned into *E. coli* expression vectors encoding either no tag (UC Berkeley Macrolab vector 2A-T, Addgene ID 29665) or an N-terminal TEV protease-cleavable His_6-tag (UC Berkeley Macrolab vector 2B-T, Addgene ID 29666). For coexpression, multigene coexpression cassettes were generated by PCR and cloned into UC Berkeley Macrolab vector 2B-T to generate an N-terminal TEV protease-cleavable His_6-tag on one protein. Targeted mutants were generated by PCR-based mutagenesis.

Vectors were transformed into *E. coli* Rosetta 2 pLysS (EMD Millipore), and 1 L cultures were grown at 37°C in 2XYT media to an OD_{600} of 0.6 before induction with 0.25 mM IPTG at 20°C for 16–18 hours. Cells were harvested by centrifugation, resuspended in buffer A (25 mM Tris-HCl pH 8.5, 300 mM NaCl, 5 mM $MgCl_2$, 10% glycerol and 5 mM mercaptoethanol) containing 5 mM imidazole, then lysed by sonication (Branson Sonifier). Lysates were clarified by centrifugation, then supernatants were passed over a Ni-NTA Superflow column (Qiagen) in resuspension buffer. The column was washed in wash buffer (buffer A containing 20 mM imidazole), then eluted in elution buffer (buffer A containing 400 mM imidazole). Eluates were concentrated by ultrafiltration (Amicon Ultra; EMD Millipore), then passed over a Superdex 200 Increase size exclusion column (Cytiva) in size

exclusion buffer (25 mM Tris-HCl pH 8.5,300 mM NaCl, 5 mM $MgCl_2$, 10% glycerol and 1mM DTT). Peak fractions were concentrated by ultrafiltration and stored at 4°C.

For denatured purification of $E1^{BilD}$:$E2^{BilB}$:His_6-Ubl^{BilA}, cells were harvested by centrifugation, resuspended in buffer A, then lysed by sonication (Branson Sonifier). Lysates were clarified by centrifugation, then supernatants were passed over a Ni-NTA Superflow column (Qiagen) in resuspension buffer. The column was washed in buffer B (100 mM Tris-HCl pH 8, 8 M Urea and 100 mM NaH_2PO_4), followed by wash buffer (buffer A containing 20 mM imidazole), then eluted in elution buffer (buffer A containing 400 mM imidazole) . Eluates were concentrated by ultrafiltration (Amicon Ultra; EMD Millipore) and stored at 4°C.

For characterization of oligomeric state by size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS), 100 µl of purified protein complexes at a concentration of 5 mg/ml were injected onto a size exclusion column (Superdex 200 Increase 10/300 GL, Cytiva) in size exclusion buffer, then light scattering and differential refractive index (dRI) profiles were collected using miniDAWN TREOS and Optilab T-rEX detectors (Wyatt Technology). SEC-MALS data were analyzed using ASTRA software version 8 and visualized with Prism version 10 (GraphPad Software).

DUB^{BilC} activity assays

To generate a model substrate for DUB^{BilC} cleavage, Ubl^{BilA} was cloned into a vector encoding a C-terminal GFP tag (UC Berkeley Macrolab vector H6-msfGFP, Addgene ID 29725) and purified as above. The Ubl^{BilA}-GFP fusion protein (10 µg) was mixed with 5 µg of DUB^{BilC} (wild type or mutants) in 20 µL reaction buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, 20 mM $MgCl_2$, 20 µM $ZnCl_2$ and 1 mM DTT, then incubated 30 minutes at 37°C. Reactions were analyzed by SDS-PAGE with Coomassie blue staining. For protein N-terminal sequencing by Edman degradation, cleavage products were separated by SDS-PAGE, transferred to a Bio-Rad ImmunBlot PVDF membrane, and visualized by staining with Coomassie Blue R-250. The band representing the C-terminal cleavage product of Ubl^{BilA}-GFP was excised, the membrane was washed with methanol and deionized water, then loaded onto a Shimadzu PPSQ-53A instrument for analysis. Five rounds of Edman degradation were performed, and the “evaluated value” scores for each round were analyzed to obtain the likely N-terminal sequence.

Western blotting

For western blots of His_6-Ubl^{BilA} conjugates, proteins were transferred to PVDF membranes using a Trans-Blot Turbo RTA Mini 0.2 µm PVDF Transfer Kit (Bio-Rad) according to the manufacturer’s instructions, using a Trans-Blot Turbo Transfer System (Bio-Rad). Membranes were blocked with 5% Non-Fat Dry Milk in TBST, then incubated with mouse anti-His tag primary antibody (Millipore Sigma SAB1305538) at 1:1,000 dilution, followed by HRP-conjugated secondary antibody (HRP Goat anti-mouse IgG, Millipore Sigma AP128P) at 1:30,000 dilution. After washing, HRP signal was detected with Amersham ECL Select Western Blotting Detection Reagent (Cytiva) using a ChemiDoc Imaging System (Bio-Rad) in Protein Blot - Chemiluminescence setting.

Crystallography

For crystallization of the *E. aridi* DUB^{BilC}(E33A):Ubl^{BilA} complex (Form 1), purified protein at 30 mg/mL in crystallization buffer (25 mM Tris-HCl pH 8.5, 200 mM NaCl, 5 mM $MgCl_2$, and 1 mM TCEP (tris(2-carboxyethyl)phosphine)) was mixed 1:1 with well solution containing 100 mM HEPES pH 7.5, 0.2 M $MgCl_2$, and 25% PEG 3350 in hanging drop format. Crystals were harvested into cryoprotectant solution containing an additional 10% glycerol and frozen in liquid nitrogen. Diffraction data were collected at the Advanced Photon Source (Argonne National Lab) NE-CAT beamline 24ID-E on April 7, 2023 (collection temperature 100 K; x-ray wavelength 0.97918 Å) (Extended Data Table 1). Data were processed with the RAPD data-processing pipeline (https://github.com/RAPD/RAPD), which uses XDS[49] for data indexing and reduction, POINTLESS[50] for space group assignment, and AIMLESS[51] for scaling. The structure was determined by molecular replacement in PHASER[52] using a predicted structure from AlphaFold 2 [30] as a search model. The model was manually rebuilt in COOT[53] and refined in phenix.refine[54] using positional and individual B-factor refinement. B-factors for all atoms except waters were refined anisotropically. The final model has good geometry, with 98.75% of residues in favored Ramachandran space, 1.25% allowed, and 0% outliers. The overall MolProbity score is 1.18, and the MolProbity clash score is 3.98.

For crystallization of the *E. aridi* DUB^{BilC}(E33A):Ubl^{BilA} complex (Form 2), purified protein at 30 mg/mL in crystallization buffer (25 mM Tris-HCl pH 8.5, 200 mM NaCl, 5 mM $MgCl_2$, and 1 mM TCEP (tris(2-carboxyethyl)phosphine)) was mixed 1:1 with well solution containing 100 mM MES pH 6.5 and 1M sodium/potassium tartrate. Crystals were harvested into cryoprotectant solution containing an additional 30% glycerol and frozen in liquid nitrogen. Diffraction data were collected at the Advanced Photon Source (Argonne National Lab) NE-CAT beamline 24ID-E on April 7, 2023 (collection temperature 100 K; x-ray wavelength 0.97918 Å) (Extended Data Table 1). Data were processed with the RAPD data-processing pipeline, and the structure was determined by molecular replacement in PHASER using the Form 1 BilC^{E33A}:BilA structure (with ligands and waters excluded) as a search model. The model was manually rebuilt in COOT and refined in phenix.refine using positional and individual isotropic B-factor refinement. The final model has good geometry, with 99.58% of residues in favored Ramachandran space, 0.42% allowed, and 0% outliers. The overall MolProbity score is 1.08, and the MolProbity clash score is 1.86.

For crystallization of the *E. aridi* E1^{BilD}:E2^{BilB}:Ubl^{BilA} complex (Form 1), purified protein at 10 mg/mL in crystallization buffer was mixed 1:1 with well solution containing 100 mM HEPES pH 7.5, 100 mM sodium citrate, 5% isopropanol, and 10% PEG 3350 in hanging drop format. Crystals were harvested into cryoprotectant solution containing an additional 20% glycerol and frozen in liquid nitrogen. Diffraction data were collected at the Advanced Photon Source (Argonne National Lab) NE-CAT beamline 24ID-E on April 4, 2023 (collection temperature 100 K; x-ray wavelength 0.97918 Å) (Extended Data Table 1). Data were processed with the RAPD data-processing pipeline, and the structure was determined by molecular replacement in PHASER using predicted structures of each protein (E1^{BilD} CYS domain excluded) from AlphaFold2 as a search model. One copy of Ubl^{BilA}, two copies of E2^{BilB}, and two copies of E1^{BilD} were located. The model was manually

rebuilt in COOT and refined in phenix.refine using positional and individual isotropic B-factor refinement. The final model has good geometry, with 96.23% of residues in favored Ramachandran space, 3.47% allowed, and 0.30% outliers. The overall MolProbity score is 1.64, and the MolProbity clash score is 5.86.

For crystallization of the *E. aridi* $E1^{BilD}$:$E2^{BilB}$:Ubl^{BilA} complex (Form 2), purified protein at 10 mg/mL in crystallization buffer was mixed 1:1 with well solution containing 100 mM imidazole pH 8.0, 200 mM $MgCl_2$, and 10% PEG 3350 in hanging drop format. Crystals were harvested into cryoprotectant solution containing an additional 20% glycerol and frozen in liquid nitrogen. Diffraction data were collected at the Advanced Photon Source (Argonne National Lab) NE-CAT beamline 24ID-E on April 7, 2023 (collection temperature 100 K; x-ray wavelength 0.97918 Å) (Extended Data Table 1). Data were processed with the RAPD data-processing pipeline, and the structure was determined by molecular replacement in PHASER using predicted structures of each protein ($E1^{BilD}$ CYS domain excluded) from AlphaFold2 as a search model. The model was manually rebuilt in COOT and refined in phenix.refine using positional and individual isotropic B-factor refinement. The final model has good geometry, with 96.05% of residues in favored Ramachandran space, 3.53% allowed, and 0.42% outliers. The overall MolProbity score is 1.82, and the MolProbity clash score is 5.62.

In-Gel Digestion

For mass spectrometry identification of Ubl^{BilA}-conjugated proteins, *E. aridi* $E1^{BilD}$ and $E2^{BiLB}$ were coexpressed with His_6-Ubl^{BilA}(V95K) in *E. coli*, and purified by Ni-NTA chromatography in denatured solution as above. Purified proteins were separated by SDS-PAGE and visualized with Coomassie blue staining, then mass spectrometry was performed by in-gel trypsin digestion as previously described[55]. In brief, gel slices were cut into ~1 mm x 1 mm x 1 mm cubes using a clean razor blade in a glass dish. Gel slices were reduced in 100 µl of 10 mM DTT for 30 minutes at 37°C. Cysteine alkylation was performed by incubating the samples at room temperature in the dark for 20 minutes following the addition of 6 µl of 0.5 M iodoacetamide (30 mM iodoacetamide final concentration). To digest proteins, 100 µl of 10 ng/µl trypsin (Promega #V511A) in 20 mM ammonium bicarbonate (pH 8) was added to submerge the gel pieces, then incubated on ice for 30 minutes until fully swollen. An additional 20–50 µl of ammonium bicarbonate buffer was added to ensure the gel slices were fully submerged prior to overnight incubation at 37°C. The next day, trypsin digested peptides were extracted from the sample via multiple extractions using 50% acetonitrile/5% formic acid, dried under vacuum and reconstituted in 20 µl of 0.1% trifluoroacetic acid (pH 2).

LC-MS/MS Analysis

Peptides were analyzed by LC-MS/MS on a Vanquish Neo high performance liquid chromatography system coupled to a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific). Data-dependent analysis (DDA) was performed using a Top 10 approach using a linear gradient of 2 – 30% mobile phase B (Mobile Phase A: 0.1% formic acid in H_2O, Mobile Phase B: 0.1% formic acid in ACN) at a flow rate of 75 µL/min on a 15-cm C_{18} column maintained at 40°C (Acclaim Pepmap 1 mm I.D., 2 µm particle size, 100 Å pore

size; Thermo Fisher Scientific). The LC method for DDA starts at 2% B and is held constant for 1 min, followed by a change to 30% B across 20 min, followed by a change to 70% B across 3 min, and then finally increased to 95% B across 0.1 min. Column washing and equilibration is performed across 4.1 min making the total method time 28.2 min. Parallel reaction monitoring (PRM) was performed in DIA mode utilizing a global inclusion list with a full scan using a linear gradient of 2 – 40% mobile phase B at a flow rate of 1.5 μL/min on an in-house packed 15-cm pulled-tip column emitter (100 μm I.D. with 2.2-μm C_{18} beads, 120 Å pore size; Sepax Technologies) maintained at 50 °C (PRSO-V2; Sonation GmbH). The LC method starts at 2% B and is held constant for 5 min, followed by a change to 40% B across 25 min, and then finally increased to 95% B across 1.5 min. Column washing and equilibration is performed across 13.5 min making the total method time 45 min. Targeted PRM acquisition of the Ubl^{BilA} (V95K) C-terminal peptides were performed in DIA mode using a targeted inclusion list for the +2 and +3 charge states for the modified and non-modified peptide species. Initial data review and quality control of PRM acquired transition spectra were analyzed in Skyline (v22.2.0.527). Final extracted transition ion chromatograms were generated using FreeStyle (v1.7, Thermo Fisher Scientific) using a ± 10 ppm m/z tolerance for each transition.

Peptide Sequencing Analysis

Database searching of data-dependent acquired MS spectra was performed using the Trans-Proteomic Pipeline software suite v6.3.2 Arcus (TPP, Seattle Proteome Center)[56,57]. The search parameter file and protein database used for the database search is included in Supplementary Table 2. Briefly, database search was performed using COMET and peptides were quantified using XPRESS in label-free mode. A static modification of 57.021464 Da was applied to cysteine residues. Two differential modifications were applied as follows: Oxidation of methionine (15.9949 Da) and the expected AG remnant of Ubl^{BilA} (V95K) on lysine/N-termini (128.05857 Da). Quality control was performed following data analysis and manual inspection of chromatography and MS/MS spectral assignments. Redundant peptide identifications were removed to generate the final list of unique peptides (Supplementary Table 3).

The protein database included all proteins from the *E. coli* proteome (Uniprot: ECOLX) plus the following:

```
>tr|HIS6_Ubl-BilA_delta97-V95K

MKSSHHHHHHENLYFQSNASKDSRKGDNHGGGSGKIEIIVVVNGQPTQVEANPNQPLHVV

RTKALENTQNVAQPPDNWEFKDEAGNLLDVDKKIGDFGFANTVTLFLSLKAGKAG

>tr|E1-BilD

MALANFIDRAATAASQVLTDFHLGDFKAALEKQVVAVAFDDQAISCAEGQATLDLAVRLLA

RLYPVLAILPLDSAASSQAQALERLAKSINRKIGIRRSGKSATVCLVAGATRPSLRCPTFF
```

IGSDGWAAKLSRTDPVGSGSSLLPYGAGAASCFGAANVFRTIFAAQLTGAESDENIDLSLY

SYNKSRAGDAGPIDPAVDLGETHLVGLGAIAHGALWALARQSGLSGRLHVVDHEAVELSNL

QRYVLAGQAEIGMSKAVLATTALRSTALEVEAHPLKWAEHVARRGDWIFDRVGVALDTAAD

RVAVQGALPRWIANAWTQEHDLGISRHGFDDGQACLCCMYMPSGKSKDEHQLVAEELGIPE

AHEQVKALLQTNAGVPNDFVVRVATAMGVPFEPLAPFVGQPLRSFYQQAICGGLVFQLSDG

SRLVRTVVPMAFQSALAGIMLAAELVKHSAGFPMSPTTSTRVNLLRPLGSHLHDPKAKDSS

GRCICSDEDFISAYRRKYGNGVEPLSNISAEQKRTSPLPRTGRQVCA

>tr|E2-BilB

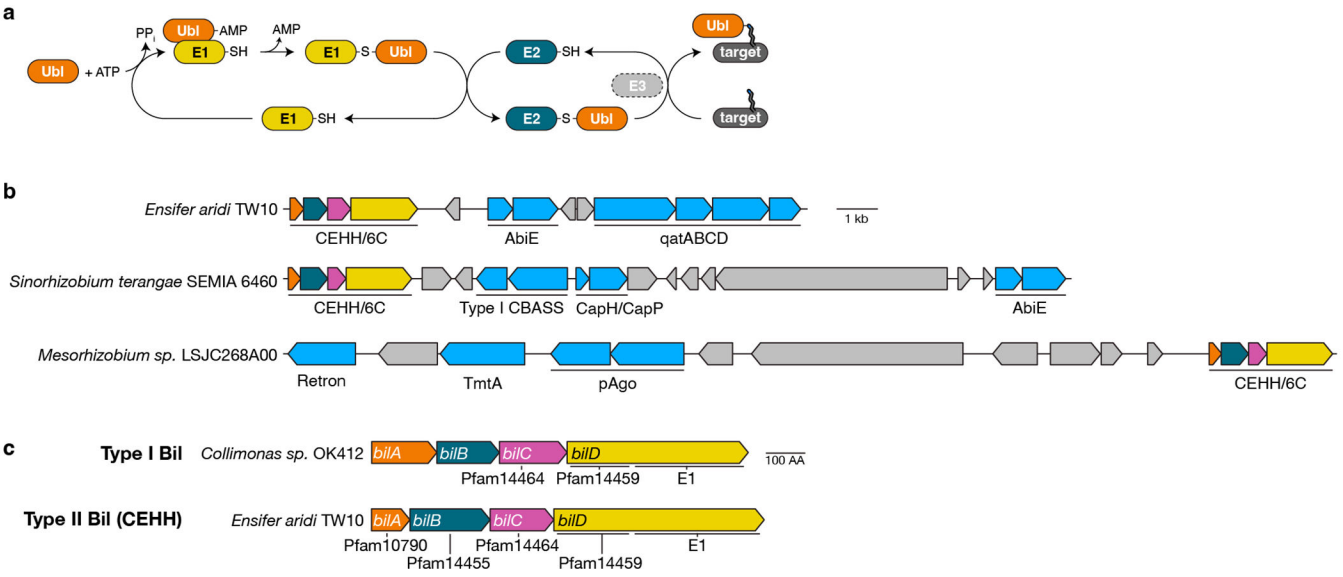
MPELQTVDPEVSRAKFDREISRFRPYADAYRMQGCFLIEESFPSAFFIFASPKVKPRVIGA

AIEIDFTNYDLRPPSVVFVDPFTRQPIARKDLPLNMLRRPQLPGTPPEMISNLIQQNAVSLT

DFIQANSLQDSPFLCMAGVREYHDNPAHSGDPWLLHRGSGEGCLAFILDKIIKYGTGPVEQL

HIQLQYAVGLLVPPQAIPE

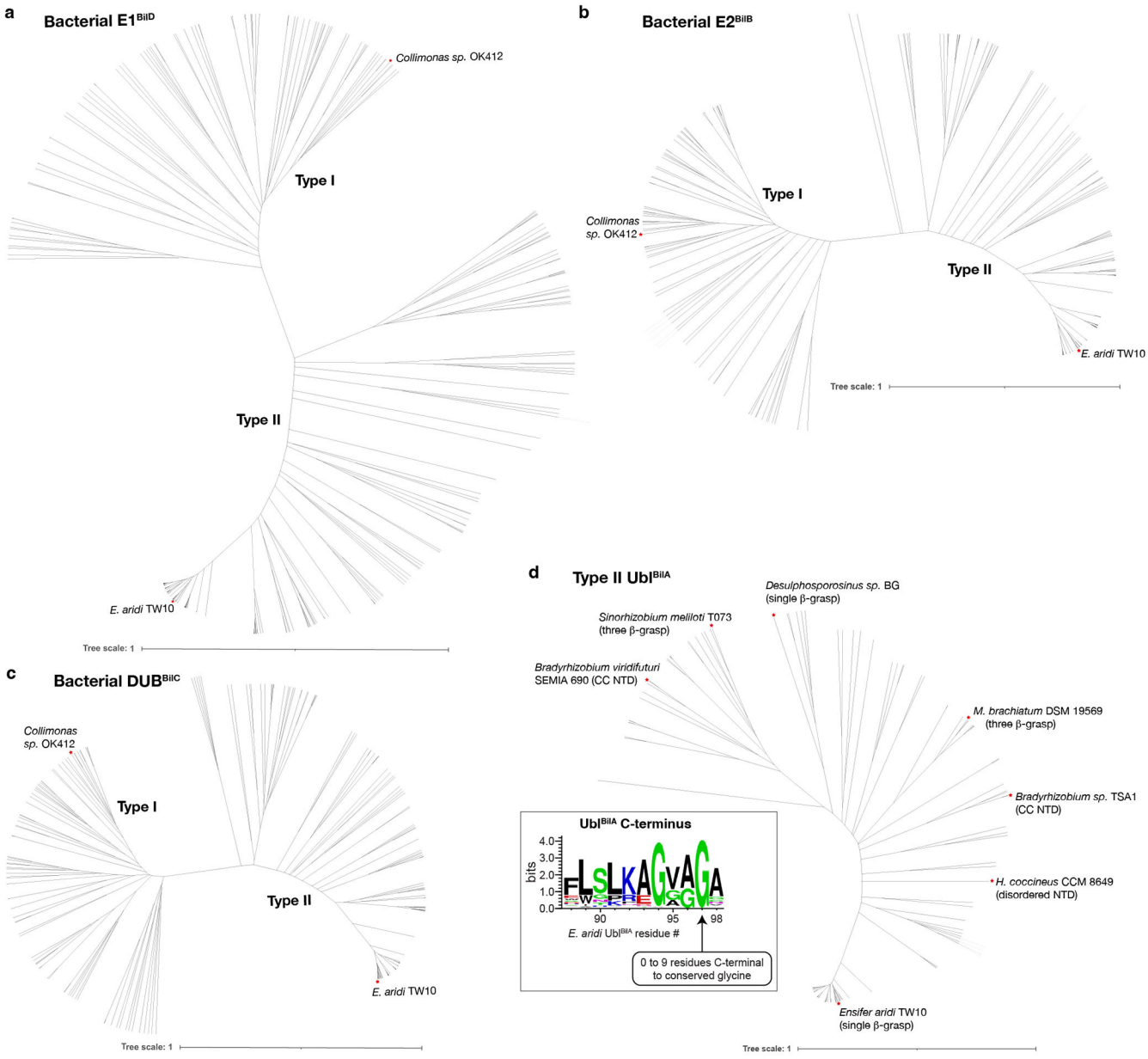
Extended Data



Extended Data Figure 1. Comparison of Type I and Type II BilABCD operons.

(a) Schematic of a typical ubiquitination pathway, with ubiquitin-like protein (Ubl) in orange, E1 in yellow, E2 in blue, optional E3 in light gray, and target in gray. An E1 protein mediates adenylation of the Ubl C-terminus followed by generation of an E1~Ubl thioester linkage. This thioester linkage is transferred to E2, then to a lysine residue on a target

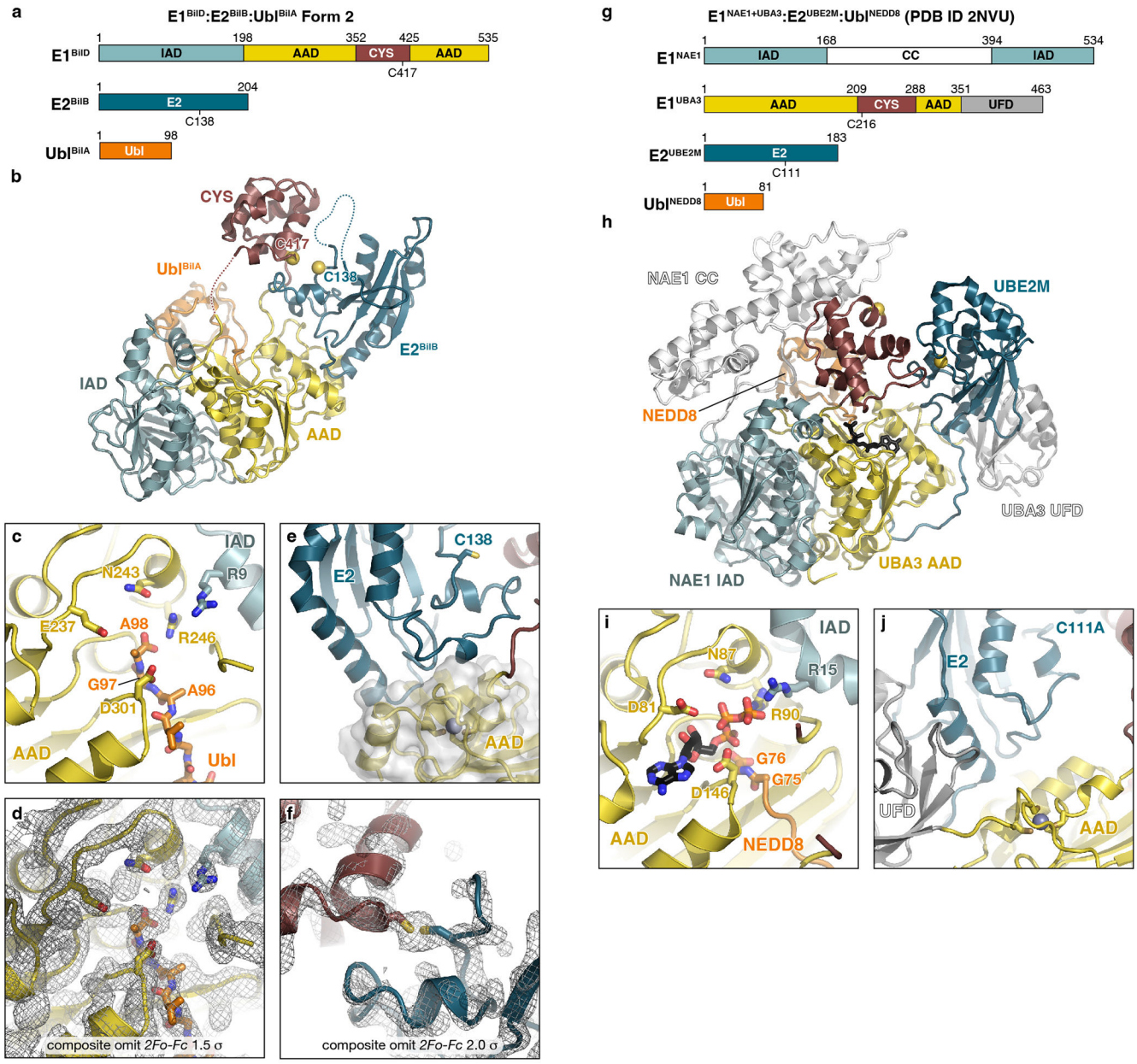
(optionally with the help of an E3). **(c)** Schematics of CEHH/6C operons in defense islands identified by the PADLOC server[31]. See Supplementary Table 1 for additional information. **(c)** Operon schematics of a Type I BilABCD operon from *Collimonas sp.* OK412 [24], and a Type II BilABCD (CEHH) operon from *E. aridi* TW10 (Supplementary Table 1). Noted under each gene are the conserved PFAM domain annotations for that gene.



Extended Data Figure 2. Sequence analysis of BilABCD proteins.

(a) Unrooted average distance tree assembled from E1^{BilD} proteins from Type I BilABCD operons[24] and Type II BilABCD operons (Supplementary Table 1). Branches representing proteins from Type I and Type II operons are marked. Specific E1^{BilD} proteins from *Collimonas sp.* OK412 (Type I) and *E. aridi* TW10 (Type II) are indicated with red stars

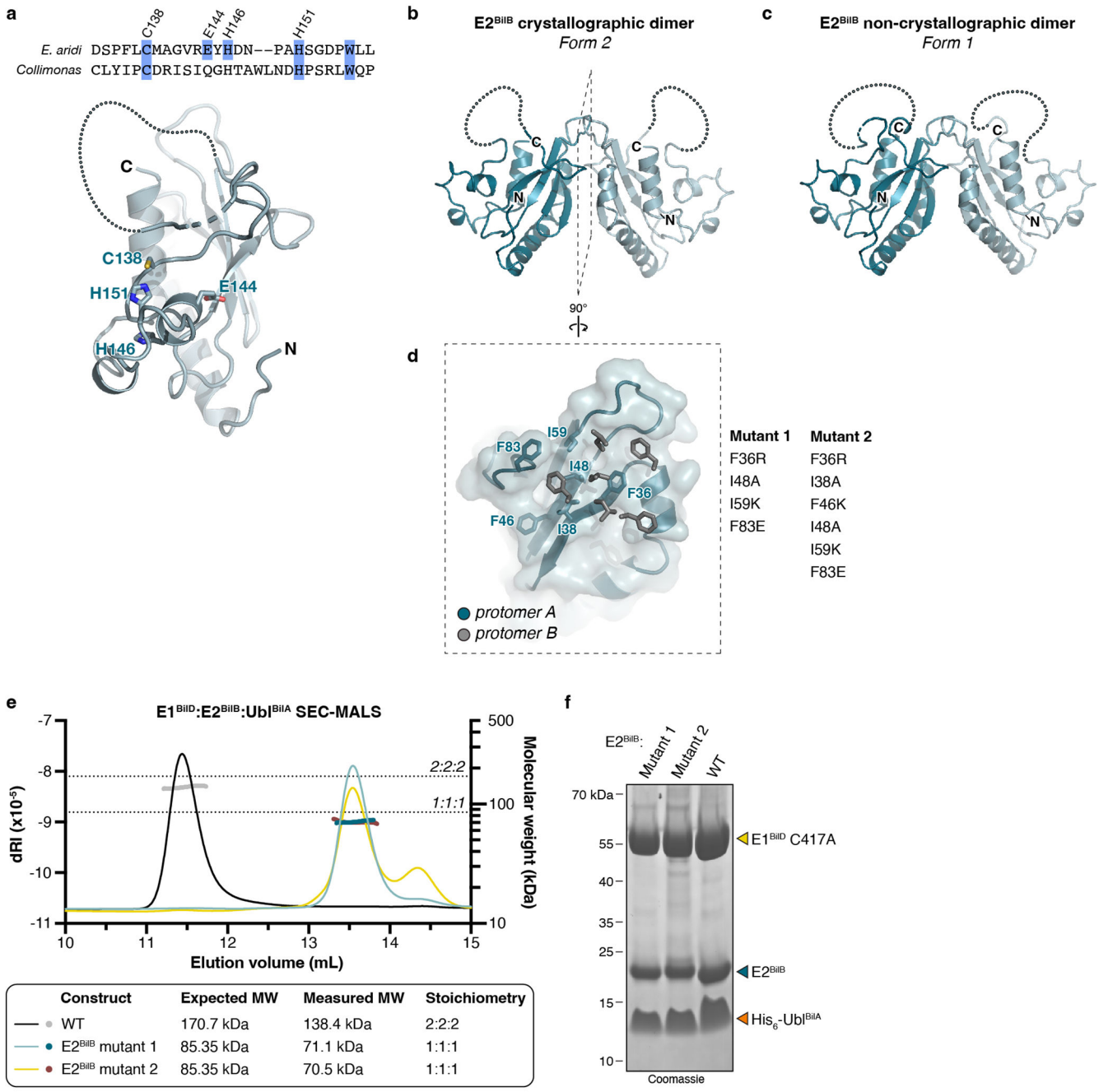
and labeled. Scale bar: 1 substitution per site. **(b)** Unrooted average distance tree assembled from $E2^{BilB}$ proteins from Type I BilABCD operons[24] and Type II BilABCD systems (Supplementary Table 1). Branches representing proteins from Type I and Type II operons are marked. Specific $E2^{BilB}$ proteins from *Collimonas sp.* OK412 (Type I) and *E. aridi* TW10 (Type II) are indicated with red stars and labeled. Scale bar: 1 substitution per site. **(c)** Unrooted average distance tree assembled from DUB^{BilC} proteins from Type I BilABCD operons[24] and Type II BilABCD operons (Supplementary Table 1). Scale bar: 1 substitution per site. Branches representing proteins from Type I and Type II operons are marked. Specific DUB^{BilC} proteins from *Collimonas sp.* OK412 (Type I) and *E. aridi* TW10 (Type II) are indicated with red stars and labeled. Scale bar: 1 substitution per site. **(d)** Unrooted evolutionary tree of all Ubl^{BilA} proteins in Type II BilABCD operons (Supplementary Table 1). Specific examples are labeled and their domain architectures (inferred from sequence analysis) noted. LLPS NTD: N-terminal domain predicted to undergo liquid-liquid phase separation; CC NTD: N-terminal domain predicted to form a coiled-coil. *Inset:* Sequence logo from bacterial Type II Ubl^{BilA} proteins (Supplementary Table 1). Type II Ubl^{BilA} homologs possess up to nine residues C-terminal to the highly conserved glycine (G97 in *E. aridi* BilA).



Extended Data Figure 3. Structural parallels between bacterial and eukaryotic ubiquitination machinery.

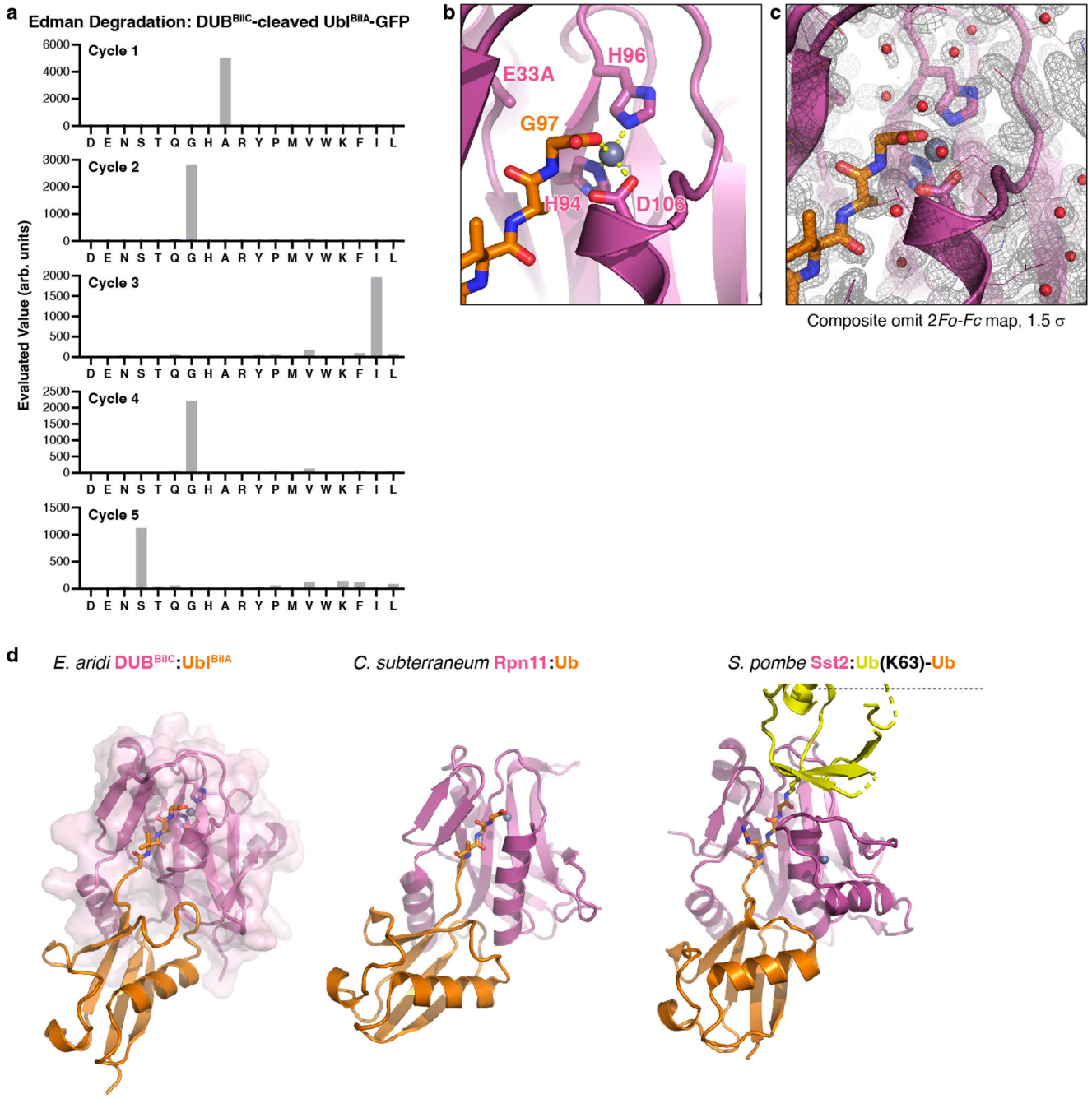
(a) Domain architecture of *E. aridi* $E1^{BilD}$, $E2^{BilB}$, and Ubl^{BilA}. IAD: inactive adenylation domain; AAD: active adenylation domain; CYS: E1 catalytic cysteine-containing domain; Ubl: Ubiquitin-like domain. **(b)** Crystal structure (Form 2) of the *E. aridi* $E1^{BilD}$:$E2^{BilB}$:Ubl^{BilA} complex, with domains colored as in (a). **(c)** Closeup view of the $E1^{BilD}$ adenylation active site with bound Ubl^{BilA} C-terminus. Conserved active site residues are shown as sticks and labeled. **(d)** View equivalent to (c) showing $2F_o$-F_c composite-omit electron density at 1.5σ. **(e)** Closeup view of the $E1^{BilD}$-$E2^{BilB}$ binding interface. The zinc ion (gray sphere) is coordinated by $E1^{BilD}$ residues C340, C343, C491, and C493. **(f)** Closeup view of the catalytic cysteine residues of $E1^{BilD}$ (C417; brown) and

$E2^{BilB}$ (C138; blue) with $2F_o$-F_c composite-omit electron density at 2.0σ. **(g)** Domain architecture of *H. sapiens* $E1^{NAE1\text{-}UBA3}$, $E2^{UBE2M}$, and Ubl^{NEDD8} (PDB ID 2NVU)[37]. CC: coiled-coil domain; UFD: ubiquitin-fold domain. **(h)** Crystal structure of the *H. sapiens* $E1^{NAE1\text{-}UBA3}$:$E2^{UBE2M}$:Ubl^{NEDD8} complex (PDB ID 2NVU)[37], with domains colored as in panel (e). **(i)** Closeup view of the $E1^{UBA3}$ adenylation active site with bound Ubl^{NEDD8} C-terminus. Conserved active site residues are shown as sticks and labeled, and bound ATP is shown as sticks. **(j)** Closeup view of the $E1^{BilD}$-$E2^{BilB}$ binding interface. The UBA3 UFD is shown in gray. The zinc ion (gray sphere) is coordinated by UBA3 residues C199, C202, C343, and C346.



Extended Data Figure 4. Type II Bil E2^{BilB} protein structure and homodimer formation.
(a) *Top:* Sequence alignment of *E. aridi* E2^{BilB} (residues 133–158) and *Collimonas sp.* OK412 E2^{BilB} (residues 108–135), with the four residues of the originally identified CEHH motif in 6C/CEHH operon proteins shown in blue highlights. Of the four residues, only C138 and H151 are highly conserved across both Type I and Type II E2^{BilB} proteins. Bottom: Structure of *E. aridi* E2^{BilB}, with the four residues of the CEHH motif shown as sticks and labeled. **(b)** Crystallographic dimer of E2^{BilB} in the Form 2 structure. **(c)** Closeup of the non-crystallographic E2^{BilB} dimer in the Form 1 structure, oriented equivalently to panel (b). **(d)** Closeup view of the E2^{BilB} dimer interface (rotated 90° from panel (b)).

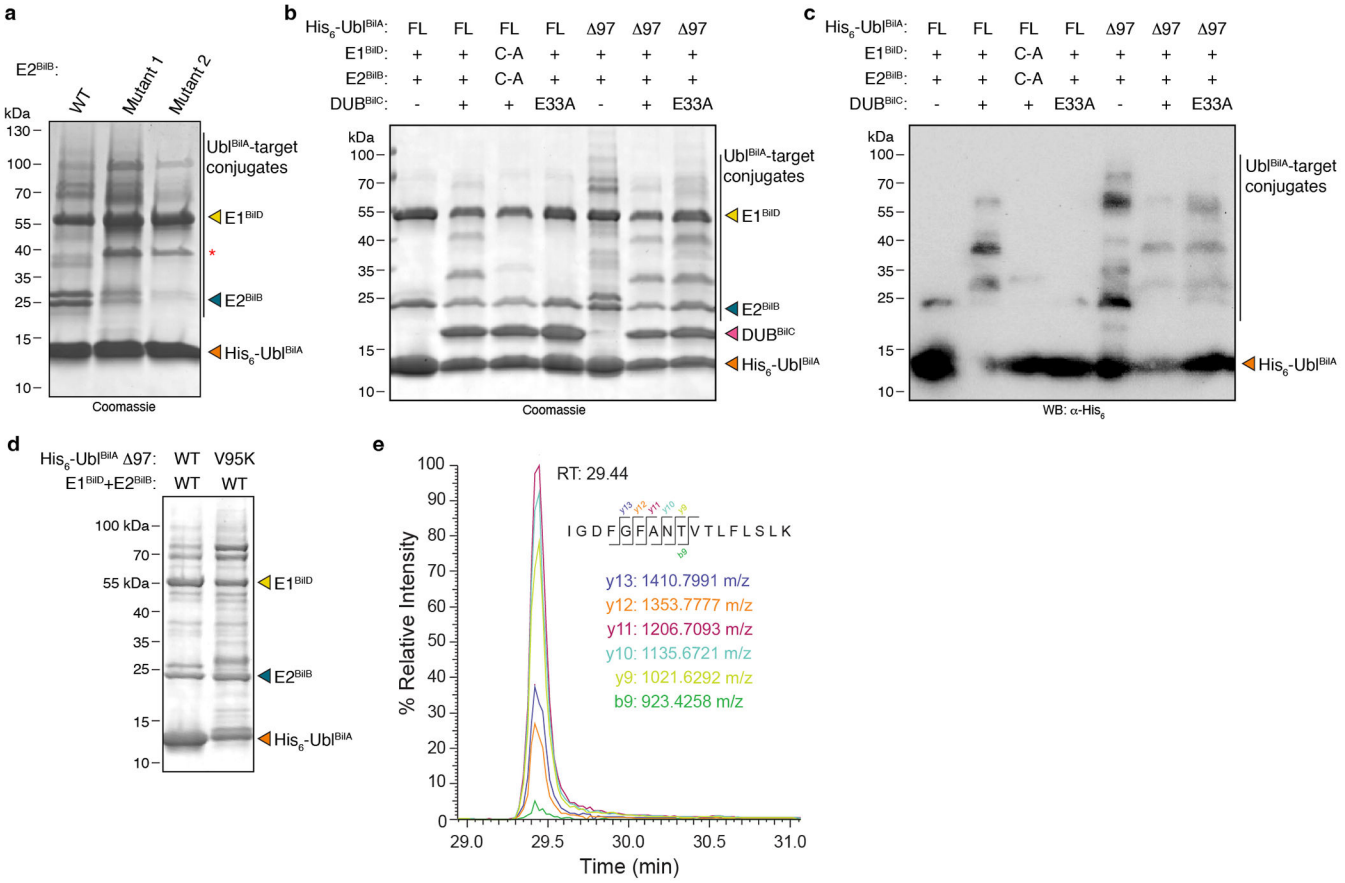
Residues involved in the interface from one protomer are colored blue, shown in sticks, and labeled. The same residues from the dimer-related protomer are shown as sticks and colored gray. Mutant 1 and Mutant 2 are two multi-site mutants of $E2^{BilB}$ used in panels (e)-(f). **(e)** Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) analysis of purified *E. aridi* $E1^{BilD}$ (1–507, C417A):$E2^{BilB}$ (1–181, wild type or mutant):Ubl^{BilA} (17–97) complexes. Thin lines indicate protein concentration as measured by differential refractive index (dRI; left axis), and thick lines indicate molecular weight in kDa (right axis). Dotted horizontal lines indicate the expected molecular weight of a 1:1:1 complex (85.35 kDa) and a 2:2:2 complex (170.7 kDa). Data for three constructs is shown: wild-type E2 (black/gray), mutant 1 (F36R/I48A/I59K/F83E; light blue/dark blue), and mutant 2 (F36R/I38A/F46K/I48A/I59K/F83E; yellow/brown). **(f)** SDS-PAGE analysis of purified complexes analyzed by SEC-MALS. For gel source data, see Supplementary Figure 1. This experiment was independently performed three times, with consistent results.



Extended Data Figure 5. Biochemical and structural analysis of *E. aridi* DUB^{BilC}.

(a) N-terminal sequencing (Edman degradation) of DUB^{BilC}-cleaved Ubl^{BilA}-GFP fusion (C-terminal fragment), showing the evaluated value from each of five cycles of degradation. The inferred N-terminal sequence of the fragment is AGIGS. This analysis was performed once. **(b)** Closeup view of the DUB^{BilC}(E33A):Ubl^{BilA} (Form 1) active site, with proteins colored as in Figure 3d. Active site residues of DUB^{BilC} and glycine 97 of Ubl^{BilA} are labeled. **(c)** View equivalent to panel (b), showing 2*Fo-Fc* composite omit map density at 1.5 σ. **(d)** Comparison of *E. aridi* DUB^{BilC}:Ubl^{BilA} (left) to two similar structures,

Caldiarchaeum subterraneum Rpn11-homolog bound to ubiquitin-homolog (center)[62] and *Schizosaccharomyces pombe* Sst2 bound to a ubiquitin K63-linked ubiquitin (right)[63]. Overall Cα r.m.s.d. values for DUB^{BilC} versus its homolog are in Table 1.



Extended Data Figure 6. Identification of Ubl^{BilA} targets.

(a) SDS-PAGE analysis of Ni^{2+} affinity-purified His$_6$-Ubl^{BilA} Δ97 and associated proteins in native conditions, after coexpression with E1^{BilD} and E2^{BilB}. Bands representing His$_6$-Ubl^{BilA}, E1^{BilD}, and E2^{BilB} are marked. E2^{BilB} WT: wild type; Mutant 1: F36R/I48A/I59K/F83E; Mutant 2: F36R/I38A/F46K/I48A/I59K/F83E. Red asterisk indicates a likely Ubl^{BilA}-E2^{BilB} conjugate that is more abundant when E2^{BilB} is mutated. For gel source data for panels (a)-(d), see Supplementary Figure 1. **(b)** SDS-PAGE analysis of Ni^{2+} affinity-purified His$_6$-Ubl^{BilA} (full-length (FL) or Δ97) and associated proteins in native conditions, after coexpression with E1^{BilD} (wild type or C417A mutant; indicated as "C-A"), E2^{BilB} (wild type or C138A mutant; indicated as "C-A"), and DUB^{BilC} (wild type or E33A mutant). Bands representing His$_6$-Ubl^{BilA}, E1^{BilD}, E2^{BilB}, and DUB^{BilC} are marked. **(c)** Anti-His$_6$ western blot analysis of the experiment shown in panel (b), showing His$_6$-Ubl^{BilA}-target conjugates. **(d)** SDS-PAGE gel (visualized by Coomassie Blue staining) of His$_6$-tagged Ubl^{BilA} Δ97 (wild type or V95K mutant) coexpressed with E1^{BilD} and E2^{BilB}, then purified in denaturing conditions. **(e)** Extracted Ion Chromatograms (EICs) of transition ions of the properly cleaved, unmodified Ubl^{BilA}(Δ97, V95K) residues 76–92, using a 10 ppm m/z tolerance. RT: retention time.

Extended Data Table 1.
Crystallographic data collection and structure determination

Each dataset was collected from an individual crystal. Values in parentheses are for the highest resolution shell. Accession numbers for final refined coordinates, structure factors, and raw diffraction datasets are provided in the Data Availability statement.

	E. aridi **DUB^{BilC}(E33A): Ubl^{BilA} Form 1**	***E. aridi*** **DUB^{BilC}(E33A): Ubl^{BilA} Form 2**	***E. aridi*** **E1^{BilD}:E2^{BilB}: Ubl^{BilA} Form 1**	***E. aridi*** **E1^{BilD}:E2^{BilB}: Ubl^{BilA} Form 2**
Data collection				
Space group	I222	F432	$P2_1$	$P4_22_12$
Cell dimensions				
a, *b*, *c* (Å)	53.34, 64.69, 141.69	214.89, 214.89, 214.89	73.85, 89.66, 131.60	199.93, 199.93, 66.88
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 94.93, 90	90, 90, 90
Resolution (Å)	70.85-1.36(1.38-1.36)	124.07-1.68 (1.71-1.68)	89.66-2.47 (2.53-2.47)	199.93-2.68(2.80-2.68)
R_{merge}	0.058 (0.481)	0.097 (0.956)	0.093 (0.812)	0.101 (3.776)
$I/\sigma I$	25.0 (4.0)	35.8 (5.2)	11.9 (1.5)	19.2 (0.7)
Completeness (%)	99.9 (98.0)	99.9 (100.0)	98.9 (95.0)	100 (100)
Redundancy	12.7 (12.5)	40.3 (41.1)	3.5 (3.3)	13.7 (14.4)
Refinement				
Resolution (Å)	70.85-1.36	50.00-1.68	74.01-2.47	70.68-2.68
No. reflections	53036	48715	60794	38499
R_{work} / R_{free}	0.1534/0.1799	0.1461/0.1628	0.1963/0.2439	0.2075/0.2446
No. atoms				
Protein	3768	3767	20,307	5,452
Ligand/ion	1 (Zn^{2+})	1 (Zn^{2+})	2 (Zn^{2+})	1 (Zn^{2+})
Water	443	381	129	18
B-factors				
Protein	20.90	19.66	69.58	89.39
Ligand/ion	14.37	14.48	57.51	125.06
Water	33.03	36.92	56.99	71.19

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank the staff at The Protein Facility of the Iowa State University Office of Biotechnology for assistance with Edman degradation, R. Sorek for sharing information prior to publication, and members of the Corbett laboratory for helpful discussions. This work was funded by NIH R35 GM144121 (to K.D.C.); the Howard Hughes Medical Institute Emerging Pathogens Initiative (to K.D.C.); the NIH Director's New Innovator Award DP2 AT012346 (to A.T.W.), a Mallinckrodt Foundation Grant (to A.T.W.), the Boettcher Foundation's Webb-Waring Biomedical Research Program (to A.T.W.), and the Pew Biomedical Scholars Program (to A.T.W.); NIH R01 GM116897 (to H.Z. and R.T.S.), R01 GM151191 (to H.Z. and R.T.S.), and S10 OD023498 (to H.Z.). L.R.C. is

supported by the UCSD Molecular Biophysics Training Grant (T32 GM139795); J.C. is supported by a Pfizer-Cell Signaling San Diego graduate fellowship; and H.E.L. is supported as a fellow of the Jane Coffin Childs Memorial Fund for Medical Research. This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). The Eiger 16M detector on 24-ID-E is funded by a NIH-ORIP HEI grant (S10 OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

Data Availability

Final coordinates and structure factors for all structures have been deposited at the RCSB Protein Data Bank (https://www.rcsb.org) under accession codes 8TYX (DUB^{BilC}(E33A):Ubl^{BilA} Form 1), 8TYY (DUB^{BilC}(E33A):Ubl^{BilA} Form 2), 8TZ0 ($E1^{BilD}$:$E2^{BilB}$:Ubl^{BilA} Form 1), and 8TYZ ($E1^{BilD}$:$E2^{BilB}$:Ubl^{BilA} Form 2). Raw diffraction images have been deposited at the SBGrid Data Bank (https://data.sbgrid.org) under accession codes 1039 (DUB^{BilC}(E33A):Ubl^{BilA} Form 1), 1040 (DUB^{BilC}(E33A):Ubl^{BilA} Form 2), 1041 ($E1^{BilD}$:$E2^{BilB}$:Ubl^{BilA} Form 1), and 1042 ($E1^{BilD}$:$E2^{BilB}$:Ubl^{BilA} Form 2). Sequence data were downloaded from the IMG database of bacterial genomes (https://img.jgi.doe.gov/). Source data for the graphs shown in Extended Data Figures 4e and 5a accompany this manuscript.

References

1. Oh E, Akopian D & Rape M Principles of Ubiquitin-Dependent Signaling. Annu Rev Cell Dev Biol 34, 137–162 (2018). 10.1146/annurev-cellbio-100617-062802 [PubMed: 30110556]
2. van der Veen AG & Ploegh HL Ubiquitin-like proteins. Annu Rev Biochem 81, 323–357 (2012). 10.1146/annurev-biochem-093010-153308 [PubMed: 22404627]
3. Perng YC & Lenschow DJ ISG15 in antiviral immunity and beyond. Nat Rev Microbiol 16, 423–439 (2018). 10.1038/s41579-018-0020-5 [PubMed: 29769653]
4. Cappadocia L & Lima CD Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. Chem Rev 118, 889–918 (2018). 10.1021/acs.chemrev.6b00737 [PubMed: 28234446]
5. Lake MW, Wuebbens MM, Rajagopalan KV & Schindelin H Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex. Nature 414, 325–329 (2001). 10.1038/35104586 [PubMed: 11713534]
6. Lehmann C, Begley TP & Ealick SE Structure of the Escherichia coli ThiS-ThiF complex, a key component of the sulfur transfer system in thiamin biosynthesis. Biochemistry 45, 11–19 (2006). 10.1021/bi051502y [PubMed: 16388576]
7. Scaglione KM et al. The ubiquitin-conjugating enzyme (E2) Ube2w ubiquitinates the N terminus of substrates. J Biol Chem 288, 18784–18788 (2013). 10.1074/jbc.C113.477596 [PubMed: 23696636]
8. Bernier-Villamor V, Sampson DA, Matunis MJ & Lima CD Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. Cell 108, 345–356 (2002). 10.1016/s0092-8674(02)00630-x [PubMed: 11853669]
9. Stewart MD, Ritterhoff T, Klevit RE & Brzovic PS E2 enzymes: more than just middle men. Cell Res 26, 423–440 (2016). 10.1038/cr.2016.35 [PubMed: 27002219]
10. Mevissen TET & Komander D Mechanisms of Deubiquitinase Specificity and Regulation. Annu Rev Biochem 86, 159–192 (2017). 10.1146/annurev-biochem-061516-044916 [PubMed: 28498721]
11. Pearce MJ, Mintseris J, Ferreyra J, Gygi SP & Darwin KH Ubiquitin-like protein involved in the proteasome pathway of Mycobacterium tuberculosis. Science 322, 1104–1107 (2008). 10.1126/science.1163885 [PubMed: 18832610]

12. Burns KE, Liu WT, Boshoff HIM, Dorrestein PC & Barry CE 3rd. Proteasomal protein degradation in Mycobacteria is dependent upon a prokaryotic ubiquitin-like protein. J Biol Chem 284, 3069–3075 (2009). 10.1074/jbc.M808032200 [PubMed: 19028679]
13. Striebel F et al. Bacterial ubiquitin-like modifier Pup is deamidated and conjugated to substrates by distinct but homologous enzymes. Nat Struct Mol Biol 16, 647–651 (2009). 10.1038/nsmb.1597 [PubMed: 19448618]
14. Guth E, Thommen M & Weber-Ban E Mycobacterial ubiquitin-like protein ligase PafA follows a two-step reaction pathway with a phosphorylated pup intermediate. J Biol Chem 286, 4412–4419 (2011). 10.1074/jbc.M110.189282 [PubMed: 21081505]
15. Sutter M, Damberger FF, Imkamp F, Allain FH & Weber-Ban E Prokaryotic ubiquitin-like protein (Pup) is coupled to substrates via the side chain of its C-terminal glutamate. J Am Chem Soc 132, 5610–5612 (2010). 10.1021/ja910546x [PubMed: 20355727]
16. Maculins T, Fiskin E, Bhogaraju S & Dikic I Bacteria-host relationship: ubiquitin ligases as weapons of invasion. Cell Res 26, 499–510 (2016). 10.1038/cr.2016.30 [PubMed: 26964724]
17. Berglund J, Gjondrekaj R, Verney E, Maupin-Furlow JA & Edelmann MJ Modification of the host ubiquitome by bacterial enzymes. Microbiol Res 235, 126429 (2020). 10.1016/j.micres.2020.126429 [PubMed: 32109687]
18. Vozandychova V, Stojkova P, Hercik K, Rehulka P & Stulik J The Ubiquitination System within Bacterial Host-Pathogen Interactions. Microorganisms 9 (2021). 10.3390/microorganisms9030638
19. Iyer LM, Burroughs AM & Aravind L The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like β-grasp domains. Genome Biol 7, R60 (2006). 10.1186/gb-2006-7-7-r60 [PubMed: 16859499]
20. Burroughs AM, Balaji S, Iyer LM & Aravind L Small but versatile: the extraordinary functional and structural diversity of the beta-grasp fold. Biol Direct 2, 18 (2007). 10.1186/1745-6150-2-18 [PubMed: 17605815]
21. Burroughs AM, Jaffee M, Iyer LM & Aravind L Anatomy of the E2 ligase fold: implications for enzymology and evolution of ubiquitin/Ub-like protein conjugation. J Struct Biol 162, 205–218 (2008). 10.1016/j.jsb.2007.12.006 [PubMed: 18276160]
22. Burroughs AM, Iyer LM & Aravind L Natural history of the E1-like superfamily: Implication for adenylation, sulfur transfer, and ubiquitin conjugation. Proteins: Structure, Function and Bioinformatics 75, 895–910 (2009). 10.1002/prot.22298
23. Burroughs AM, Iyer LM & Aravind L Functional diversification of the RING finger and other binuclear treble clef domains in prokaryotes and the early evolution of the ubiquitin system. Mol Biosyst 7, 2261–2277 (2011). 10.1039/c1mb05061c [PubMed: 21547297]
24. Millman A et al. An expanded arsenal of immune systems that protect bacteria from phages. Cell Host Microbe 30, 1556–1569 e1555 (2022). 10.1016/j.chom.2022.09.017 [PubMed: 36302390]
25. Hör J, Wolf SG & Sorek R Bacteria conjugate ubiquitin-like proteins to interfere with phage assembly. bioRxiv (2023). 10.1101/2023.09.04.556158
26. Grau-Bove X, Sebe-Pedros A & Ruiz-Trillo I The eukaryotic ancestor had a complex ubiquitin signaling system of archaeal origin. Mol Biol Evol 32, 726–739 (2015). 10.1093/molbev/msu334 [PubMed: 25525215]
27. Hennell James R et al. Functional reconstruction of a eukaryotic-like E1/E2/(RING) E3 ubiquitylation cascade from an uncultured archaeon. Nat Commun 8, 1120 (2017). 10.1038/s41467-017-01162-7 [PubMed: 29066714]
28. Ledvina HE et al. An E1-E2 fusion protein primes antiviral immune signalling in bacteria. Nature 616, 319–325 (2023). 10.1038/s41586-022-05647-4 [PubMed: 36755092]
29. Jenson JM, Li T, Du F, Ea CK & Chen ZJ Ubiquitin-like Conjugation by Bacterial cGAS Enhances Anti-phage Defence. Nature (2023). 10.1038/s41586-023-05862-7
30. Jumper J et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). 10.1038/s41586-021-03819-2 [PubMed: 34265844]
31. Payne LJ et al. PADLOC: a web server for the identification of antiviral defence systems in microbial genomes. Nucleic Acids Res 50, W541–550 (2022). 10.1093/nar/gkac400 [PubMed: 35639517]

32. Makarova KS, Wolf YI, Snir S & Koonin EV Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. J Bacteriol 193, 6039–6056 (2011). 10.1128/JB.05535-11 [PubMed: 21908672]
33. Doron S et al. Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359, eaar4120 (2018). 10.1126/science.aar4120 [PubMed: 29371424]
34. Tesson F et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat Commun 13, 2561 (2022). 10.1038/s41467-022-30269-9 [PubMed: 35538097]
35. Lau RK, Enustun E, Gu Y, Nguyen JV & Corbett KD A conserved signaling pathway activates bacterial CBASS immune signaling in response to DNA damage. EMBO J 41, e111540 (2022). 10.15252/embj.2022111540 [PubMed: 36156805]
36. Yuan L et al. Crystal structures reveal catalytic and regulatory mechanisms of the dual-specificity ubiquitin/FAT10 E1 enzyme Uba6. Nat Commun 13, 4880 (2022). 10.1038/s41467-022-32613-5 [PubMed: 35986001]
37. Huang DT et al. Basis for a ubiquitin-like protein thioester switch toggling E1-E2 affinity. Nature 445, 394–398 (2007). 10.1038/nature05490 [PubMed: 17220875]
38. Olsen SK, Capili AD, Lu X, Tan DS & Lima CD Active site remodelling accompanies thioester bond formation in the SUMO E1. Nature 463, 906–912 (2010). 10.1038/nature08765 [PubMed: 20164921]
39. Olsen SK & Lima CD Structure of a ubiquitin E1-E2 complex: insights to E1-E2 thioester transfer. Mol Cell 49, 884–896 (2013). 10.1016/j.molcel.2013.01.013 [PubMed: 23416107]
40. Yan Y et al. Ubiquitin-like cGAS chain formation by a super enzyme activates anti-phage response. bioRxiv (2023). 10.1101/2022.05.25.493364
41. Liu S et al. Insights into the evolution of the ISG15 and UBA7 system. Genomics 114, 110302 (2022). 10.1016/j.ygeno.2022.110302 [PubMed: 35134494]
42. Dzimianski JV, Scholte FEM, Bergeron E & Pegan SD ISG15: It's Complicated. J Mol Biol 431, 4203–4216 (2019). 10.1016/j.jmb.2019.03.013 [PubMed: 30890331]
43. Freitas BT, Scholte FEM, Bergeron E & Pegan SD How ISG15 combats viral infection. Virus Res 286, 198036 (2020). 10.1016/j.virusres.2020.198036 [PubMed: 32492472]

Additional References

44. Katoh K & Standley DM MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution 30, 772–780 (2013). 10.1093/molbev/mst010 [PubMed: 23329690]
45. Waterhouse AM, Procter JB, Martin DMA, Clamp M & Barton GJ Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191 (2009). 10.1093/bioinformatics/btp033 [PubMed: 19151095]
46. Mirdita M et al. ColabFold: making protein folding accessible to all. Nat Methods 19, 679–682 (2022). 10.1038/s41592-022-01488-1 [PubMed: 35637307]
47. Evans R et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv (2022). 10.1101/2021.10.04.463034
48. Steinegger M & Soding J MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 35, 1026–1028 (2017). 10.1038/nbt.3988 [PubMed: 29035372]
49. Kabsch W XDS. Acta crystallographica Section D, Biological crystallography 66, 125–132 (2010). 10.1107/S0907444909047337 [PubMed: 20124692]
50. Evans P Scaling and assessment of data quality. Acta crystallographica Section D, Biological crystallography 62, 72–82 (2006). 10.1107/S0907444905036693 [PubMed: 16369096]
51. Evans PR & Murshudov GN How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr 69, 1204–1214 (2013). 10.1107/S0907444913000061 [PubMed: 23793146]
52. McCoy AJ et al. Phaser crystallographic software. Journal of applied crystallography 40, 658–674 (2007). 10.1107/S0021889807021206 [PubMed: 19461840]

53. Emsley P, Lohkamp B, Scott WG & Cowtan K Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66, 486–501 (2010). 10.1107/S0907444910007493 [PubMed: 20383002]
54. Afonine PV et al. Towards automated crystallographic structure refinement with phenix.refine. Acta crystallographica Section D, Biological crystallography 68, 352–367 (2012). 10.1107/S0907444912001308 [PubMed: 22505256]
55. Zhou W, Ryan JJ & Zhou H Global analyses of sumoylated proteins in Saccharomyces cerevisiae. Induction of protein sumoylation by cellular stresses. J Biol Chem 279, 32262–32268 (2004). 10.1074/jbc.M404173200 [PubMed: 15166219]
56. Deutsch EW et al. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. Proteomics Clin Appl 9, 745–754 (2015). 10.1002/prca.201400164 [PubMed: 25631240]
57. Deutsch EW et al. Trans-Proteomic Pipeline: Robust Mass Spectrometry-Based Proteomics Data Analysis Suite. J Proteome Res 22, 615–624 (2023). 10.1021/acs.jproteome.2c00624 [PubMed: 36648445]
58. Suzuki N et al. Crystallization of small proteins assisted by green fluorescent protein. Acta Crystallogr D Biol Crystallogr 66, 1059–1066 (2010). 10.1107/S0907444910032944 [PubMed: 20944239]
59. Streich FC Jr. & Lima CD Capturing a substrate in an activated RING E3/E2-SUMO complex. Nature 536, 304–308 (2016). 10.1038/nature19071 [PubMed: 27509863]
60. Scott DC et al. A dual E3 mechanism for Rub1 ligation to Cdc53. Mol Cell 39, 784–796 (2010). 10.1016/j.molcel.2010.08.030 [PubMed: 20832729]
61. Juang YC et al. OTUB1 co-opts Lys48-linked ubiquitin recognition to suppress E2 enzyme function. Mol Cell 45, 384–397 (2012). 10.1016/j.molcel.2012.01.011 [PubMed: 22325355]
62. Fuchs ACD, Maldoner L, Wojtynek M, Hartmann MD & Martin J Rpn11-mediated ubiquitin processing in an ancestral archaeal ubiquitination system. Nat Commun 9, 2696 (2018). 10.1038/s41467-018-05198-1 [PubMed: 30002364]
63. Shrestha RK et al. Insights into the mechanism of deubiquitination by JAMM deubiquitinases from cocrystal structures of the enzyme with the substrate and product. Biochemistry 53, 3199–3217 (2014). 10.1021/bi5003162 [PubMed: 24787148]

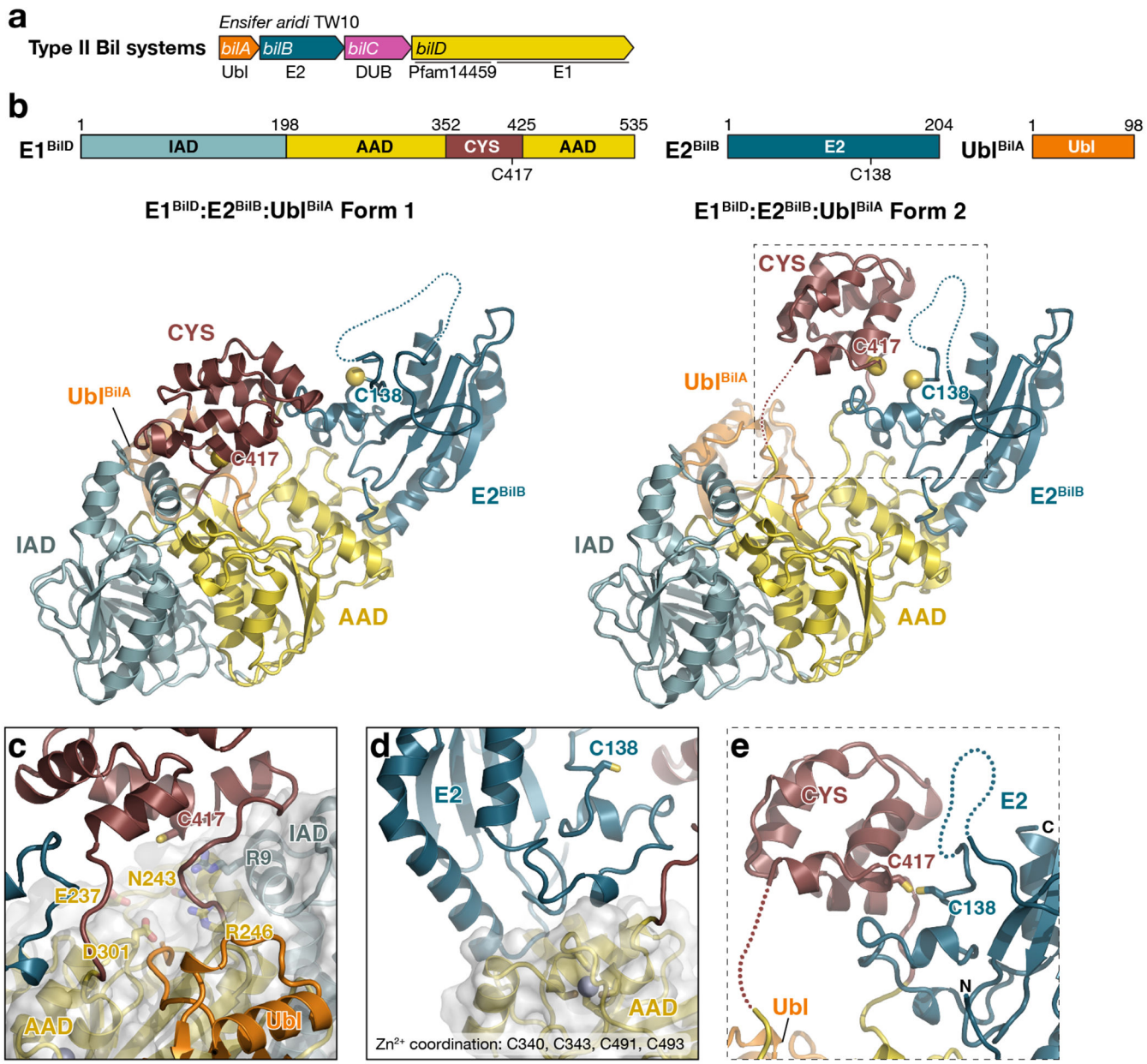


Figure 1. Structure of a bacterial E1-E2-Ubl complex

(a) Schematic of a Type II BilABCD operon from *Ensifer aridi* TW10. See Supplementary Table 1 for a full list of Type II BilABCD operons, and Extended Data Figure 1 for comparison with Type I BilABCD operons. **(b)** *Top:* Domain schematic of *E. aridi* $E1^{BilD}$, $E2^{BilB}$, and Ubl^{BilA}. The inactive adenylation domain (IAD) of $E1^{BilD}$ is colored light blue, the active adenylation domain (AAD) yellow, and the α-helical insertion containing the putative catalytic cysteine (CYS domain) brown. Putative catalytic cysteine residues in $E1^{BilD}$ (C417) and $E2^{BilB}$ (C138) are noted. *Left: E. aridi* $E1^{BilD}$:$E2^{BilB}$:Ubl^{BilA} complex Form 1 crystal structure, with domains colored as in the schematic. Putative catalytic cysteine residues are shown as yellow spheres. *Right: E. aridi* $E1^{BilD}$:$E2^{BilB}$:Ubl^{BilA} complex Form 2 crystal structure. See Extended Data Figure 3 for comparison to a canonical

eukaryotic E1(NAE1-UBA3):E2(Ubc12):Ubl(NEDD8) complex. **(c)** Closeup view of the Ubl^{BilA} C-terminus (orange) docked in the E1BilD adenylation active site (yellow, with conserved active-site residues shown as sticks and labeled). The E1 catalytic cysteine residue (C417) is positioned 14 Å away from the Ubl C-terminus. See Extended Data Figure 3c, g for comparison of bacterial and eukaryotic E1 adenylation active sites. **(d)** Closeup view of the $E1^{BilD}$-$E2^{BilB}$ binding interface, including the structural Zn^{2+} ion coordinated by four cysteine residues in $E1^{BilD}$. Extended Data Figure 3d, h for comparison of bacterial and eukaryotic E1-E1 interfaces. **(e)** Closeup view of the $E1^{BIlD}$ and $E2^{BilB}$ catalytic cysteine positions in the Form 2 structure. Compared to Form 1, the E1 CYS domain is rotated upward and the $E1^{BilD}$ catalytic cysteine (C417) is positioned within 2 Å of the E2 catalytic cysteine (C138), mimicking the structural state adopted during an $E1^{BilD}$-to-$E2^{BilB}$ transthioesterification reaction.

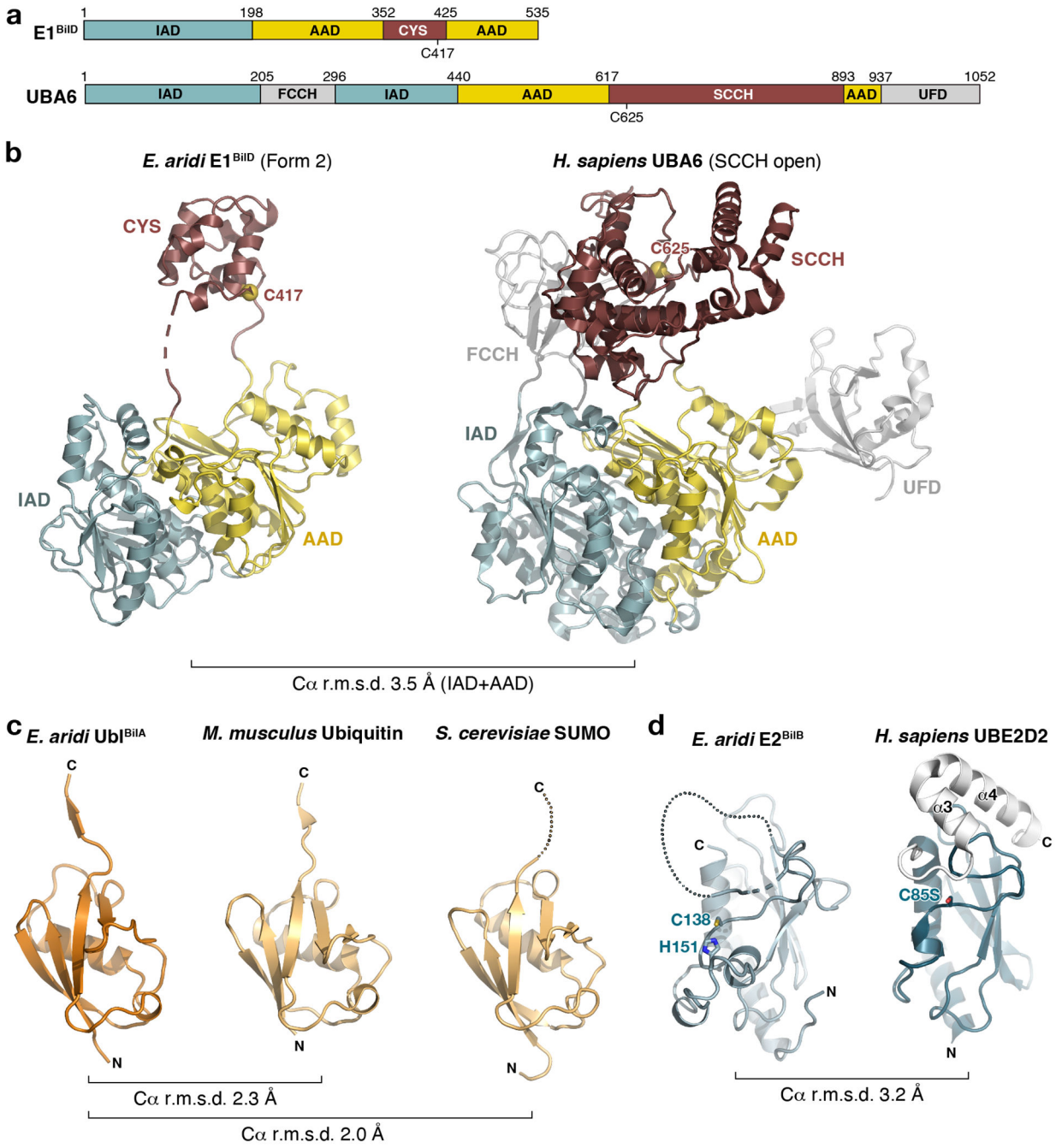


Figure 2. *E. aridi* E1^{BilD}, E2^{BilB}, and Ubl^{BilA} resemble canonical eukaryotic ubiquitination machinery

(a) Domain schematic of *E. aridi* E1^{BilD} and the *H. sapiens* ubiquitin/FAT10 E1 UBA6 (PDB ID 7SOL)[36]. IAD: inactive adenylation domain; AAD: active adenylation domain; CYS: catalytic cysteine-containing domain; FCCH: first catalytic cysteine half-domain; SCCH: second catalytic cysteine half-domain (equivalent to CYS); UFD: ubiquitin fold domain. **(b)** Structures of *E. aridi* E1^{BilD} (left) and *H. sapiens* UBA6 (PDB ID 7SOL)[36], with domains colored as in panel (a) and catalytic cysteines shown as spheres and colored

yellow. The Cα r.m.s.d. is 3.5 Å over 401 residue pairs spanning the two proteins' IAD and AAD domains. **(c)** Structures of *E. aridi* Ubl^{BilA} (left), *M. musculus* ubiquitin (center; PDB ID 4NQL)[63], and *S. cerevisiae* SUMO/Smt3 (right; PDB ID 5JNE)[59]. **(d)** Structures of *E. aridi* $E2^{BilB}$ (left) and the *H. sapiens* E2 UBE2D2 (right; PDB ID 4DGG)[61]. Each protein's catalytic cysteine residue is shown as sticks and labeled (Cys85 is mutated to serine in the *H. sapiens* UBE2D2 structure. *E. aridi* $E2^{BilB}$ His151 is also shown as sticks. See Extended Data Figure 4a for an equivalent view showing all four conserved residues of the originally identified CEHH motif. For UBE2D2, the C-terminal α-helices not shared by $E2^{BilB}$ are shown in white.

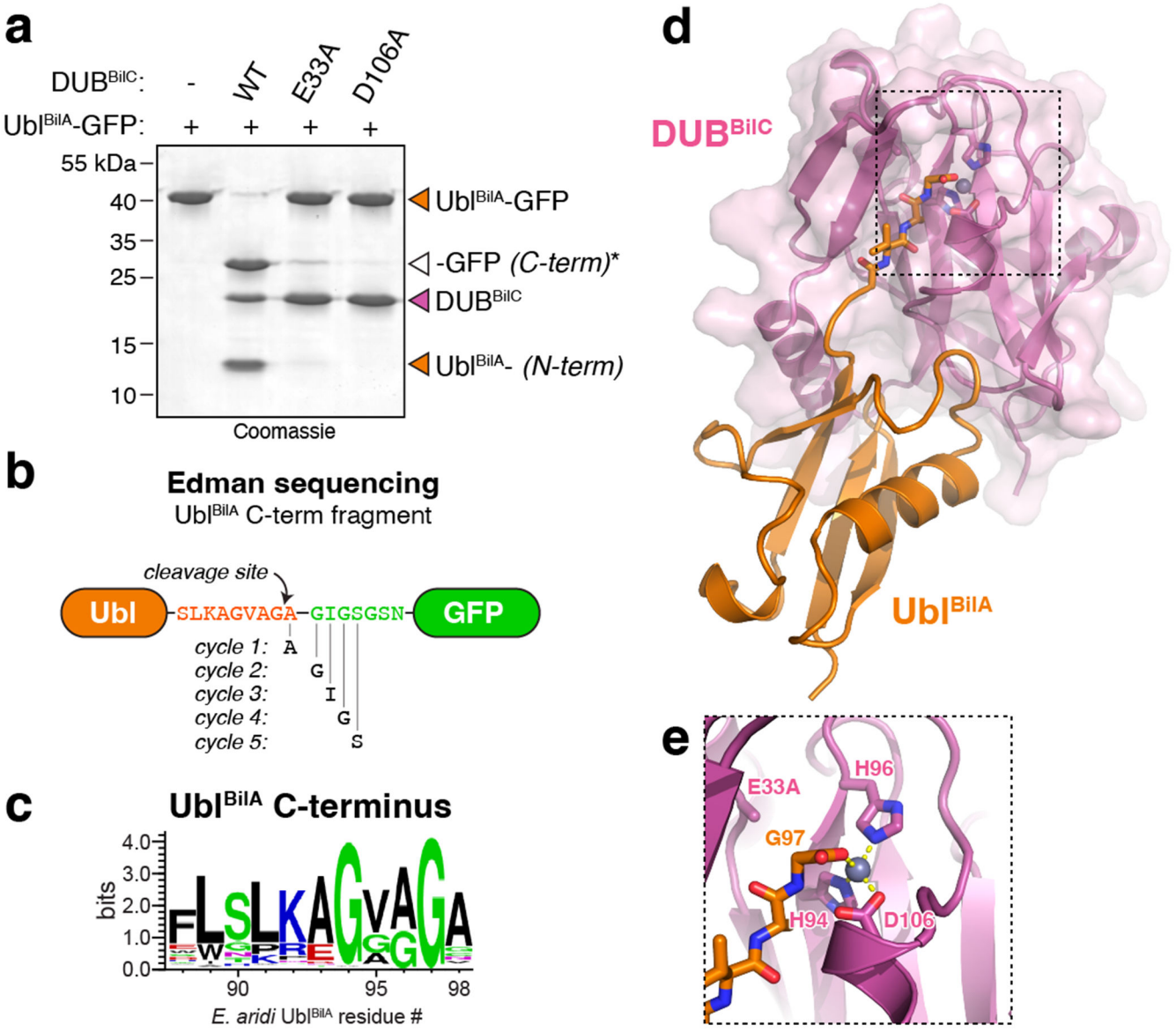


Figure 3. *E. aridi* DUB^{BilC} specifically cleaves Ubl^{BilA}

(a) Biochemical analysis of *E. aridi* DUB^{BilC} cleaving a Ubl^{BilA}-GFP fusion protein. DUB^{BilC} constructs used were wild-type (WT), E33A active site mutant, or D106A active site mutant. For gel source data, see Supplementary Figure 1. This experiment was independently performed three times, with consistent results. **(b)** Schematic of the cleavage reaction shown in panel (a) and analysis by N-terminal sequencing (Edman degradation) of the C-terminal fragment (marked with an asterisk in panel (a)). See Extended Data Figure 5a for N-terminal sequencing data showing the sequence AGIGS, pinpointing the DUB^{BilC} cleavage site as C-terminal to Ubl^{BilA} Gly97. **(c)** Sequence logo from bacterial Type II Ubl^{BilA} proteins (Supplementary Table 1). Type II Ubl^{BilA} homologs possess up to nine residues C-terminal to the highly conserved glycine (Gly97 in *E. aridi* Ubl^{BilA}). **(d)** X-ray crystal structure of DUB^{BilC}(E33A) (purple) bound to Ubl^{BilA} (orange). A composite omit map (Extended Data Figure 5b) shows that Ubl^{BilA} is cleaved at Gly97. See Extended

Data Figure 5d for structural comparisons with eukaryotic JAMM-family peptidases. **(e)** Closeup view of the DUB^{BilC}(E33A) active site with bound zinc ion (gray) and the Ubl^{BilA} C-terminus (orange).

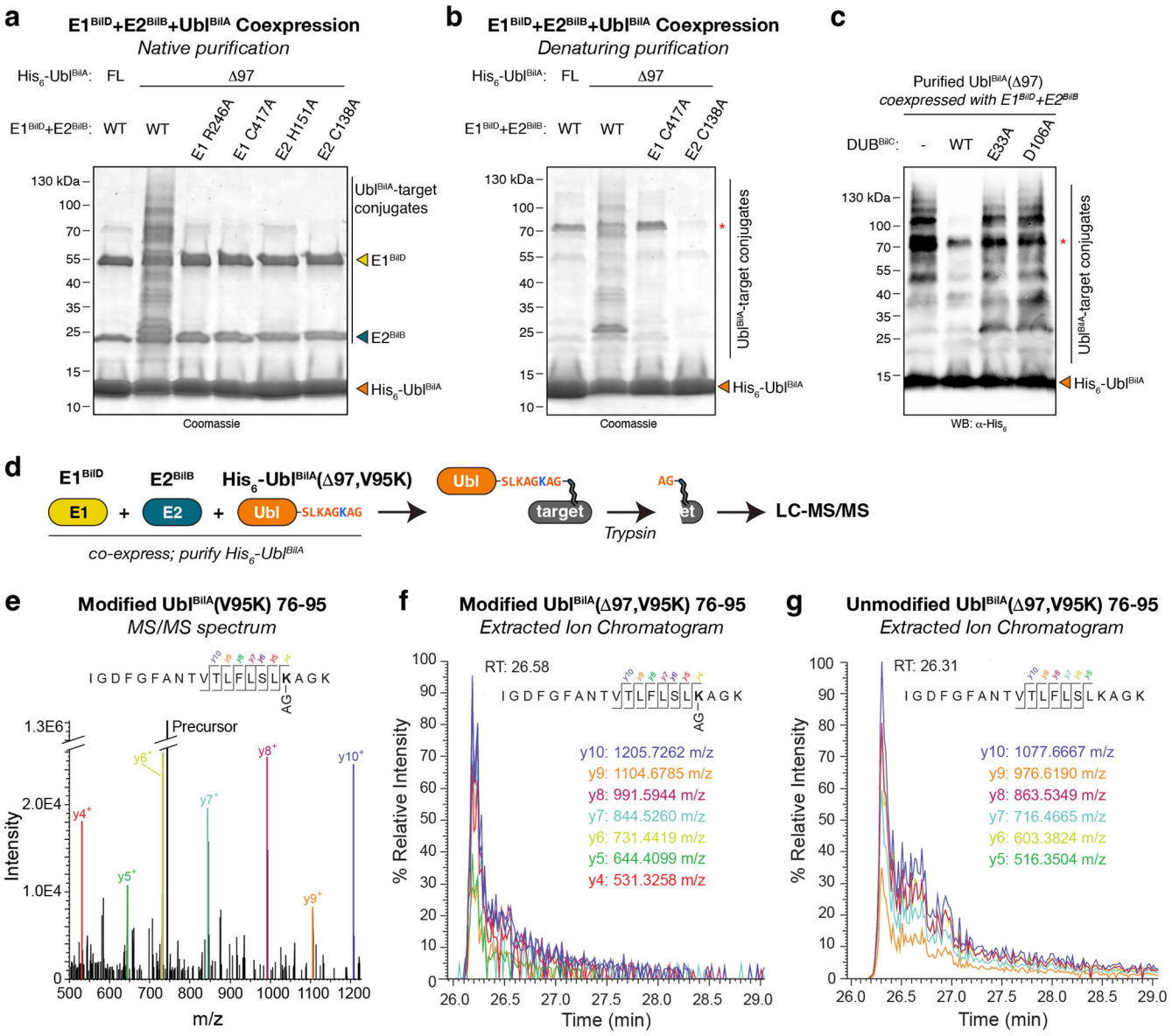


Figure 4. Bacterial BilABCD systems catalyze ubiquitination

(a) SDS-PAGE analysis of Ni^{2+} affinity-purified His_6-Ubl^{BilA} and associated proteins in native conditions, after coexpression with $E1^{BilD}$ and $E2^{BilB}$. Bands representing His_6-Ubl^{BilA}, $E1^{BilD}$, and $E2^{BilB}$ are marked. Ubl^{BilA}-target conjugates are visible in the second lane. Ubl^{BilA} FL: full-length, unreactive Ubl^{BilA}. Ubl^{BilA} Δ97: truncated at glycine 97 to enable E1-mediated catalysis; E1 R246A: adenylation active site mutant; E1 C417A: catalytic cysteine mutant; E2 C138A: catalytic cysteine mutant. WT: wild type. For gel source data for panels (a)-(c), see Supplementary Figure 1. Experiments shown in panels (a)-(c) were independently performed three times, with consistent results. **(b)** SDS-PAGE analysis of Ni^{2+} affinity-purified His_6-Ubl^{BilA} and associated proteins in denaturing conditions, after coexpression with $E1^{BilD}$ and $E2^{BilB}$. Bands representing His_6-Ubl^{BilA} and His_6-Ubl^{BilA}-target conjugates are marked. Red asterisk: contaminant. **(c)** SDS-PAGE and

anti His_6-tag western blot analysis of affinity-purified His_6-Ubl^{BilA}-target conjugates (from denaturing purifications) treated with DUB^{BilC}. **(d)** Experimental scheme for identification of Ubl^{BilA} targets. See Extended Data Figure 6d for purification of His_6-Ubl^{BilA}(Δ97, V95K)-target complexes. **(e)** MS/MS spectrum of modified Ubl^{BilA}(Δ97, V95K) residues 76-95. Y-ions shifted by the dipeptide Ala-Gly (AG) remnant are annotated in color. **(f)** Extracted Ion Chromatograms (EICs) of transition ions of the mis-cleaved, modified Ubl^{BilA}(Δ97, V95K) residues 76-95, using a 10 ppm m/z tolerance. RT: retention time. **(g)** Extracted Ion Chromatograms (EICs) of transition ions of the mis-cleaved, unmodified Ubl^{BilA}(Δ97, V95K) residues 76-95, using a 10 ppm m/z tolerance. RT: retention time. See Extended Data Figure 6e for EIC of the properly cleaved, unmodified peptide spanning Ubl^{BilA}(Δ97, V95K) residues 76-92.

Table 1.

Structural similarity between *Ensifer aridi* BilABCD proteins and homologs

Root mean squared deviation (r.m.s.d.) for Cα atoms when comparing each *Ensifer aridi* protein to homologous eukaryotic proteins.

Protein	Homolog	Cα r.m.s.d.	Reference
$E1^{BilD}$	*H. sapiens* UBA6	3.5 Å (401 Cα)	36
$E1^{BilD}$	*H. sapiens* NAE1-UBA3	NAE1: 3.6 Å (221 Cα) UBA3: 3.7 Å (250 Cα)	37
Ubl^{BilA}	*M. musculus* ubiquitin	1.7 Å (72 Cα)	58
Ubl^{BilA}	*S. cerevisiae* Smt3	1.7 Å (73 Cα)	59
Ubl^{BilA}	*E. coli* ThiS	3.3 Å (57 Cα)	6
Ubl^{BilA}	*E. coli* MoaD	3.3 Å (64 Cα)	5
$E2^{BilB}$	*S. cerevisiae* UBC12	3.0 Å (94 Cα)	60
$E2^{BilB}$	*H. sapiens* UBE2D2	3.2 Å (99 Cα)	61
DUB^{BilC}	*C. subterraneum* Rpn11	2.5 Å (110 Cα)	62
DUB^{BilC}	*S. pombe* Sst2	1.2 Å (79 Cα)	63