The Molecular Basis for Ubiquitin and Ubiquitin-like Specificities in Bacterial Effector Proteases

Highlights
- Bacterial CE proteases exhibit distinct ubiquitin/ubiquitin-like specificities
- Substrate specificity is acquired through variability in three common regions
- Structural and functional data redefine CE clan relationships across kingdoms
- CE effectors are fitted with accessory domains that modulate function

In Brief
Focusing on examples from pathogenic bacteria, Pruneda et al. examine a family of proteases that displays remarkably distinct specificities toward ubiquitin and ubiquitin-like modifications. Leveraging structural and functional data, the authors derive mechanisms through which substrate specificity is achieved and redefine relationships within the enzyme family across kingdoms of life.

Accession Numbers
- 5HAF
- 5HAG
- 5HAM
- 5JP3
- 5JP1
The Molecular Basis for Ubiquitin and Ubiquitin-like Specificities in Bacterial Effector Proteases

Jonathan N. Pruneda,1 Charlotte H. Durkin,2 Paul P. Geurink,3 Huib Ovaa,3 Balaji Santhanam,4 David W. Holden,2 and David Komander1,*

1Division of Protein and Nucleic Acid Chemistry, MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK
2Section of Microbiology, MRC Centre for Molecular Bacteriology and Infection, Imperial College London, London SW7 2AZ, UK
3Division of Cell Biology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 Amsterdam, the Netherlands
4Division of Structural Studies, MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK
*Correspondence: dk@mrc-imb.cam.ac.uk
http://dx.doi.org/10.1016/j.molcel.2016.06.015

SUMMARY

Pathogenic bacteria rely on secreted effector proteins to manipulate host signaling pathways, often in creative ways. CE clan proteases, specific hydrolases for ubiquitin-like modifications (SUMO and NEDD8) in eukaryotes, reportedly serve as bacterial effector proteins with deSUMOylase, deubiquitinase, or, even, acetyltransferase activities. Here, we characterize bacterial CE protease activities, revealing K63-linkage-specific deubiquitinases in human pathogens, such as Salmonella, Escherichia, and Shigella, as well as ubiquitin/ubiquitin-like cross-reactive enzymes in Chlamydia, Rickettsia, and Xanthomonas. Five crystal structures, including ubiquitin/ubiquitin-like complexes, explain substrate specificities and redefine relationships across the CE clan. Importantly, this work identifies novel family members and provides key discoveries among previously reported effectors, such as the unexpected deubiquitinase activity in Xanthomonas XopD, contributed by an unstructured ubiquitin binding region. Furthermore, accessory domains regulate properties such as subcellular localization, as exemplified by a ubiquitin-binding domain in Salmonella Typhimurium SseL. Our work both highlights and explains the functional adaptations observed among diverse CE clan proteins.

INTRODUCTION

Bacterial colonization and proliferation within a eukaryotic host depends on processes that antagonize immune responses, ensure bacterial survival, and promote replication. For this, bacteria employ a repertoire of effector proteins, ranging in number from dozens to hundreds, which are directly delivered into the host cell by sophisticated secretion machineries (Figueira and Holden, 2012). Once inside, these effector proteins can hijack host factors and catalyze chemical modifications that are, in some cases, entirely foreign to the eukaryotic system (Salomon and Orth, 2013).

Bacterial effectors often target kinase cascades involved in inflammation (Salomon and Orth, 2013) or the eukaryotic ubiquitin (Ub) system (Corn and Vucic, 2014) to promote bacterial pathogenicity. Ub regulates countless cellular processes by forming structurally and functionally distinct polymeric chains that can be used independently or in concert during complex signaling cascades (Swatek and Komander, 2016). In addition, ubiquitin-like (Ubi) proteins, such as NEDD8, SUMO, and ISG15, play similarly important roles in host cell biology (van der Veen and Ploegh, 2012) and invasion response pathways (Liu et al., 2013; Radosevich et al., 2015).

Ub and Ubi modifications are tightly regulated by discrete families of proteases. Humans encode ~80 active deubiquitinases (DUBs) that are unable to hydrolyze SUMO or NEDD8, although there are a few exceptions (Clague et al., 2013). The CE protease clan in humans consists of six SUMO-specific SENPs and the NEDD8-specific NEDP1/SENP8 that are collectively termed Ubl proteases (ULPs) (Ronau et al., 2016). Mechanisms dictating the exquisite specificities of these SUMO and NEDD8 proteases have been characterized in detail (Reverter and Lima, 2004; Reverter et al., 2005; Shen et al., 2005).

Bacteria and viruses also encode CE clan enzymes as effectors to interrupt eukaryotic host response processes (Ronau et al., 2016). Importantly, the characterized members from bacteria display diverse enzymatic activities and include not only ULPs (Kim et al., 2008; Orth et al., 2000) but also DUBs (Rytkönen et al., 2007; Misaghi et al., 2006; Zhou et al., 2005; Ye et al., 2007). In addition, some CE effectors, such as Yersinia pestis YopJ, display an unusual Ser/Thr acetyltransferase activity (Mittal et al., 2006; Mukherjee et al., 2006; Jones et al., 2008). Hence, it appears that the CE fold is highly versatile and adaptable and has evolved in bacterial pathogens to accommodate a range of enzymatic activities. However, the proteolytic activities and level of cross-reactivity in CE clan effectors are currently unclear since comprehensive comparisons have not been performed, and the molecular basis for achieving distinct target specificities from a single enzyme fold is unknown. Published
structures of individual bacterial (Chosed et al., 2007; Sheedlo et al., 2015) and viral examples (Ding et al., 1996) show a similarity to eukaryotic ULPs, but fall short of providing generalizable themes to explain diverse substrate specificities. The uncertainties regarding activity within this diverse family of enzymes have hindered a detailed understanding of the potential role(s) carried out during bacterial invasion.

To address this shortcoming, we cloned and expressed a panel of bacterial CE enzymes for in-depth biochemical and structural analysis. We identify several Ub-specific CE proteases, including the first DUBs in Shigella (ShiCE) and Rickettsia (RickCE), and show that most prefer K63-linked chains. Ub and K63 specificity is explained by a set of crystal structures, which show immense diversity in their convergent adaptations of a Ub recognition site. By comparing these bacterial examples with their Ub-specific counterparts in eukaryotes, we identify and define three regions of variability within the CE fold that can be tailored to suit Ub or Ubl specificity. With this knowledge we revisited the Xanthomonas effector protein XopD, a reported deSUMOylase in plants, to identify an extended construct that targets not only SUMO but, surprisingly, also Ub modifications. Structures of Ub- and SUMO-bound XopD reveal striking plasticity in substrate recognition, due, in large part, to an N-terminal low-complexity Ub binding region (LC-UBR) that binds the Ub Ile44 patch but lacks secondary structure. The cumulative structural and functional data enable construction of a robust family dendrogram for CE clan enzymes across kingdoms. Finally, we uncover the functional importance for accessory domains outside the CE fold that, in the case of SseL, serve to target the effector to Salmonella-containing vacuoles (SCVs) during infection. Our comprehensive analysis of bacterial proteases redefines the roles they play during infection and establishes a framework for cross-kingdom relationships among the entire CE clan.

RESULTS
Selection and Properties of Bacterial CE Clan Effectors
While the presence of CE clan effectors in bacteria and biochemical functions for some members have been described, a comprehensive comparison of proteolytic specificities has not been performed. To fill this gap, we selected a divergent set of enzymes from human bacterial pathogens (Figure 1A), including putative DUBs SseL (Salmonella enterica serovar Typhimurium, hereafter S. Typhimurium [Rytkönen et al., 2007]), ChlaDUB1 (Chlamydia trachomatis [Misaghi et al., 2006]), and EIoD (Escherichia coli [Catic et al., 2007]), as well as Ser/Thr acetyltransferases YopJ (Yersinia pestis [Mukherjee et al., 2006; Mittal et al., 2006]) and AvrA (S. Typhimurium [Jones et al., 2008]). To this, we added entirely uncharacterized CE clan proteins RickCE (Rickettsia bellii, GenBank: ABEO4279.1), LegCE (Legionella pneumophila ssp. pneumophila, GenBank: AAU28953.1), and ShiCE (Shigella flexneri; GenBank: EGK20985.1) (Figure 1A). In all of the selected enzymes, the catalytic CE fold is present in the context of additional domains of unknown function (Figures 1A and S1A), only few of which bear sequences similar to domains found in eukaryotes (see below). Following cloning and the subsequent E. coli expression of suitable constructs (Figure 1A), the panel of CE enzymes was purified to homogeneity (Figure 1B). The CE fold contains a conserved catalytic Cys (Figure S1B), which was mutated in each enzyme to Ala (e.g., SseL<sup>A</sup>) and purified analogously (Figure 1B).

Activities of CE Clan Bacterial Effectors
Cys-based DUBs can be covalently modified by Ub-based suicide probes, in which the C terminus has been modified with an electrophilic trap, such as propargylamine (Ub-PA; Ekebus et al., 2013). SseL, ChlaDUB1, and EIoD, as well as the previously uncharacterized ShiCE and RickCE, were all covalently modified by Ub-PA in a catalytic Cys-dependent manner, while the remaining enzymes, including YopJ, AvrA, and LegCE, remained unmodified (Figure 1C).

To assess all Ub and Ubl proteolysis in a simple, parallel assay, we used fluorescence polarization (FP) to measure hydrolysis of Ub, SUMO1, NEDD8, and ISG15 substrates that were isopeptide linked to a fluoroently labeled Lys-Gly peptide (Figure 1D) (Geurink et al., 2012; Basters et al., 2014). Isopeptide linkage imposes a more chemically precise environment for a Ub/Ubl modification. When tested against the full panel, human SENP1 and NEDP1 exclusively cleaved SUMO1- and NEDD8-modified peptides, respectively (Figure S1C). YopJ, AvrA, and LegCE showed no proteolytic activity in our assays, regardless of the presence of the cofactor inositol hexakisphosphate (IP6) (Mittal et al., 2010) (Figures 1E and S1D–S1F). Furthermore, AvrA showed no sign of an interaction with Ub when measured by nuclear magnetic resonance (NMR) titration (Figure S1G). YopJ and AvrA were, however, highly active in an in vitro autoacetylation assay with IP6, and LegCE, likewise, showed modest acetylation activity compared to its catalytically inactive mutant (Figure S1H), indicating that these effectors are dedicated acetyltransferases.

Interestingly, the remaining bacterial effectors SseL, EIoD, and ShiCE are Ub-specific proteases, and ChlaDUB1 and RickCE cleave both Ub- and, to a lesser extent, NEDD8-modified peptides (Figure 1E). Like NEDP1, ChlaDUB1 can remove the C-terminal five amino acids from pro-NEDD8 with peptidase activity (Figure S1I), but does not cleave the regulatory NEDD8 modification from the cullin RING ligase adaptor Cul1 (Figure S1J).

As there are many examples of interesting polyUb chain specificity in eukaryotic DUBs (Mevissen et al., 2013), it was important to test if any bacterial CE DUBs showed a preference for particular polyUb substrates. Early studies on SseL indicated higher activity for K63-linked over K48-linked chains (Rytkönen et al., 2007), and Legionella pneumophila effector SdeA prefers K63 linkages over K11 and K48 linkages (Sheedlo et al., 2015), but preference among all eight possible Ub-Ub linkages for these and all other enzymes have remained unstudied. Strikingly, we found that CE DUBs encoded by human pathogens showed strong preference for K63-linked chains, only targeting K48 and K11 chains at later time points or higher enzyme concentrations (Figures 1F and S1K–S1M), indicating significant pressure to remove this important infection-associated post-translational modification during invasion (Corn and Vucic, 2014).

Structural Analysis of the CE Proteases
The new functionalities in CE clan effectors prompted structural characterization to reveal how this enzyme fold evolved to
accommodate such divergent proteolytic activities. We determined high-resolution crystal structures of the CE clan effectors SseL, RickCE, and ChlaDUB1 (Figures 2A and S2A–S2C; Table 1). Due to low-sequence similarities, all structures were experimentally phased (see Experimental Procedures). The crystallized full-length SseL construct comprised an additional N-terminal ~135-residue helical domain (see below). The catalytic CE fold bears similarity to eukaryotic CE proteases (Figure 2B), and the conserved active-site residues are required for DUB activity (Figures S2D and S2E). However, several regions of high variability were immediately apparent.

A comparison to the previously determined SENP2-SUMO2 and NEDP1-NEDD8 complexes (Figure 2C) (Reverter and Lima, 2004; Reverter et al., 2005; Shen et al., 2005) revealed differences among bacterial CE DUBs in three variable regions (VRs; Figure 2D) near the substrate binding (S1) site that were...
arranged around a structurally conserved helix (constant region, CR). Variations within these three regions are found in all examples of the CE fold (Figures 2E-2I), as detailed below.

The first variable region, VR-1, is found at the beginning of the CE fold (Figure 2D), which in NEDP1 and the SENP family is a β-hairpin that forms electrostatic interactions with the Ubl (Figures 2C, 2H, and 2I). In bacterial examples, the analogous region is either disordered in the absence of substrate (SseL; Figures 2A and 2E) or not part of the crystallized construct (RickCE; Figures 2A and 2G). A corresponding VR-1 region is absent in ChlaDUB1; interestingly, the ChlaDUB1 C terminus occupies this space instead (Figures 2A and 2F).

The second variable region, VR-2, is located between the first two strands of the central β sheet, and in the eukaryotic examples forms hydrophobic contacts with the Ubl β sheet (Figure 2D). Intriguingly, this region is similar among the Ubl proteases (Figures 2H and 2I), but diverse among the CE-fold DUBs (Figures 2E-2G). In fact, RickCE has an ~75-residue insertion at this site that forms a helical arm in the crystal structure (Figure 2A).

The third variable region, VR-3, is an insertion between the second and third strands of the β sheet, just preceding the catalytic His (Figure 2D). NEDP1 and ChlaDUB1 contain an insertion of 7 and 24 residues, respectively, in what is a short β-turn in the other structures (Figures 2A, 2C, 2F, and 2I). In NEDP1, VR-3 forms an extended loop that guides the NEDD8 C terminus into the active site through backbone hydrogen bonding (Figures 2C and 2I). In contrast, ChlaDUB1 forms a helix that may serve a similar purpose in recognition of the Ub C terminus (Figures 2A and 2F).

Overall, the dramatic differences in substrate recognition imparted by CE-fold variable regions suggest that bacteria have adopted different strategies of altering the CE-fold to evolve DUB activity.

Re-evaluation of XopD Ub/Ubl Specificity

During our analysis of the CE fold, we noticed that a putative VR-1 region had been omitted in the crystallized construct of the Xanthomonas campestris effector XopD (Figure 3A) (Chosed et al., 2007). Given the importance of all three variable regions in the eukaryotic Ub-bound structures (Figures 2C, 2H, and 2I), we compared the activity of XopD without (ΔVR-1; amino acids [aa] 335–515) and with the analogous VR-1 region (+VR-1, aa 298–515). In a suicide probe assay, XopD ΔVR-1 reacted solely with Solarum lycopersicum tomato SUMO (tSUMO), while the plant pathogen clearly prioritizes distinct linkages, while the plant pathogen clearly prioritizes distinct linkages and SUMO, which may, hence, mediate anti-bacterial signaling in plants.

Molecular Analysis of XopD Cross-Specificity

To understand what structural adaptations enable XopD cross-reactivity, we determined the crystal structures of covalent XopD–Ub (2.9 Å; Figures 3E and S3B; Table 1) and XopD–tSUMO (2.1 Å; Figures 3F and S3C; Table 1) complexes. The XopD–Ub structure contained four copies in the asymmetric unit, all of which were similar (0.62 Å root-mean-square deviation [RMSD] over XopD and Ub; Figure S3D). Moreover, the core fold of XopD itself was highly similar in the Ub-bound, tSUMO-bound, and the published apo structure (PDB: 2OIV, 0.61 Å RMSD over aa 338–515; Figure S3E). To our surprise, the VR-1 responsible for much of the observed XopD protease activity was not a β-hairpin as had been observed in eukaryotic Ubl proteases (Figure 2D), but, instead, was extended and devoid of secondary structure. Furthermore, the extended VR-1 conformation was completely different between the Ub- and tSUMO-bound complexes. A comparison of the two structures clearly explains the observed VR-1 dependence in activity. While the Ub S1 site primarily consists of contacts to VR-1 and some to VR-2, tSUMO binding displays the opposite trend, with the majority of interactions taking place at VR-2 (Figures 3E-3G).

In the Ub-bound structure, XopD VR-1 threads underneath the Ub Ile44 hydrophobic patch, with residues Pro322 and Val325 making contacts to Ub Ile44 and His68, and also coordinates Ub Arg72 with Asp327 (Figure 3H, foreground). Additionally, Met374 of VR-2 contacts the LeuB loop of Ub (Figure 3H, background). Incorporation of either I44A or R72A mutations into Ub suicide probes resulted in a dramatic loss in reactivity to XopD (Figure 3J).

Figure 2. Structural Analysis of Bacterial CE Deubiquitinases

(A) Cartoon representations of SseL (2.7 Å, teal), ChlaDUB1 (2.1 Å, yellow), and RickCE (2.4 Å, violet) aligned on the catalytic triad (ball and stick).

(B) Superposition of the CE core based on the catalytic triad for structures in (A) and human SENP2 (PDB: 1TH0).

(C) Representations of the human SENP2-SUMO2 noncovalent complex (PDB: 2I00, left) and NEDP1–NEDD8 covalent complex (PDB: 2BKR, right) illustrating a common Ubi binding site.

(D) Schematic of the CE fold based on the NEDP1 structure (PDB: 2BKR), highlighting the constant region (CR) and variable regions (VR) that form the Ub/Ubi-binding S1 site (blue).

See also Table 1 and Figure S2.
<table>
<thead>
<tr>
<th>Data Collection and Refinement Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SseL 24-340</strong></td>
</tr>
<tr>
<td><strong>Data Collection</strong></td>
</tr>
<tr>
<td><strong>Beamline</strong></td>
</tr>
<tr>
<td><strong>Space group</strong></td>
</tr>
<tr>
<td><strong>Cell dimensions</strong></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
</tr>
<tr>
<td>a, β, γ (°)</td>
</tr>
<tr>
<td><strong>Wavelength (Å)</strong></td>
</tr>
<tr>
<td><strong>Observed reflections</strong></td>
</tr>
<tr>
<td><strong>Unique reflections</strong></td>
</tr>
<tr>
<td><strong>Resolution (Å)</strong></td>
</tr>
<tr>
<td><strong>Rmerge</strong></td>
</tr>
<tr>
<td><strong>I/sI</strong></td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
</tr>
<tr>
<td><strong>Phasing</strong></td>
</tr>
<tr>
<td><strong>Method</strong></td>
</tr>
<tr>
<td><strong>Resolution</strong></td>
</tr>
<tr>
<td><strong>Anom completeness</strong></td>
</tr>
<tr>
<td><strong>Anom multiplicity</strong></td>
</tr>
<tr>
<td><strong>&lt;FOM&gt;</strong></td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
</tr>
<tr>
<td><strong>Reflections in test set</strong></td>
</tr>
<tr>
<td><strong>Number of atoms</strong></td>
</tr>
<tr>
<td>Protein</td>
</tr>
<tr>
<td>Ligand/ion</td>
</tr>
<tr>
<td>Water</td>
</tr>
<tr>
<td><strong>B factors</strong></td>
</tr>
<tr>
<td>Wilson B</td>
</tr>
<tr>
<td>Protein</td>
</tr>
<tr>
<td>Ligand/ion</td>
</tr>
<tr>
<td>Water</td>
</tr>
<tr>
<td><strong>Rmsd</strong></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
</tr>
<tr>
<td>Bond angles (°)</td>
</tr>
<tr>
<td>Ramachandran statistics (favored/allowed/outliers)</td>
</tr>
</tbody>
</table>

Values in parentheses are for the highest resolution shell. Anom, anomalous; rmsd, root-mean-square deviations; FOM, figure of merit. See also Figures S2 and S3.
The VR-1 interface to the Ub Ile44 patch was highly unusual and unprecedented among published ubiquitin-binding domains, because such a low-complexity Ub binding region (LC-UBR) had not been previously described. Verifying the interaction in solution, NMR titration experiments following 15N-labeled Ub revealed a number of select resonances that displayed small chemical shift perturbations and/or line broadening with increasing concentrations of the XopD +VR-1 construct (Figure S3G), all of which depended on the VR-1 region (Figure S3H). When mapped to the XopD–Ub crystal structure, the changes observed in the NMR titration experiment nicely agree with contacts made between XopD VR-1 and the Ub Ile44 patch (Figure S3I).

In contrast, the bulk of contacts made in the tSUMO-bound structure are through VR-2, which forms multiple ionic interactions (XopD Arg385 and Asp368 contacting tSUMO Glu86/Asp88 and Arg69, respectively) surrounding a hydrophobic interface (XopD Met374 contacting tSUMO Met90) (Figure 3I). The tSUMO A62R or R69E mutation both resulted in a decrease in tSUMO suicide probe reactivity (Figure 3J). Surprisingly, a M90A mutation had much smaller effects (Figure 3J), suggesting that the small hydrophobic interface may play a lesser role in tSUMO recognition than does the extensive charge complementarity. Although highly divergent in sequence, the Ub and tSUMO C termini threaded into the active site similarly, with no apparent tertiary. Although highly divergent in sequence, the Ub and tSUMO recognize than does the extensive charge complementarity. The importance of variable regions in Ub recognition extends to each crystallized DUB (Figures 2E–2G) and also to the Legionella effector SdeA, which uses yet another variation of VR-1 to recognize Ub (Sheedlo et al., 2015) (Figures S4A and S4B). Mutations in the S1 site of SseL confirmed the Ub binding mode predicted based on analogous Ub/Ubl-bound structures (Figure 4A). A conserved hydrophobic (Trp or Phe) at the start of the core CE fold is present in eukaryotic CE ULPs and corresponds to Met159 in SseL. SseL M159A mutation significantly reduced K63 diUb hydrolysis (Figures 4A and 4B). Likewise, solvent-exposed hydrophobic residues within SseL VR-2 (Tyr183 and Ile196) abolished or reduced DUB activity, respectively.

Understanding Ub/Ubl Specificity in CE Clan DUBs
The importance of variable regions in Ub recognition extends to each crystallized DUB (Figures 2E–2G) and also to the Legionella effector SdeA, which uses yet another variation of VR-1 to recognize Ub (Sheedlo et al., 2015) (Figures S4A and S4B). Mutations in the S1 site of SseL confirmed the Ub binding mode predicted based on analogous Ub/Ubl-bound structures (Figure 4A). A conserved hydrophobic (Trp or Phe) at the start of the core CE fold is present in eukaryotic CE ULPs and corresponds to Met159 in SseL. SseL M159A mutation significantly reduced K63 diUb hydrolysis (Figures 4A and 4B). Likewise, solvent-exposed hydrophobic residues within SseL VR-2 (Tyr183 and Ile196) abolished or reduced DUB activity, respectively.

Figure 3. Molecular Analysis of XopD Ub/Ubl Specificity
(A) Topology diagram, as in Figure 2E, for the XopD crystal structure (PDB: 2OIV). The crystallized construct lacks a potential VR-1 region.
(B) XopD constructs with VR-1 (+VR-1: 298–515) or without VR-1 (ΔVR-1: 335–515) were tested against a panel of Ub/Ubl suicide probes for 1 hr at room temperature (propargylamine-derived probes). Open arrow, unmodified; closed arrow, modified.
(C) FP assays as in Figure 1D, including a substrate derived from S. lycopersicum SUMO (tS, tomato SUMO). Left: Ub/Ubl specificity of XopD +VR-1. Right: comparison of XopD +VR-1 cleaving the Ub and tSUMO substrates (shown in black) with the ΔVR-1 construct.
(D) Linkage specificity analysis as in Figure 1F for XopD +VR-1.
(E) Cartoon representation of the 2.9 Å XopD–Ub covalent complex crystal structure.
(F) Cartoon representation of the 2.1 Å XopD–tSUMO covalent complex crystal structure.
(G) Overlay of structures shown in (E) and (F), illustrating differences in Ub/tSUMO conformation and usage of XopD variable regions. For clarity, VR-1 is not shown.
(H) Close-up of the XopD–Ub interaction. In the foreground, VR-1 coordinates Ub Arg72 and the Ile44 hydrophobic patch. In the background, VR-2 contacts the Ub Leu73 loop.
(I) Close-up of the XopD–tSUMO interaction. VR-2 forms the basis of tSUMO binding and makes a number of polar and salt-bridge interactions surrounding the hydrophobic contact of tSUMO Met90.
(J) Series of Ub and tSUMO suicide probe reactions performed on ice, testing interactions observed in the Ub/tSUMO domain, as well as in their C termini (propargylamine-derived probes). Open arrow, unmodified; closed arrow, modified.
(K) As in (J), for XopD mutations encompassing VR-1 and VR-2. Open arrow, unmodified; closed arrow, modified. Asterisk, contamination in tSUMO suicide probe.
(L) FP-based cleavage assays testing the XopD variants used in (K) against the Ub and tSUMO KG-modified substrates. Wild-type XopD data from (C) are shown for reference.
(M) S. lycopersicum protein extract blotted for total Ub (Ubi-1; Novus Biologicals) following a 1 hr room temperature treatment with the XopD variants used in (K) at 1 μM final concentration.
(N) Schematic summarizing the activities displayed by wild-type and VR-1/VR-2 mutant XopD variants. See also Table 1 and Figure S3.
Interestingly, deletion of the putative substrate-binding VR-1 region alone (SseL \(\text{D}_{138-147}\); SseL \(\text{D}_{\text{N}}\): aa 158–340) or removal of the N-terminal domain including VR-1 (SseL \(\text{D}_{\text{N}}\)) abolished diUb cleavage, suggesting that the disordered sequence indeed contributes to substrate binding (Figures 2E, 4A, and 4B).

The importance of unique variable regions in ChlaDUB1 and RickCE was also tested using Ub/Ubl suicide probes as a
coordinated in the XopD

tution at position 72 in the Ub/Ubl C terminus. Arg72 of Ub is well-

A key difference between Ub and NEDD8 is an Arg-to-Ala substi-

Ub and NEDD8 are highly similar (58% identical, 85% similar), yet

Linkage

Changing Specificities for Ub/Ubl and polyUb Chain

Ub and NEDD8 are highly similar (58% identical, 85% similar), yet

most CE-clan proteases, including SseL and XopD, distinguish

between them with high specificity (Figures 1E, S1C, and S3C).

A key difference between Ub and NEDD8 is an Arg-to-Ala substitu-
tion at position 72 in the Ub/Ubl C terminus. Arg72 of Ub is well-
coordinated in the XopD–Ub structure (Figure 3H), as well as in other
Ub-bound DUB structures (Ye et al., 2011). Analysis of a modeled
SseL–Ub complex highlighted two acidic residues, Asp163 and Glu164, that may also coordinate Ub Arg72 (Figure 4D). Importantly, a
NEDD8 suicide probe with an A72R mutation (Ye et al., 2011) covalently modified SseL, as efficiently as the wild-type Ub probe (Figure 4E). Likewise, incorporation of the
A72R mutation imparts reactivity of the NEDD8 suicide probe to-
ward XopD (Figure S4C). Thus, in addition to large-scale changes in the variable regions of the S1 site, proper coordination of the Ub/
Ubl C terminus contributes an added level of substrate specificity.

Unlike XopD (Figure 3D), CE effectors from human pathogens strongly prefer K63-linked chains (Figure 1F). This suggests there must be a linkage specificity-imposing S1’ Ub binding site that participates in Ub recognition (Mevissen et al., 2013), though such an S1’ site in CE DUBs has not been annotated. Examination of the SseL, ChlaDUB1, and RickCE structures re-

vealed a conserved hydrophobic region near the catalytic Cys (Figures 4F and S4D), and mutating residues in this patch on SseL simultaneously reduced activity toward K63-linked diUb, while improving hydrolysis of K48-linked diUb (Figure 4G). This was quantified by a fluorescence polarization-based diUb cleavage assay (Keusekotten et al., 2013). Based on the Michaelis-Menten kinetic parameters, wild-type SseL displays an approximately 85-fold preference for K63 chains over K48; this is predominantly the result of a difference in $k_{cat}$ (Figure 4H).

The Y244A mutation causes an approximately 30-fold reduction in $k_{cat}/K_m$ toward K63-linked chains, while simultaneously increasing $k_{cat}/K_m$ toward K48-linked chains 10-fold. Hence, SseL Y244A displays a ~1.5-fold preference for K48 over K63 diUb (Figure 4H); this preference is also confirmed when tested against a panel of all eight Ub chain types (Figure 4I).

Together, mutagenesis work establishes the location of the S1’ Ub binding site in CE clan DUBs, and the ability to modulate Ub/Ubl, as well as linkage specificity with point mutations, high-

lights the versatility and adaptability of the CE protease fold.

Bioinformatic Analysis of the CE Clan Highlights Distinct Functional Families

Using our new CE protease structures as a guide, we expanded and refined the sequence alignment of CE clan members in
eukaryotes, viruses, and bacteria (Figure S5). In addition to the
active site, we identified several conserved positions important for domain structure and substrate recognition, including the semi-conserved Trp at the start of the CE fold and a highly conserved Trp following the catalytic His.

Rewardingly, reconstruction of the dendrogram for the entire CE clan shows an interesting segregation of families that agrees with the demonstrated functional properties for these enzymes (Figure 5). In agreement with their dual Ub/SUMO protease activity, the XopD-like family, including XopD and examples from other plant-associated bacteria, are most closely related to the eukaryotic ULP families. As dedicated DUBs, the SseL-
lke examples (including ElaD and ShiCE) also segregate near the
XopD/ULP subfamily. The Ub/NEDD8 cross-reactive en-
zymes ChlaDUB1 and RickCE fall into their own clade. Interest-
ingly, the YopJ family, that also includes Legionella LegCE, lacks, e.g., conserved Trp residues (Figure S5) and is most diver-
ent, consistent with their roles as dedicated acetyltransferases (Figures 1E and S1H).

CE fold proteins from viruses constitute three additional groups. The first group (viral group I) includes examples from Ad-
enviridae and clusters near the Ub/Ubl proteases. This group includes the adenovirus L3 23K proteinase, which, in addition to general protease activity, also possesses DUB activity (Balakev et al., 2002).

Viral groups II and III are more divergent and also include bacterial proteins (Figure 5). The Legionella effector SidE (Figures S4A and S4B) was independently identified by our sequence analysis as a part of viral group II and has since been shown to display mixed activities toward Ub, NEDD8, and, interestingly, ISG15 (Sheedlo et al., 2015). Viral group III, which contains Vaccinia virus protein I7L, has been shown to target an Ala-

Gly-Xaa motif instead of the Gly-Gly-Xaa motif of Ub/Ubl modi-

fiers, and hence, members within the viral group III might be more general proteases dedicated to cleavage events necessary for viral maturation (Byrd et al., 2003). Therefore, our structural and functional analysis of divergent CE clan members has

enabled the refinement of relationships across kingdoms, has

enabled identification of further examples in other species, and

allows a distinction between dedicated DUBs, dedicated

ULPs, and specialized acetyltransferases.

CE Clan Proteases Are Further Functionalized with Accessory Domains

In addition to the catalytic fold, bacterial CE clan effectors contain additional N- or C-terminal domains of high diversity and unknown functions. These range from predicted transmembrane regions, protein-protein interaction or regulatory domains, and, even, additional enzymatic folds (Figures 1A and 6A) and likely represent an additional layer of regulation, similar to eu-

karyotic ULP enzymes (Ronau et al., 2016).

The crystal structure of full-length SseL revealed an unannotated N-terminal domain, in which eight helices form a superhelical structure (Figure 6A). Comparison with known structures using the Dali server (Holm and Rosenerstrom, 2010) indicated similarity with the VHS (VPS-27, Hrs, and STAM) domains of
GGA1 (PDB: 1PY1, [He et al., 2003], Dalil Z score of 5.0, RMSD
4.0 Å), and STAM1 (PDB: 3LDZ [Ren and Hurley, 2010], Dali
Z score of 4.2, RMSD 3.8 Å (Figure 6B). In eukaryotes, VHS domains are involved in vesicular trafficking, either through direct interaction with a phosphorylated receptor (e.g., GGA1 [Shiba et al., 2002]) or by binding Ub attached to cargo (e.g., STAM1 [Ren and Hurley, 2010]).

VHS domains, so far, are unknown to occur in bacteria, and the functional peptide- and Ub-binding interfaces differ in the SseL VHS domain. SseL did not interact with phosphorylated or unphosphorylated mannose-6-phosphate receptor (MPR) peptides (Figure S6B). In contrast, NMR titration experiments of the isolated SseL VHS domain revealed low micromolar interactions with the Ile36 and Ile44 hydrophobic patches of Ub (Figures 6C–6E and S6C). Identical Ub binding was also observed with an SseL construct including the CE fold (aa 24–340) (Figure S6D), and this was unaffected by mutations in the S1 Ub binding site or when a covalent SseL–Ub complex was used in the titration (Figures S6E and S6F), suggesting that the VHS domain constitutes the highest affinity Ub binding site.

Inspection of the SseL VHS structure identified two solvent-exposed hydrophobic patches centered on Leu46 and Trp105 as potential Ub-interacting surfaces (Figures 6F and S6G). Ub binding was unaffected with a SseL L46R mutant (Figure S6H), but abrogated in a SseL W105A mutant (Figure 6G). This Ub binding site on SseL is distinct from that on the STAM1 VHS domain and remote from the catalytic center (Figure S6I). The SseL W105A mutant had no effect on Ub chain hydrolysis for K63-linked di-, tri-, or tetra-Ub (Figures 6H and S6J).

**Figure 5. Bioinformatic Analysis of the CE Protease Clan**
Structure-informed dendrogram for the CE clan, with representative members from eukaryotes, bacteria, and viruses. Clustered families are highlighted and labeled according to the archetypal examples. Eukaryotic family members are labeled with numbers corresponding to their SENP nomenclature. Characterized members are highlighted in bold. See also Figure S5.
suggested that the added Ub binding site outside of the SseL catalytic domain does not contribute to the recognition or hydrolysis of longer Ub chains, but may instead serve alternative roles in regulating SseL during Salmonella infection.

**The SseL VHS Domain Dictates Subcellular Localization of the Effector**
Following translocation from the Salmonella-containing vacuole (SCV), SseL robustly localizes to the outer leaflet of the vacuolar membrane, as well as to Salmonella-induced filaments (Sifs) that emanate from the SCV (Rykönen et al., 2007). To determine if this localization is determined by the Ub-binding activity of the VHS domain, we reintroduced HA-tagged versions of wild-type (SseL-2HA WT), S1 site mutant (SseL-2HA Y183A), and VHS mutant (SseL-2HA W105A) into the *Salmonella* Typhimurium SseL strain to be expressed from a plasmid under the native promoter. HeLa cells were infected with the complemented strains for 16 hr, then fixed, and immune-labeled for

---

**Figure 6. CE Catalytic Domains Are Fitted with Accessory Domains**

(A) SseL crystal structure (active-site Cys in a ball-and-stick representation), focusing on the superhelical N-terminal VHS domain.

(B) Superposition of the SseL N-terminal domain (cyan) with the VHS domains of STAM1 (PDB: 3LDZ, dark blue), and GGA1 (PDB: 1PY1, orange).

(C) $^1$H,$^15$N-HSQC TROSY spectra of 80 μM $^15$N-labeled Ub alone (black) and in the presence of 0.25 (20 μM, red) or 0.5 (40 μM, blue) molar equivalents of the SseL VHS domain (see the Supplemental Experimental Procedures). Numbers refer to the assigned Ub residues.

(D) Calculated ratio in peak intensity observed between $^15$N-labeled Ub alone and in the presence of 0.5 molar equivalent SseL VHS domain. A red dashed line marks the level of significance used in (E). Gray bars indicate prolines or missing resonances.

(E) Surface representation of the Ub structure (PDB: 1UBQ) with red painted surface corresponding to regions significantly affected by SseL VHS binding.

(F) Exposed hydrophobic patch within the SseL VHS domain: W105 was chosen for mutation.

(G) $^1$H,$^15$N-HSQC TROSY spectra of 80 μM $^15$N-labeled Ub alone (black) and in the presence of 0.5 molar equivalents of SseL (24–340) W105A (40 μM, orange).

(H) Time course assays monitoring cleavage of K63- and K48-linked tetraUb chains with the SseL VHS domain Ub-binding mutant.
See also Figure S6.
SseL-2HA and CSA-1 (a marker for bacteria). Wild-type SseL localized to SCVs and Sifs in a SPI-2 type 3 secretion system (T3SS)-dependent manner (compared to ∆ssaV T3SS-deficient control), as expected (Figure 7A). The localization of S1 site mutant Y183A was indistinguishable from wild-type, indicating that Ub binding to the catalytic domain does not dictate subcellular localization (Figures 7A–7C and S7A). The W105A VHS mutant, however, showed a severe localization defect and was distributed diffusely throughout the host cytoplasm (Figures 7A–7C and S7A). Therefore, the N-terminal VHS domain of SseL serves to localize its catalytic DUB function through an additional Ub binding site (Figure 7D). Both ElaD and ShiCE encode similar N-terminal VHS domains and the Ub binding site is conserved (Figures 1A and S1A). Hence, we would predict that this subfamily of effectors has adopted analogous strategies of regulation.

DISCUSSION

During infection, bacteria secrete effector proteins, such as CE clan proteins, which inactivate inflammatory signaling cascades and promote bacterial growth. CE clan effectors were reported to be deSUMOylases, deubiquitinases, and, even, acetyltransferases, contrasting the specific and important roles of these enzymes in removal of a small subset of Ub-like modifications (SUMO and NEDD8) in eukaryotes. This ambiguity triggered the here presented comprehensive characterization of bacterial CE effectors, revealing dedicated DUBs and dedicated acetyltransferases, but also mixed Ub/Ubl proteases (Ub/NEDD8 and strikingly also Ub/SUMO). All this is performed by a single, highly adaptable catalytic fold. Complementary bioinformatic analysis based on structural data expands the family and reveals some of the functional
differences. For example, we have annotated and characterized the first DUB of Shigella, an activity that has been inconspicuously lacking considering its large repertoire of effector E3 ligases (Ashida et al., 2014). However, the functional variation observed among CE clan members is staggering, making some enzymatic features unpredictable, especially when mixed activities are suspected.

DUB activity in CE effectors has been suggested to exist in various bacteria, and while we now understand strategies for adaptation of Ub specificity in molecular detail, it is surprising that bacteria have adopted such distinct structural approaches to generate a Ub-specific S1 site. This suggests the convergent evolution of distinct bacterial DUBs. Based on our structural work, we identify three common regions of variability that, together, encode Ub/Ubl substrate specificity. In our minds, the most surprising insight is the mixed activity of XopD being (tomato–) SUMO and Ub isopeptidase, since these modifiers are quite distinct and the capacity to efficiently recognize both substrates suggests a large evolutionary pressure to become a multifunctional protease. Interestingly, our structural analysis of XopD bound to either tSUMO or Ub demonstrates that XopD’s flexibility in substrate recognition primarily arises from incorporation of an intrinsically disordered VR-1, which can be tailored to suit either binding event.

Moreover, we now understand the preference of most CE DUBs for K63-linked chains because of the presence of a defined, linkage-specificity-generating S1’ site. This site on CE proteases was unanticipated and is unknown for the CE ULPs; whether the six SENP members in humans preferentially target distinct SUMO chains is unclear. Our mutagenesis data further show that the observed K63 specificity hinges on several key contacts, and changes in these can switch specificity. Hence, we would anticipate that CE DUBs could target other chain types, if advantageous for infection (see XopD; Figure 3D).

Finally, the CE folds of most effectors are accessorized with additional domains, and we have shown that a Ub-binding VHS domain in SseL (and by sequence similarity in ElaD and ShiCE) mediates localization of the effector. The VHS domains of ESCRT-0 components STAM and Hrs bind ubiquitinated cargo to direct the lysosomal maturation pathway (Ren and Hurley, 2010). Similarly, the VHS domain of SseL binds Ub (albeit via a distinct surface) and functions in localizing it to Sifs and to the cytosolic face of the SCV, where the catalytic CE domain can remove K63-linked polyUb. This could block recruitment of xenophagy adaptors, such as p62 (Mesquita et al., 2012), through general removal of K63-linked chains, or directly target specific host proteins, such as oxysterol-binding protein (OSBP) (Auweter et al., 2012). Furthermore, given that many of these accessory domains are unrelated across CE clan members (Figures 1A and S6A), the breadth of possibilities through which CE enzymatic activities may be regulated is likely extensive. Our study serves as a framework to further investigate this intriguing class of effectors and, in particular, highlights several interesting mechanisms by which bacterial effectors encode Ub/Ubl specificity.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification of the CE Effectors
Protein sequences were either cloned from bacterial or synthesized DNA or were received as gifts from colleagues. Constructs were generated as outlined in Figure 1A in the pOPIN-B or pOPIN-K vectors and purified using affinity, anion exchange, and size exclusion chromatography.

DUB Activity Assays
Qualitative DUB assays, suicide probe assays, TAMRA-based Ub/Ubl FP assays, and FASt-based dUb cleavage assays are described in detail in the Supplemental Experimental Procedures.

Protein Crystallization and Structure Determination
Crystallization conditions were optimized based on initial hits obtained from commercial screens in sitting-drop format. Structure determination was performed using either single wavelength anomalous dispersion (SAD) experimental phasing of selenomethionine derivatives or molecular replacement (see Table 1).

Construction of CE Dendrogram
Novel and existing CE structures from bacteria, eukaryotes, and viruses were used to create a “seed” structure-based sequence alignment, which was iteratively refined and expanded on with additional examples. This alignment was used to identify more divergent examples of the CE fold and to construct a dendrogram of the CE clan using the UPGMA method.

NMR Spectroscopy
NMR titration spectra of uniformly labeled 15N-Ub were recorded at 25°C on either a Bruker Avance III 600 MHz or an Avance2+ 700 MHz spectrometer, equipped with cryogenic triple-resonance TCI probes. Data processing and analysis were performed in Topspin (Bruker) and NMRView (One Moon Scientific).

S. Typhimurium Infection Assays
Infection assays were performed as described previously (Rytkönen et al., 2007), with changes as discussed in the Supplemental Experimental Procedures.

ACCESSION NUMBERS
The accession numbers for the coordinates and structure factors for SseL, ChiaDUB1, RickGE, XopD–Ub, and XopD–tSUMO reported in this paper have been uploaded to PDB: SHAF, SHAG, SHAM, SJP3, and SJP1.

SUPPLEMENTAL INFORMATION
Supplemental Information includes Supplemental Experimental Procedures and seven figures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2016.06.015.

AUTHOR CONTRIBUTIONS
D.K. conceived the project, and together, J.N.P. and D.K. designed the project. J.N.P. performed all of the biochemical and structural experiments, with input from D.K. B.S. performed the bioinformatics analysis. C.H.D. performed the Salmonella infection experiments, with input from D.W.H. P.P.G. and H.O. provided Ub/Ubl reagents. J.N.P. and D.K. wrote the manuscript, with input from all authors.

ACKNOWLEDGMENTS
We would like to thank R. Mittal (MRC Laboratory of Molecular Biology), R. Hay (University of Dundee), B. Schulman (St. Jude Children’s Research Hospital), and A. Canto-Pastor and D. Baulcombe (University of Cambridge) for providing reagents, and the beamline staff at ESRF ID23-1 and DLS I03, I04,
and I04-1 for their assistance. Access to DLS was supported, in part, by the EU FP7 infrastructure grant BIOSTRUCT-X (contract no. 283570). The graphical abstract uses images adapted from Servier Medical Art. We also thank members of the D.K. lab for reagents and advice and K. Swatek for critical comments on the manuscript. Work in the D.K. lab was funded by the Medical Research Council (U105192732), the European Research Council (309756), and the Lister Institute for Preventive Medicine. J.N.P. was supported by an EMBO Long-term Fellowship. Work in the H.O. lab was supported by a Netherlands Organization for Scientific Research VICI grant (724.013.002) and the European Research Council (281699), C.H.D. and D.W.H. were funded by a Medical Research Council grant (WMNN_P45464) (to D.W.H.). B.S. was funded by the Medical Research Council (U105158559), D.K. and H.O. are part of the DUB Alliance that includes Cancer Research Technology and FORMA Therapeutics. H.O. is founder and shareholder of the company UbiQ that markets reagents in the ubiquitin field.

Received: April 7, 2016
Revised: May 2, 2016
Accepted: June 8, 2016
Published: July 14, 2016

REFERENCES


Molecular Cell 63, 261–276, July 21, 2016 275


Supplemental Information

The Molecular Basis for Ubiquitin and Ubiquitin-like Specificities in Bacterial Effector Proteases

Jonathan N. Pruneda, Charlotte H. Durkin, Paul P. Geurink, Huib Ovaa, Balaji Santhanam, David W. Holden, and David Komander
Figure S1: CE effector proteins demonstrate mixed proteolytic activities

A

B

C

D

E

F

G

H

I

J

K

L

M

SENP2 541 LGLNGD - CGMFTCKYAD
NEDP1 155 ACGNVGDY - CSVMVICNIE
ShiCE 277 LSNVPNGSLGCFCYHFLQ
ChlaDub1 337 IGRGSSEQCGWCGQDEI
Elad 309 LLQYLQSGQAFVCAAQ
ShiCE 307 LLQYLQSGQAFVCAAQ
RickCE 621 OTDOQAGFCQTVGONGI
LegCE 256 LHSRAGCATETDCAS
YopJ 165 DRSSECGIFSLA
AvrA 165 DRSSECGIFSLA

SENP2 541 LGLNGD - CGMFTCKYAD
NEDP1 155 ACGNVGDY - CSVMVICNIE
ShiCE 277 LSNVPNGSLGCFCYHFLQ
ChlaDub1 337 IGRGSSEQCGWCGQDEI
Elad 309 LLQYLQSGQAFVCAAQ
ShiCE 307 LLQYLQSGQAFVCAAQ
RickCE 621 OTDOQAGFCQTVGONGI
LegCE 256 LHSRAGCATETDCAS
YopJ 165 DRSSECGIFSLA
AvrA 165 DRSSECGIFSLA

**C**

**D**

**E**

**F**

**G**

**H**

**I**

**J**

**K**

**L**

**M**
Figure S1: CE effector proteins demonstrate mixed proteolytic activities; related to Figure 1
A) Sequence alignment for the N-terminal accessory domains of SseL, ElaD, and ShiCE. Agreement in secondary structure prediction calculated for each individual sequence is shown above. B) Sequence alignment used to identify the putative catalytic Cys in bacterial CE enzymes (marked by an asterisk) by similarity to eukaryotic CE examples. C) Normalized FP measured as a function of time, as in Figure 1E, following addition of the well-studied SUMO-specific human SENP1 and NEDD8-specific human NEDP1. D) As in C for YopJ and AvrA in the presence of excess activator IP6. E) Suicide probe reaction, as in Figure 1C, monitoring reactivity of untagged YopJ and AvrA toward the indicated Ub/Ubl probes following 1 h incubation at room temperature (propargylamine-derived probes). F) As in C for untagged YopJ. G) $^1$H,$^15$N-HSQC TROSY spectra of 80µM $^15$N-labeled Ub alone (black) and in the presence of 1 molar equivalent of AvrA (80µM, red). Inset, enlarged region of the spectral overlay with Ub resonance assignments labeled. H) Autoradiography monitoring isotope incorporation following a 2 h incubation of each protein with $^{14}$C-labeled Acetyl-CoA in the presence of IP6. Compare signal in wild-type over inactive mutant background (see Supplemental Experimental Procedures). I) Time course monitoring cleavage of pro-NEDD8 (81 aa) to the mature form (76 aa), for the dual-specific ChlaDUB1, NEDD8-specific NEDP1, and Ub-specific SseL. This demonstrates that, like other CE proteases, ChlaDUB1 possesses both peptidase and isopeptidase activities. J) Time course monitoring cleavage of NEDDylated Cul1 C-terminal domain (CTD, modified at Lys720) in the presence of ChlaDUB1, and NEDP1 or COP9 signalosome (CSN) as negative and positive controls, respectively. Although ChlaDUB1 and NEDP1 can remove some NEDD8 modifications (e.g., CTD-(NEDD8)2), only the specialized CSN protease can remove the regulatory NEDD8 modification from Cul1 (CTD-NEDD8). K) Linkage specificity analysis for all diUb substrates, as in Figure 1F, for SseL and ChlaDUB1 at 10-fold higher enzyme concentration. L) Linkage specificity analysis for all diUb substrates, as in Figure 1F, for LegCE, YopJ, and AvrA. M) As in L for YopJ and AvrA in the presence of activator IP6.
Figure S2: Structural analysis of bacterial CE deubiquitinases

A. SseL (24-340) 2.7Å

B. ChlaDUB1 (130-401) 2.1Å

C. RickCE (420-691) 2.4Å

D. Active Site

E. WT, C285A, H223A, N242A, Q278A min

250nM SseL

Ub

K63 dUb -Ub
Figure S2: Structural analysis of bacterial CE deubiquitinases; related to Figure 2, Table 1
A) Full asymmetric unit (ASU) of the 2.7Å SseL (24-340) crystal structure showing a 2|Fo|-[Fc] electron density map contoured at 1σ, and superposition of the two copies in the ASU. Crystallization of this construct depended upon an N-terminal His-tag, which makes crystal contacts in one molecule and forms a short helix against the N-terminal domain in another. Most likely resulting from these crystal contacts at the N-terminus, the two molecules in the ASU show deviation within the N-terminal domain (1.99Å Cα RMSD for residues 24-136), but overlay well in the catalytic domain (0.39Å Cα RMSD for residues 154-340). SseL is colored in shades of teal, ordered regions of the His-tag are colored in grey. B) Full ASU of the 2.1Å ChlaDUB1 (130-401) structure showing 2|Fo|-[Fc] electron density contoured at 1σ. The first ~20 residues form a highly extended structure via crystal contacts. C) Full ASU of the 2.4Å RickCE (420-691) crystal structure, along with a 2|Fo|-[Fc] electron density map contoured at 1σ, and an alignment of the two copies in the ASU. The two identical molecules in the ASU (0.27Å Cα RMSD) are interlaced by a large insertion forming a helical arm. D) Active site Cys, His, Asn, and Gln of SseL (teal) are shown. E) K63 diUb cleavage assay using wild-type and catalytic mutants of SseL.
Figure S3: Molecular analysis of XopD Ub/Ubl specificity

A. 

B. XopD–Ub 2.9Å

C. XopD–tSUMO 2.1Å

D. 

E. 

F. 

G. 

H. 

I. 

J. 

K. 

L. 

M. 

N. 

O. 

P. 

Q. 

R. 

S. 

T. 

U. 

V. 

W. 

X. 

Y. 

Z. 

Probe: 

Ub- 

XopD 

tSUMO 

XopD

D429

H409

C470

H409

C470

XopD (mol2)

XopD (mol1)

XopD (mol3)

XopD (mol4)

Ub

tSUMO

Ub

tSUMO

Ub: VLRLRGG

tSUMO: MLHQTGG

L73 / Q93

R72 / H92

R74 / T94

Active Site

VR-1

VR-2

VR-1

VR-2

Active Site

Active Site

Active Site

Active Site

Active Site

Active Site

Active Site

Active Site

Active Site

Active Site

Active Site

Active Site
**Figure S3: Molecular analysis of XopD Ub/Ubl specificity**; related to Figure 3, Table 1

**A)** Catalytically inactive XopD + VR-1 C470A tested against the Ub and tSUMO suicide probes following a 1 h incubation at room temperature (propargylamine-derived probes). Open arrow, unmodified. **B)** Close-up of the XopD–Ub covalent linkage, showing XopD catalytic triad residues and 2/|Fo|−|Fc| electron density of the Ub C-terminus, contoured at 1σ. **C)** As in **B**, for the XopD–tSUMO complex. **D)** Superposition of the four XopD–Ub molecules present in the ASU. **E)** Superposition of XopD from its apo (PDB 2OIV), Ub-bound, and tSUMO-bound structures. VR-1 is not shown for clarity. **F)** Close-up of the tract leading into the XopD active site. C-terminal residues for Ub and tSUMO are shown above, and their positions within the XopD-bound structures is depicted below. **G)** $^1$H,$^{15}$N-HSQC TROSY spectra of 80µM $^{15}$N-labeled Ub alone (black) and in the presence of 0.5 (40µM, red), 1 (80µM, blue), or 1.8 (144µM, green) molar equivalents of XopD + VR-1 (298-515). **H)** As in **G**, for 1.8 (144µM, magenta) molar equivalents of XopD ΔVR-1 (335-515). **I)** Resonances significantly perturbed in **G** by either line broadening or chemical shift perturbation mapped onto the XopD–Ub crystal structure (red). Prolines and other resonances missing from the NMR spectrum are shown in black.
Figure S4: Structure-based manipulation of CE DUB activities and specificities

A. SdeA-Ub (5CRA)

B. SdeA (5CRA)

C. XopD (298-515)

D. ChiaDUB1 and RickCE
**Figure S4: Structure-based manipulation of CE DUB activities and specificities; related to Figure 4**

A) Cartoon representation of SidE family member SdeA in a covalent complex with Ub (PDB 5CRA). B) Topology diagram, as in **Figure 2E**, for the SdeA crystal structure (PDB 5CRA). C) Suicide probe assay showing XopD +VR-1 (298-335) reaction with Ub, NEDD8, and NEDD8 A72R probes at room temperature for 1 h (chloroethylamine-derived probes) Open arrow, unmodified; closed arrow, modified. D) Close-up of the conserved hydrophobic S1’ sites of ChlaDUB1 and RickCE.
Figure S5: Bioinformatic analysis of the CE protease clan

<table>
<thead>
<tr>
<th>Organism</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>X. campestris XopD</td>
<td>HWSLLVVVRDR AYHDSDMAQ ALSDGYS - - CGDVGVETG</td>
</tr>
<tr>
<td>A. citrulli</td>
<td>HWSLLVVDQ AFHYSDLSVS ARGDDGHS - - CDGHVLQA</td>
</tr>
<tr>
<td>R. etli</td>
<td>HWSLLFVODL AYHYDSLGT AQGDNYSY - - CGGVVVDATA</td>
</tr>
<tr>
<td>R. loti</td>
<td>HWSLLLVRDR AYHYDSIQQ AQGKNADV - - CGGVVVDG</td>
</tr>
<tr>
<td>S. cryophilus</td>
<td>HWSMVVSCEF YFYSMSMN QQRNYG - - CGVHIAAF</td>
</tr>
<tr>
<td>H. sapiens 1</td>
<td>HWCLAVVDV IYYDSMSG QQMNQSD - - CQFMACKY</td>
</tr>
<tr>
<td>D. melanogaster 1</td>
<td>HWCMAIHHL IRYDSDKG PRLDGSQD - - CGIFSCMF</td>
</tr>
<tr>
<td>S. purpuratus 1</td>
<td>HWCLAVFDV IVYDSDMG TQLLNGSD - - CQFMACKY</td>
</tr>
<tr>
<td>C. gigas 1</td>
<td>HWCLAVFDV IRYFDSDMG PQQMNQSD - - CQFMACKY</td>
</tr>
<tr>
<td>H. robusta 1</td>
<td>HWCLAVFDV IRYFDSDMG PQQMNQSD - - CQFMACKY</td>
</tr>
<tr>
<td>R. loti</td>
<td>HWSLVVLVDL KYLDSMGQ PQQLNGSD - - CQFMACKY</td>
</tr>
<tr>
<td>H. sapiens 2</td>
<td>HWSLVVLVDL KYLDSRQG PQQLNGSD - - CQFMACKY</td>
</tr>
<tr>
<td>M. musculus 2</td>
<td>HWSLVVMDL KLYLDSMGQ PQQLNGSD - - CQFMACKY</td>
</tr>
<tr>
<td>D. melanogaster 8</td>
<td>HWSLLVFSR FYHDSYGN LQANGY - - CGIHVICM</td>
</tr>
<tr>
<td>H. sapiens 8</td>
<td>HWSLLYLYQ FFYSHSRS PAQNSYD - - CGMVYICN</td>
</tr>
<tr>
<td>L. pneumophila LegCE</td>
<td>HWSLI ELYQ ILYLSDGA KLDHRASRG - - CATFTMDA</td>
</tr>
<tr>
<td>L. longbeachae</td>
<td>HISFLEDF FILLCDPLGF VLGNCGRG - - CAYFTDDG</td>
</tr>
<tr>
<td>S. purpuratus 8</td>
<td>HWSLLVSRF PHYDSUGS TQENYCD - - GGLPVCICN</td>
</tr>
<tr>
<td>C. gigas 8</td>
<td>HWSLLYVIR FRHHSDRD PQGSNSD - - CGVFAIAI</td>
</tr>
<tr>
<td>HAdV 2 L3 23K</td>
<td>HMSKAFWNP CYLFPEPGF SVGNPSASA CGLFCMCF</td>
</tr>
<tr>
<td>HA/F</td>
<td>HYLALAWNP CYDLFPDFGTV GPFPSCASA CGLFCMCF</td>
</tr>
<tr>
<td>OAdV 1</td>
<td>HIIILALESP LFIFDPGLWSV QTCTCAGS CGLFCIFF</td>
</tr>
<tr>
<td>FrAdV 1</td>
<td>HWIIFppelin FFMPDFPGW AVCCCTOASA CGLFCGLF</td>
</tr>
<tr>
<td>ASPV</td>
<td>HWAIIFVDNM CWISIYEYNS RHORSDTE - - CGYPSLFY</td>
</tr>
<tr>
<td>CroV</td>
<td>HWAVOLINL YYFDFSFAK KIAGHRDNECG VYNSIF</td>
</tr>
<tr>
<td>APMV</td>
<td>HWAMFYRD YLYCDSNGK SYKCDGSE - - CGVYSCNF</td>
</tr>
<tr>
<td>Fowlpox virus</td>
<td>HWKCAIGD YCFYDSGGN VNLKLES - - CGMFTCLF</td>
</tr>
<tr>
<td>Myxoma virus</td>
<td>HWKCAIFDK VCFYDSGGN VNLKLES - - CMFISVF</td>
</tr>
<tr>
<td>MCV</td>
<td>HWCLFVFDR VAFYDSGGN VNLKLES - - CMFISSL</td>
</tr>
<tr>
<td>Orf virus</td>
<td>HWKCIIFDK VAFYDSGGN VNLKLES - - CMFISSL</td>
</tr>
<tr>
<td>YMTV</td>
<td>HWKCAIKNK VAFYDSGGN VNLKLES - - CMFISSL</td>
</tr>
<tr>
<td>Variola virus</td>
<td>HWKCVIYDK VSFYDSGGN VNLKLES - - CMFISSL</td>
</tr>
<tr>
<td>L. longbeachae</td>
<td>HWTIALN VGYTDSLNA WTDGSS - - CGPSLIN</td>
</tr>
<tr>
<td>A. asiaticus</td>
<td>HWGIL EVAYEDSNT LKODDDTS - - CGAITYEEN</td>
</tr>
<tr>
<td>C. trachomatis ChlaDUB1</td>
<td>HWLIVDLY FYDFLYSN VIGRSGSS - - CGAMCQF</td>
</tr>
<tr>
<td>S. negevensis</td>
<td>HWLFLDDYRE VYDYDKIN KLPGDYQ - - CGPVALYF</td>
</tr>
<tr>
<td>S. typhimurium SseL</td>
<td>HWSLLFYK CLIFNTYDD NLONVPNG - - CGLFYHT</td>
</tr>
<tr>
<td>E. coli</td>
<td>HWLLLVSLLQ CVIFSNLRA DLDQYLSQG - - CGAVCMCA</td>
</tr>
<tr>
<td>S. flexneri ShiCE</td>
<td>HWLHLQLSV CVIFSNLRA DLDQYLSQG - - CGAVCMCA</td>
</tr>
<tr>
<td>R. bellii RickCE</td>
<td>HWTLVAKY LTFTDSLGN KQTDOSA - - CGVFTVDN</td>
</tr>
<tr>
<td>R. bellii</td>
<td>HWATSIRKLYYNPDNGH HQSSTQS - - CGAFTAED</td>
</tr>
<tr>
<td>O. tsutsugamushi</td>
<td>HWAMAIKVK VSNTDEMPG VQTNVYDD - - CGPVPVDN</td>
</tr>
<tr>
<td>M. xanthus</td>
<td>HWACEVIEL VLFDFDPLG AVOGDHGS - - CGTMWAIC</td>
</tr>
<tr>
<td>Y. pestis YopJ</td>
<td>HFEVDDKIL LIFPEA - N DIDRSSSE - - CGFSFAL</td>
</tr>
<tr>
<td>S. typhimurium Avar</td>
<td>HI SVOFVR VLFPEPA - A DIDRSSSE - - CGFSFAL</td>
</tr>
<tr>
<td>B. claridgeiae</td>
<td>HSVDIHDQT LIFPEV - S DIDRSSSE - - CGFSFAL</td>
</tr>
<tr>
<td>L. massiliensis</td>
<td>HWTAMIDEV LGAKEFDK GIYDGDAS - - CSRFAIDI</td>
</tr>
<tr>
<td>R. gyll</td>
<td>HFSSVYDI LLLVIAAGL QICCCFEH - - GSFTLDH</td>
</tr>
<tr>
<td>V. paraeaeolaylucidus</td>
<td>HGCIAVDCAI LIGIPY VDMORSGE - - CLMSFL</td>
</tr>
<tr>
<td>L. longbeachae SidE</td>
<td>HWMLIKK SSLDPLGE PNLGAHLNLGLGYYWASA</td>
</tr>
<tr>
<td>L. pneumophila SidE</td>
<td>HWMLIKK SSLDPLGE PNLGAHLNLGLGYYWASA</td>
</tr>
<tr>
<td>L. hackelia</td>
<td>HWTLAI- AVTISFTDSLSS WRPDGS - - CGPYSLAN</td>
</tr>
</tbody>
</table>
**Figure S5: Bioinformatic analysis of the CE protease clan;** related to Figure 5
A) Structure-guided sequence alignment of representative members from the CE clan, used to construct the dendrogram shown in Figure 5. For clarity, only regions immediately surrounding the catalytic center are shown. Coloring is based on similarity. One conserved Trp that is typically lacking in dedicated acetyltransferases is marked with an arrow. Catalytic residues are marked below with an asterisk.
Figure S6: CE catalytic domains are fitted with accessory domains

A

B

C

D

E

F

G

H

I

J

<table>
<thead>
<tr>
<th>0</th>
<th>5</th>
<th>20</th>
<th>60</th>
<th>0</th>
<th>5</th>
<th>20</th>
<th>60</th>
<th>0</th>
<th>5</th>
<th>20</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>L46R</td>
<td>W105A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 250mM SseL | dUb | K83 | Ub |

| N (ppm) | 122.5 |

| 15N-Ub + SseL (24-340) |

| 15N-Ub + SseL (24-340) Y183A |

| 15N-Ub + SseL (24-340)~Ub |

| 15N-Ub + SseL (24-340) L46R |

| 15N-Ub + SseL (24-340) L46R |

| 15N-Ub + SseL (24-340)~Ub |

| STAM (STAM:Ub, 3LDZ) |

| Ub (STAM:Ub, 3LDZ) |

| STAM (STAM:Ub, 3LDZ) |

| SseL, (VHS domain) |

| SseL, (CE domain) |

| dUb |

| K83 |

| Ub |

| Active site |

| W105 |
Figure S6: CE catalytic domains are fitted with accessory domains; related to Figure 6
A) Domain annotation of select CE enzymes from bacteria. TM, transmembrane helix; PDZ, PDZ protein interaction domain; S41, S41 peptidase domain; TPR, tetratricopeptide repeat. B) GST pull-down assay testing interaction of SseL to the cation independent mannose-6-phosphate receptor (CI-MPR) C-terminal peptide, as well as its LL-AA double mutant form and CK2-phosphorylated form. C) Full spectrum for the NMR titration shown in Figure 6C. 80 µM 15N-labeled Ub alone (black) and in the presence of 0.25 (20µM, red) or 0.5 (40µM, blue) molar equivalents of SseL VHS domain. Line broadening effects suggest an interaction in the low-to-mid micromolar range. D) 1H,15N-HSQC TROSY titration experiment showing 80 µM 15N-labeled Ub alone (black) and in the presence of 0.25 (20 µM, red) or 0.5 (40 µM, blue) molar equivalents of SseL (24-340). E) As in C, showing 0.5 molar addition of the S1 site Y183A mutant (green). F) As in C, showing 0.5 molar addition of the covalently-linked SseL~Ub complex (magenta). G) Exposed hydrophobic patch within the SseL VHS domain; Leu46 was chosen for mutation. H) 1H,15N-HSQC TROSY spectrum of 80 µM 15N-labeled Ub alone (black) and in the presence of 0.5 molar equivalents of SseL (24-340) L46R (40µM, cyan). I) Overlay of the SseL catalytic and VHS domains (teal and cyan, respectively) with the STAM:Ub crystal structure (PDB 3LDZ, blue and red, respectively). As a conventional VHS:Ub interaction would clash with the position of the SseL catalytic domain, an unconventional site centered around Trp105 is utilized. J) Time course assays monitoring cleavage of K63-linked diUb with the SseL VHS domain mutants.
Figure S7: SseL accessory VHS domain dictates subcellular localization

A

pWSK SseL-2HA

ΔssaV

ΔsseL

ΔssaV

ΔsseL

WT

WT

YW3A

YW105A

WT

WT

YW3A

YW105A

SseL-2HA

SseB

DnaK

Pellet

Secreted
Figure S7: SseL accessory VHS domain dictates subcellular localization; related to Figure 7
A) Induced SPI-2 T3SS secretion of cultured S. Typhimurium strains expressing 2HA-tagged wild-type or mutant SseL, in the ΔsseL or ΔssaV (SPI-2 T3SS null) background. Secreted and pellet fractions were separated by SDS-PAGE and immunoblotted for HA (SseL), SseB (SPI-2 T3SS translocon protein), and DnaK (intracellular control protein).
Supplemental Experimental Procedures:

Cloning and molecular biology
SseL, ElaD, and LegCE were cloned from the genomic DNA of Salmonella Typhimurium (strain LT2), Escherichia coli (serotype O157:H7), and Legionella pneumophila (strain Philadelphia, kind gift from C. Buchrieser, Institut Pasteur) using Phusion DNA polymerase (NEB). Genes encoding Chlamydia trachomatis ChlaDUB1, Shigella flexneri ShiCE, Rickettsia bellii RickCE, and Xanthomonas campestris XopD were obtained by gene synthesis (Life Technologies) and subcloned using KOD polymerase (Novagen). Amplified products were cloned into pOPIN vectors (Berrow et al., 2007) using the In-Fusion HD system (Clontech). Mutagenesis was performed using the QuikChange protocol. Constructs for YopJ in pMW and AvrA in pGEX6P-1 were kind gifts from R. Mittal (MRC LMB Cambridge). Constructs for SENP1 and NEDP1 in pHISTEV were kind gifts from R. Hay (University of Dundee). tSUMO was amplified from a Solanum lycopersicum cDNA preparation (kind gift from A. Canto-Pastor and D. Baulcombe, University of Cambridge) and ligated into pTXB1 with conventional methods.

Protein expression and purification
SseL (24-340), ChlaDUB1 (130-401), ElaD (2-407), ShiCE (2-405), RickCE (378-691 or 420-691), LegCE (141-360), and XopD (298-515 or 335-515) were expressed from the pOPIN-B vector (N-terminal His-3C), AvrA (1-288) from the pGEX6P-1 vector (N-terminal GST-3C), YopJ (1-288) from the pMW vector (C-terminal 3C-GST), and SENP1 (415-643), NEDP1 (2-212) from the pHISTEV vector (N-terminal His-TEV). The SseL 24-340 construct may represent the full-length sequence secreted by S. Typhimurium, depending on usage of an alternative GTG start codon at position 24. The SseL VHS domain construct, cloned into pOPIN-B, included residues 24-137 as well as an ‘EQQVGENLWRD’ C-terminal sequence designed using the crystal structure to reconstitute key stabilizing interactions made by the C-terminal catalytic domain. With the exception of YopJ, SENP1, NEDP1, and SseL VHS domain, all tags were cleaved as part of the purification. All constructs were transformed into the E. coli Rosetta2 pLacI strain. 2-12 L cultures were grown at 37°C in 2xTY medium to an OD600 of 0.8-1.0, at which time protein expression was induced with 0.2 mM IPTG for 16-20 h at 18°C. Cells were resuspended in 25 mM Tris (pH 7.4), 200 mM NaCl, 2 mM β-mercaptoethanol (Buffer A). Following freeze-thaw and addition of EDTA-free Complete protease inhibitor tablets (Roche), DNase, and Lysozyme, cells were lysed by sonication or using an EmulsiFlex C3. Lysates were clarified by centrifugation at 35000 x g for 25 min, and applied to Talon (Clontech) or Glutathione Sepharose 4B (GE Healthcare) resin for gravity flow purification according to manufacturer recommendations. His-tagged proteins were washed thoroughly with Buffer A, eluted in Buffer A with 250 mM imidazole, and (with the exception of SENP1 and NEDP1) the tag was cleaved overnight with His-3C protease during dialysis back to Buffer A. Cleaved protein was reapplied to Talon resin and flow-through was collected for further purification. GST-tagged proteins were washed thoroughly with Buffer B (25 mM Tris (pH 8.5), 5 mM DTT) plus 500 mM NaCl, then Buffer B plus 50 mM NaCl, prior to elution with 10 mM GSH (in the case of YopJ) or overnight cleavage with addition of GST-3C protease. When required, cleaved proteins were exchanged into Buffer B plus 50 mM NaCl and applied to anion exchange chromatography (ResourceQ 6 mL, GE Healthcare), followed by elution with a 50-500 mM NaCl gradient. The SseL VHS domain was purified from inclusion bodies. Following solubilization in Buffer A plus 8 M Urea, 10 mM NH4Cl, SseL VHS domain was bound to gravity flow Ni2+ resin (Qiagen), refolded using a gradient to Buffer A without Urea, and eluted in Buffer A plus 250 mM imidazole. All proteins were subjected to final gel filtration (Superdex75, GE Healthcare) equilibrated in 25 mM HEPES (pH 8.0), 150 mM NaCl, 5 mM DTT. Protein-containing fractions were concentrated, aliquoted, and flash-frozen prior to storage at -80°C.

TAMRA-based cleavage assays
All Ub/Ubl-TAMRA-KG reagents were prepared as previously described (Geurink et al., 2012; Basters et al., 2014), with the exception of the tSUMO-TAMRA-KG reagent, which was prepared as follows:

Native chemical ligation
tSUMO-MesNa (aa 1-96, 4.0 mg, 0.37 µmol), prepared as previously described (Borodovsky et al., 2002), and TMR-Lys(thioG) (4.0 mg, 5.4 µmol) were dissolved in 200 µL of aqueous buffer containing 6.0 M Gm HCl, 0.15 M Na2HPO4 and 0.25 M MPAA at pH 7.2 and shaken overnight at 37°C. The product was purified by RP-HPLC (Shimadzu, C18, 10-70% v/v ACN/H2O gradient, 0.05% v/v TFA). The appropriate fractions were pooled and lyophilized.

Desulfurization
The resulting product (ca. 4 mg) was dissolved in aqueous buffer containing 6.0 M Gm HCl, 0.15 M Na2HPO4 and 0.25 M TCEP at pH 7.0 to a concentration of 1 mg/mL protein. Reduced glutathione (GSH) was added to the solution to a concentration of 100 mM. The pH of the solution was adjusted to pH 7.0 by addition of 1 M NaOH. VA-044 was added to the solution to a final concentration of 75 mM. The reaction mixture was flushed with argon and shaken for 4 h at 37°C after which LC-MS analysis indicated a complete reaction. The product
was purified by RP-HPLC (Shimadzu, C18, 10-70% v/v ACN/H2O gradient, 0.05% v/v TFA). The appropriate fractions were combined, lyophilized, dissolved in H2O/ACN/formic acid (65/35/10; v/v/v; 10 mL) and lyophilized again. Finally, the product was purified on a Superdex75 size exclusion column (GE Healthcare) equilibrated in 50 mM Tris (pH 7.6), 100 mM NaCl. This yielded 0.85 mg (75 nmol, 21%) of the tSUMO-TAMRA-KG substrate.

Fluorescence polarization (FP) assays monitoring cleavage of the Ub/Ubl substrates were performed as described previously (Geurink et al., 2012). Briefly, enzymes and Ub/Ubl substrates were diluted in 25 mM Tris (pH 7.4), 100 mM NaCl, 5 mM β-mercaptoethanol, 0.1 mg/mL BSA (FP buffer). Enzyme concentrations were varied to capture a representative activity profile, up to a maximum concentration of 150 nM. Substrate concentrations were held at 150 nM. Enzyme and substrate were mixed in black 384-well plates (Corning) immediately before the assay began. FP was measured in a Pherastar plate reader (BMG Labtech) equipped for 550 nm excitation and 590 nm emission. Cleavage measurements were normalized to independently measured negative (Ub/Ubl substrate only) and positive (25 nM TAMRA-KG peptide) control samples. All comparative data (Ub vs. Ubl, wild-type vs. mutant) are presented from one representative replicate.

**Qualitative cleavage assays**

Deubiquitination specificity assays were performed as described previously (Licchesi et al., 2012). Enzymes were diluted to a 2x stock concentration in 25 mM Tris (pH 7.4), 150 mM NaCl, 10 mM DTT and combined 1:1 with 6 μM diUb prepared in 100 mM Tris (pH 7.4), 100 mM NaCl, 10 mM DTT. Reactions were incubated at 37°C, time points were resolved by SDS-PAGE and silver stained (BioRad).

**Acetylation assays**

Enzymes were diluted to 5 μM in 25 mM HEPES (pH 8.0), 50 mM NaCl, 200 mM inositol hexakisphosphate (IP6), 0.5 mM DTT and incubated with 60 μM [1-14C] AcCoA (60 mCi/mmol, PerkinElmer) at 37°C for 2 h. Reactions were resolved by SDS-PAGE. The gel was dried and exposed to a Phosphor screen, which was then imaged on a Typhoon scanner (GE Healthcare). In our assays, catalytically inactive mutants, as well as unrelated protein standards (not shown), can be non-enzymatically modified by AcCoA. Therefore, only an increased signal in wild-type assays compared to the catalytically inactive controls should be considered true acetylation activity.

**Suicide probe assays**

Ub- and Ubl-PA probes were generated as described previously (Ekkebus et al., 2013), with exception of tSUMO-PA and all Ub/tSUMO-PA point mutants, which were prepared according to (Borodovsky et al., 2002) with propargylamine. Assays testing Ub/NEDD8 specificity were performed using suicide probes prepared from 2-chloroethyamine, as described previously (Borodovsky et al., 2002). Enzymes were diluted to 10 μM in 25 mM Tris (pH 7.4), 150 mM NaCl, 10 mM DTT, and combined 1:1 with 100 μM probe. Reactions were incubated at room temperature or on ice (as specified), time points were resolved by SDS-PAGE and Coomassie stained.

**Protein crystallization**

Initial hits were obtained using commercial screens in the sitting-drop vapor diffusion format. Native and SeMet His-SseL (24-340) were prepared in 25 mM Tris (pH 7.4), 200 mM NaCl, 2 mM β-mercaptoethanol and crystallized at 14 mg/mL in 0.1 M Tris, 1.6 M K2HPO4 (final pH 8.7), with a 1 μL sitting drop at 2:1 protein:precipitant ratio. SseL crystals were cryoprotected in ParatoneN (native crystals) or paraffin oil (SeMet crystals).

Native ChlaDUB1 (130-401) was prepared in 25 mM Tris (pH 7.4), 125 mM NaCl, 4 mM DTT and crystallized at 12 mg/mL in 0.1 M HEPES (pH 7.5), 20% PEG 8000, with a 400 nL sitting drop at 1:1 protein:precipitant ratio. SeMet ChlaDUB1 was crystallized at 8 mg/mL in a 1 μL sitting drop of the matched condition, using seeds of native crystals. Crystals were cryoprotected using mother liquor containing 22-25% glycerol.

Native RickCE (420-691) was prepared in 25 mM Tris (pH 7.4), 125 mM NaCl, 4 mM DTT and crystallized by hanging drop, mixing 2 μL of 13 mg/mL protein with 1 μL of 0.1 M Tris (pH 7.5), 2 M ammonium citrate. SeMet RickCE was crystallized the same manner, using native RickCE crystals as seeds. Crystals either had no cryoprotectant (SeMet RickCE), or were cryoprotected with mother liquor containing 22-25% glycerol. The XopD–Ub complex was prepared in 25 mM Tris (pH 7.4), 125 mM NaCl, 4 mM DTT and crystallized at 12 mg/mL in 0.1 M CHES (pH 9.5), 1.0 M sodium citrate, with a 200 nL sitting drop at 1:1 protein:precipitant ratio. Crystals were cryocooled with no protectant. The XopD–tSUMO complex was prepared similarly, and
crystallized at 8 mg/mL in 0.1 M bicine (pH 9.0), 1.6 M ammonium sulfate. XopD~tSUMO crystals were cryoprotected in 3.4 M malonate.

**Data collection, structure determination, and refinement**

Diffraction data were collected at the European Synchrotron Radiation Facility (ESRF) beam line ID23-1, and the Diamond Light Source (DLS) beam lines I03, I04, and I04-1 (see Table 1). Images were integrated using either MOSFLM (Battye et al., 2011) or XDSC (Kabsch, 2010) and scaled using Aimless (Evans and Murshudov, 2013). Structures were solved experimentally using SeMet SAD datasets in PHENIX AutoSol and AutoBuild (Adams et al., 2010; Terwilliger et al., 2009; Terwilliger et al., 2008) (Table 1). XopD~Ub and XopD~tSUMO were solved using molecular replacement in Phaser (McCoy et al., 2007) using the apo XopD structure (PDB 2O1V) and Ub (PDB 1UBQ) or SUMO1 structures (PDB 1TGZ). Model building and refinement were performed using COOT (Emsley et al., 2010) and PHENIX (Adams et al., 2010). The XopD~Ub structure was refined with the higher resolution XopD structure (XopD~tSUMO) and Ub (PDB 1UBQ) as reference models. See Table 1 for final statistics.

All figures were generated using PyMOL (www.pymol.org). Secondary structure topology diagrams are based on output from the Pro-origami server (Stivala et al., 2011).

**Plant extract protease assays**

*S. lycopersicum* leaves (kind gift from A. Canto-Pastor and D. Baulcombe, University of Cambridge) were flash frozen in liquid nitrogen and ground by mortar and pestle in a lysis buffer containing 50 mM HEPES (pH 8.0), 150 mM NaCl, 25 mM sucrose, 5 mM DTT, 1 mM EDTA, and 1 mM PMSF. Insoluble material was removed following 16000 x g centrifugation at 4°C for 10 min. The clarified lysate was incubated with recombinant XopD constructs at a 1 μM final concentration for 1 h at room temperature. Thirty micrograms of the resulting samples were separated by SDS-PAGE, transferred to nitrocellulose, and blotted for total ubiquitin (Ubi-1, Novus Biologicals) following Ponceau staining of total protein.

**Quantitative cleavage assays**

FlAsH-tagged diUb substrates were assembled enzymatically (UBE2N/UBE2V1 for K63 chains, UBE2R1 for K48 chains) by ligation of C-terminally FlAsH tagged Ub to ‘blocked’ Ub mutants (K63R for K63 chains, K48R for K48 chains). Unlabeled diUb chains were assembled analogously, replacing FlAsH-tagged Ub with a construct lacking its C-terminus (Ub aa 1-72). FlAsH-labeled diUb was held at a constant concentration of 75nM in all assays, and the remaining substrate consisted of unlabeled diUb. Assays were performed in 20 μL using black 384-well plates (Corning). FP was measured in a Pherastar plate reader (BMG Labtech) equipped for 485 nm excitation and 520 nm emission. SseL wild-type was used at 100 nM and 10 μM final concentration for K63 and K48 measurements, respectively. SseL Y244A was used at 1 μM final concentration in all assays. For each substrate concentration, initial rates, measured in triplicate, were subtracted from a substrate-only control and converted into velocities of substrate conversion. Kinetic analysis was performed using Graphpad Prism 6.

**Construction of CE dendrogram**

Given the significant sequence and structural diversities within the CE clan, we adopted a multipronged approach to build an accurate sequence alignment as described below. 1) Bonafide members of the CE superfamily of proteases were obtained from MEROPS, Pfam, and Interpro databases (Rawlings et al., 2014; Finn et al., 2014; Mitchell et al., 2015). 2) A “seed” structure-based sequence alignment was constructed based on available structures of CE superfamily domains such as SENP1, SENP2, NEDP1, L3 23K of human adenovirus 2 and XopD (PDB ID: 2XPH, 3Z05, 2BKZ, 1AVP, and 2O1X, respectively) (Rimsa et al., 2011; Alegre and Reverter, 2014; Shen et al., 2005; Ding et al., 1996; Chosed et al., 2007) and the presented structures of SseL, ChiDaUB1, and RickCE. The alignment was constructed using MUSTANG (Konagurthu et al., 2006) and DaliLite (Holm and Park, 2000) programs and was further manually refined based on residue conservation patterns and assignment of secondary structures from the above PDB structures. 3) The alignment was then expanded to include sequence data from other families within the CE clan such as YopJ and the Poxviridae proteases. We also included more sequence data from the families already represented in the seed alignment. An initial version of the expanded alignment was constructed using MUSCLE program (Edgar, 2004) with the option of profile-to-sequence alignment, which does not perturb the original seed alignment. This expanded alignment was further manually refined based on residue conservation patterns and secondary structure prediction by JPRED (Drozdzetskyi et al., 2015) to obtain a final expanded alignment. 4) Once an accurate version of the expanded alignment was constructed, we performed database searches using HMMER, PSI-BLAST and HHpred (Finn et al., 2011; Altschul et al., 1997; Söding et al., 2005) to find more divergent homologs of known CE superfamily members. This process identified e.g. the SiDE family as part of the CE clan. Representative members of SiDE family were then included in the expanded alignment. 5) MEGA software (Tamura et al., 2007) was used to construct a dendrogram based on the expanded alignment of the CE clan using
the UPGMA method with a bootstrap procedure for 1000 trials. A maximum likelihood method was also employed, but did not reliably produce dendrograms with cohesive functional relationships, likely due to high levels of dissimilarity among branchpoints.

**GST pull-down assays**

GST-tagged CI-MPR peptide was prepared by ligating sequence for the ‘FHDDSDELLHI’ peptide (CI-MPR aa 2480-2491) into the pOPIN-K vector, which encodes an N-terminal GST-3C tag. The wild-type sequence, as well as the ‘LL-AA’ double mutant were expressed and purified alongside the empty pOPIN-K vector, which contains the ‘GTVDPTGKRATSAQML’ sequence at its C-terminus as a control. 50 µg of GST or GST-MPR were phosphorylated by incubation with 250 units of CK2, 5 mM MgCl$_2$, and 1 mM ATP at 30°C for 30 min. For the pull-down assay, 50 µg of each GST-tagged protein was bound to 25 µL of Glutathione Sepharose 4B resin equilibrated in 20 mM Tris (pH 7.4), 100 mM NaCl, 5 mM ß-mercaptoethanol, and washed 3x with 250 µL additional cold buffer. 50 µg of SseL was added to a final concentration of 25 µM, and allowed to bind on ice for 30 min. Samples were washed 7x with 250 µL of cold buffer, and eluted with 25 µL buffer containing 10 mM GSH. Samples of SseL loading and GST elutions were resolved by SDS-PAGE and visualized by Coomassie staining.

**NMR spectroscopy**

Uniformly labeled $^{15}$N-Ub was expressed in M9 minimal media supplemented with $^{15}$N-NH$_4$Cl, and purified as described previously (Pickart and Raasi, 2005). All proteins were exchanged into 25 mM sodium phosphate (pH 7.0), 150 mM NaCl, 1 mM DTT. NMR spectra were recorded at 298K on either a Bruker Avance III 600 MHz or Avance2+ 700 MHz spectrometer, equipped with cryogenic triple resonance TCI probes. Data processing and analysis were performed in Topspin (Bruker) and NMRView (One Moon Scientific).

**S. Typhimurium infection assays**

The S. Typhimurium $\Delta$sseL strain was transformed with pWSK29 plasmids containing SseL-2HA mutants (see table of strains below). SPI-2 dependent secretion of SseL was tested using the pH shift secretion assay (Yu et al., 2010). Briefly, overnight cultures were diluted 1:50 into Mg-MES (pH 5) and grown for 4 hours, then transferred to Mg-MES (pH 7.2) for 90 minutes. Secreted protein and bacterial cells were collected and resolved by SDS-PAGE and immunoblotted with DnaK (Salmonella intracellular chaperone, Enzo ADI-SPA-880-D), SseB (SPI-2 T3SS translocon protein, generated in Beuzôn et al., 2002) and HA (SseL, HA.11 Cambridge Biosciences) antibodies.

HeLa cells were infected for 16 hours at a multiplicity of infection (MOI) of 200 with late exponential phase S. Typhimurium strains. Cells were collected in SDS sample buffer for total cell lysate or collected into Phosphate-Buffered Saline (PBS), pelleted at 300 x g and lysed in PBS + 0.1% triton. The lysate was clarified at 3,000 x g and the supernatant collected as the cytosolic fraction. Samples were resolved by SDS-PAGE and immunoblotted with HA.11 (SseL), DnaK (S. Typhimurium intracellular chaperone), or Actin antibodies. HeLa cells seeded on coverslips were infected as above and fixed with 3% Paraformaldehyde at 16 hpi. 0.1% saponin was used to permeabilize HeLa cells membrane but not bacterial membranes. Translocated SseL and S. Typhimurium were detected using HA (3F10 Roche) and CSA-1 (Insight Biotechnology) antibodies, respectively.

Strains used in this study:

<table>
<thead>
<tr>
<th>Strain</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta$sseL</td>
<td>$\Delta$sseL:Km in 12023</td>
<td>(Rytkönen et al., 2007)</td>
</tr>
<tr>
<td>$\Delta$sseL pSseL 2HA</td>
<td>pWSK29$sseL$-2HA in $\Delta$sseL</td>
<td>(Rytkönen et al., 2007)</td>
</tr>
<tr>
<td>$\Delta$saV pSseL 2HA</td>
<td>pWSK29$sseL$-2HA in $\Delta$saV</td>
<td>(Rytkönen et al., 2007)</td>
</tr>
<tr>
<td>$\Delta$sseL pSseL 2HA Y183A</td>
<td>pWSK29$sse$LY183A$-2HA in $\Delta$sseL</td>
<td>This study</td>
</tr>
<tr>
<td>$\Delta$sseL pSseL 2HA W105A</td>
<td>pWSK29$sse$LW105A$-2HA in $\Delta$sseL</td>
<td>This study</td>
</tr>
</tbody>
</table>
Supplemental References: