Chemical probes of surface layer biogenesis in *Clostridium difficile*

T. H. Tam Dang,[1,†] Lucia de la Riva,[1,2,†] Elisabeth M. Storck,[1] Robert P. Fagan,[2] William P. Heal,[1] Claire Janoir,[3] Neil F. Fairweather,[2] and Edward W. Tate[1*]

[1]Department of Chemistry, Imperial College London, Exhibition Rd., London SW7 2AZ, UK; [2] Centre for Molecular Microbiology and Infection, Department of Life Sciences, Imperial College London, Exhibition Rd., London SW7 2AZ; [3] EA 4043, Université Paris-Sud 11, Faculté de Pharmacie, 92296 Châtenay-Malabry, France. [†]These authors contributed equally to this work. *Author for correspondence: e.tate@imperial.ac.uk

Abstract: *Clostridium difficile*, a leading cause of hospital-acquired infection, possesses a dense surface layer (S-layer) that mediates host-pathogen interactions. The key structural components of the S-layer result from proteolytic cleavage of a precursor protein, SlpA, into high- and low-molecular-weight components. Here we report the discovery and optimization of the first inhibitors of this process in live bacteria, and their application for probing S-layer processing. We also describe the design and *in vivo* application of activity-based probes that identify the protein Cwp84 as the cysteine protease that mediates SlpA cleavage. This work provides novel chemical tools for the analysis of S-layer biogenesis, and for the potential identification of novel drug targets within clostridia and related bacterial pathogens.

INTRODUCTION

Clostridium difficile is a Gram-positive antibiotic-resistant anaerobe that can cause severe gastrointestinal disease in humans (*1, 2*) and livestock (*3*), and is a leading cause of hospital-acquired infection (*4*). The *C. difficile* cell is completely coated by the surface layer (S-layer), a paracrystalline proteinacious array that mediates host-pathogen interactions (*5, 6*). Previously we have shown that the key structural components of the S-layer are the result of proteolytic cleavage of a precursor protein, SlpA, into high- and low-molecular-weight components, namely the HMW and LMW S-layer proteins (SLPs) (*5*). These proteins reassemble to form the hexagonal S-layer lattice, and recently we reported the structure of the resultant HMW/LMW SLP complex (*7*). It is estimated that during exponential growth up to 400 molecules of SlpA per cell per second are synthesized, translocated to the S-layer and cleaved (*8*), indicating an acute demand for LMW and HMW SLPs in the expanding S-layer. The cleavage site in SlpA is widely conserved across many clinical strains of *C. difficile*, suggesting a similarly conserved protease is responsible for the processing event. Whilst constitutive gene knockouts can be achieved in *C. difficile* using the elegant ClosTron system (*9*), there is no reliable technique for achieving conditional knockouts, and despite widespread efforts it has not proven possible to generate a SlpA knockout strain to date. We hypothesized that a chemical genetic approach exploiting specific protease inhibitors and activity-based probes (ABPs) could be a powerful

approach for the study of SlpA processing *in vivo*. Here we report the discovery of the first inhibitors of S-layer processing in live *C. difficile* cells, and the development of activity-based probes that enable detection and identification of the protease responsible for S-layer processing.

^ SCHEME 1

RESULTS AND DISCUSSION

The cleavage products of SlpA may be observed directly by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) followed by Western blot analysis of the total S-layer fraction against an anti-LMW SLP antibody (*7*). Screening a panel of commercially-available protease inhibitors in this *in vivo* assay against the *C. difficile* reference strain 630 revealed that E-64 (Fig. 1A) has a modest inhibitory effect on SlpA processing at 100 µM, as shown by the appearance of unprocessed SlpA at 74 kDa apparent molecular weight (Fig. 1B). E-64 represents the first known inhibitor of S-layer processing in *C. difficile*; it is a natural product first isolated from *Aspergillus japonicas*, and is a broad-spectrum irreversible inhibitor of papain cysteine proteases (*10*). Further investigations demanded facile access to inhibitor analogues, and a flexible solid-phase synthetic route (Scheme S1) was therefore used to generate epoxide **1**, a peptidic analogue of E-64 (*11*). **1** retains the N-terminal *trans*-(*S,S*)-epoxysuccinyl moiety, which was appended to a Leu-Arg dipeptide amide, recapitulating the key stereochemistry and functionality of the natural product (Fig. 1A). **1** was found to possess identical activity to E-64 in the *in vivo* assay (Fig. 1B), as does ethyl ester **2**, the synthetic precursor to **1**. Analogues in which the warhead epoxide is in the (*R,R*) configuration or in which the epoxysuccinyl group is replaced with an acetyl group are inactive *in vivo* (data not shown), demonstrating a highly stereoselective mode of action and confirming the requirement for the electrophilic warhead. The activity of **1** was further measured across a panel of seven clinical *C. difficile* strains in addition to 630 with a broad range of sequence variation around the cleavage site of SlpA (Table S1). Activity was closely comparable (Fig. S1A), strongly suggesting a conserved mode of action across all these strains.

^ FIGURE 1

With this evidence in hand we sought to explore the structure-activity relationship for this class of inhibitors, as a precursor to identifying the target enzyme. Little is known about expression and activity of proteases in *C. difficile* and to date only a handful has been characterized to any degree. Activity based probes (ABPs) are a powerful tool for the *de novo* identification of enzymes, particularly in organisms that lack facile genetic tools (*12, 13*). However, there are relatively few examples of the application of ABPs to bacterial systems (*14-19*), and their use in *C. difficile* is without precedent to date. A series of ABPs based on **1** were designed by appending L-propargyl glycine (Pra) to an expanded library of inhibitors with the objective of identifying their target and, by

extension, the protease that mediates S-layer processing. The alkyne motif in Pra provides a small chemical tag that is expected to have minimal effect on distribution and binding to the target enzyme, and can be addressed by a selective bioorthogonal ligation reaction to enable subsequent addition of labels for detection (*12, 20-22*). A panel of probes **3-17** was synthesized, each featuring a different amino acid motif in the targeting element (Fig. 1A), and tested *in vivo* against 630 at 100 µM, revealing a striking pattern of inhibitory activity (Fig. 2A). **3** (LR) had identical activity to E-64 and the parent inhibitor **1**, as expected. However, **5** (F) had equal activity, whilst **9** (Y) and **13** (R) showed significantly improved potency over E-64 and **1**; the remaining analogues **4**, **6-8**, **10-12** and **14-16** (A, I, L, V, S, H, K, E, G, P) had no detectable activity under the conditions tested. This interesting structure-activity relationship in which both positively charged and large aromatic residues are permitted can be rationalized by examining the conservation of the SlpA cleavage site across *C. difficile* strains (Table S1). On the basis of our previously published data (*23, 24*), the P_2 position (i.e. the second residue upstream of the cleavage site) is either Y or K, and we conjecture that these inhibitors target a papain class cysteine protease in the canonical mode for E-64 (*25, 26*), whereby the epoxide warhead occupies the P_1 position (S or G in most known cleavage sites) and the first amino acid in the inhibitor occupies P_2. Further testing of **4** (A) and **12** (K) supports this hypothesis, with no activity observed for **4** up to 500 µM, and moderate activity at concentrations of **12** above 250 µM (Fig. S2).

We next investigated the dependence of *in vivo* inhibition on inhibitor concentration (Fig. 2B). For **3**, little effect on S-layer processing was observed below 50 µM. However, **9** (Y) retains significant inhibition down to a concentration of 10 µM, indicating an order of magnitude improvement in *in vivo* activity over E-64 and **3**. Full-length SlpA continues to accumulate in the S-layer with increasing concentration of **3**, but reaches a plateau at around 500 µM (Fig. 2B). This puzzling result was readily rationalized upon examination of the supernatant of inhibited cultures: at inhibitor concentrations of 250 µM and above SlpA is shed from cells, eventually reaching levels in excess of the total S-layer protein retained on the cell wall (Fig. 2B). This remarkable result indicates that unprocessed SlpA is not well-tolerated within the tight matrix of the S-layer, and that inhibited cultures upregulate production of SlpA. This may be related to a general stress response, as increased *slpA* expression has been reported in *C. difficile* exposed to a variety of stresses (*27, 28*). Notably, although inhibition of S-layer processing does not appear to have a direct impact on growth rate in the rich media used in these experiments, it has a dramatic effect on cell wall integrity. The cell wall of *C. difficile* grown under normal culture conditions is exceptionally stable, highlighted by the fact that boiling cells in a strong anionic detergent (0.7% SDS, 100 °C) does not induce lysis; however, inhibited cells are highly sensitive to lysis under these conditions (Fig. S3).

^ FIGURE 2

As shown in Scheme 1, at least two proteases are expected to be involved in the maturation of SlpA – the signal peptidase that cleaves the signal sequence responsible for directing translocation to the cell surface, and the putative protease that separates the LMW and HMW SLPs. If translocation is a prerequisite for subsequent processing, in principle inhibition of either of these proteases could result in accumulation of full-length SlpA. However, full-length SlpA is found in the S-layer and supernatant fractions in inhibited cells, strongly suggesting that translocation precedes inhibition and that the inhibitors described here target the downstream processing step (step B in Scheme 1). We next sought to determine the target labeled by these ABPs in live cells, with the aim of identifying the protease specific to S-layer processing in *C. difficile*. Samples from each of the inhibition experiments above were visualized by direct in-gel fluorescence following bioorthogonal ligation to a fluorescent reporter (Fig. S4). This reagent bears an azide for bioorthogonal ligation using 3+2 copper catalyzed azide-alkyne cycloaddition (CuAAC) and a TAMRA fluorophore for sensitive detection of labeled proteins by in-gel fluorescence (*29, 30*). The results from this experiment are striking (Fig. 2C): a strongly labeled fluorescent band is observed in direct correlation with ABP activity. A second band is observed at slightly higher apparent molecular weight, and is particularly strongly labeled by **13** (R). Interestingly, weak bands were observed for **6** (I), **8** (V), **11** (H) and **12** (K), suggesting large hydrophobic and positivity charged residues may be tolerated to some extent. A number of lower molecular-weight bands were more weakly labeled in an inhibition-independent manner, presumably due to background CuAAC ligation (Fig. S5).

Having established that putative targets of these probes reside in the S-layer of *C. difficile* and should be readily accessible to ABPs in the medium we considered that ABPs directly incorporating a biotin label would simplify the process of target identification, removing the requirement for two-step labeling and eliminating any interference from background CuAAC ligation. Biotin was appended to the C-terminus *via* a short PEG linker by performing solid phase synthesis on Biotin NovaTag resin, providing PEG-Biotin labeled ABPs **17** (LR) and **18** (R) (Fig. 1A). These ABPs were found to possess identical activity to the parent inhibitors in live cell assays (Fig. 2D), and NeutrAvidin-HRP blot of S-layer fractions showed the same strong bands for labeled protein (Fig. 2E). The activity and labeling properties of **17** were also conserved across multiple strains of *C. difficile* (Fig. S1B). Labeling experiments on strain 630 were scaled up to provide sufficient protein for identification, and affinity purification on NeutrAvidin-agarose followed by elution into loading buffer and SDS PAGE enabled the direct visualization of bands by coomassie staining (Fig. 2E). Whilst the labeled bands were not visible by coomassie staining prior to enrichment, they were readily identified after pull-down. The highly abundant LMW and HMW SLPs remain largely in the supernatant, whilst the labeled protein co-purifies with a number of unlabelled proteins that have markedly increased abundance in enriched samples from cells treated with **17** or **18**.

MALDI-ToF fingerprint analysis against the *C. difficile* 630 genome revealed that the labeled proteins are all members of the cell wall protein family of which the cysteine protease Cwp84 is a member. Cwp84 is one of the few proteases previously characterized in *C. difficile*, and was previously reported to have proteolytic activity against various matrix proteins (*31, 32*). Genomic analysis reveals that Cwp84 is encoded in the same genomic region as SlpA (*33, 34*), and shows strong secondary structural homology to papain class cysteine proteases (*35*), consistent with the results described above. Recombinant Cwp84 is known to self-process (*31*), and the two labeled proteins may be intermediates on this pathway. CD1751 is a close homolog of Cwp84 (63% identity at the amino acid level), but the high sequence coverage achieved in these analyses (45%) rules out this protease as a candidate target. Three bands that are not labeled but that co-purify with labeled Cwp84 were identified as Cwp66, CD2767 and CD2797. *cwp66* and CD2797 are part of the same gene cluster as *slpA*; Cwp66 has previously been described as a temperature-dependent adhesin (*36*), whilst CD2767 and CD2797 have no known function.

The data described above provide convincing evidence that Cwp84 is the key target of the inhibitors and ABPs described herein. Furthermore, the close linkage between inhibition activity and labeling supports the hypothesis that Cwp84 is the protease responsible for S-layer processing in *C. difficile*. To confirm the on-target specificity of Cwp84 for SlpA we designed a construct ('LHG') containing SlpA residues 1-362 fused to a C-terminal GST tag (Fig. 3A). SlpA[1-362] includes the LMW SLP and the N-terminal part of the HMW SLP sufficient for interaction with the LMW SLP (termed the HMW interaction domain, HID, Fig. S6A) (*7*), and thus encompasses the SlpA cleavage site. Co-expression of His-tagged Cwp84 and LHG in *E. coli* on a dual expression plasmid resulted in cleavage of LHG to the LMW SLP and a GST-tagged fragment. These bands are not produced when LHG is co-expressed in the presence of a Cwp84 mutant in which the active site Cys identified by sequence homology was mutated to Ala (C116A) (Fig. 3B and Fig. S6B). N-terminal sequencing of each fragment demonstrated that Cwp84 cleaves this substrate at Ser321, i.e. at ...ETKS↓AND..., the same location observed *in vivo* in *C. difficile* strain 630 (Table S1) (*33*). The substrate selectivity of Cwp84 was then probed directly using this assay: seven LHG constructs bearing an alanine point mutation at the P_6' to P_1' or A322G at the P_1 positions flanking the cleavage site were co-expressed in the presence of Cwp84 and tested for cleavage (Fig. 3C). A clear trend was observed: the only mutant that completely lacks cleavage is K320A (the P_2' position), confirming the key importance of this site for substrate recognition. Interestingly, T319A appears to reduce cleavage efficiency relative to the wild type site, and E318A appears to enhance it somewhat. Cwp84 was expressed at equivalent levels across these experiments (Fig. S7), strongly suggesting that these effects are directly due to differences in substrate affinity rather than variations in Cwp84 activity.

^ FIGURE 3

CONCLUSION

Starting from the observation that SlpA is cleaved *in vivo* to generate the major structural components of the *C. difficile* surface layer, we have identified E-64 and **1** as the first inhibitors of this process, and reported the development of improved inhibitors **5**, **9** and **13** with *in vivo* activity ranging down to low-µM concentration. These chemical genetic probes of S-layer formation have been used to demonstrate that full-length SlpA is poorly tolerated in the S-layer, leading to a marked reduction in cell integrity. Our results also point to a regulatory mechanism that drives increased SlpA expression in response to chemical inhibition of Cwp84, a remarkable observation in view of the already very high rate of S-layer synthesis in uninhibited cells. On the basis of the results presented here it is plausible that S-layer processing is regulated and concerted, and we speculate that it may be coordinated by a conserved protein complex. Alkyne-tagged and directly-labeled activity-based probes have been developed, enabling the target of these inhibitors to be identified as Cwp84 and providing the first reported example of the application of ABPs in *C. difficile*; the SlpA cleavage site has also been confirmed by an *in vivo* assay reconstituted in *E. coli*. The lack of a direct effect on cell growth even at high inhibitor concentrations suggests that Cwp84 may not be an essential gene under optimal growth conditions, and by extension that S-layer processing might not be essential for viability, at least in rich media. This hypothesis has recently been confirmed by a Cwp84 knockout strain of *C. difficile*, reported whilst this paper was in preparation (*37*). *C. difficile* 630 Δ*cwp84* is indeed viable in rich medium, but grows more slowly than wild-type 630 and displays significantly different colony morphology. Furthermore, the S-layer of this strain contains significant quantities of unprocessed SlpA, and the LMW and HMW SLPs are not observed. Taken together with the results we report here these data provide very strong evidence that the major role for Cwp84 *in vivo* lies in cleavage of SlpA during surface layer biogenesis, though our data do not preclude possible co-activity against extracellular matrix proteins (*31*). We have recently shown that another S-layer protein, CwpV, is cleaved post-translationally in strain 630 (*38*), but it is currently unknown whether Cwp84 or another as yet unidentified protease is responsible for this processing.

Cwp84 is highly conserved in *C. difficile (39, 40)*, and our studies across multiple clinical strains demonstrate that Cwp84 expression, activity and function also appear to be highly conserved. However, CD1751, a homologue of Cwp84, was not detected in labeled samples, and does not appear to play a role in S-layer processing under standard culture conditions. The apparently canonical binding mode of E-64 derivatives for Cwp84 and our substrate specificity analyses provide *in vivo* evidence for classification of Cwp84 as a papain-class cysteine protease that possesses a key substrate recognition determinant at the P_2' position. It is interesting to note that this enzyme appears to accept substrates with residues of contrasting chemical character (either aromatic or charged) at P_2 (*23, 24*), whilst most other residues are excluded on the basis of our inhibition and labeling data. Furthermore, certain ABPs described here (**13** and **18**) appear to intercept Cwp84 at multiple stages of self-

processing in a structure-dependent fashion, presenting a potential tool for analysis of Cwp84 activation *in vivo*.

The marked effect of Cwp84 inhibition on cell integrity as judged by susceptibility to lysis might be expected to lead to strong effects on viability under conditions of stress (e.g. antibiotic treatment, attack by the host immune system), and warrants further investigation of S-layer processing as a potential therapeutic target. The data presented here also demonstrate that Cwp84 is readily accessible to inhibitors in the medium, circumventing issues of intracellular uptake through the dense S-layer. Weakening of the S-layer by this mechanism might also result in improved penetration of co-administered antibiotics, and it will be interesting to see whether Cwp84 inhibitors can act synergistically with such agents. The inhibitors and ABPs described here represent useful and novel tools for studying cell wall biogenesis in *C. difficile*, and provide a unique approach for the detection of active Cwp84 in live cells. Work is ongoing in our labs to exploit these probes for the further investigation of these processes, and their extension to related pathogenic organisms.

METHODS

C. difficile **strains**. *C. difficile* 630 has been fully sequenced (*35*). R8366, R7404, Y, CDKK167 and CDKK959 have been described previously (*23, 33*). R13541 was obtained from Deʹ irdre Nıʹ Eidhin (Trinity College, Dublin) and SE528 from Dr Michelle Delemee, Universiteʹ Catholique de Louvain, Brussels.

C. difficile **culture conditions.** *C. difficile* was routinely cultured either on blood agar base II (Oxoid Ltd, Basingstoke, UK) supplemented with 7% horse blood (TCS Biosciences, Botolph Claydon, UK); or in brain-heart infusion (BHI) broth (Oxoid). Culture was undertaken in an anaerobic cabinet (Don Whitley Scientific, Shipley, UK) at 37 ºC in a reducing anaerobic atmosphere (10% CO_2, 10% H_2, 80% N_2). Glycine extracts were prepared as described previously (*33*).

In vivo **inhibition assay.** A single colony of *C. difficile* was inoculated in BHI broth and grown anaerobically for 1 hour. The culture was aliquoted, and bacteria were grown in the presence of each inhibitor overnight (16-18 hours). The cells were harvested by centrifugation, resuspended in SDS-PAGE loading buffer, boiled and analyzed by SDS-PAGE and western blotting. All assays used strain 630, unless otherwise specified.

SlpA cleavage assay in *E. coli*. A derivative of SlpA was co-expressed in *E. coli* with Cwp84 to assay cleavage activity. Residues 1-362 of SlpA were cloned as a fusion with glutathione-S-transferase in MCS2 of pACYCDuet-1 (Novagen). Residues 30-803 of Cwp84 were then cloned in the MCS1 of the same plasmid to form the co-expression plasmid. Derivatives of Cwp84 and LHG containing amino acid substitutions were created using inverse PCR. *E. coli* cells were grown in LB-

broth with chloramphenicol (30 µg ml^{-1}) and glucose (0.5%) to OD 0.7, IPTG added to 120 µM and induction continued for 24 hours. Cell lysates were prepared and analyzed by SDS-PAGE and western blotting using anti-LMW SLP (1 in 400,000) and anti-GST antibodies (1 in 4,000). Pull-downs using glutathione sepharose 4B matrix was performed according to the manufacturer's protocol (Amersham).

ACKNOWLEDGEMENTS

Supporting Information Available: supplementary scheme, tables and figures, full experimental details, including the synthesis and characterization of **1-18**. This material is free *via* the Internet.
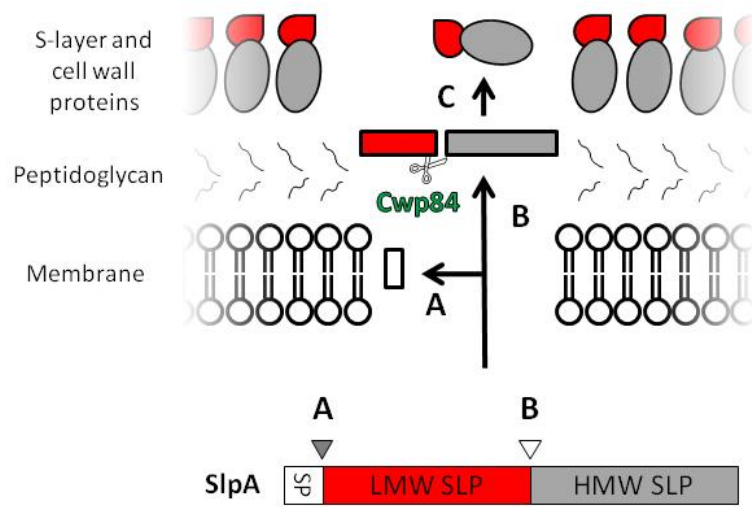
REFERENCES

1. Bartlett, J. G. (2007) *Clostridium difficile*: old and new observations, *Journal of clinical gastroenterology 41 Suppl 1*, S24-29.
2. Kuipers, E. J., and Surawicz, C. M. (2008) *Clostridium difficile* infection, *The Lancet 371*, 1486-1488.
3. Songer, J. G. (2004) The emergence of Clostridium difficile as a pathogen of food animals, *Anim. Health Res. Rev. 5*, 321-326.
4. Rupnik, M., Wilcox, M. H., and Gerding, D. N. (2009) *Clostridium difficile* infection: new developments in epidemiology and pathogenesis, *Nat. Rev. Microbiol. 7*, 526-536.
5. Calabi, E., Calabi, F., Phillips, A. D., and Fairweather, N. (2002) Binding of *Clostridium difficile* surface layer proteins to gastrointestinal tissues, *Infect. Immun. 70*, 5770-5778.
6. Masuda, K., Itoh, M., and Kawata, T. (1989) Characterization and reassembly of a regular array in the cell wall of *Clostridium difficile* GA I4131, *Microbiol. Immunol. 33*, 287-298.
7. Fagan, R. P., Albesa-Jove, D., Qazi, O., Svergun, D. I., Brown, K. A., and Fairweather, N. F. (2009) Structural insights into the molecular organization of the S-layer from *Clostridium difficile*, *Mol. Microbiol. 71*, 1308-1322.
8. Sleytr, U. B., and Messner, P. (1983) Crystalline surface layers on bacteria, *Annu. Rev. Microbiol. 37*, 311-339.
9. Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P., and Minton, N. P. (2007) The ClosTron: A universal gene knock-out system for the genus *Clostridium*, *Journal of microbiological methods 70*, 452-464.
10. Schirmeister, T., and Klockow, A. (2003) Cysteine protease inhibitors containing small rings, *Mini. Rev. Med. Chem. 3*, 585-596.
11. Bogyo, M., Verhelst, S., Bellingard-Dubouchaud, V., Toba, S., and Greenbaum, D. (2000) Selective targeting of lysosomal cysteine proteases with radiolabeled electrophilic substrate analogs, *Chem. Biol. 7*, 27-38.
12. Heal, W. P., Wickramasinghe, S. R., and Tate, E. W. (2008) Activity based chemical proteomics: profiling proteases as drug targets, *Curr. Drug Discov. Technol. 5*, 200-212.
13. Fonovic, M., and Bogyo, M. (2007) Activity based probes for proteases: applications to biomarker discovery, molecular imaging and drug screening, *Curr. Pharm. Des. 13*, 253-261.
14. Stubbs, K. A., Scaffidi, A., Debowski, A. W., Mark, B. L., Stick, R. V., and Vocadlo, D. J. (2008) Synthesis and use of mechanism-based protein-profiling probes for retaining beta-D-glucosaminidases facilitate identification of Pseudomonas aeruginosa NagZ, *J. Am. Chem. Soc. 130*, 327-335.

15. Shen, A., Lupardus, P. J., Albrow, V. E., Guzzetta, A., Powers, J. C., Garcia, K. C., and Bogyo, M. (2009) Mechanistic and structural insights into the proteolytic activation of Vibrio cholerae MARTX toxin, *Nat. Chem. Biol. 5*, 469-478.
16. Puri, A. W., and Bogyo, M. (2009) Using small molecules to dissect mechanisms of microbial pathogenesis, *ACS Chem. Biol. 4*, 603-616.
17. Bottcher, T., and Sieber, S. A. (2009) Structurally refined beta-lactones as potent inhibitors of devastating bacterial virulence factors, *ChemBioChem 10*, 663-666.
18. Pitscheider, M., and Sieber, S. A. (2009) Cinnamic aldehyde derived probes for the active site labelling of pathogenesis associated enzymes, *Chem. Commun.*, 3741-3743.
19. Staub, I., and Sieber, S. A. (2009) Beta-lactam probes as selective chemical-proteomic tools for the identification and functional characterization of resistance associated enzymes in MRSA, *J. Am. Chem. Soc. 131*, 6271-6276.
20. Hang, H. C., Loureiro, J., Spooner, E., van der Velden, A. W., Kim, Y. M., Pollington, A. M., Maehr, R., Starnbach, M. N., and Ploegh, H. L. (2006) Mechanism-based probe for the analysis of cathepsin cysteine proteases in living cells, *ACS Chem. Biol. 1*, 713-723.
21. Heal, W. P., Wickramasinghe, S. R., Bowyer, P. W., Holder, A. A., Smith, D. F., Leatherbarrow, R. J., and Tate, E. W. (2008) Site-specific N-terminal labelling of proteins *in vitro* and *in vivo* using N-myristoyl transferase and bioorthogonal ligation chemistry, *Chem. Commun.*, 480-482.
22. Speers, A. E., and Cravatt, B. F. (2004) Profiling enzyme activities in vivo using click chemistry methods, *Chem. Biol. 11*, 535-546.
23. Calabi, E., and Fairweather, N. (2002) Patterns of sequence conservation in the S-layer proteins and related sequences in *Clostridium difficile*, *J. Bacteriol. 184*, 3886-3897.
24. Eidhin, D., Ryan, A., Doyle, R., Walsh, J. B., and Kelleher, D. (2006) Sequence and phylogenetic analysis of the gene for surface layer protein, *slpA*, from 14 PCR ribotypes of *Clostridium difficile*, *J Med Microbiol 55*, 69-83.
25. Bhattacharya, S., Ghosh, S., Chakraborty, S., Bera, A. K., Mukhopadhayay, B. P., Dey, I., and Banerjee, A. (2001) Insight to structural subsite recognition in plant thiol protease-inhibitor complexes : understanding the basis of differential inhibition and the role of water, *BMC Struct. Biol. 1*, 4.
26. Schaschke, N., Assfalg-Machleidt, I., Machleidt, W., and Moroder, L. (1998) Substrate/propeptide-derived endo-epoxysuccinyl peptides as highly potent and selective cathepsin B inhibitors, *FEBS Lett. 421*, 80-82.
27. Deneve, C., Bouttier, S., Dupuy, B., Barbut, F., Collignon, A., and Janoir, C. (2009) Effects of Subinhibitory Concentrations of Antibiotics on Colonization Factor Expression by Moxifloxacin-Susceptible and Moxifloxacin-Resistant Clostridium difficile Strains, *Antimicrob. Agents Chemother.*
28. Deneve, C., Delomenie, C., Barc, M. C., Collignon, A., and Janoir, C. (2008) Antibiotics involved in Clostridium difficile-associated disease increase colonization factor gene expression, *J Med Microbiol 57*, 732-738.
29. Tate, E. W. (2008) Recent advances in chemical proteomics: exploring the post-translational proteome, *J Chem Biol 1*, 17-26.
30. Heal, W. P., Wickramasinghe, S. R., Leatherbarrow, R. J., and Tate, E. W. (2008) N-Myristoyl transferase-mediated protein labelling in vivo, *Org Biomol Chem 6*, 2308-2315.
31. Janoir, C., Pechine, S., Grosdidier, C., and Collignon, A. (2007) Cwp84, a surface-associated protein of *Clostridium difficile*, is a cysteine protease with degrading activity on extracellular matrix proteins, *J. Bacteriol. 189*, 7174-7180.
32. Poilane, I., Karjalainen, T., Barc, M. C., Bourlioux, P., and Collignon, A. (1998) Protease activity of Clostridium difficile strains, *Can. J. Microbiol. 44*, 157-161.
33. Calabi, E., Ward, S., Wren, B., Paxton, T., Panico, M., Morris, H., Dell, A., Dougan, G., and Fairweather, N. (2001) Molecular characterization of the surface layer proteins from *Clostridium difficile*, *Mol. Microbiol. 40*, 1187-1199.
34. Karjalainen, T., Waligora-Dupriet, A.-J., Cerquetti, M., Spigaglia, P., Maggioni, A., Mauri, P., and Mastrantonio, P. (2001) Molecular and genomic analysis of genes encoding surface-anchored proteins from *Clostridium difficile*, *Infect. Immun. 69*, 3442-3446.
35. Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., Thomson, N. R., Roberts, A. P., Cerdeno-Tarraga, A. M., Wang, H., Holden, M. T., Wright, A., Churcher, C., Quail, M. A., Baker, S., Bason, N., Brooks, K., Chillingworth, T., Cronin, A., Davis, P., Dowd, L., Fraser, A., Feltwell, T., Hance, Z., Holroyd, S., Jagels, K., Moule, S., Mungall, K., Price, C., Rabbinowitsch, E., Sharp, S., Simmonds, M., Stevens, K., Unwin, L., Whithead, S., Dupuy, B., Dougan, G., Barrell, B., and Parkhill, J. (2006) The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome, *Nat. Genet. 38*, 779-786.

36. Waligora, A. J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A., and Karjalainen, T. (2001) Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties, *Infect. Immun. 69*, 2144-2153.

37. Kirby, J. M., Ahern, H., Roberts, A. K., Kumar, V., Freeman, Z., Acharya, K. R., and Shone, C. C. (2009) Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of Clostridium difficile, *J. Biol. Chem. 284*, 34666-34673.

38. Emerson, J. E., Reynolds, C. B., Fagan, R. P., Shaw, H. A., Goulding, D., and Fairweather, N. F. (2009) A novel genetic switch controls phase variable expression of CwpV, a Clostridium difficile cell wall protein, *Mol. Microbiol. 74*, 541-556.

39. Savariau-Lacomme, M. P., Lebarbier, C., Karjalainen, T., Collignon, A., and Janoir, C. (2003) Transcription and analysis of polymorphism in a cluster of genes encoding surface-associated proteins of *Clostridium difficile*, *J. Bacteriol. 185*, 4461-4470.

40. Stabler, R. A., Gerding, D. N., Songer, J. G., Drudy, D., Brazier, J. S., Trinh, H. T., Witney, A. A., Hinds, J., and Wren, B. W. (2006) Comparative phylogenomics of *Clostridium difficile* reveals clade specificity and microevolution of hypervirulent strains, *J. Bacteriol. 188*, 7297-7305.

Scheme 1 - Dang et al.

Scheme 1. Overview of the three-stage maturation of SlpA in *C. difficile* to generate the S-layer protein complex. **A**: removal of the signal peptide by signal peptidase; **B**: cleavage by a protease (identified herein as Cwp84), to generate the high and low molecular weight surface layer proteins (HMW and LMW SLPs); **C**: re-association of LMW and HMW SLPs to form the S-layer matrix.
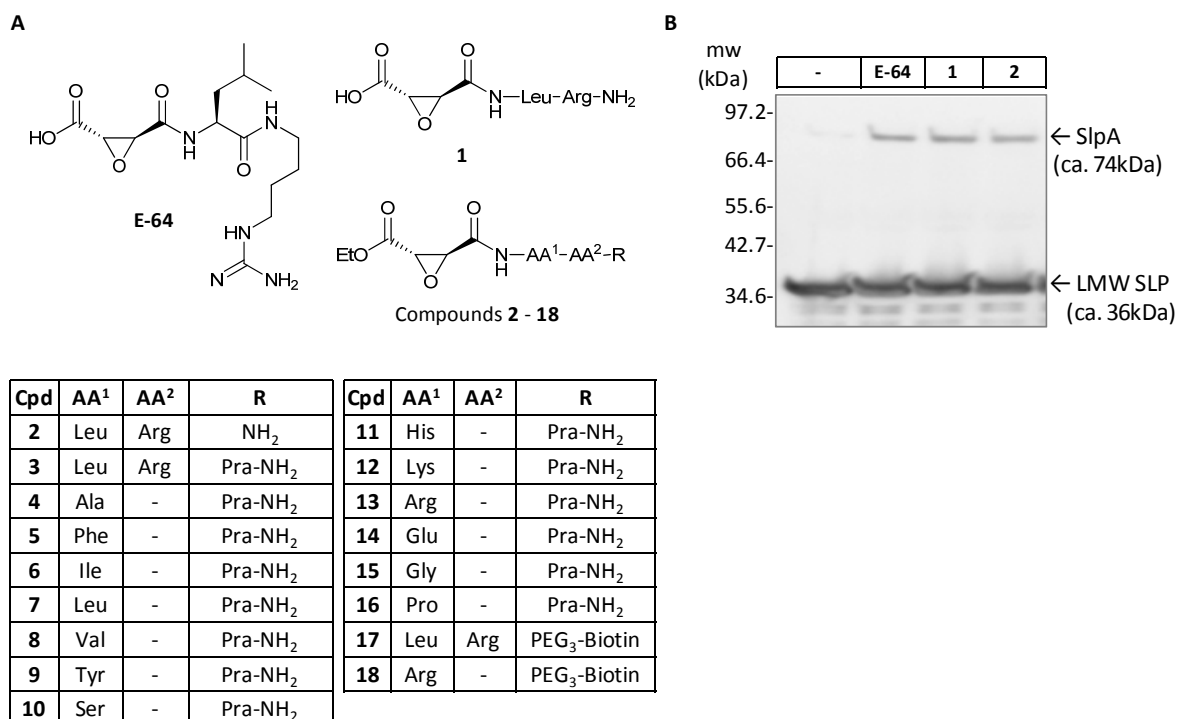
Figure 1 - Dang et al.

A

E-64

1

Compounds **2** - **18**

Cpd	AA¹	AA²	R
2	Leu	Arg	NH_2
3	Leu	Arg	$Pra-NH_2$
4	Ala	-	$Pra-NH_2$
5	Phe	-	$Pra-NH_2$
6	Ile	-	$Pra-NH_2$
7	Leu	-	$Pra-NH_2$
8	Val	-	$Pra-NH_2$
9	Tyr	-	$Pra-NH_2$
10	Ser	-	$Pra-NH_2$

Cpd	AA¹	AA²	R
11	His	-	$Pra-NH_2$
12	Lys	-	$Pra-NH_2$
13	Arg	-	$Pra-NH_2$
14	Glu	-	$Pra-NH_2$
15	Gly	-	$Pra-NH_2$
16	Pro	-	$Pra-NH_2$
17	Leu	Arg	PEG_3-Biotin
18	Arg	-	PEG_3-Biotin

B

Fig. 1 Structures of inhibitors and ABPs synthesized in this study

A: structures of E-64, synthetic inhibitors and ABPs **1-18** used in this study; **B**: activity of E-64, **1** and **2**, as measured by Western blot of S-layer fractions using anti-LMW SLP. *C. difficile* 630 cultures were treated with compound at 100 µM during overnight growth under anaerobic conditions in Brain Heart Infusion (BHI) medium, followed by isolation of the S-layer fraction as described (*7*).
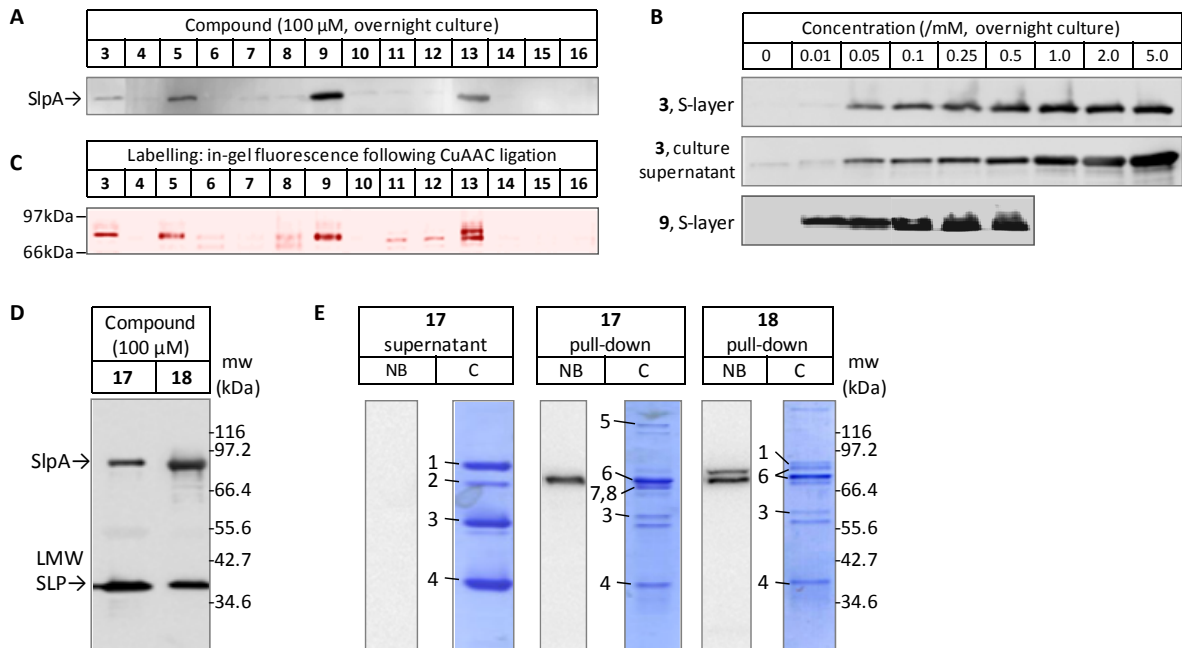
Figure 2 - Dang et al.

Fig. 2 *In vivo* activity and labeling by ABPs in *C. difficile* S-layer processing

A: activity across ABP series **3-16**, visualized by Western blot using anti-LMW SLP; **B**: concentration series for **3** (S-layer and culture supernatant) and **9** (S-layer), SlpA visualized as for A; **C**: labeling across ABP series **3-16**, as revealed by in-gel fluorescence after CuAAC ligation to a fluorescent reporter; **D**: activity of biotinylated ABPs **17** and **18**; **E**: NeutrAvidin-HRP blots (NB) and coomassie-stained gels (C) of glycine extracts labeled by **17** and **18**, showing proteins after affinity-purification (pull-down) with NeutrAvidin-agarose; a representative supernatant after pull-down following treatment of cells with **17** is also shown. The marked bands were identified as: 1, SlpA; 2, CD2791; 3, HMW SLP; 4. LMW SLP; 5. CD2797; 6. Cwp84; 7. Cwp66; 8. CD2767. See supporting information for further details.
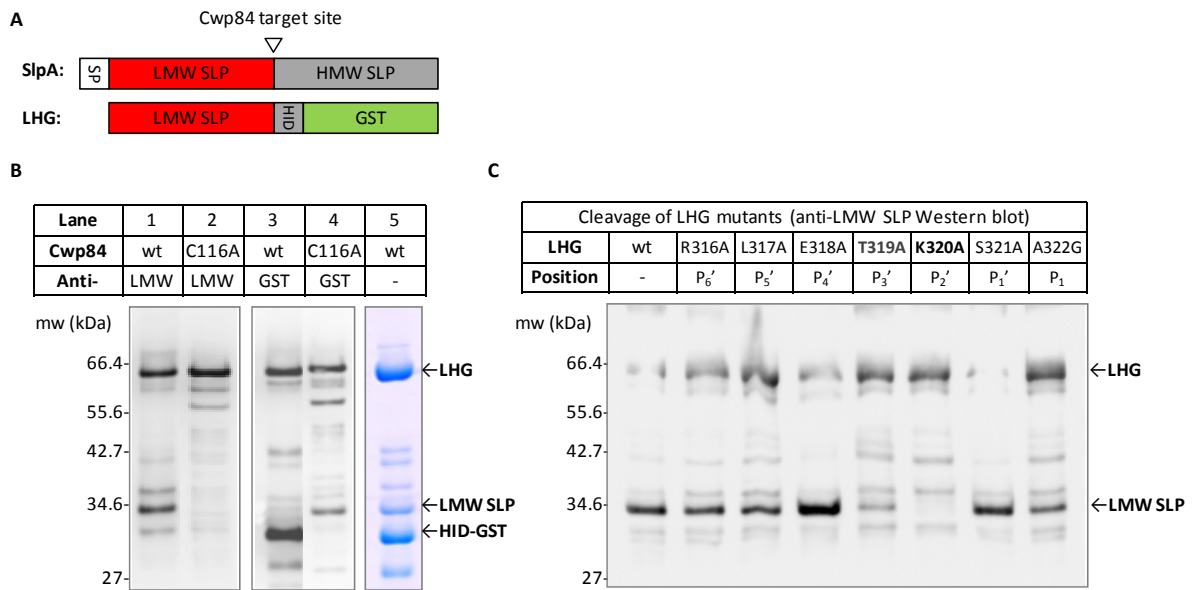
Figure 3 - Dang et al.

Fig. 3 Model cleavage assay for SlpA reconstituted in *E. coli*.

A: LHG construct; **B**: LHG processing in *E. coli* co-expressing LHG (all lanes) and either Cwp84 (lanes 1, 3, 5) or Cwp84[C116A] (lanes 2, 4); cells were lysed 24 hours post-induction, and soluble protein visualized by Western blot using anti- αLMW SLP (lanes 1, 2) or anti-GST (lanes 3, 4). Lane 5 shows coomassie-stained proteins following affinity purification by pull-down onto glutathione sepharose beads; N-terminal sequencing of the HID-GST band excised from the gel shown in lane 5 confirmed the cleavage site as ...ETKS↓ANDTIA...; **C**: *in vivo* cleavage assay for LHG constructs by Cwp84 containing a point mutations at the P_6' to P_1 sites flanking the scissile bond.

Graphical Abstract - Dang et al.